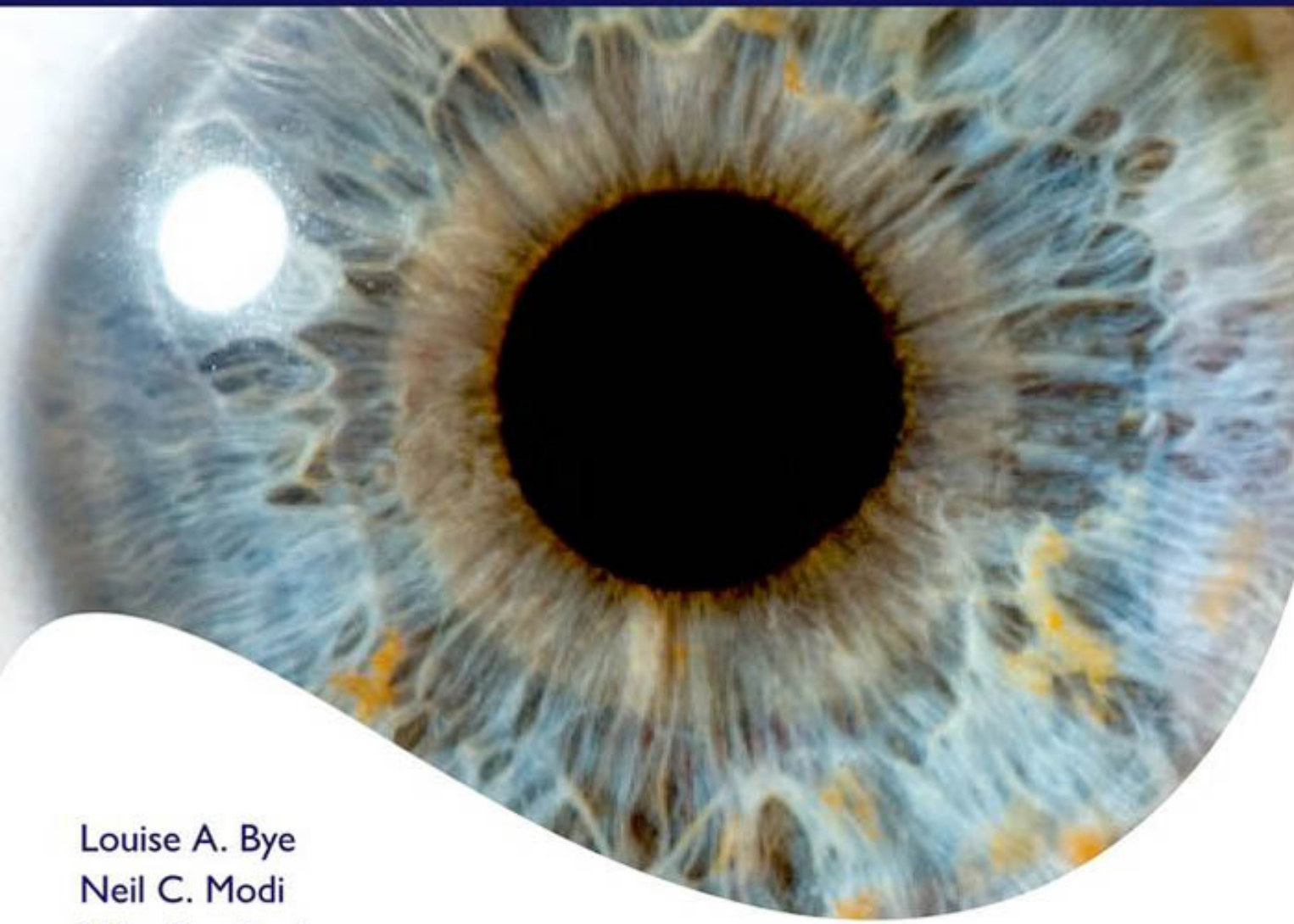
OST OXFORD SPECIALTY TRAINING



Louise A. Bye
Neil C. Modi
Miles Stanford

BASIC SCIENCES FOR OPHTHALMOLOGY

Contents

Symbols and abbreviations

5′-GMP	guanosine-5′-monophosphate
5HT$_2$	5-hydroxytryptamine
11-Ral	11-*cis*-retinaldehyde
11-Rol	11-*cis*-retinol
ABCR	ATP-binding cassette
AC/A	accommodative convergence/accommodation
ACF	anterior cranial fossa
AD	autosomal dominant
ADP	adenosine diphosphate
adRP	autosomal dominant retinitis pigmentosa
AIDS	acquired immune deficiency syndrome
AMD	age-related macular degeneration
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
ANOVA	analysis of variance
ANP	atrial natriuretic peptide
ANS	autonomic nervous system
APC	antigen-presenting cell
AR	autosomal recessive
ARMD	age-related macular degeneration
arRP	autosomal recessive retinitis pigmentosa
ATP	adenosine triphosphate
BCG	bacilli Calmette–Guerin
BINO	bilateral internuclear ophthalmoplegia
BSV	binocular single vision
CA	carbonic anhydrase
CALT	conjunctival-associated lymphoid tissue
CAM	cell adhesion molecules
cAMP	cyclic adenosine monophosphate
CD/DS	chondroitin sulphate/dermatan sulphate
CDC	Communicable Disease Centre
cGMP	cyc;ic guanosine monophosphate
CMV	cytomegalovirus
CN	cranial nerve
CNS	central nervous system
CNV	choroidal neovascularization
CRAO	central retinal artery occlusion
CRBP	cellular retinol-binding protein
CRVO	central retinal vein occlusion
CSF	cerebrospinal fluid
CSLO	confocal scanning laser ophthalmoscope
CTR	common tendinous ring
DAG	diacylglycerol
DC	dendritic cell
DCR	dacrocystorhinostomy
DISC	death-inducing signalling complex
DNA	deoxyribonucleic acid
DS	dioptre sphere
ECM	extracellular matrix
EDTA	ethylene-diamine-tetra-acetic acid
ELISA	enzyme-linked immunosorbent assay
ELM	external limiting membrane
EMZL	extranodal marginal zone lymphoma
EOG	electro-oculogram
EPV	Epstein–Barr virus
ER	endoplasmic reticulum
ERG	electroretinogram
FasL	Fas ligand
FGF	fibroblast growth factor
FISH	fluorescence *in situ* hybridization
FTA-Abs	fluorescent treponemal antibody absorption
GABA	gamma-aminobutyric acid
GAG	glycosaminoglycan
GDP	guanosine diphosphate
GTP	guanosine triphosphate
GVHD	graft-versus-host disease
H&E	haematoxylin and eosin
HAART	highly active antiretroviral therapy
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HPV	human papillomavirus
HSV-1	herpes simplex virus type 1
HTLV	human T-cell lymphotrophic viruses
IDP	inner dense plaque
IFN	interferon
Ig	immunoglobin
IGF-1	insulin-like growth factor
IL	interleukin
ILM	inner limiting membrane
INO	internuclear ophthalmoplegia
InsP$_3$	inositol 1,4,5-triphosphate
IOF	inferior orbital fissure
IOL	intraocular lens
IOP	intraocular pressure
IPD	inter-pupillary distance
IPL	inner plexiform layer
IPM	interphotoreceptor matrix
IR	infrared
IRBP	interphotoreceptor retinoid-binding protein
IRMA	intraretinal microvascular abnormality

ISA intrinsic sympathomimetic activity
ITT intention to treat
KS keratin sulphate
LASIK laser intrastromal keratomileusis
LGIC ligand-gated ion channel
LGN lateral geniculate nuclei
LOSD line of sight point for distance
LOSN line of sight point for near
LPS lipopolysaccharide
LRAT lecithin/retinol acetyltransferase
MALT mucosa-associated lymphoid tissue
MAPK mitogen-activated protein kinase
MAR minimum angle of resolution
MCF middle cranial fossa
MHC major histocompatibility complex
MIC minimal inhibitory concentration
MIP major intrinsic polypeptide, macrophage inflammatory protein
MLF medial longitudinal fasciculus
MPL medial palpebral ligament
MRI magnetic resonance imaging
mRNA messenger RNA
MRSA meticillin-resistant *Staphylococcus aureus*
NAD nicotinamide-adenine dinucleotide
NAD^+ oxidized form of NAD
NADH reduced nicotinamide adenine dinucleotide
NADPH reduced NAD phosphate
Nd:YAG neodymium doped yttrium aluminium garnet
NICE National Institute for Health and Clinical Excellence
NK natural killer
NMDA *N*-methyl-D-aspartate
NPE non-pigmented epithelium
NSAID non-steroidal anti-inflammatory drug
OCT optical coherence tomography
ODP outer dense plaque
OPL outer plexiform layer
PAMPS pathogen-associated molecular patterns
PAS periodic acid-Schiff
PBL peripheral blood lymphocyte
PCF posterior cranial fossa
PCR polymerase chain reaction
PDE phosphodiesterase
PKC protein kinase C
PMN polymorphonuclear neutrophil
POAG primary open-angle glaucoma
POHS presumed ocular histoplasmosis syndrome
PPC power progression corridor
PPRF paramedian pontine reticular formation
pRB retinoblastoma protein
PRK photorefractive keratectomy
PRM pathogen recognition molecule
PRR pattern recognition receptor
$PtdInsP_2$ phosphatidylinositol 4,5-bisphosphate
PTK phototherapeutic keratectomy
PVD posterior vitreous detachment
PVFL posterior vertex power of the lens
R rhodopsin
R* metarhodopsin II
RAG recombination activating gene
RAPD relative afferent pupillary defect
Rb retinoblastoma
RBP retinol-binding protein
RCT randomized controlled trial
RDH all-*trans*-retinol dehydrogenase
RF recombination fraction
RFLP restriction fragment length polymorphism
RI refractive index
RIA radioimmunoassay
RNA ribonucleic acid
RNFL retinal nerve fibre layer
RNI reactive nitrogen intermediate
ROI reactive oxygen intermediate
ROS reactive oxygen species, rod outer segment
RP retinitis pigmentosa
RPE retinal pigment epithelium
RR relative risk
RSM relative spectacle magnification
RT reverse transcriptase
RTK receptor tyrosine kinase
SE standard error
SEM standard error of the mean
SITA Swedish Interactive Threshold Algorithm
SM spectacle magnification
SNP single nucleotide polymorphisms
SOF superior orbital fissure
SSPE subacute sclerosing panencephalitis
$T_{\alpha\beta\gamma}$ transducin
TAC transit amplifying cells
TB tuberculosis
Tc cytotoxic T cells
TCA tricarboxylic acid
TCR T-cell receptor
TGF transforming growth factor
Th 'helper' T cell
Th2 T helper 2 cell
TI thymus- independent
TIGR trabecular meshwork inducible glucocorticoid response
TLR Toll-like receptor
TNF tumour necrosis factor
***t*-Ral** all-*trans*-retinal
***t*-RE** all-*trans*-retinyl ester
Treg regulatory T cell
***t*-Rol** all-*trans*-retinol
TTR transthyretin
UTI urinary tract infection
UTR untranslated region
UV ultraviolet
VDRL Venereal Diseases Research Laboratory
VEGF vascular endothelial growth factor
VEP visual evoked potential
VF visual field
VIP vasointestinal protein
VZV varicella zoster virus
XLPR X-linked retinitis pigmentosa
XR X-linked disorders
YAG yttrium aluminium garnet

CHAPTER 1

Anatomy

Cranial cavity

The skull

The skull is divided into two groups of bones:

- those that define the cranium
- those that define the face.

The cranium itself is divided into the vault and the base of the skull.

There are six bones that constitute the cranium:

1. frontal
2. parietal (paired)
3. occipital
4. temporal (paired)
5. sphenoid
6. ethmoid.

The following are the eight bones that constitute the facial bones:

1. zygomatic (paired)
2. maxillae (paired)
3. nasal (paired)
4. lacrimal (paired)
5. vomer
6. palatine (paired)
7. inferior conchae (paired)
8. mandible.

Lateral aspect of the skull

The lateral aspect of the skull is shown in Fig. 1.1.

- The frontal bones form the anterior part of the side of the skull and articulate with the parietal bone at the coronal suture.
- The parietal bones form the side and roof of the cranium and articulate with each other in the midline at the saggital suture. They articulate with the occipital bone at the lambdoid suture.
- The side of the skull is completed by the squamous part of the occipital bone, parts of the temporal bone (in particular the squamous, tympanic, mastoid process, styloid process, and zygomatic process), and the greater wing of the sphenoid.
- The thinnest area of the lateral wall of the skull is known as the pterion. It resides at the unison between the parietal bone and the greater wing of the sphenoid. The pterion is essentially the lateral bony housing for the anterior division of the middle meningeal artery and vein within the cranial cavity. Blows or fractures of the pterion have been known to rupture these vessels, leading to significant haemorrhage.

The temporal fossa lies between the temporal lines (which are an extension of the posterior curve of the frontozygomatic ridge) and the infratemporal crest of the greater wing of sphenoid (above the zygomatic arch).

The inferotemporal fossa lies beneath the infratemporal crest on the greater wing of sphenoid.

The pterygomaxillary fissure is a vertical fissure that lies within the fossa between the pterygoid process of the sphenoid bone and the maxilla.

This fissure transmits the terminal part of the maxillary artery and alveolar branches of the maxillary nerve. It leads medially into the pterygopalatine fossa, which is a small triangular cavity behind and below the apex of the orbital cavity. It communicates:

- laterally with the infratemporal fossa (through the pterygomaxillary fissure)
- medially with the nasal cavity (through the sphenopalatine foramen)
- superiorly with the skull (through the foramen rotundum)
- anteriorly with the orbit (through the inferior orbital fissure (IOF)).

The pterygopalatine fossa contains the pterygopalatine ganglion (suspended by nerve roots from the maxillary nerve), the second branch of the maxillary nerve (V2), the nerve of the pterygoid canal (a continuation of the facial nerve and

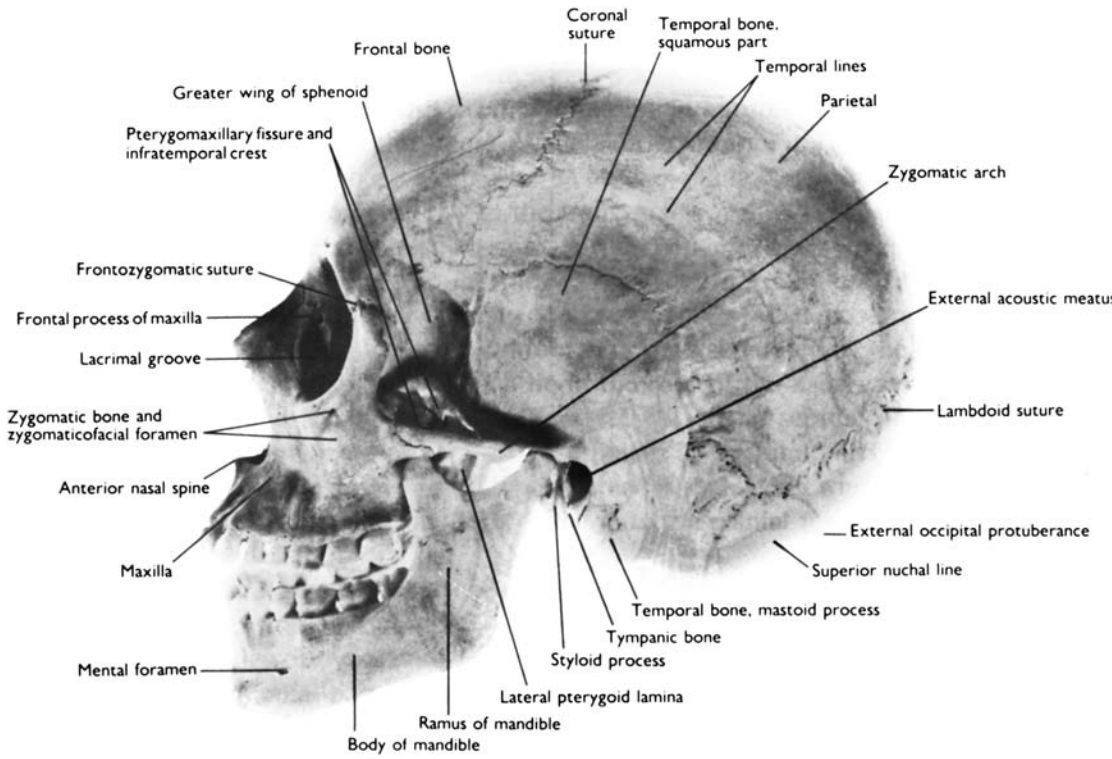


Fig. 1.1 The lateral aspect of the skull.

Reproduced from G.J. Romanes, *Cunningham's Manual of Practical Anatomy, Head and Neck and Brain*, Volume Three, Fifteenth Edition, Figure 5, Page 5, 1986, with permission from Oxford University Press.

carrying parasympathetic fibres), and the terminal third of the maxillary artery.

Anterior aspect of the skull

The anterior aspect of the skull is shown in Fig. 1.2.

The frontal bone

The frontal bone forms the curve of the forehead. It articulates medially with the nasal bones and the frontal processes of the maxilla. The frontal bones articulate laterally with the zygomatic bone and posteriorly with both parietal bones. Within the frontal bone itself are the two hollow spaces lined with mucous membrane, the frontal air sinuses.

The maxillary bones

The maxillary bones exist as a pair. They join medially to form:

- the upper jaw
- the anterior part of the hard palate (the posterior is formed by the palatine bone)
- a contribution to the lateral wall of the nasal cavity
- part of the orbital floor.

Within the maxillary bones lie the maxillary sinuses. These sinuses also share a close relationship with the orbital floor in both eyes. In the event of maxillary sinusitis, it is possible that infection may spread through the floor of the orbit into the orbit itself to give rise to orbital cellulitis.

The zygoma

The zygoma articulates medially with the maxilla and laterally with the zygomatic process of the temporal bone to give rise to the zygomatic arch. Superiorly the zygoma interacts with frontal bone at the fronto-zygomatic suture. Here, these two bones make up the lateral orbital rim.

Each zygoma contains two foramina. They house:

- zygomaticofacial nerves
- zygomaticotemporal nerves.

Orbital margins

The orbital margin/rim is a quadrilateral spiral structure and is bound by several bones.

For more detail see the separate section on the orbit (p. 25).

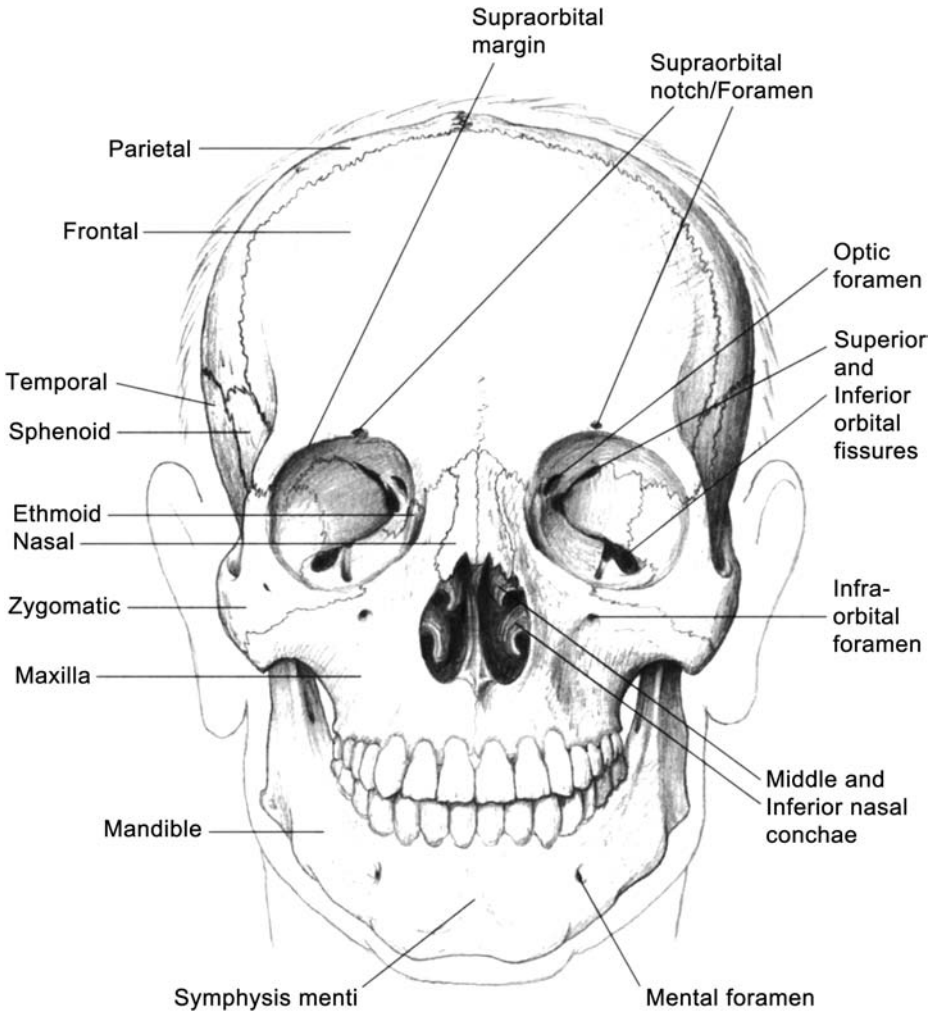


Fig. 1.2 Anterior view of the skull.

Reproduced from Pamela MacKinnon and John Morris, *Oxford Textbook of Functional Anatomy, Head and Neck*, 2005, Figure 6.1.8, Page 46, with permission from Oxford University Press.

The nasal apparatus

Anteriorly, two nasal bones form the bridge of the nose. These nasal bones are the superior boundaries to the opening of the nasal cavity (the anterior nasal aperture). Inferiorly and laterally, the boundaries of the nasal aperture are formed by the maxillae.

The nasal cavity is bisected by the nasal septum, which is mostly formed by the vomer (the bony extension of the nasal septum). On the lateral walls of the cavity lie the conchae. There are three sets of conchae to each lateral wall. The superior and middle conchae extend from the ethmoid. The inferior conchae are separate bones and extend from the maxilla.

Superior aspect of the skull

The superior aspect of the skull is shown in Fig. 1.3.

Superiorly, the frontal bone unites with the two parietal bones at the coronal suture.

The sagittal suture is formed between the two parietal bones.

Posterior aspect of the skull

The posterior aspect of the skull is shown in Fig. 1.4.

The lambdoid suture is formed by the unison of (the squamous part of the) occipital bone and both parietal bones.

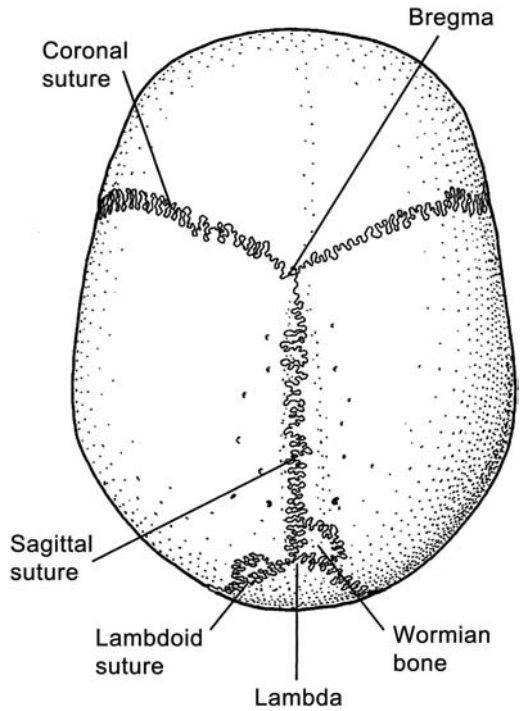


Fig. 1.3 Superior view of the skull.

Reproduced from Pamela MacKinnon and John Morris, *Oxford Textbook of Functional Anatomy, Head and Neck*, 2005, Figure 6.1.9b, Page 46, with permission from Oxford University Press.

In the midline of the occipital bone, there is an elevation called the external occipital protuberance and the occipital crest. From this, there arise attachments to muscles and to the ligamentum nuchae.

Inferior surface of the base of the skull

Fig. 1.5 shows the inferior surface of the skull with the mandible (lower jaw) removed.

The two maxillae form the upper jaw, the anterior part of the hard palate, part of the lateral walls of the nasal cavities, and part of the floor of the orbital cavities. The two maxillary bones join in the midline to form the intermaxillary suture, which forms the lower margin of the nasal aperture.

The palatal processes of the maxillae and the horizontal plates of the palatine bones can be identified. Anteriorly in the midline, the incisive fossa and foramen can be seen. Posterolaterally are the greater and lesser palatine foramina.

Above the posterior edge of the hard palate, the choanae (posterior nasal apertures) can be seen. These are separated medially by the vomer and laterally by the medial pterygoid plates of the sphenoid bone. Posterolateral to the lateral pterygoid plate lie the foramen ovale and the foramen spinosum.

Superior surface of the base of the skull

The internal surface of the base of the skull is shown in Fig. 1.6.

The base of the skull is divided into three large fossae. Each fossa lies at a successively lower level than the previous one:

- anterior cranial fossa (ACF)
- middle cranial fossa (MCF)
- posterior cranial fossa (PCF).

The anterior cranial fossa

The ACF houses the frontal lobes of the cerebral hemispheres. It is bound anteriorly and anterolaterally by the frontal bone.

The posterior border is marked by the lesser wing of sphenoid. The medial part of the lesser wing of sphenoid forms the anterior clinoid process. The floor of the ACF is bound by cribriform plates of the ethmoid medially and the ridged orbital plates of the frontal bone laterally. Medially, the crista galli is a sharp upward projection of the ethmoid bone, for the attachment of the falx cerebri.

The middle cranial fossa

There are several key structures that lie within the MCF. These are summarized in Table 1.1.

The lesser wing of the sphenoid and the anterior margin of the sulcus chiasmatis bind the MCF anteriorly. Other anterior boundary structures are the anterior clinoid processes, which overhang the MCF and form attachments with the free edge of the tentorium cerebelli.

The posterior border is formed by the petrous temporal bones. The posterior border also contains the posterior clinoid processes to which the attached edge of the tentorium cerebelli are fixed.

The lateral boundaries are formed by the squamous parts of the temporal bone, the greater wings of the sphenoid, and a small portion of the parietal bones.

The posterior cranial fossa

The contents of the PCF are shown in Table 1.2.

The PCF is the deepest of the cranial fossa. The ridge of the petrous portion of the temporal bone forms the anterior boundary of the PCF.

The posterior border of the PCF occurs at the internal surface of the squamous portion of the occipital bone.

The roof of the fossa is formed by the tentorium cerebelli.

The floor consists of the basilar and squamous portions of the occipital bone coupled with temporal bone (mastoid portion) laterally.

The meninges and venous sinuses

The meninges

The brain and spinal cord are surrounded by three membranes or meninges:

- the dura mater
- the arachnoid mater
- the pia mater.

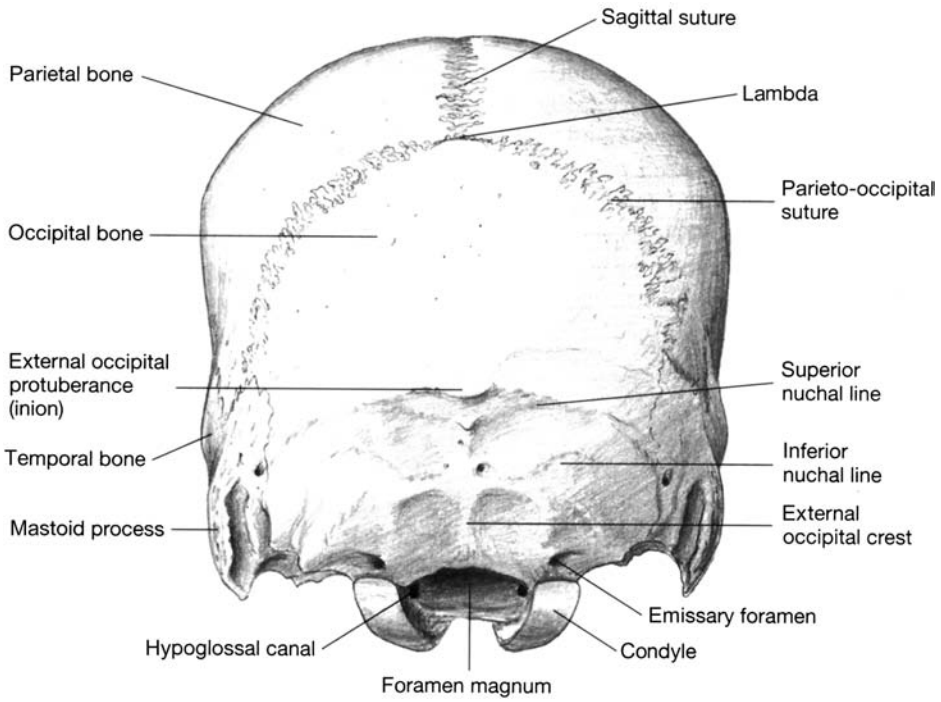


Fig. 1.4 The posterior aspect of the skull.

Reproduced from Pamela MacKinnon and John Morris, *Oxford Textbook of Functional Anatomy, Head and Neck*, 2005, Figure 6.1.10, Page 46, with permission from Oxford University Press.

The endosteum does not extend through the foramen magnum to remain continuous with the dura mater of the spinal cord. It does, however, become continuous with the periosteum around the skull.

The dura mater is continuous through the foramen magnum and remains continuous with the dura mater of the spinal cord.

Inside the skull, the meningeal layer sends up four septa, which divide the cranial cavity into freely communicating spaces. These septa prevent rotator displacement of the cranial contents. They are as follows:

- The falx cerebri—this is a sickle-shaped fold of dura mater that lies in the midline between the two cerebral hemispheres. Its narrow frontal end is attached to the crista galli and its broad posterior end is attached to the upper surface of the tentorium cerebella. The superior saggital sinus runs in its upper fixed margin and the inferior saggital sinus runs in its lower free margin. The straight sinus runs in its inferior attachment to the tentorium cerebella.
- The tentorium cerebelli—this is a crescent-shaped fold of dura mater that lies over the upper surface of the cerebellum and provides support for the occipital lobes of the cerebral hemispheres. Anteriorly, there is a notch—the tentorial notch—for the passage of the midbrain.
- The falx cerebelli—this is a small fold of dura mater that is attached to the internal occipital crest and passes between the two cerebellar hemispheres.
- The diaphragm sellae—this is a small fold of dura mater that forms the roof of the sella turcica. There is a small opening in the centre for passage of the stalk of the hypophysis cerebri.

The dura mater is sensitive to stretch, producing the sensation of headache, because of its innervations by the trigeminal, vagus, and first three cervical nerves as well as branches from the sympathetic system. The dura has a number of arteries that supply it. The most important is the middle meningeal artery, which arises from the maxillary artery in the infratemporal fossa. It passes through the foramen spinosum and passes along the squamous part of the temporal bone, lying between the meningeal and endosteal layers of the dura. It may be damaged in head injuries, especially to the temporal area.

The arachnoid mater acts as an impermeable membrane. It is separated from the dura by the subdural space and from the pia by the subarachnoid space, which is filled with cerebrospinal fluid (CSF). In certain areas, the arachnoid projects as arachnoid granulations into the venous sinuses. These are the sites at which the CSF drains into the bloodstream. The arachnoid mater is attached to the pia mater across the fluid-filled

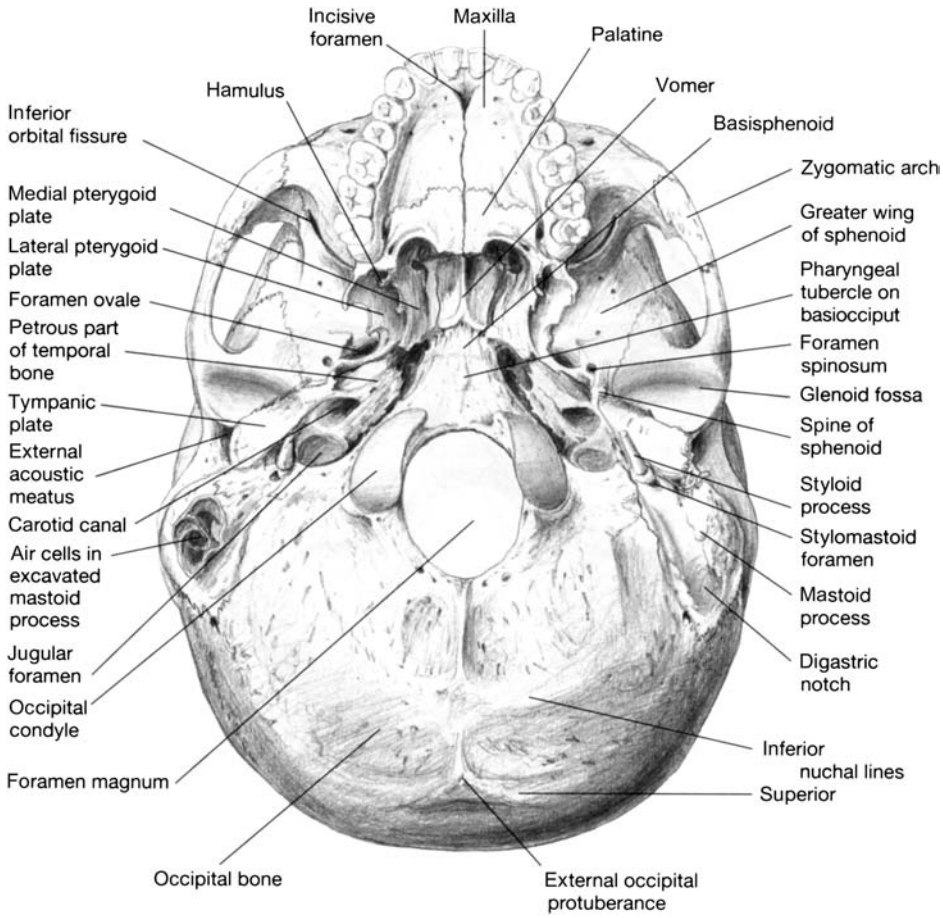


Fig. 1.5 The inferior surface of the base of the skull.

Reproduced from Pamela MacKinnon and John Morris, *Oxford Textbook of Functional Anatomy, Head and Neck*, 2005, Figure 6.1.12, Page 47, with permission from Oxford University Press.

subarachnoid space by fine strands of fibrous tissue. All structures passing through, both to and from the brain, must pass through the subarachnoid space. All the cerebral arteries, veins, and cranial nerves lie in this space. The optic nerve is surrounded by a sheath of arachnoid mater that extends into the orbital cavity.

Pia mater is a vascular membrane that covers the surface of the brain. It also covers and gives a sheath to the arteries and cranial nerves.

CSF removes waste products of neuronal activity and provides a medium in which the brain is bathed, thus protecting it from trauma. It is produced by the choroidal plexuses in the lateral, third, and fourth ventricles of the brain. It leaves the ventricular system of the brain through formamina in the roof of the fourth ventricle. From here, it circulates upwards over the surface of the cerebral hemispheres and downwards over the spinal cord. The spinal subarachnoid space extends down to the second sacral vertebra.

The venous sinuses of the cranial cavity

The venous sinuses lie within the layers of the dura mater (Fig. 1.7). The sinus walls do not contain a muscular layer but do comprise an endothelial layer.

The superior sagittal sinus lies medially and takes up the fixed, superior border of the falx cerebri. It originates from an anterior direction starting above the foramen caecum to move backwards towards the internal occipital protruberance. Here it is redirected horizontally to the right (usually) to give the transverse sinus. The superior sagittal sinus constitutes the major venous drainage site from middle and posterior portions of the two cerebral hemispheres.

The inferior sagittal sinus lies parallel and deep to the superior sagittal sinus. It lies medially within the free folds of the falx cerebri and consequently drains the lower portions of the medial cerebral hemispheres. The inferior sagittal sinus drains into the straight sinus, which

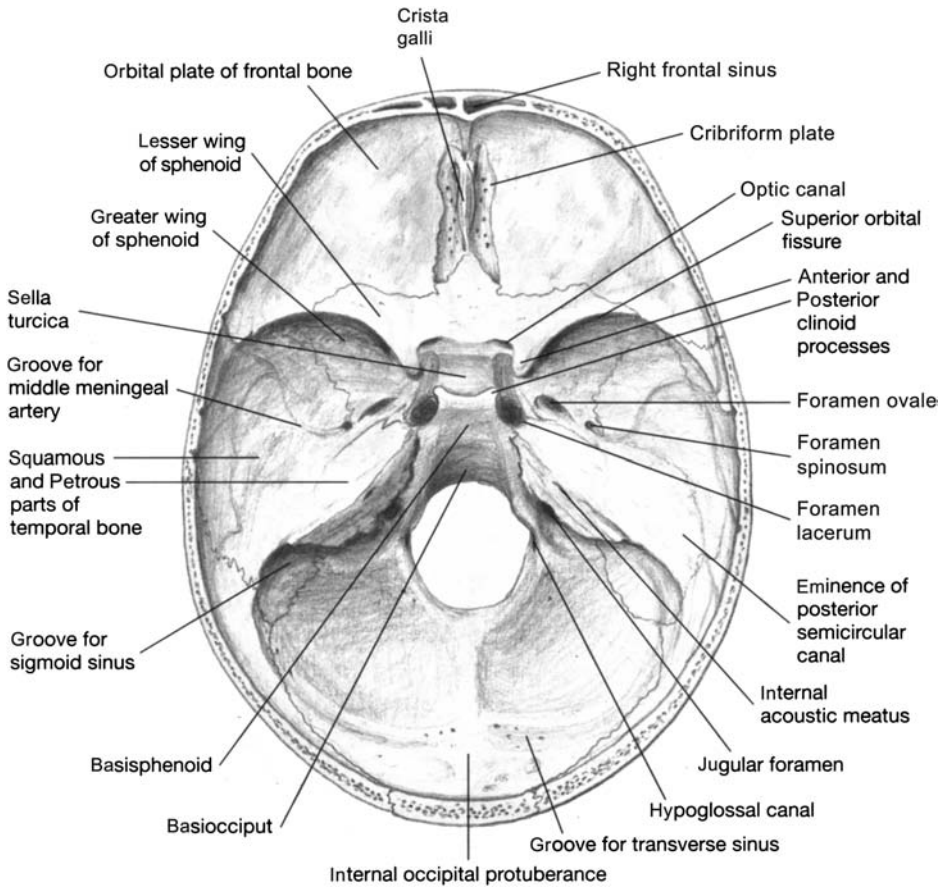


Fig. 1.6 The internal surface of the base of the skull.

Reproduced from Pamela MacKinnon and John Morris, *Oxford Textbook of Functional Anatomy, Head and Neck*, 2005, Figure 6.1.14, Page 48, with permission from Oxford University Press.

Table 1.1 Structures that lie within the middle cranial fossa

Structure	Location	Contents and conveyed structures
Sulcus chiasmatis	The most anterior structure of the MCF that lies between the optic canals, just anterior to the tubercle of the sella (tuberculum sella)	Very occasionally the optic chiasm comes into contact with this tranverse groove
Sella turcica (Turkish saddle)	Central flattened ridge of the sphenoid bone between the two optic canals In the sagittal plane the sella turcica lies between the anterior and posterior clinoid processes	The sella embodies the sulcus chiasmatis and hypophyseal fossa The hypophyseal fossa accommodates the pituitary gland
Optic canals	Anterolateral to the sella turcica	Optic nerve Ophthalmic artery
SOF	The SOF lies at the anterior border of the cavernous sinuses and between the greater and lesser sphenoid wings	Conveys: • lacrimal, frontal, and nasociliary nerves • cranial nerves (CN) III, IV, and VI • superior ophthalmic vein
Foramen rotundum	Greater wing of sphenoid posterior to the SOF	Conveys the maxillary nerve (CN V)

(*continued*)

Table 1.1 (*continued*)

Structure	Location	Contents and conveyed structures
Foramen ovale	Greater wing of sphenoid posterior to the foramen rotundum	Conveys: • O—the otic ganglion • V—the mandibular division of the maxillary nerve (V) • A—accessory meningeal nerve • L—lesser petrosal nerve • E—an emissary vein
Foramen spinosum	Greater wing of sphenoid posterior to the foramen ovale	Conveys: • the middle meningeal artery • the middle meningeal vein
Foramen lacerum	Located at the medial junction of the petrous temporal bone, the basilar part of the occipital bone, and greater wing of sphenoid	Conveys the internal carotid artery Provides a conduit for the internal carotid artery to travel from the carotid canal to the cavernous sinus
Trigeminal impression	Anterior part of the petrous portion of the temporal bone	Houses the trigeminal ganglion
Tegmentum tympani	Lies ahead of the anterior ridge of the petrous portion of the temporal bone	Forms the roof for: • the tympanic antrum • the mastoid antrum

MCF, middle cranial fossa; SOF, superior orbital fissure.

Table 1.2 Structures that lie in the posterior cranial fossa

Structure	Location	Contents and conveyed structures
Foramen magnum	Posterior to the clivus (Latin for slope), which is the central basilar portion of the occipital bone	Medulla and its meninges Ascending accessory nerves Vertebral arteries
Hypoglossal canal	Anterolateral wall of the foramen magnum	Conveys the hypoglossal (XII) nerve
Jugular foramen	Lies on the medial third of the junction between the petrous temporal bone and the occipital bone	Conveys the inferior petrosal nerve and the sigmoid sinus Conveys the IX, X, and XI cranial nerves
Internal acoustic meatus	Lies on the posterior of petrous portion of the temporal bone between the foramen lacerum and the jugular foramen	Conveys the VII and VIII cranial nerves
Transverse sinus groove	Lateral aspect of the internal occipital bone	Houses the transverse sinus

resides in the junction of the tentorium cerebelli and the falx cerebri. At this junction the straight sinus receives the two basilar veins and the cerebral vein of Galen, after which the straight sinus becomes continuous with the left (usually) transverse sinus. The tranverse sinuses reflect forward within the attached margins of the tentorium cerebelli.

The transverse sinuses receive blood from the superior petrosal sinuses, and inferior cerebral and cerebellar veins. On each side, the transverse sinuses tip downwards to become the right and left sigmoid sinuses, respectively. In turn, the right and left sigmoid sinuses continue forwards into the superior bulb of the internal jugular vein. The jugular veins exit the cranial cavity via the jugular foramen.

Central nervous system

Neuroanatomy

Major features

The two cerebral hemispheres are incompletely separated by the great longitudinal fissure. This is normally filled by the falx cerebri. Deep between the two hemispheres lies the corpus callosum, which contains commissural fibres that unite corresponding areas of the two hemispheres. The brainstem is the site of origin of the cranial nerves. Behind (dorsal) to the brainstem lies the cerebellum. The

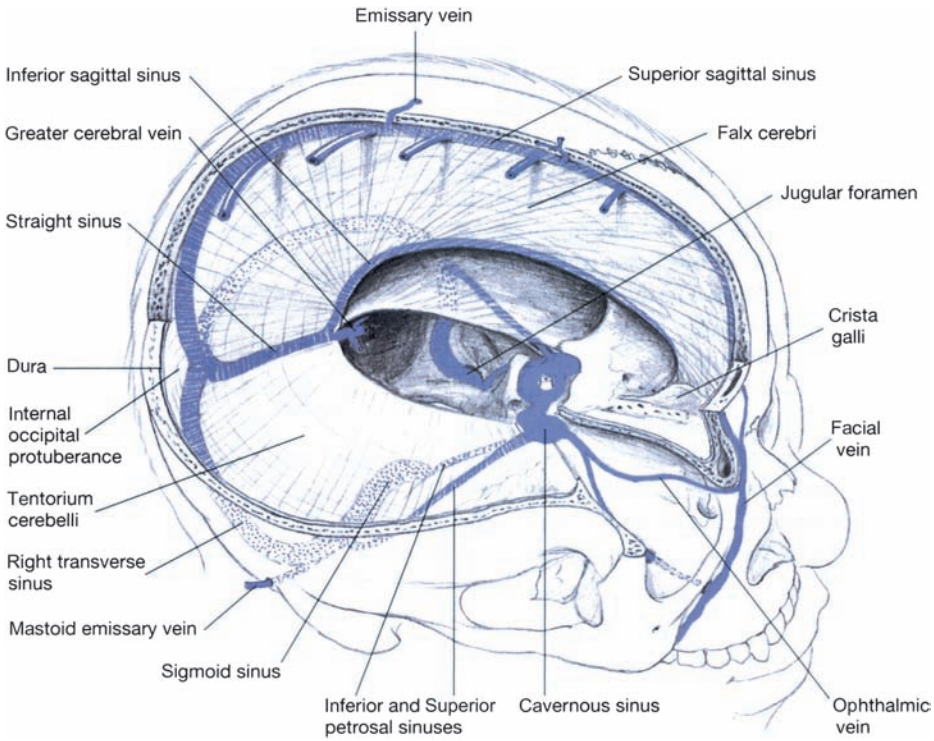


Fig. 1.7 The venous sinuses.

Reproduced from Pamela MacKinnon and John Morris, *Oxford Textbook of Functional Anatomy, Head and Neck*, 2005, Figure 6.9.2, Page 144, with permission from Oxford University Press.

cerebellum is separated from the occipital lobes above by the tentorium cerebelli.

Ventricular system

The various compressions and developments from the neural canal create a network of cavities within the central nervous system known as the ventricular system (Fig. 1.8).

The ventricular system is the site of production of CSF, which is secreted by the choroid plexus.

The choroid plexus, which produces the vast majority (70%) of CSF, is formed from a combination of ependymal cells (these line the ventricular system), the pia mater, and capillaries. Other cerebral capillaries form the remaining 30%.

The ventricular system is made up of four cerebral ventricles:

- the paired lateral ventricles
- the midline third ventricle
- and, below the third, the fourth ventricle.

The two lateral ventricles, located within each cerebral hemisphere, are the largest of the four ventricles. They are C-shaped and lie loosely around the dorsal aspects of the basal ganglia.

Both of the lateral ventricles extend into the frontal, occipital, and temporal lobes via the anterior, posterior, and temporal horns, respectively.

The lateral ventricles both exchange CSF via the interventricular foramina of Monro with the third ventricle.

The third ventricle is found along the midline within the diencephalon. It is situated within the midbrain and communicates via the cerebral aqueduct of Sylvius with the fourth ventricle, which is located within the hindbrain.

The three foramina to the subarachnoid space are found within the fourth ventricle:

- medially—the foramen of Magendie, leading to the cerebellomedullary cistern
- laterally—the two (left and right) foramen of Luschka, leading to the pontine cistern.

This permits CSF produced in the ventricles to escape into the subarachnoid space to bathe the brainstem, cerebellum, and cerebral cortex.

The fourth ventricle is also continuous with the central canal of the spinal cord, allowing CSF to bathe the spinal cord from within.

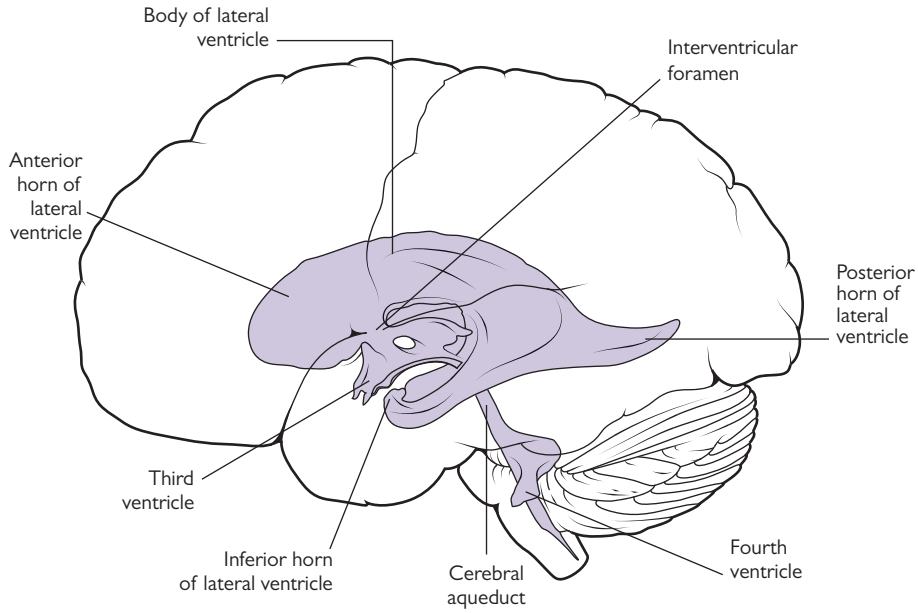


Fig. 1.8 The cerebral ventricular system.

This figure was published in *Neuroanatomy: an illustrated colour text*, A. Crossman and D. Neary, Figure 1.19 The cerebral ventricular system, p. 14, Copyright Elsevier 1995.

Relations of the ventricles

The caudate nucleus lies beneath the C-shaped concavity on each of the lateral ventricles. The head of the caudate nucleus lies next to the anterior horn, whereas the tail curves backwards and down before reflecting forward to form part of the roof of the inferior horn of the lateral ventricle.

The thalamus is situated inferiorly and medial to the caudate nucleus. The thalamus is the largest structure in the diencephalons and is located on top of the brainstem in a position to send nerve fibres out to the cerebral cortex. The two halves of the thalamus are prominent bulb-shaped masses, located symmetrically on the lateral walls of the third ventricle.

The thalamus is related to the following structures:

- the floor of the body of the lateral ventricles
- the roof of the inferior horn of the lateral ventricles.

The hypothalamus is located below the thalamus, just above the brainstem and behind the pituitary stalk. The superior surface of the hypothalamus forms the floor of the third ventricle. The groove in the lateral wall of the third ventricle demarcates the thalamus–hypothalamus junction.

The floor of the fourth ventricle denotes a landmark for several structures, including the sixth and seventh nerve nuclei.

Brainstem

Ascending and descending fibres that pass between the brain and spinal cord pass through the brainstem. It is the site of origin of many cranial nerves (discussed below) and is also responsible for controlling vital cardiovascular and respiratory function.

Cerebellum

The cerebellum is responsible for maintenance of equilibrium, posture, and muscle tone, and coordinates movement. It is composed of two laterally located hemispheres, joined in the midline by the vermis. The superior surface lies under the tentorium cerebelli. It is attached to the brainstem by fibres that lie laterally to the fourth ventricle on both sides. These fibres are split into three parts:

- inferior peduncle
- middle peduncle
- superior cerebellar peduncle.

These carry nerve fibres between the medulla, pons, and midbrain. The cerebellum consists of an outer layer of grey matter (the cerebellar cortex) and a central core of white matter. The surface is convoluted into a regular pattern of folia or parallel folds. The white matter consists of fibres that run to and from the cerebellar cortex.

Afferents of the cerebellum

Afferents originate in the spinal cord (spinocerebellar tracts), inferior olivary nucleus (spino-olivary tracts), vestibular nuclei (vestibule-cerebellar tracts), and pons (ponto-cerebellar tracts). Afferent axons travel through one of the cerebellar peduncles and terminate in the cerebellar cortex.

There are two main groups of cerebellar afferents:

- climbing fibres
- mossy fibres.

The vast majority of climbing fibres start from the contralateral inferior olivary nucleus. These fibres insert into the cerebellum through the inferior cerebellar peduncle and give rise to excitatory synapses with the Purkinje cells in the cerebellar cortex.

The inferior olivary nucleus accepts afferents from:

- cortical fibres
- spinal fibres
- vestibulorecticular fibres
- red nucleus fibres
- the superior colliculus
- the interstitial nucleus of Cajal.

The largest provider of cerebellar afferents comes from the contralateral pontine nuclei via the mossy fibres. These contralateral fibres pass through the middle cerebellar peduncle to reach the granular layer of the cerebellum. The reason why the pontine nuclei contain so many afferent fibres is probably due to its inputs. It receives cerebral information from the motor and somatosensory regions as well as the visual association cortex.

Information regarding proprioception is conveyed to the cerebellum via two main pathways:

1. the dorsal spinocerebellar tracts
2. the ventral spinocerebellar tracts.

The dorsal spinocerebellar tracts take their origin in the nucleus dorsalis, which lies in the thoracic and lumber areas lower down the spinal cord. These fibres run ipsilaterally through the spine and insert into the cerebellum via the inferior peduncle. This track carries data on proprioreception from the legs and lower trunk.

The upper limb equivalent of the dorsal spinocerebellar tract is the cuneocerebellar tract, which gives proprioreceptive data from the upper limbs, the upper trunk, and the neck.

The ventral spinocerebellar tracts have nuclei situated within the thoracic spine, usually at or below T4–5. In comparison to the dorsal counterpart, these fibres decussate twice, once at the immediate level within the spine and again just before entering the cerebellum through the superior peduncle. The main purpose of this tract is to relay data concerning the afferent limbs of the withdrawal reflexes and the state of the spinal cord. As with the dorsal spinocerebellar tracts the ventral spinocerebellar tracts convey information on one half of the body, in this case the lower half. The equivalent upper half is maintained by the rostral spinocerebellar tracts.

In terms of evolution, the oldest area within the cerebellum is the floculonodular node. This receives afferents from the ipsilateral vestibular nuclei (via the inferior peduncle) and naturally relays information regarding balance and eye movements.

Cerebellar efferrents

The functions of the efferents in the cerebellum are concerned with posture, balance, and fine muscle control.

There are no direct connections between the cerebellar nuclei and the spinal cord or the cranial nerves. These nuclei instead influence smooth coordination via a complicated process of feed back and feed forward systems.

There are three main groups of cerebellar efferents:

- the vermal and floculonodular efferent system
- the paraventral efferents
- the lateral zone of the cerebellar cortex.

The vermal and floculonodular efferent system begins with fibres originating from the Purkinje cells in the midcerebellum travelling to the fastigial nuclei. Coordination, posture, balance, and vestibular eye movements are all dependent on this system.

The paraventral efferents are thought to be involved with fine-tuning initiated movements and predominantly affect the movement of flexor muscle groups.

The lateral zone of the cerebellar cortex contributes to the preparation of rapid and skilled voluntary muscle movement. These axons, which aim for the thalamus, travel via the superior cerebellar peduncle. The vast majority of the thalamic fibres end up in the nucleus ventralis lateralis, which then passes on information to the premotor cortex.

Cerebral hemispheres

Rostral to the brainstem is the forebrain, which consists of the diencephalon and cerebral hemispheres. The two sides of the diencephalon are separated by the lumen of the third ventricle, whose lateral walls they constitute. The thalamus is the largest (of four) part of the diencephalon, forming a major part of the wall of the third ventricle. The hypothalamus is also part of the diencephalon and this forms part of the lower wall of the third ventricle. The pituitary gland arises from the hypothalamus and is attached to it by the infundibulum or pituitary stalk.

The cerebral hemispheres are the largest part of the brain, and consist of a cortex and an inner mass of white matter. Buried within the white matter are several large masses called basal ganglia. The hemispheres are separated by the great longitudinal fissure, which accommodates the falx cerebri. At the base of the fissure lies the corpus callosum, which is a mass of commissural fibres that run transversely and link corresponding areas of the two cerebral cortices.

The cerebral cortex has convolutions (gyri) and furrows (sulci). Particular gyri and sulci are constant features, common to every human brain, which have specific functions (Fig. 1.9a).

The frontal lobe is demarcated posteriorly by the central sulcus and inferiorly by the lateral fissure. Posterior to the central sulcus is the parietal lobe. The posterior part of the parietal and temporal lobe is the occipital lobe, ending in the occipital pole. This boundary can be seen only on saggital section (Fig. 1.9b).

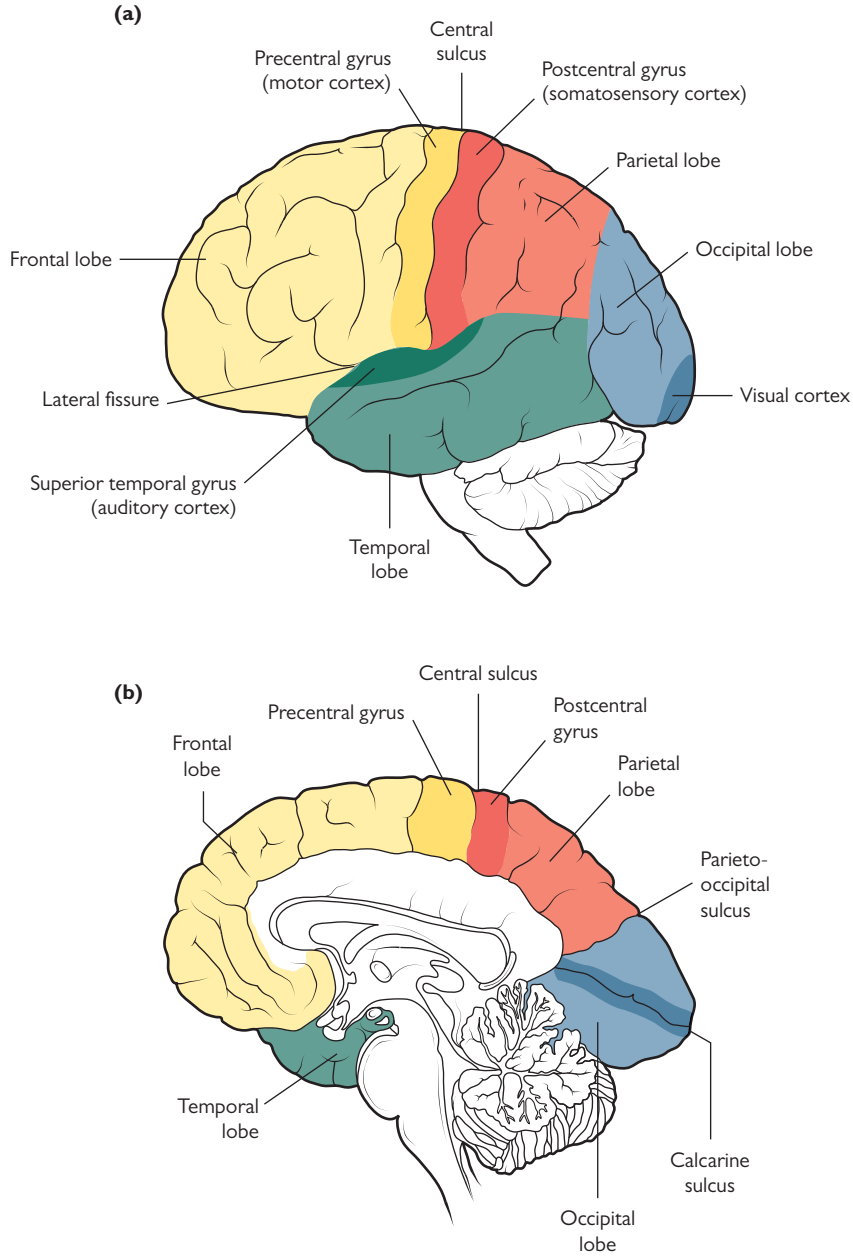


Fig. 1.9 (a) The lateral aspect of the brain. (b) Saggital section of the brain.

This figure was published in *Neuroanatomy: an illustrated colour text*, A. Crossman and D. Neary, Figure 1.23 The lateral aspect of the brain, and Figure 1.24 Saggital section of the brain, p. 18, Copyright Elsevier 1995.

- In the frontal lobe, the gyrus in front of the central sulcus is the precentral gyrus. This contains the primary motor cortex. It receives inputs from the cerebellum and the thalamus. The primary motor cortex contains a somatotopic representation of the different body parts called a motor homunculus (Latin for 'little man'). With the exception of the face (which is upright), the body is represented upside down, and the hand, having a relatively large representation, is found above the face followed by the arm, torso, and leg. The supplementary motor

area is situated medially on the surface of the frontal lobe just in front of the primary motor cortex. This area is associated with postural mechanisms (Fig. 1.9).

- In the parietal lobe, the postcentral gyrus (or primary somatosensory cortex) is the site of termination of pathways that carry sensation, i.e. pain, pressure, temperature, and touch, from the opposite side of the body. The special senses have their own representation in other areas and these are served by the cranial nerves (Fig. 1.9).
- The visual field is represented in the occipital lobe. The organization of this will be detailed much more thoroughly (Fig. 1.9 and pp. 26–7).
- Afferent and efferent fibres which pass between the cerebral cortex and subcortical structures, e.g. brainstem, thalamus, spinal cord are arranged in a regular pattern (corona radiata). Deeper inside the hemisphere, the fibres are arranged in a dense sheet of white matter, known as the internal capsule.
- The fibres of the optic radiation are some of the most posterior fibres of the internal capsule. They run horizontally and posterior to the lentiform nucleus.
- Inside the cortex lie masses of grey matter (basal ganglia), the largest of which is the corpus striatum. These are responsible for muscle tone, posture, and movement.

Broca's area (Brodmann's area 44 and 45)

This is located on the inferior frontal gyrus of the dominant hemisphere (usually left). It is the anterior area responsible for the motor component of language production. Pathology in this region may result in expressive dysphasia.

Wernicke's area

Traditionally associated with receptive dysphasia, this area is considered to consist of the posterior section of the superior temporal gyrus in the dominant cerebral hemisphere (which is the left hemisphere in about 90% of people).

CLINICAL TIP

Occipital lobe lesions

Occipital lobe lesions cause:

- partial seizures—visual hallucinations of unformed nature, e.g. lights and colours (simple partial seizures)
- sensory/motor deficit—contralateral homonymous hemianopia.

Bilateral occipital lobe lesions cause cortical blindness of which the patient may be unaware (Anton's syndrome).

Bilateral occipito-parietal lesions can spare elementary vision but prevent the recognition and depiction of objects (apperceptive visual agnosia).

Frontal eye fields

The frontal eye fields lie in front of the premotor cortex in the middle frontal gyrus, on the lateral surface of the hemisphere (Brodmann's area 8). They are associated with:

- voluntary eye movement
- the accommodation pathway.

Damage to this area causes conjugate deviation of the eyes toward the side of the lesion.

This saccadic movement is a fast eye movement. It may be compared with the quick phase of nystagmus, which is an involuntary fast saccade. The quick phase of nystagmus has been shown to originate at the paramedian pontine reticular formation (PPRF).

The mechanisms that intiate, allow, and control voluntary smooth pursuit are less well understood. Smooth pursuit is thought to be initiated at the parieto-occipital-temporal junction, where fibres travel down to terminate in the ipsilateral PPRF.

The vestibular nuclei of the vestibulo-cochlear system also have connections with the contralateral PPRF. This creates slow eye movements in the opposite direction of the side in which the horizontal canal is stimulated.

Supranuclear gaze pathways

The frontal eye fields and cells within the superior colliculus are thought to be the orignators of saccades. These are rapid eye movements that occur when a person wishes to switch quickly from one target to another.

The fibres (fronto-mesencephalic pathway) from this area travel within the anterior limb of the internal capsule to pass though the thalamus before decussating at the level of the lower midbrain to terminate at the contralateral PPRF, otherwise known as the horizontal gaze centre. There are two PPRFs, which lie lateral to each sixth nerve nuclei and act as functional centres responsible for horizontal eye movements. They act as a final common pathway for conjugate horizontal movements initiated by higher centres, including:

- quick phase of nystagmus
- coordination of saccadic movements
- smooth pursuit movements
- vestibular nuclei-related smooth movements.

The right PPRF mediates conjugate horizontal movements to the right, and vice versa, by coordinating the nuclei of cranial nerves III, IV, and VI (Fig. 1.10).

- Fibres from the PPRF relay with the ipsilateral sixth nerve nuclei.
- One cell group within the sixth nerve nuclei controls the lateral rectus.
- The second cell group has fibres that travel within the contralateral medial longitudinal fasciculus (MLF) to reach the subdivision of the third nerve nuclei, which controls the contralateral medial rectus.
- The MLF is a tract of fibres that travel the length of the midbrain and pons. These include connections between the three ocular motor nuclei (CN III, IV, and VI) and

fibres from the vestibular nuclei in the medulla involved in vestibule-ocular reflexes.

- Thus this system is able to create conjugate eye movements with abduction in one eye and simultaneous adduction in the fellow eye.

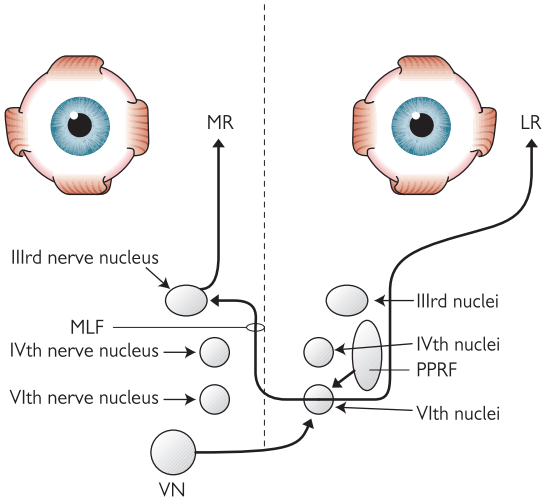


Fig. 1.10 Control of horizontal eye movements. LR, lateral rectus; MLF, medial longitudinal fasciculus; MR, medial rectus; PPRF, paramedian pontine reticular formation; VN, vestibular nucleus.

Reproduced from Venki Sundaram, Allon Barsam, Amar Alwitry, and Peng Khaw, *Training in Ophthalmology*, 2009, Figure 9.26, p. 389, with permission from Oxford University Press.

CLINICAL TIP

Internuclear ophthalmoplegia

Internuclear ophthalmoplegia (INO) is a clinical manifestation of intrinsic brainstem disease. It is usually caused by demyelination or ischaemia. It has distinctive clinical features that indicate a lesion in the medial longitudinal fasciculus.

INO is seen as a lag of the adducting eye, with nystagmus of the abducting eye occurring on completion of a conjugate eye movement. The impairment of adduction is due to interruption of fibres in the MLF on the ipsilateral side. The nystagmus of the contralateral abducting eye is a jerk type, with fast phase away from the midline. The mechanism of nystagmus is not fully understood. Convergence may also be affected in INO.

In some patients, adduction may be normal on versions and slowing of adducting saccades may be the only clinical finding (saccades require higher frequency discharge).

In multiple sclerosis, the INO may be bilateral (BINO). In this case, both eyes may have limitation of adduction and nystagmus on abduction.

HELPFUL HINT

Each cranial nerve has a motor or sensory (or both) nucleus in the brain, and peripheral nerves emerge from the brain and exit from the skull to reach their targets.

Cranial nerves I, II, and VIII are entirely sensory.

Cranial nerves III, IV, VI, XI, and XII are entirely motor.

Cranial nerves V, VII, IX, and X have both sensory and motor components.

There is no vertical equivalent of the PPRF, although the construction of vertical saccades is thought to be localized to the rostral interstitial nucleus of the MLF. This structure also receives instructions and information from the frontal eye fields, the vestibular nuclei, and the PPRF.

Instructions to create upward saccadic movement travel through the posterior commissure to the division of the third nerve, which controls the muscles of upgaze. In downward gaze the instructions travel to the third and fourth (superior oblique) nuclei.

Cranial nerves that pass through the orbit

There are 12 pairs of cranial nerves, as summarized in Table 1.3.

The oculomotor nerve (III)

The third cranial nerve supplies every extraocular muscle except the superior oblique and the lateral rectus. It also supplies the sphincter pupillae and the ciliary muscle with parasympathetic fibres.

The third nerve nuclei are situated in the midbrain at the level of the superior colliculus (Fig. 1.11). They contain two motor nuclei:

- the main motor nucleus is situated in the anterior grey matter surrounding the cerebral aqueduct of the midbrain
- the accessory parasympathetic nucleus (Edinger–Westphal nucleus) lies posterior to the main motor nucleus. Preganglionic nerves travel to the orbit to synapse in the ciliary ganglion. Postganglionic fibres travel along short ciliary nerves to the ciliary muscles and sphincter pupillae of the iris. The accessory nucleus also receives:
  - corticonuclear fibres for accommodation
  - fibres from the pretectal nucleus for the direct and consensual light reflex.

The third nerve emerges anteriorly from the midbrain, medial to the cerebral peduncle, and enters the subarachnoid space. It continues anteriorly between the posterior cerebral artery and the superior cerebellar arteries. The nerve then runs parallel with the posterior communicating artery (Fig. 1.12).

The nerve perforates dura mata on the lateral side of the posterior clinoid process and enters the cavernous sinus wall running above the trochlear nerve (IV). The nerve then passes forward, receiving two branches:

Table 1.3 Components, function, and skull opening for the 12 cranial nerves

Number	Name	Components	Function	Skull opening
I	Olfactory	Sensory	Smell	Openings in cribriform plate
II	Optic	Sensory	Vision	Optic canal
III	Oculomotor	Motor (somatic)	Elevates upper lid, turns eye up, down, and medially	SOF
	Oculomotor	Motor (visceral)	Constricts pupil, accommodates eye	
IV	Trochlear	Motor (somatic)	Turns eye down and medially, Incyclotorsion	SOF
V	Trigeminal ophthalmic division	Sensory	Cornea, forehead, scalp, eyelids, and nose Also mucous membranes of paranasal sinuses and nasal cavity	SOF
	Trigeminal maxillary division	Sensory	Skin of face over maxilla, teeth of upper jaw, mucous membrane of nose, maxillary sinus, and palate	Foramen rotundum
	Trigeminal mandibular division	Motor	Muscles of mastication, belly of digastric, mylohyoid, tensor velli palatni, tensor tympani	Foramen ovale
		Sensory	Skin of cheek, over mandible, and side of head, teeth of lower jaw and temporomandibular joint, mucous membrane of mouth and anterior part of tongue	
VI	Abducens	Motor	Turns eye outwards	SOF
VII	Facial	Sensory	Taste from anterior two-thirds of tongue	Internal acoustic meatus, facial canal
		Motor	Muscle of facial expression	
VIII	Vestibulocochlear	Sensory	From utricle and saccule and semicircular canals	Internal acoustic meatus
IX	Glossopharyngeal	Motor	Stylopharyngeus muscle	Jugular foramen
		Secretomotor	Parotid salivary gland	
		Sensory	General sensation and taste from posterior third of tongue and pharynx, carotid sinus (baroreceptor), carotid body (chemoreceptor)	
X	Vagus	Motor Sensory	Heart, great thoracic blood vessels, larynx, trachea, bronchi, and lungs Liver, kidneys, pancreas, and gut (pharynx to splenic flexure)	Jugular foramen
XI	Accessory: cranial root	Motor	Muscles of soft palate, pharynx, and larynx in branches of vagus	Jugular foramen
	Accessory: spinal root	Motor	Sternocleidomastoid and trapezius	
XII	Hypoglossal	Motor	Muscles of tongue	Hypoglossal canal

SOF, superior orbital fissure.

- the sensory communicating branch from the ophthalmic division of the trigeminal nerve
- the sympathetic branch from the nerve plexus around the internal carotid artery.

The oculomotor nerve now divides into a small superior division and a large inferior division, which enter the orbit through the superior orbital fissure (SOF) within the tendinous ring.

The superior division of the oculomotor nerve passes upwards and lateral to the optic nerve and enters the superior rectus muscle at the junction of its proximal and middle thirds. The nerve passes through the superior rectus muscle and terminates in the levator palpebrae superioris muscle.

The inferior division of the oculomotor nerve divides into three branches, which supply the medial and inferior recti, and the inferior oblique muscles. The branch to the medial rectus passes medially below the optic nerve to enter the lateral surface of the muscle between the proximal and middle third. The branch to the inferior rectus runs forwards on its upper surface and enters the muscle between

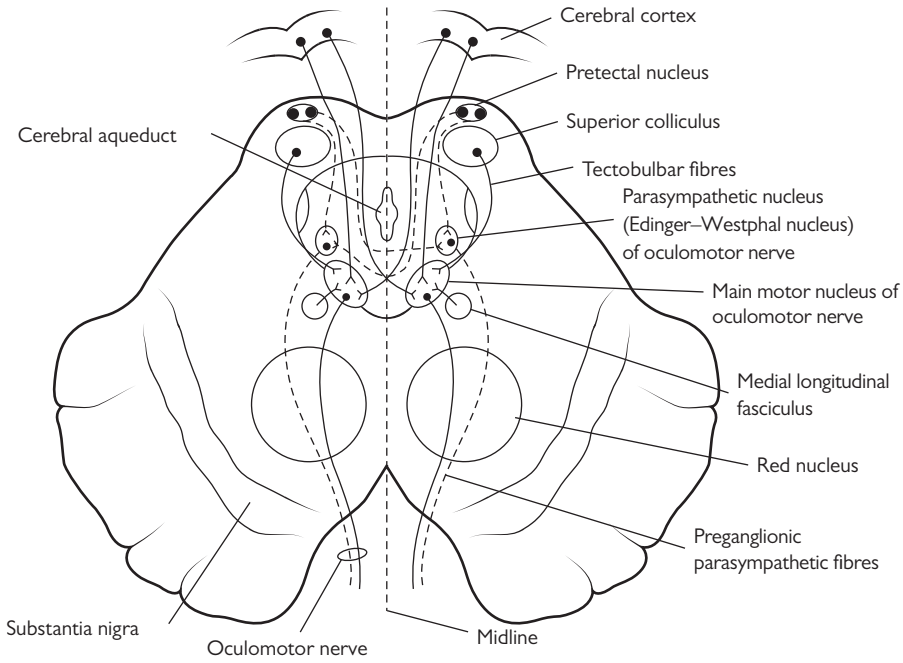


Fig. 1.11 The oculomotor nuclei and their central nervous system connections.

Reproduced from Snell RS and Lemp MA, *Clinical anatomy of the eye*, Second edition, Figure 10.7, p. 305, Wiley. Copyright 1998, with permission.

the proximal and middle third. The branch to the inferior oblique is the longest of the three, and passes forwards close to the orbital floor and lateral to the inferior rectus. It enters the posterior border of the oblique muscle. The nerve to the inferior oblique gives a short branch to the ciliary ganglion. This contains parasympathetic fibres that synapse in the ciliary ganglion. The postganglionic fibres travel in the short ciliary nerves to supply the sphincter pupillae and ciliary muscles.

Three points of interest surround the muscles that are innervated by the third nerve nuclei:

1. The levator is supplied by a single central group of cells (central caudal nucleus) via both third nerve nuclei.
2. The superior rectus remains the only muscle of those innervated by the third nerve to be supplied by the contralateral third nerve nucleus.
3. The other muscles supplied by the third cranial nerve are supplied by the ipsilateral oculomotor nerve nucleus.

CLINICAL TIP

The proximity of the third nerve to the posterior communicating arteries and the superior cerebellar arteries means that any aneurysm arising in these vessels can impinge on the nerve, producing a palsy affecting the actions of this nerve.

The trochlear nerve (IV)

As the thinnest of the cranial nerves, the fourth nerve supplies the contralateral superior oblique muscle.

The fourth cranial nerve nuclei are located beneath the third nuclei at the level of the inferior colliculus (Fig. 1.13). The trochlear nucleus is situated in the anterior part of the grey matter surrounding the cerebral aqueduct of the midbrain. This nucleus is entirely motor and receives fibres as follows:

- corticonuclear fibres from both cerebral hemispheres
- tectobulbar fibres, which connect it to the visual cortex via the superior colliculus
- fibres from the medial longitudinal fasciculus, which connect it to the nucleus of the third, sixth, and eighth cranial nerves.
- the fibres of the fourth nerve decussate before exiting the midbrain and are the only cranial nerves to exit posteriorly (Fig. 1.13).
- the nerve runs forward in the subarachnoid space around the surface of the midbrain around the cerebellar peduncles to eventually run anteriorly (Fig. 1.13).
- the nerve then enters the cavernous sinus wall in a position inferior to the third nerve, and exits the sinus above the third nerve.
- the nerve passes through the SOF, outside the common tendinous ring (CTR) and medial to the frontal nerve in order to access the orbit.

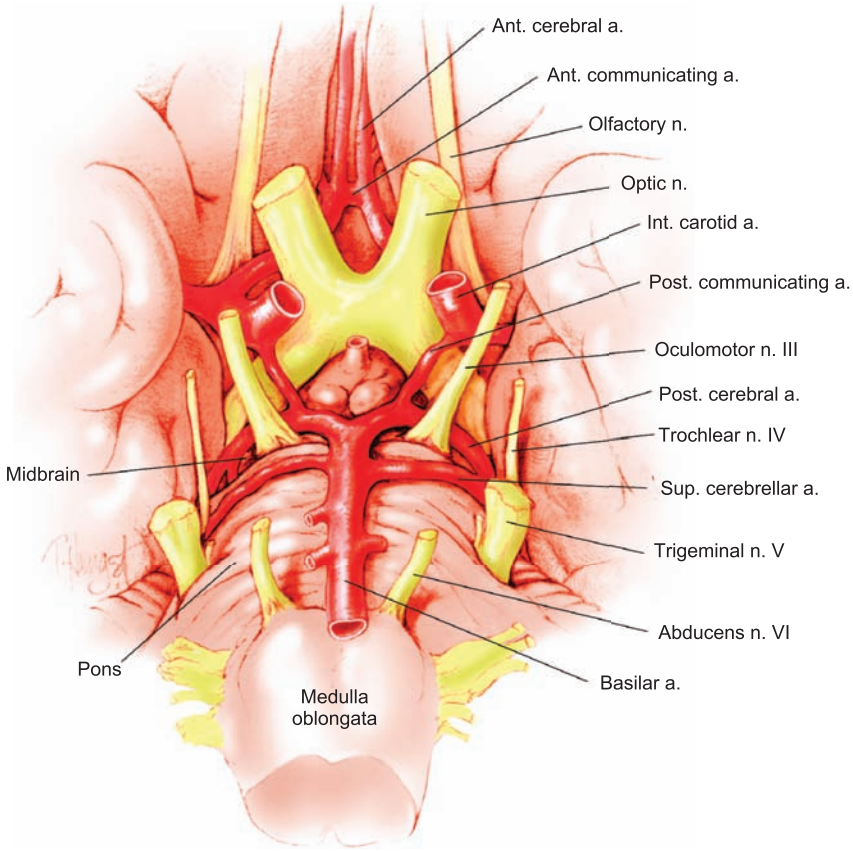


Fig. 1.12 Cranial nerves III, IV, V, and VI.

Reproduced from David R. Jordan, Louise Mawn, and Richard L. Anderson, *Surgical Anatomy of the Ocular Adnexa, A Clinical Approach*, Second Edition, Figure 5.2, Page 130, 2012 with permission from Oxford University Press.

- The nerve then passes medially, superior to the root of the levator palpebrae, before supplying the superior oblique.

The abducens (VI) nerve

The sole motor supply to the lateral rectus is the sixth cranial nerve.

Course

- The nucleus of the sixth cranial nerve lies near the floor of the fourth ventricle. It exits the brainstem anteriorly between the pons and medulla and lies within the subarachnoid space (Fig. 1.14).
- The nerve runs anteriorly, superiorly, and laterally to reach the dorsum sellae of the sphenoid bone.
- From the dorsum sellae, the nerve travels anteriorly over the petrous apex of the temporal bone to enter the cavity of the cavernous sinus. Here it runs inferonasally to the internal carotid artery.
- In the cavernous sinus, a sympathetic branch travels with the nerve from the internal carotid plexus, which later leaves the abducens nerve to join the oculomotor nerve.
- The sixth nerve enters the orbit through the SOF, within the CTR, between the superior and inferior division of the third nerve.
- The sixth nerve finally innervates the lateral rectus by moving anteriorly and laterally within the orbit.
- Fig. 1.15 is a schematic diagram of cranial nerves III, IV, and VI anatomy.

The trigeminal nerve (V)

The trigeminal nucleus is divided into four nuclei (Fig. 1.16):

- main sensory nucleus—lies in the posterior pons
- spinal nucleus—continuous superiorly with the main sensory nucleus in the pons and extends inferiorly through the whole length of the medulla oblongata and down to the second cervical vertebra (C2) of the spinal cord

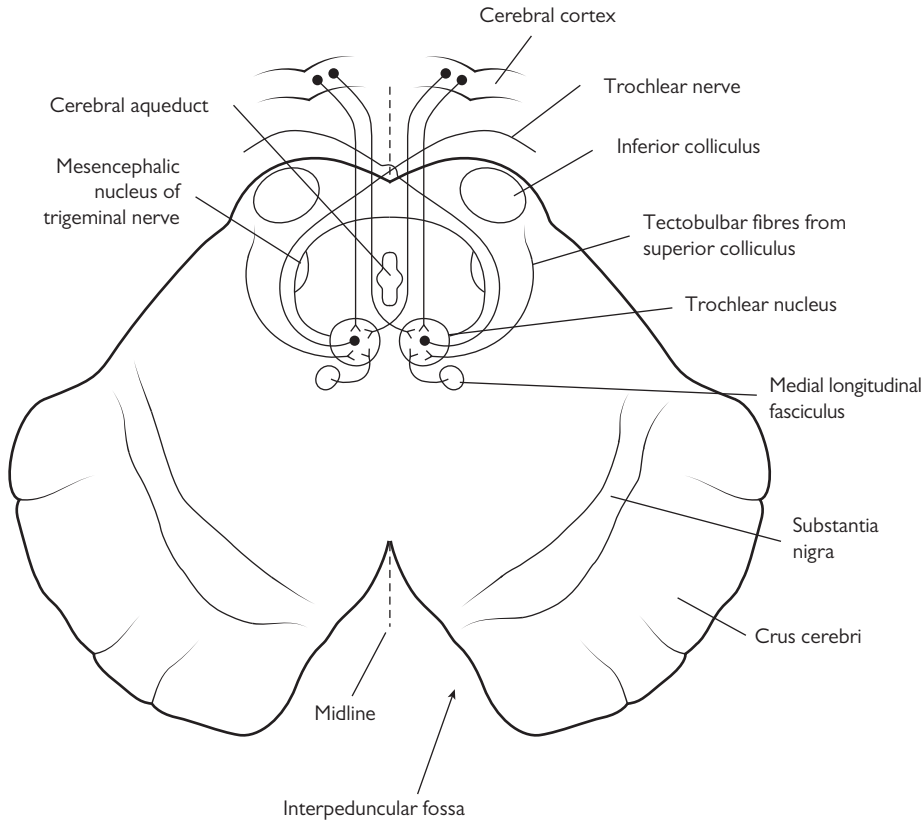


Fig. 1.13 The trochlear nuclei and their central nervous system connections.

Reproduced from Snell RS and Lemp MA, *Clinical anatomy of the eye*, Second edition, Figure 10.11, p. 307, Wiley. Copyright 1998, with permission.

- mesencephalic nucleus—a collection of nerve cells in the lateral grey matter around the cerebral aqueduct
- motor nucleus—medial to the main sensory nucleus.

Sensations of pain, temperature, touch, and pressure are communicated along axons with cell bodies in the trigeminal ganglion. This lies on the apex of the petrous part of the temporal bone. The preganglionic cells enter the pons as ascending or descending branches. The ascending branches terminate in the main sensory nucleus (touch and pressure) and the descending cells terminate in the spinal nucleus (pain and temperature). The motor component of the trigeminal nerve travels in V_3.

CLINICAL TIP

Only the sixth nerve runs within the cavernous sinus and it becomes one of the first cranial nerves to be affected in cavernous sinus thrombosis.

The sixth and seventh cranial nerves are closely linked not only because the facial nerve (VII) wraps around the sixth nerve nuclei to give rise to the facial colliculus, but also because of the closeness in the area of the petrous part of the temporal bone. In middle ear infections, the petrous part of the temporal bone may be breached, forming a combined sixth and seventh nerve palsy (Gradenigo's syndrome). The three cranial nerves that supply eye movement are closely related, as shown in Fig. 1.15.

The ophthalmic division of the trigeminal nerve (V_1)

V_1 arises from the anteromedial surface of the trigeminal ganglion and moves anteriorly in order to enter the lateral wall of the cavernous sinus beneath the fourth nerve but above the maxillary nerve (V_2).

Before accessing the orbit via the SOF V_1 divides into the following branches:

- the frontal nerve
- the lacrimal nerve
- the nasociliary nerve.

The lacrimal nerve

Both the lacrimal and frontal nerves enter the orbit through the lateral part of the SOF, external to the CTR. Once within the orbit, the lacrimal nerve accepts parasympathetic fibres from the zygomaticotemporal branch of V_2 (see

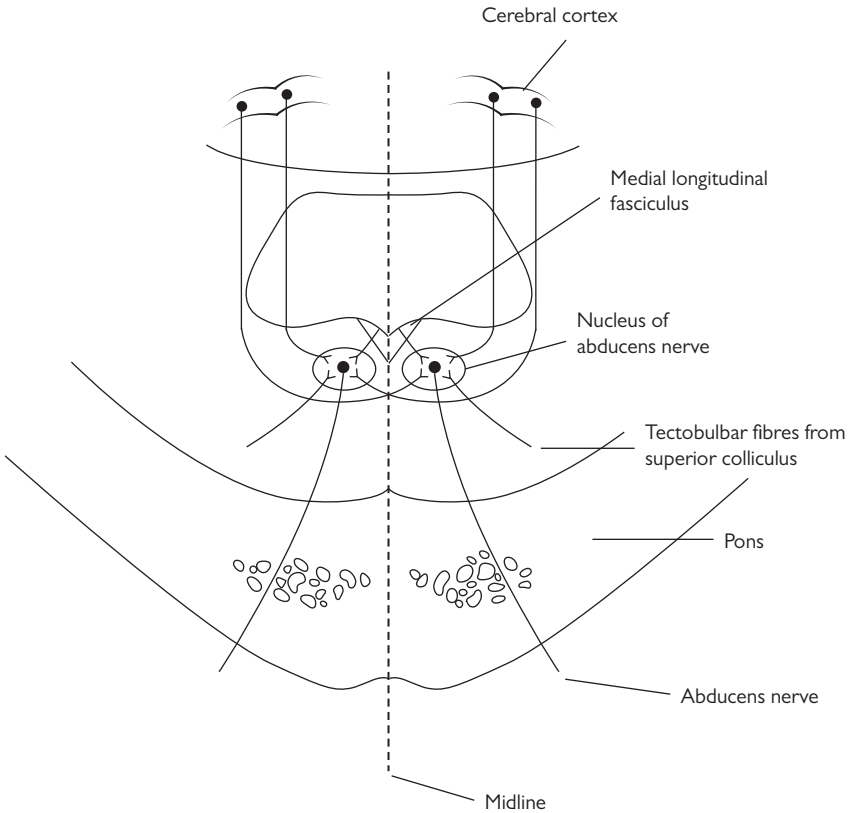


Fig. 1.14 The abducens nerve nuclei and their central nervous system connections.
Reproduced from Snell RS and Lemp MA, *Clinical anatomy of the eye*, Second edition, Figure 10.12, p. 308, Wiley. Copyright 1998, with permission.

The lacrimal system, p. 39). The lacrimal fibres then terminate within the lacrimal gland. The nerve also provides sensory innervation to the conjunctiva and the skin of the upper eyelid (Fig. 1.17).

The frontal nerve

As the thickest branch of V_1, the nerve enters the orbit between the lacrimal and the fourth nerve within the SOF, external to the CTR. The frontal nerve travels in the superior aspect of the orbit and divides into a large supraorbital branch and a supratrochlear branch.

The supraorbital branch supplies the upper lid, conjunctiva, and the skin of the forehead as far back as the vertex. As the nerve passes through the supraorbital notch, it supplies a small branch to the mucous membrane of the frontal sinus.

The supratrochlear nerve travels above the trochlea of the superior oblique and moves upwards to pierce the orbital septum to eventually innervate the medial upper lid and the medial skin of the forehead.

The nasociliary nerve

The nasociliary nerve enters the orbit through the SOF within the CTR and lies between the superior and inferior branches of the third nerve on the lateral aspect of the optic nerve.

The nasociliary nerve then wraps anteriorly around the optic nerve in a superomedial fashion across the optic nerve's upper surface to supply the medial wall with its posterior ethmoidal nerves. After giving off the posterior ethmoidal nerve (often missing, supplies ethmoidal and sphenoidal air sinuses) it passes into the anterior ethmoidal foramen, where it is then named the anterior ethmoidal nerve, which supplies the mucous membranes of the ethmoidal air cells and enters the cranial cavity. It later also supplies the skin on the dorsum of the nose, including the tip and vestibule. Thus, if the skin of the nose is affected by shingles (Hutchinson's sign), it is likely that there will be ocular involvement.

Other branches of the nasociliary nerve include:

- the ramus communicans (to the ciliary ganglion)—this arises from the nasociliary nerve as soon as it enters the orbit and contains sensory fibres (which travel along the ciliary nerves) from the eyeball
- the long ciliary nerves (usually two) pass forward from the ciliary ganglion to pierce the eyeball near the optic

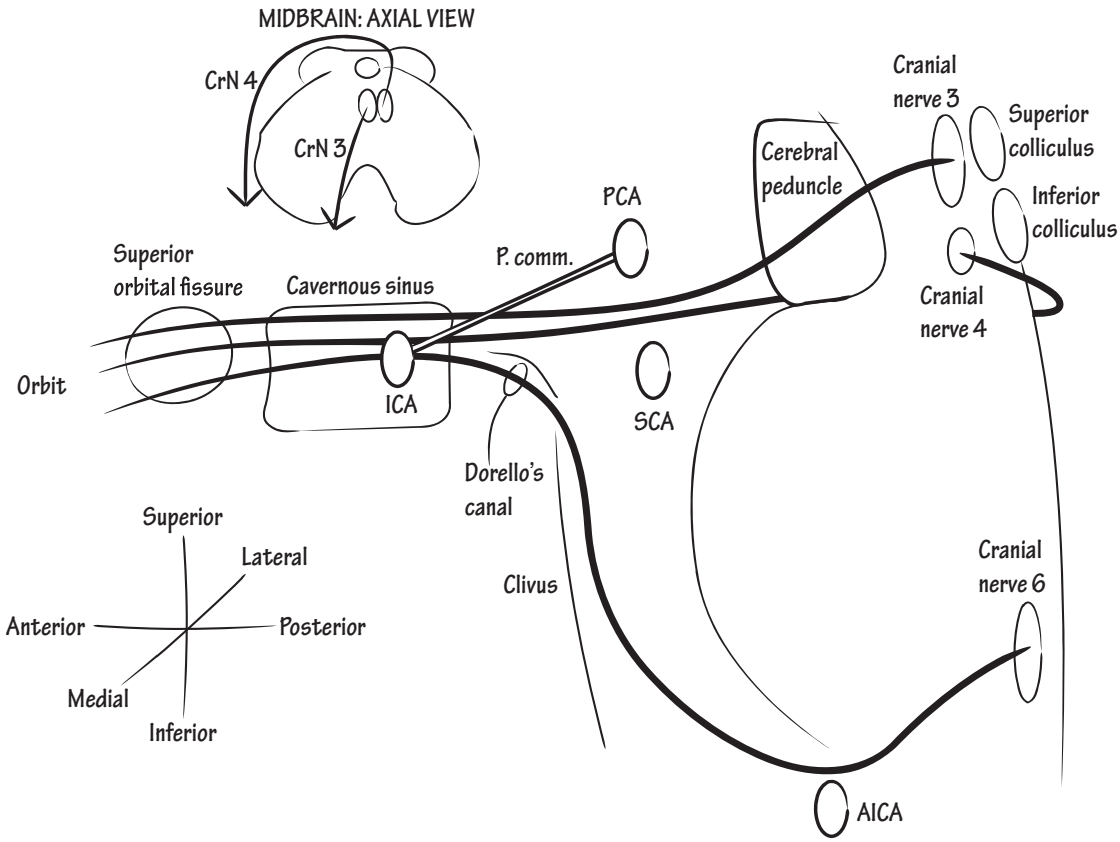


Fig. 1.15 Cranial nerves III, IV, and VI: anatomy. AICA: anterior inferior cerebellar artery; ICA: internal carotid artery; PCA: posterior cerebral artery; P. comm: posterior communicating artery; SCA: superior cerebellar artery.

Reproduced from Adam Fisch, *Neuroanatomy: Draw it to know it*, Drawing 20.9, Page 201, 2012, published by Oxford University Press, with permission from Adam Fisch.

nerve—these pass forward in the choroid to supply the ciliary body, iris, and cornea, and are composed of:

- sympathetic postganglionic fibres to the dilator pupillae
- sensory fibres from the cornea

- the infratrochlear nerve—this emerges from the nasociliary nerve and moves anteriorly underneath the trochlea and penetrates both orbital septum and orbicularis oculi to innervate the lacrimal sac, conjunctiva, the medial aspect of both eyelids, and the lateral skin of the nose.

The maxillary branch of the trigeminal (V_2)

V_2 is the second entirely sensory division of the trigeminal nerve. As V_2 leaves the trigeminal ganglion, it enters the wall of the cavernous sinus beneath V_1 as the most inferior nerve within the wall. The nerve leaves the skull through the foramen rotundum, which transmits V_2 to the pterygopalatine fossa, where it joins the pterygopalatine ganglion. The nerve passes anterolaterally through the IOF and continues as the infraorbital nerve. This runs in the infraorbital canal situated on the floor of the orbit. Whilst in this canal, the nerve is separated from the orbital contents by the overlying orbitalis muscle. It is only separated from the maxillary sinus below (as are the rest of the orbital contents) by a thin sheet of bone.

The branches of V_2 are as follows:

- meningeal—this supplies dura mater in the MCF
- ganglionic—the two ganglionic branches suspend the pterygopalatine ganglion from the lower border of the maxillary nerve. These contain sensory nerves from the orbital periosteum and mucous membranes of the nose, palate, and pharynx. They also contain postganglionic parasympathetic fibres that travel from the facial nerve to the lacrimal gland
- zygomatic—this arises from V_2 in the pterygopalatine fossa, running into the orbit through the SOF. It divides into the zygomaticotemporal and zygomaticofacial branches, which supply the skin of the forehead and cheek, respectively (Fig. 1.18).

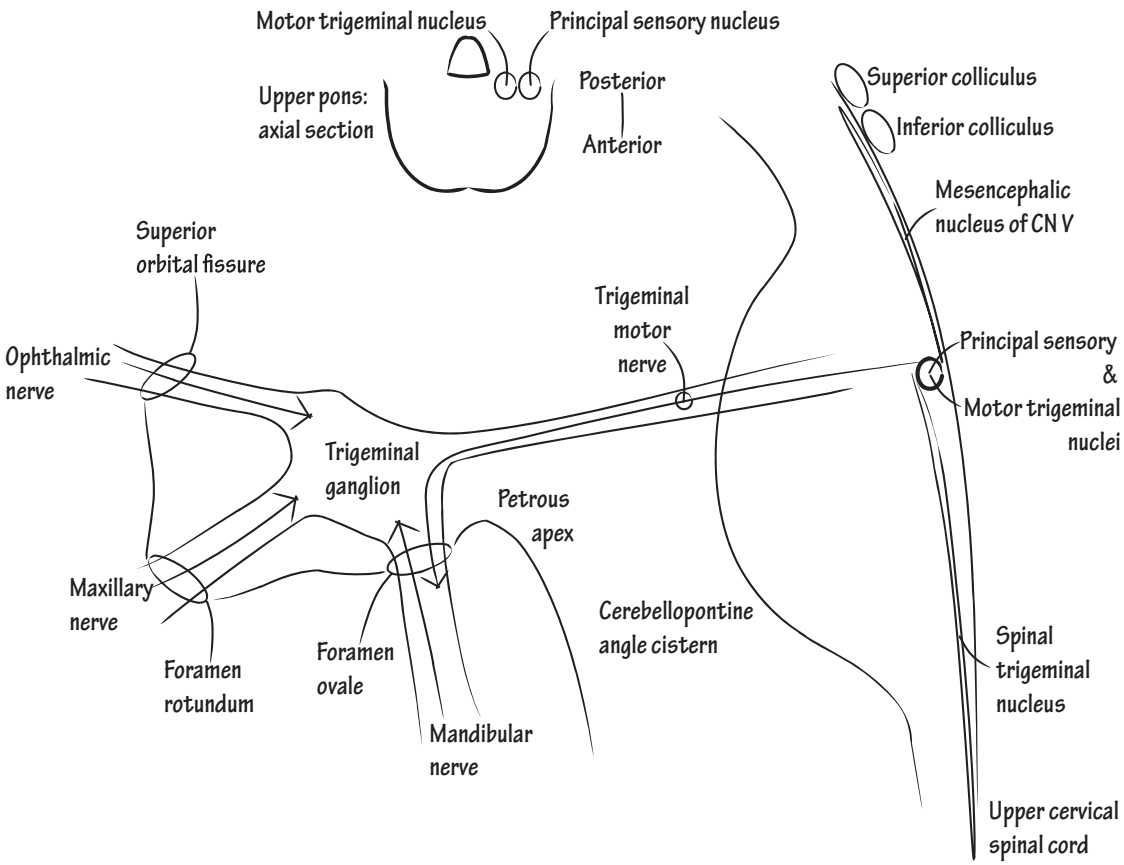


Fig. 1.16 Cranial nerve V: nuclei.

Reproduced from Adam Fisch, *Neuroanatomy: Draw it to know it*, Drawing 13.2, Page 217, 2012, published by Oxford University Press, with permission from Adam Fisch.

- posterior superior alveolar—this again arises from V_2 in the pterygopalatine fossa and supplies the mucous membranes of the maxillary sinus. It also joins the dental plexus to supply upper molar teeth, gums, and part of the mucous membrane of the cheek (Fig. 1.18)
- middle superior alveolar—this arises from the infraorbital nerve and runs down the lateral wall of the maxillary sinus to join the superior dental plexus and supply the upper molar teeth, gums, and part of the mucous membranes of the cheek (Fig. 1.18)
- anterior superior alveolar—this arises from the infraorbital nerve in the infraorbital canal and runs on the anterior wall of the maxillary sinus to supply the canine and incisor teeth (Fig. 1.18)
- facial nerves—these emerge via the infraorbital foramen to supply sensation for the skin of the cheek and mucous membrane of the upper lip.

The visual pathway

Starting anteriorly the visual pathway has the following components (Fig. 1.19):

1. neural retina
2. optic nerve
3. chiasm
4. optic tract
5. lateral geniculate nucleus
6. optic radiations
7. visual cortex.

Optic nerve

The optic nerve may be divided in terms of location:

- intraocular (1 mm)
- intraorbital (25 mm)

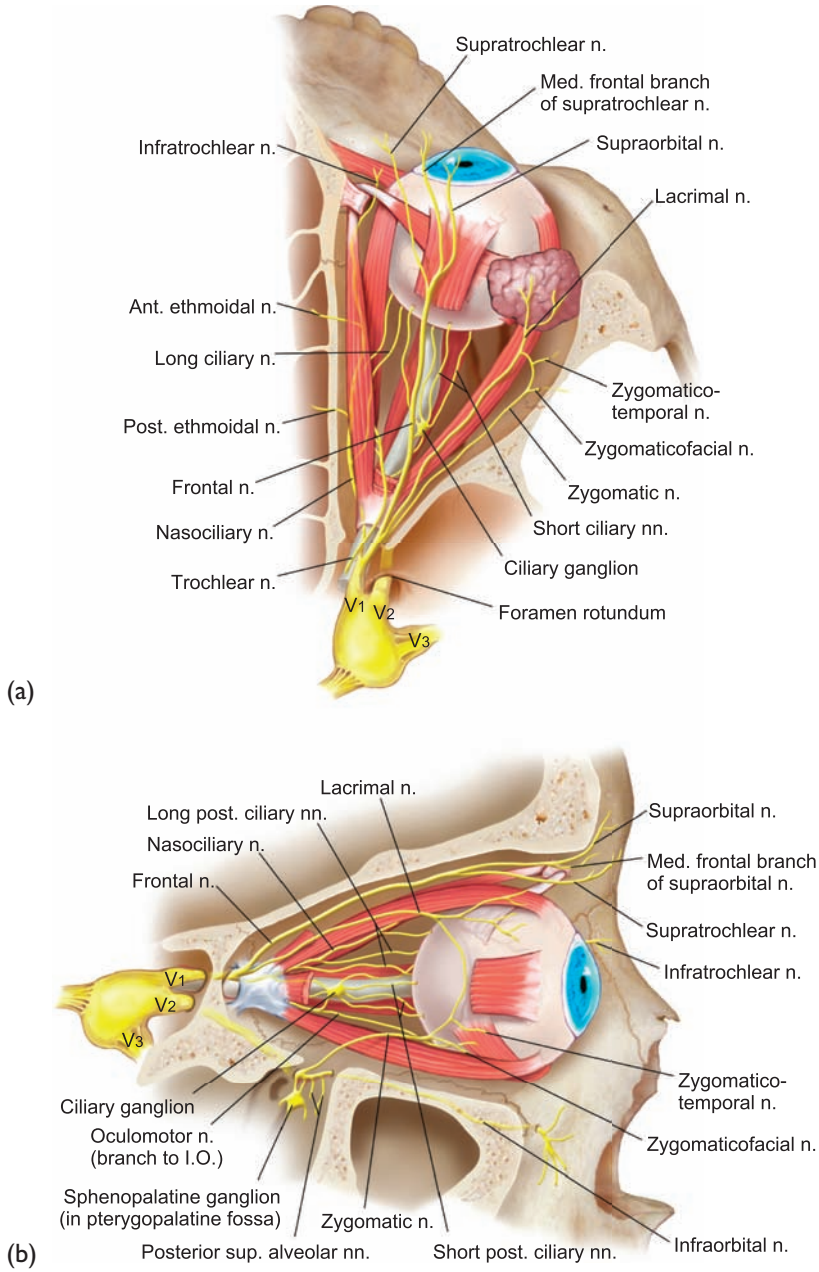


Fig. 1.17 (a) The trochlear nerve, frontal nerve, lacrimal nerve, nasociliary nerve, and ciliary ganglion as viewed from above. (b) A lateral view of the frontal nerve, nasociliary nerve, and ciliary ganglion.

Reproduced from David R. Jordan, Louise Mawn and Richard L. Anderson, *Surgical Anatomy of the Ocular Adnexa, A Clinical Approach, Second Edition*, Figure 5.8D and Figure 5.8E, Page 140, 2012, with permission from Oxford University Press.

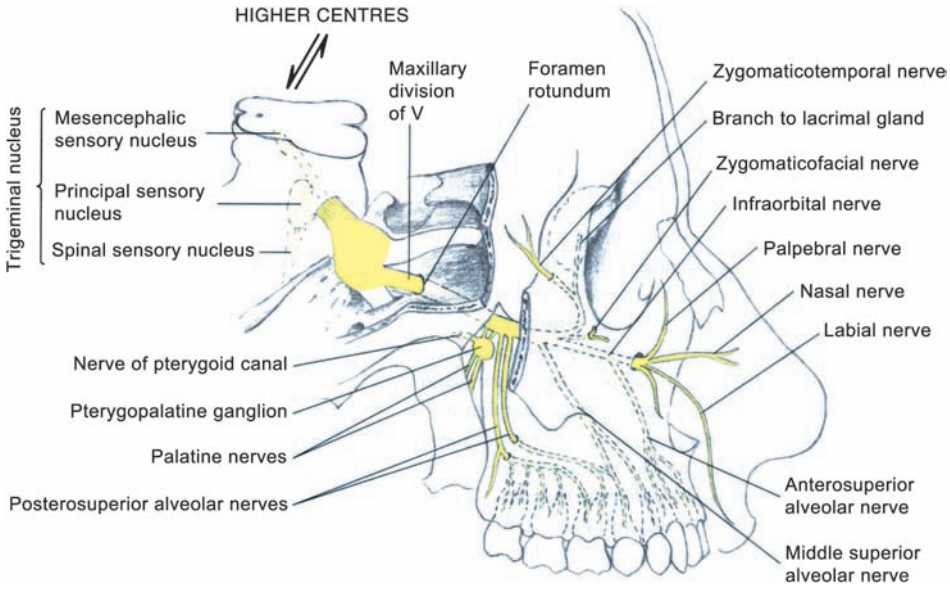


Fig. 1.18 Maxillary division of the trigeminal nerve (V).

Reproduced from Pamela MacKinnon and John Morris, *Oxford Textbook of Functional Anatomy, Head and Neck*, 2005, Figure 6.11.10, Page 176, with permission from Oxford University Press.

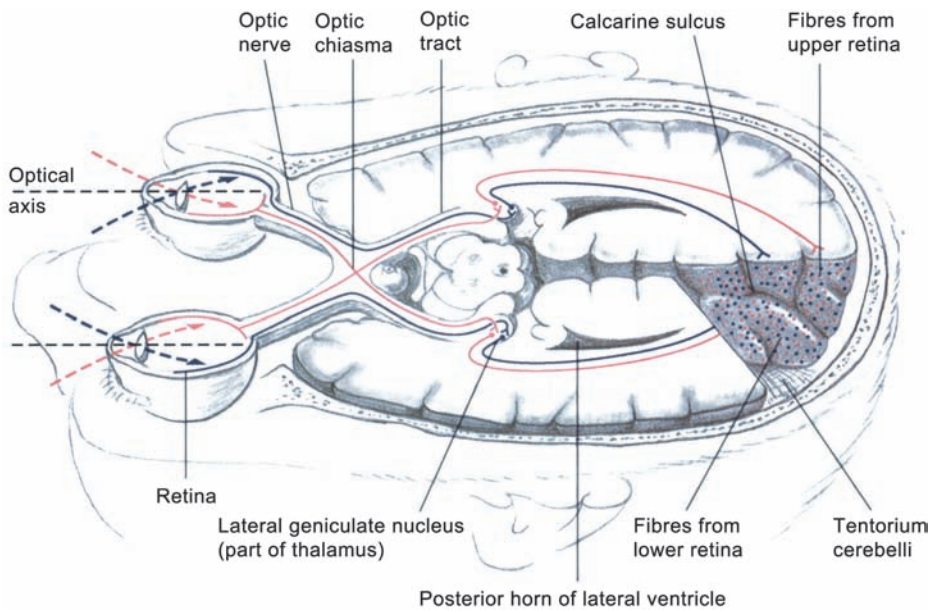


Fig. 1.19 The visual pathway and its constituent parts.

Reproduced from Pamela MacKinnon and John Morris, *Oxford Textbook of Functional Anatomy, Head and Neck*, 2005, Figure 6.11.3, Page 174, with permission from Oxford University Press.

- intracanalicular (5 mm)
- intracranial (10 mm).

The intraocular portion describes the optic nerve as it pierces the sclera. When viewed from within the optic nerve is seen *en face* as the optic disc.

Blood supply to this region of the optic nerve is supplied by branches of the anastomotic circle of Zinn in the sclera around the optic nerve. This in turn receives its blood supply from the short posterior ciliary arteries. The central retinal artery does not supply this part of the optic nerve.

The intraorbital portion is approximately 25 mm in length. This is around 6 mm longer than the eyeball and optic canal. In this portion, around 12 mm from the eyeball, the central retinal artery inserts into the nerve and the central retinal vein exits the dura.

The centre of the nerve contains a majority of ganglion cell neurons and a minority of other neural glial cells such as astrocytes. As the nerve leaves the globe, it immediately begins to thicken (from 1.5 to 3–4 mm). This is due to myelination of the axons, which begins behind the posterior lamina cribrosa.

The optic nerve sheath comprises three components, which are posteriorly continuous with the meninges of the brain:

1. dura
2. arachnoid
3. pia mater.

The blood supply to this region is from the pial plexus of vessels, which receives its arterial supply from neighbouring branches of the ophthalmic artery and a few branches of the central retinal artery.

The optic canal lies entirely within the lesser wing of the sphenoid bone. It measures 5 mm in length and transmits two structures:

- the optic nerve (and its three meningeal sheaths)
- the ophthalmic artery.

The blood supply to this portion is from the pial plexus with branches from the ophthalmic artery.

The intracranial portion of the nerve moves superiorly, posteriorly, and medially to arrive at the optic chiasm, in the floor of the third ventricle.

This region also receives its blood supply from the pial plexus, which is supplied by branches of the superior hypophyseal artery, the internal carotid artery, and the ophthalmic artery.

Optic chiasm

At the optic chiasm, nasal retinal fibres decussate to the contralateral optic tract. The temporal retinal fibres remain in the ipsilateral optic tract. Crossed fibres may loop into either the ipsilateral optic tract or the contralateral optic nerve for a short distance after they pass through the chiasm.

Nerve fibres from the inferior nasal retina cross at the anterior part of the chiasm, whereas nerve fibres from the superior nasal retina cross at the posterior part of the retina.

The chiasm usually lies above the diaphragm sellae, the seat of which holds the pituitary (Fig. 1.20). Anterior to the chiasm lie the anterior cerebral arteries and the anterior communicating artery. Posterior lies the tuber cinereum, and laterally the internal carotid arteries. Superiorly lies the third ventricle.

The optic chiasm has dimensions of 12 by 8 mm and has a covering of pia mater. It projects down into the subarachnoid space. Anterolaterally it is continuous with the optic nerves and posterolaterally with the optic tracts.

Blood supply to the nervous tissue comes from the pial vessels, which are supplied by branches of the internal carotid artery, the superior hypophyseal branch of the internal carotid artery, the anterior cerebral artery, and the anterior communicating artery.

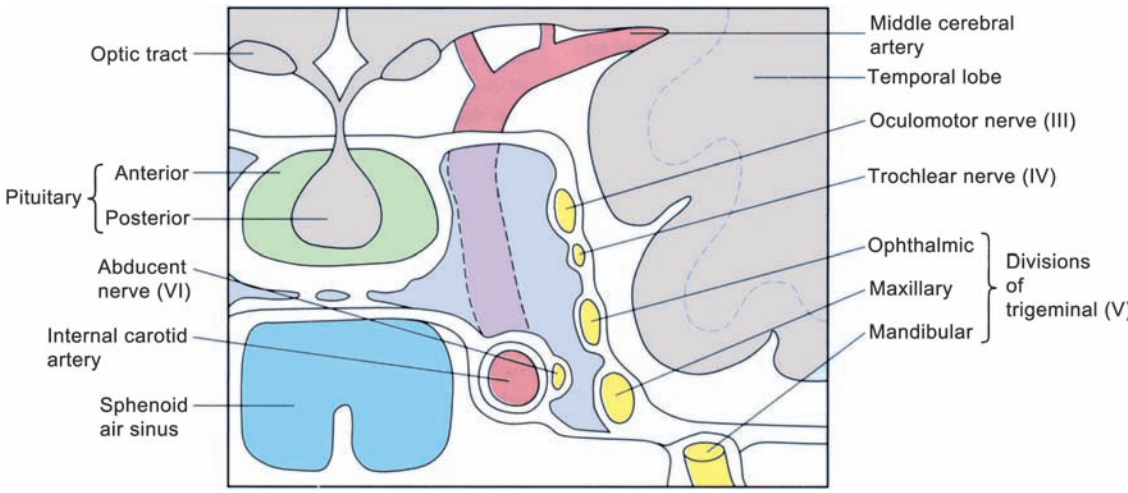


Fig. 1.20 Anatomy of the pituitary and cavernous sinus.

Reproduced from Pamela MacKinnon and John Morris, *Oxford Textbook of Functional Anatomy, Head and Neck*, 2005, Figure 6.9.8, Page 146, with permission from Oxford University Press.

Optic tracts

Two optic tracts emerge from the posterolateral aspect of the optic chiasm. Each tract passes posterolaterally, lateral to the tuber cinereum medially. The tract flattens and winds around the lateral margin of the upper part of the cerebral peduncle. Most of the nerve fibres terminate in the lateral geniculate body. These fibres are concerned with visual sensation.

Before they reach the lateral geniculate, 10% of nerve fibres pass (medially to the superior colliculus) to the pretectal nucleus. These fibres are involved in the light reflex.

Blood supply to the optic tracts is by the pial vessels, which are fed by the anterior choroidal artery, the posterior communicating artery, and the middle cerebral artery.

Lateral geniculate nucleus

Located under the pulvinar of the thalamus, the dorsal lateral geniculate nucleus is a laminated structure that is composed of six layers of cells separated from each other by white bands of optic nerve fibres (1 to 6). These nerve fibres are the myelinated axons of ganglion cells from the retina. They represent cells from the temporal half of the ipsilateral retina (layers 2, 3, and 5) and nasal half of the contralateral retina (layers 1, 4, and 6). Thus, each lateral geniculate body receives visual input from both retinas.

Blood supply to the lateral geniculate body comes from the anterior choroidal branch of the middle cerebral artery, the thalamogeniculate branches of the posterior cerebral artery, and the lateral choroidal arteries.

Optic radiations (geniculocalcarine tracts)

These are formed from the nerve fibres that originate in the lamellae of the lateral geniculate body.

Fibres from the lateral part of this lateral geniculate body receive inputs from inferior retinal quadrants (corresponds to superior visual field), and these travel laterally and inferiorly around the anterior inferior aspect of the horn of the fourth ventricle. Fibres from the periphery loop furthest away (loop of Meyer) and fibres from the macula, by contrast, loop very little. These fibres from the optic radiations continue posteriorly, joining the internal capsule and then travelling laterally to the horns of the lateral ventricle. They then turn medially to the occipital cortex.

Fibres from the medial portion of the lateral geniculate body receive inputs from the superior retinal quadrants (inferior visual field) and turn directly posteriorly to travel in the retrolentiform part of the internal capsule.

Blood supply to the anterior part of the optic radiations comes from the anterior choroidal branch of the internal carotid artery. The posterior portion is supplied by the middle and posterior cerebral arteries.

The primary visual cortex

The location of the primary visual cortex is within the walls of the calcarine sulcus and corresponds to Brodmann's area 17. The calcarine sulcus is situated on the medial aspect of the cerebral hemispheres.

The primary visual cortex begins proximally at the junction of the calcarine and parieto-occipital sulci and continues all the way to the posterior pole of the occipital cortex.

Each hemisphere receives information from its ipsilateral lateral geniculate nucleus. Hence, the right half of the visual field is represented in the left cerebral cortex.

The visual cortex located above the calcarine sulcus receives fibres related to the superior retina (ipsilateral superior temporal fibres and contralateral superior nasal fibres) and thus the inferior visual field. The cortex located below the calcarine sulcus, however, receives inputs from the inferior retina (superior field of vision).

The representation of the macula in the visual cortex is relatively large and is situated on the posterior pole of the calcarine sulcus and extends posterolaterally to the occipital cortex. This accounts for one-third of the visual cortex.

The secondary visual cortex (areas 18 and 19) surround area 17. They receive afferent fibres from the primary visual area, other cortical areas, and the thalamus. This area relates visual information from the primary area to previous visual experience. Area 18 plays a role in integrating the two halves of the visual field via commissural fibres that cross the midline. It may be of importance in sensory-motor coordination and control of the cranial nerve nuclei in their role in eye movement.

The visual cortex receives its blood supply from the posterior (and middle) cerebral artery.

Lesions of the visual pathways

Based on an understanding of the visual pathways, Fig. 1.21 represents field defects caused by lesions of the visual pathways.

Orbit and adnexa

The orbit

The orbits are a pair of large bony sockets that house the eyeballs and include the muscles, nerves, vessels, and fat as well as most of the ocular appendages. Each cavity is of pyramidal shape with its apex directed posteriorly, medially, and superiorly.

The orbit is composed of seven bones:

1. frontal
2. zygomatic
3. maxillary
4. ethmoidal
5. sphenoid
6. lacrimal
7. palatine.

Anteriorly, the orbital rim (or margin) is quadrilateral in shape. The supraorbital margin is composed of the frontal bone. The supraorbital notch (or foramen) is located at

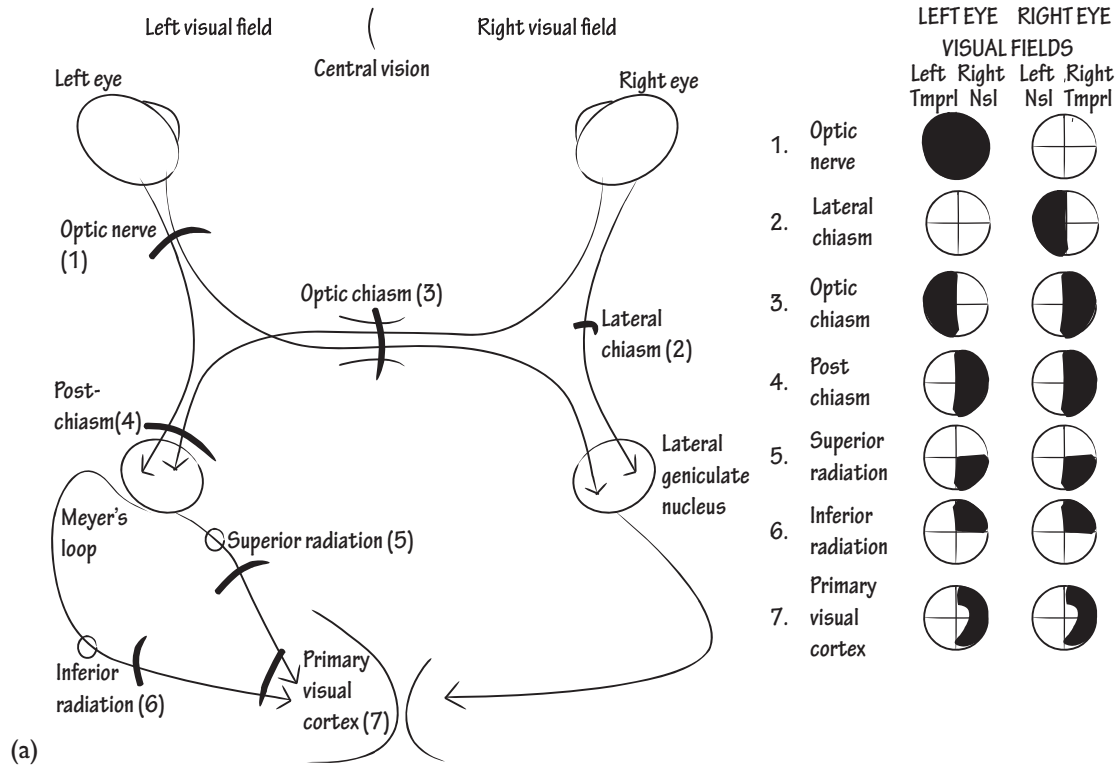


Fig. 1.21 Visual field defects and the corresponding lesions of the visual pathway.

(a) Reproduced from Adam Fisch, *Neuroanatomy: Draw it to know it*, Drawing 22.8, Page 397, and (b) Drawing 22.9, Page 399, 2012, published by Oxford University Press, with permission from Adam Fisch.

the junction of the sharp lateral two-thirds and the rounded medial third. The sharp infraorbital margin is formed laterally by the zygomatic bone and medially by the maxilla. The lateral margin is formed by the frontal process of the zygomatic bone inferiorly and superiorly by the zygomatic process of the frontal bone. The medial margin is formed by the maxillary process of the frontal bone above and by the lacrimal crest of the frontal process of the maxilla.

The orbit is divided into four walls (Fig. 1.22), each of which is lined with periosteum:

- roof
- floor
- medial wall
- lateral wall.

The orbital roof

The orbital roof comprises two bones:

- the orbital plate of the frontal bone
- the lesser wing of sphenoid bone (posterioly).

Two key features reside within the roof:

- the lacrimal fossa for the lacrimal gland—this is located anterolaterally behind the zygomatic process of the frontal bone
- the trochlea, which is the pulley for the tendon of the superior oblique muscle, lies medial to the supraorbital notch and 4 mm behind the orbital margin.

The orbital floor

The orbital roof comprises three bones:

- orbital plate of maxillary bone
- small orbital process of palatine bone
- orbital surface of zygomatic bone.

The thin floor of the orbit is mainly formed by the orbital plate of the maxilla, which separates the orbit from the maxillary sinus. The floor is continuous with the lateral wall except posteriorly, where the floor is separated by the IOF from the lateral wall. Running forward from the fissure is the infraorbital groove, which runs in an anterior direction. As this groove nears the orbital margin, it descends inferiorly into a canal that exits the skull as the infraorbital foramen just 4 mm below the inferior orbital margin on the maxillary bone. The infraorbital canal/foramen transmits

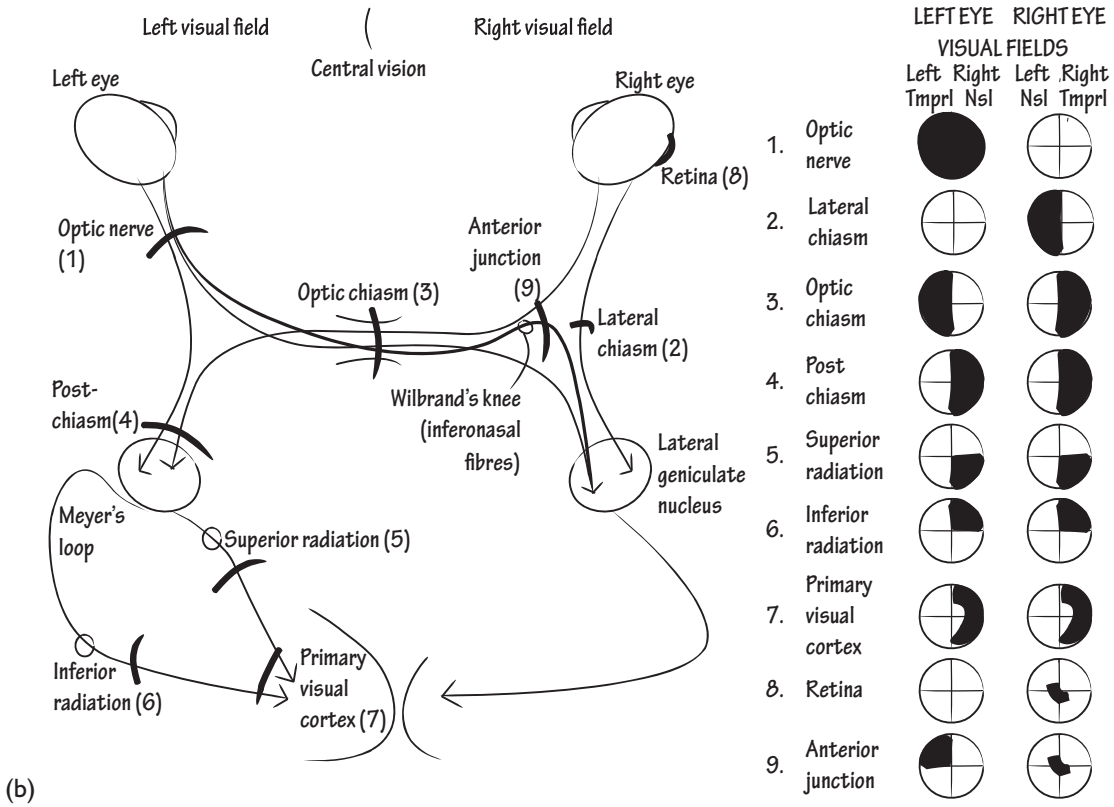


Fig. 1.21 (*continued*)

the infraorbital nerve, which is a branch of the maxillary division of trigeminal (V2).

The medial orbital wall

The medial orbital wall comprises four bones (anterior to posterior):

- the frontal process of the maxilla
- lacrimal
- ethmoid
- lesser wing of sphenoid.

The medial wall runs anteroposteriorly and parallel to the saggital plane. Anteriorly, it contains the lacrimal groove, which is the bony housing for the lacrimal sac. The groove itself is formed from lacrimal bone and the frontal process of maxilla.

The nasolacrimal canal extends inferiorly from the lacrimal fossa on the medial wall and terminates in the inferior nasal meatus in the nasal cavity.

The medial wall is the thinnest of the orbital walls and its paper-like appearance is reflected in its name: lamina papyracea.

The lateral orbital wall

This is the thickest wall and comprises two bones:

- zygomatic
- greater wing of sphenoid.

The lateral wall diverges at an angle of 45° from the saggital plane. Anteriorly, the zygoma separates this bone from the temporal fossa. Posteriorly, the greater wing of the sphenoid separates the orbit from the temporal lobe of the brain in the MCF. The lateral wall and roof are continuous anteriorly but are separated posteriorly by the SOF.

Whitnall's (lateral orbital) tubercle lies posterior to the orbital margin, on the frontal process of the zygoma. Four structures are attached to this eminence:

1. the lateral palpebral ligament
2. the suspensory ligament of the eyeball
3. the check ligament of the lateral rectus muscle
4. the aponeurosis of levator.

Openings of the orbit

The main opening of the orbit is anteriorly.

Posteriorly, the optic canal lies in the lesser wing of the sphenoid and is 4–10 mm long. It is related medially to the body of the sphenoid. It connects the MCF with the orbital cavity. The optic canal transmits the optic nerve along with its meninges, an extension of the subarachnoid space, and

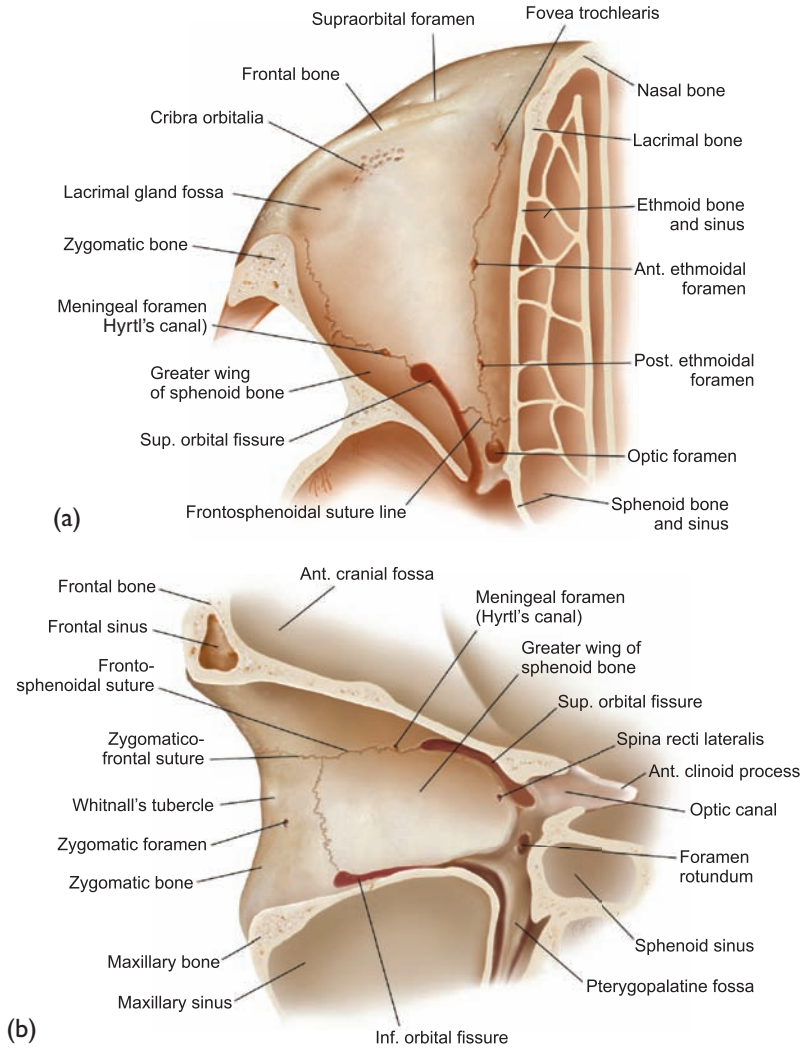


Fig. 1.22 The four orbital walls. (a) Superior orbital wall (orbital roof). (b) Lateral orbital wall. (c) Inferior orbital wall (orbital floor). (d) Medial orbital wall. (e) Orbital foramina. (f) Orbital apex.

the ophthalmic artery along with its surrounding sympathetic plexus.

The SOF lies between the greater and lesser wings of the sphenoid. It connects the MCF with the orbital cavity. It lies between the roof and lateral wall of the orbit. At the infero-medial end of the SOF lies a CTR, which is the origin for the four rectus muscles.

The following structures pass through the SOF outside the CTR (Fig. 1.23):

- lacrimal nerve
- frontal nerve
- trochlear nerve (CN IV)
- superior ophthalmic vein
- inferior ophthalmic vein.

The following structures pass through the SOF inside the CTR (Fig. 1.23):

- upper division of oculomotor nerve (CN III)
- lower division of oculomotor nerve (CN III)
- nasociliary nerve
- abducens nerve (CN VI).

The IOF lies between the greater wing of the sphenoid and the maxilla. It connects both the pterygopalatine and infero-temporal fossae with the orbital cavity. In living subjects, this is

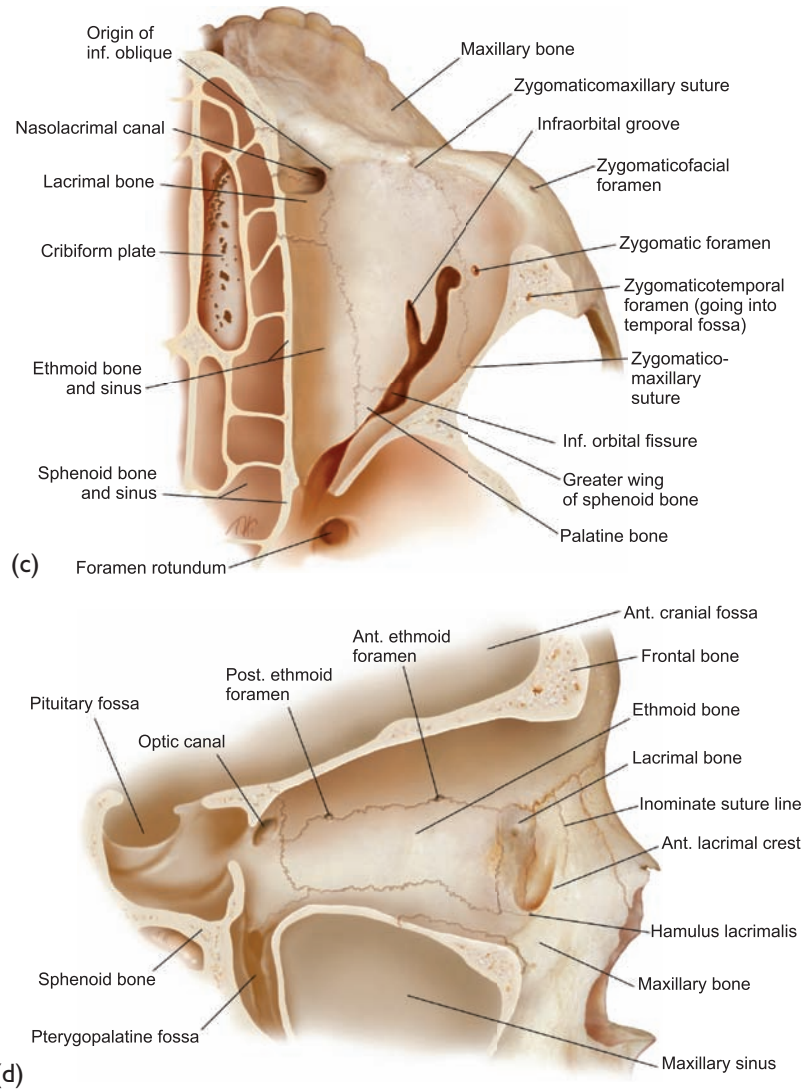


Fig. 1.22 (*continued*)

closed by the periorbita and the muscle of Müller. It transmits the maxillary nerve, which on passing through the IOF is renamed the infraorbital nerve. It also transmits the zygomatic nerve, branches of the pterygopalatine ganglion, and the inferior ophthalmic vein, which drains to the pterygoid plexus.

Two ethmoidal foramen lie in the frontal bone and the frontoethmoidal suture. They are situated where the roof joins the medial wall. The anterior ethmoidal foramen joins from the orbit to the ACF. The posterior ethmoidal foramen traverses the ethmoid bone and joins the orbit to the ethmoidal sinuses. The anterior and posterior ethmoidal nerves and arteries pass through these.

The zygomaticofacial foramen lies on the lateral wall and floor, and transmits the zygomaticofacial nerve. The zygomaticotemporal foramen lies above this, close to the sphenozygomatic suture and it transmits the zygomaticotemporal nerve.

The orbital vasculature

The ophthalmic artery

On exit from the cavernous sinus, the internal carotid artery releases its first branch, which is the ophthalmic artery (Fig. 1.24).

Within the orbit and after producing the central retinal artery, the ophthalmic artery swings medially above the optic nerve. In 15% of people, the ophthalmic artery passes below the optic nerve. The nasociliary artery follows suit.

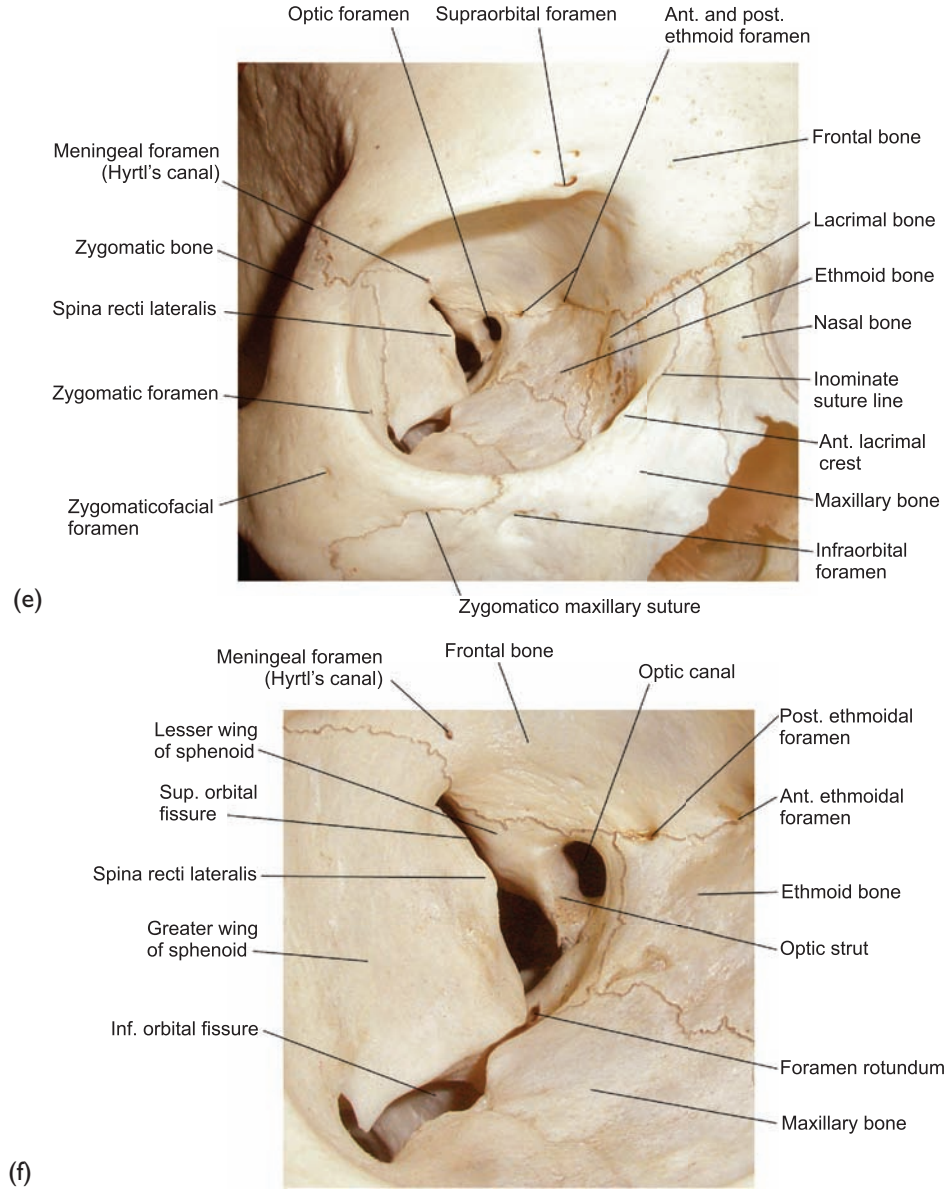


Fig. 1.22 (*continued*)

The central retinal artery lies beneath the optic nerve and pierces the dura 12 mm behind the globe from an inferomedial direction.

The ophthalmic artery then produces the following tributaries:

- the lacrimal artery
- the long and short posterior ciliary arteries
- the muscular arteries (extraocular muscle branches)
- the supraorbital artery
- the posterior and anterior ethmoidal artery
- the medial palpebral arteries.

Finally, the ophthalmic artery terminates as the supratrochlear arteries and the dorsal nasal arteries.

The central retinal artery

After penetrating the optic nerve, the central retinal artery continues anteriorly within the subarachnoid space and provides small meningeal arteries to the pial layer. Here the central retinal artery is followed by the central retinal vein.

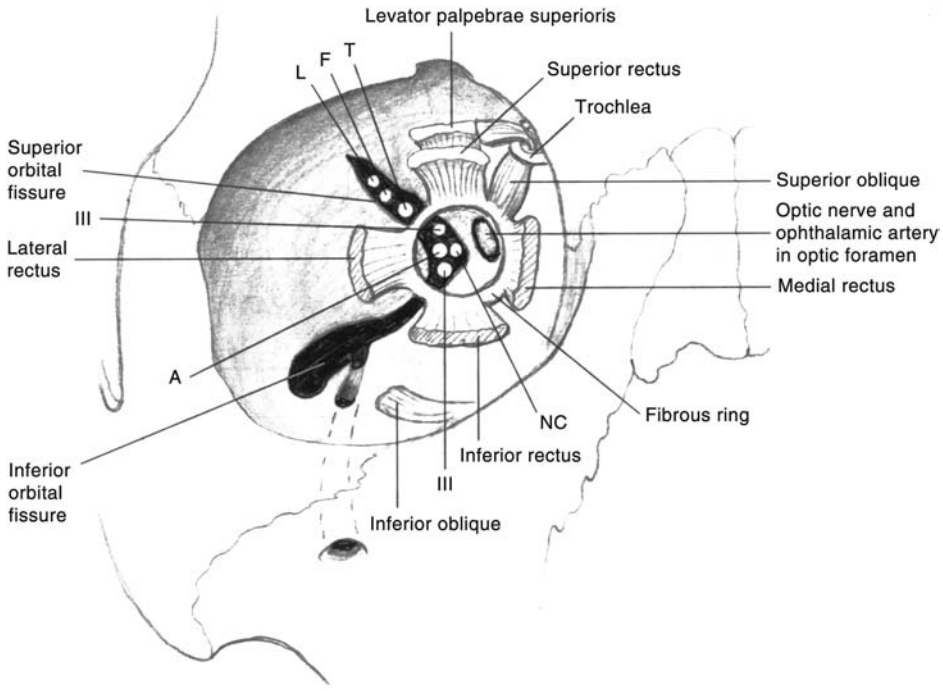


Fig. 1.23 Structures at the back of the orbit. Attachment of muscles to fibrous ring, and lacrimal (L), frontal (F), trochlear (T), nasociliary (NC), superior and inferior branches of the oculomotor nerve (III), and abducens (A) nerves.

Reproduced from Pamela MacKinnon and John Morris, *Oxford Textbook of Functional Anatomy, Head and Neck*, 2005, Figure 6.7.19, Page 123, with permission from Oxford University Press.

Passing through the lamina cribrosa, the central retinal artery performs its task as an end artery to the retina.

Despite small communications with the pial plexus and the circle of Zinn, the central retinal artery is usually considered separate to the ciliary arteries.

The cilioretinal artery

Cilioretinal arteries are present within 20% of the population and supply the retina between the macula and the optic nerve from a posterior ciliary artery source. This supply includes the nerve fibres from the foveal photoreceptors. If this artery is present, the macula often remains perfused even in the case of a central retinal artery occlusion.

The ciliary arteries

The long and short posterior ciliary arteries as well as the anterior arteries represent the vast majority of the blood supply to the globe.

The long posterior ciliary arteries

These twinned arteries arise from the ophthalmic artery as it crosses the optic nerve. They run forward with the optic nerve and penetrate the sclera outside the circle of Zinn (formed by the short posterior ciliary arteries). They then traverse the suprachoroidal (between the sclera and choroid) space to reach the ciliary body. Here, the posterior ciliary arteries join the supply from the anterior ciliary arteries (formed by the muscular branches of the ophthalmic artery) to supply the major arterial circle. The combination of the two supplies provides arterial blood to the choroid posteriorly as far as the equator of the globe. They also anastomose with the short posterior arteries to complete the choroidal circulation.

The short posterior ciliary arteries

The short posterior ciliary arteries penetrate the globe around the optic nerve head to form the arterial circle of Zinn. This provides the optic nerve head with arterial blood.

The short posterior ciliary arteries also travel anteriorly within the suprachoroidal space to supply the choroid up to the equator of the globe.

The anterior ciliary arteries

See iris circulation, p. 49.

The venous drainage of the eye

The superior ophthalmic vein arises in the medial aspect of the upper lid from a combination of the supraorbital vein and a branch from the facial vein. The superior ophthalmic vein is often larger than its inferior counterpart and passes

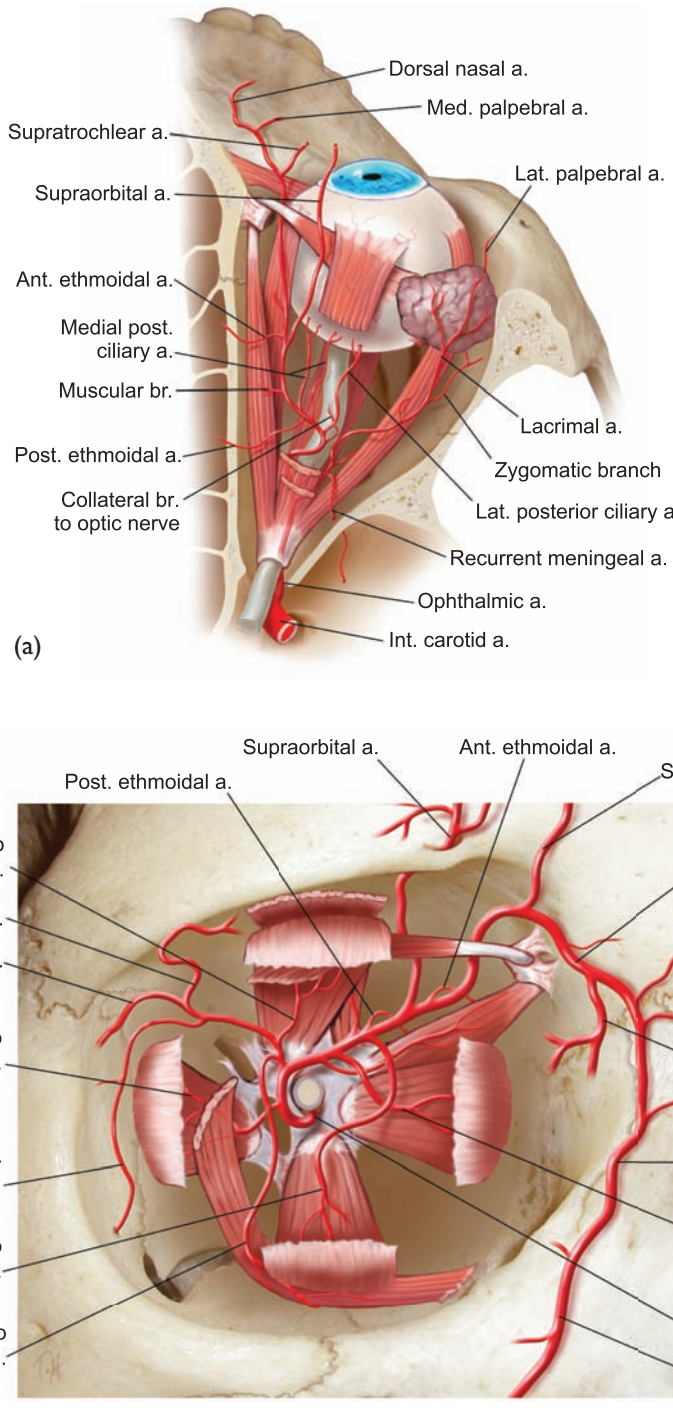


Fig. 1.24 The origin and branches of the ophthalmic artery within the orbital cavity. (a) From above, the ophthalmic artery travelling over the optic nerve. (b) Frontal view.

backward in the orbital fat, collecting venous tributaries that come from most areas supplied by the ophthalmic arteries, including the superior vortex veins and the central retinal vein.

The inferior orbital veins begin by draining a venous plexus situated at the anterior part of the orbit floor. It interacts with venous supplies such as the facial vein and the pterygoid venous plexus via the inferior orbital margin and the IOF, respectively.

As it passes backwards beneath the globe on the inferior rectus, the inferior ophthalmic vein drains the two inferior vortex veins.

Approaching the orbital apex, the inferior ophthalmic vein combines with the superior ophthalmic vein before exiting the orbit via the supraorbital fissure into the cavernous sinus. Occasionally, the two ophthalmic veins drain separately into the cavernous sinus.

Given that the pterygoid plexus is in such close proximity to the cavernous sinus, it is possible to spread infection from one to the other. This may result in a cavernous sinus thrombosis and represents an ophthalmological emergency.

The cavernous sinus

The median part of the MCF is formed by the body of the sphenoid bone. The cavernous sinus fills the majority of the inside of the body of the sphenoid bone, and is directly related to the side of this bone (see also Fig. 1.20). This large venous sinus receives venous drainage from the superior and inferior ophthalmic veins. They also receive venous blood from:

- the sphenoparietal sinus (from above)
- the inferior and superficial middle cerebral vein.

Drainage of the cavernous sinus

Although there are multiple routes of venous drainage from the cavernous sinus, the main drainage sites are as follows:

- The superior and inferior petrosal sinuses drain blood from the cavernous sinus into the internal jugular vein.
- An emissary vein that is transmitted though the foramen ovale drains blood from the cavernous sinus to the pterygoid plexus.

Relations of the cavernous sinus

Superiorly: the pituitary gland.
Inferiorly: the sphenoid and its osteal covering.
Posteriorly: the apex of the petrous part of the temporal bone.
Anteriorly: SOF and orbit.
Medially: the pituitary fossa.

The internal carotid artery enters the cranial cavity through the foramen lacerum and penetrates the posterior cavernous sinus before travelling forward near the medial walls of the sinus.

Within the sinus but outside the endothelial lining, the sixth nerve and the internal carotid artery travel alongside each other. The separation of endothelial linings means that the two blood sources do not cross over.

As the internal carotid artery exits the cavernous sinus superiorly, it reflects backwards on itself before moving towards the anterior perforate substance of the brain.

From superior to inferior, the walls of the cavernous sinuses contain the following:

1. third nerve
2. fourth nerve
3. sixth nerve
4. ophthalmic division of the fifth nerve (V1)
5. maxillary division of the fifth nerve (V2).

The extraocular muscles

Positions of the eye

Primary position pertains to when the eyeball is facing straight ahead.

The secondary positions pertain to when the eyeball is directed in horizontal and vertical deviations from the primary position (up, down, left, right).

The tertiary position of the gaze concerns oblique movements from the primary position (up and left, down, right).

These must not be confused with primary, secondary, and tertiary gaze or muscle actions.

The four rectus muscles

The four rectus muscles have their origin in the CTR at the apex of the orbit. The CTR is also known as the annulus of Zinn and is connected to the dural sheath of the optic nerve.

All four of the rectus muscles project forward within a sheath derived from Tenon's capsule and attach to the globe. The superior and inferior recti run in a direction 25° to the optical axis.

On reaching the globe, each extraocular muscle inserts at different points behind the corneal limbus (Fig. 1.25). Starting with the medial rectus and moving around to finish at the superior rectus, they form a spiral, called the spiral of Tillaux.

In order of their positions in the spiral of Tillaux, the four recti are the medial rectus, the inferior rectus, the lateral rectus, and the superior rectus.

The medial rectus

Innervation

Inferior division of the third cranial nerve. It inserts laterally at the proximal second quarter of the muscle.

Course

- From the medial CTR, the medial rectus is the largest of the four recti. It passes forward in the medial aspect of the orbit before entering the inferior Tenon's capsule to insert behind the limbus as above.

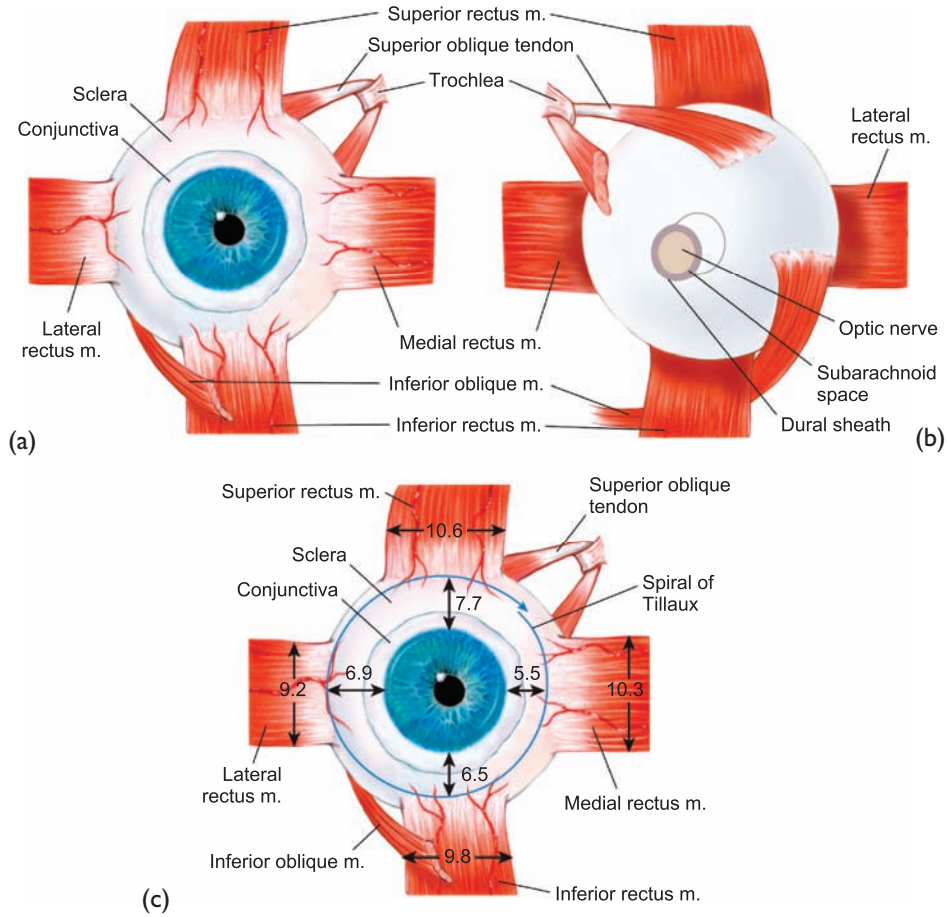


Fig. 1.25 Insertions of extraocular muscles. (a) Frontal view. (b) Back view. (c) Spiral of Tillaux.

Reproduced from David R. Jordan, Louise Mawn and Richard L. Anderson, *Surgical Anatomy of the Ocular Adnexa, A Clinical Approach*, Second Edition, Figure 4.2C, Page 116, 2012, with permission from Oxford University Press.

Relations

Above—superior oblique muscle, the ophthalmic artery and branches therein, the nasociliary nerve.
Below—floor of the orbit.

The inferior rectus

Innervation

The inferior division of the third cranial nerve. It inserts superiorly at the proximal second quarter of the muscle.

Course

- From the inferior CTR, it passes forward and slightly laterally (because of the lateral direction of the orbit from its apex) before entering the inferior Tenon's capsule to insert behind the limbus as above.
- The inferior rectus acts 23° lateral to the optical axis.
- The fascial sheath of the inferior rectus and that of the inferior oblique are connected to the suspensory ligament of the eyeball (Lockwood's ligament). Lockwood's ligament extends from the medial to lateral check ligaments and attaches the inferior rectus sheath to the lower lid.

Relations

Above—the third nerve, the optic nerve, the eye.
Below—floor of the orbit, the infraorbital nerve and vessels, maxillary sinus.

The lateral rectus

Innervation

The abducens (VI) nerve (hence abduction), which enters the muscle from the medial surface in the middle of the muscle.

Course

- The lateral rectus arises from two locations:
  1. the main origin is the lateral CTR
  2. a subsidiary arm is located outside and lateral to the CTR on the greater wing of sphenoid.
- The muscle moves forward in the lateral portion of the orbit before entering the Tenon's capsule to insert behind the limbus as above.

Relations

Above—lacrimal nerve and artery.
Below—floor of the orbit.
Medially—the abducens nerve.

The superior rectus

Innervation

Superior division of the third nerve. The nerve inserts inferiorly at the proximal second quarter of the muscle and goes on to innervate the levator palpebrae.

Course

- From the CTR, the superior rectus passes forward, entering Tenon's capsule before inserting behind the limbus.
- The superior rectus acts 23° lateral to the optical axis.
- The fascial sheath of the superior rectus and that of the levator palpebrae are connected such that they work in unison. This allows some degree of passive lid retraction on upward gaze.

Relations

Above—levator palpebrae muscle, frontal nerve, orbit roof.
Below—optic nerve, ophthalmic artery, nasociliary nerve, the superior oblique muscle.

The two oblique muscles

The oblique muscles aid the eye in rotatory movement known as in- or ex-cyclotorsion (in primary gaze).

The superior oblique

Innervation

The trochlear (IV) nerve.

Course

- The superior oblique originates external to the CTR, from the body of the sphenoid bone. Its origin is superonasal to the CTR and the optic nerve.
- The body of the muscle then passes forward and medially to bend around a connective-tissue pulley known as the trochlea (hence the innervation).
- After diversion, via the trochlea, the muscle moves laterally, downward, and slightly posteriorly, entering Tenon's capsule before passing underneath the superior rectus and attaching to the globe.
- The point of attachment is posterior to the equator and, when viewed from above, lies in the proximal lateral quadrant.
- The superior oblique achieves intorsion by acting in a direction 54° to the optical axis.

Relations

Above—superior rectus, roof of the orbit.
Below—ophthalmic artery, the nasociliary nerve.
Laterally—the supratrochlear nerve lies above and lateral to the superior oblique.

The inferior oblique

Innervation

The inferior division of the third nerve. This inserts into the middle of the muscle from above.

Course

- Of the six extraocular muscles, the inferior oblique is the only muscle to originate from the front of the orbit. It arises from the floor of orbit (near the junction of the maxillary plate and the lacrimal bone) lateral to the nasolacrimal canal and just within the orbital rim.
- It then moves laterally, backwards, and superiorly as it moves below the inferior rectus to curve upwards around the globe.
- The attachment of the inferior oblique lies posterior to the equator, in the same quadrant (when viewed from above) as the superior oblique. Overlying the attachment is the lateral rectus.
- The inferior oblique achieves excyclotorsion by acting in a direction 53° to the optical axis.

Relations

Above—inferior rectus muscle and the eye.
Below—the floor of the orbit.

Muscle, action, and movement

Primary, secondary, and tertiary muscle actions are defined by their order of amplitude, primary being the largest amplitude of movement.

Assuming the starting point of the eyeball is the primary position, the primary, secondary, and tertiary muscle actions are shown in Table 1.4.

Apart from the medial and lateral recti, different gazes will change the alignment of each muscle and thus alter the mechanics and primary action accordingly (Table 1.5).

It is important to note that muscles do not work in isolation but depend on the position of the globe. Note that:

- the superior and inferior recti pull back at 23° to the nasal side of the vertical axis with the eye in the primary position
- the tendon of the superior oblique has a forward pull at an angle of 54° to the nasal side of the vertical axis of the eyeball with the eye in the primary position

Table 1.4 Primary, secondary, and tertiary action of extraocular muscles

Muscle	Primary action	Secondary action	Tertiary action
Medial rectus	Adduction–rotation around the *y*-axis	None	None
Lateral rectus	Abduction–rotation around the *y*-axis	None	None
Superior rectus	Elevation around the *x*-axis	Adduction around the *y*-axis	Intortion around the *z*-axis
Inferior rectus	Depression around the *x*-axis	Adduction around the *y*-axis	Extortion around the *z*-axis
Superior oblique	Depression around the *x*-axis	Abduction around the *y*-axis	Intorsion around the *z*-axis
Inferior oblique	Elevation around the *x*-axis	Abduction around the *y*-axis	Extortion around the *z*-axis

All three axes are mutually perpendicular: the *z*-axis represents the anteroposterior axis and equates to the optical axis, the *x*-axis represents the lateral-medial axis, and the *y*-axis represents the superoinferior axis.

Table 1.5 Primary action in adduction and abduction

Muscle	Primary action when eye is in adduction	Primary action when eye is in abduction
Superior rectus	Intortion	Elevation
Inferior rectus	Extortion	Depression
Superior oblique	Maximal depression	Maximal intortion
Inferior oblique	Maximal elevation	Maximal extortion

- the inferior oblique pulls forward at an angle of 51° to the nasal side of the vertical axis of the eyeball with the eye in the primary position.

Once the eyeball moves, the position of the insertions relative to the origin changes and thus the muscle action changes. Hence, the careful coordination of all six muscles together is important and they all pull as a group of muscles. There must be conjugate movement of both eyes and hence 12 muscles must be coordinated through constant visual feedback to allow fine adjustments and prevent diplopia. Two principles contribute to regulation of eye movement:

- Hering's law of equal innervation states that during any conjugate eye movement, equal and simultaneous innervation flows to the yoke muscle (contralateral synergist).
- Sherrington's law of reciprocal innervation (inhibition) states that increased innervation to an extraocular muscle (e.g. right medial rectus) is accompanied by reciprocal decrease in innervation to its antagonist (i.e. right lateral rectus). This applies to both versions and vergences.

The eyelids

The eyelids protect the eyes from injury and excessive light, spread tears over the surface of the eyes, and help with the exit of tears through the tear ducts.

A horizontal furrow (the superior palpebral sulcus) divides the eyelid into an orbital part and a tarsal part. The sulcus is created because of the insertion of aponeurotic fibres from the levator palpebrae superioris.

There is a similar sulcus in the lower lid, the inferior palpebral sulcus. This is created by adhesions between the skin and the orbicularis oculi. With age, two other sulci are also formed, the naso-jugal sulcus and the lateral (or malar) sulcus.

The eyelids meet at the medial and lateral canthal angles or canthi, and the opening between the two lids is called the palpebral fissure. In Orientals, the medial angle is overlapped by a vertical skin fold, the epicanthus. The lateral angle is in direct contact with the surface of the eye, whereas the medial angle lies around 6 mm medial to the part of the lid that remains in contact with the globe.

Medially, the two eyelids are separated by the lacus lacrimalis, in the middle of which lies the caruncula lacrimalis. A semi-lunar fold, the plica semilunaris, lies medial to the caruncula (Fig. 1.26).

Each eyelid margin is around 2 mm thick and 30 mm long. About 5 mm from the medial angle there is a mound, the papilla lacrimalis, at the summit of which is the punctum lacrimale (0.4–0.8 mm diameter). The punctum leads into the lacrimal canal, which is discussed further in the section on the lacrimal system (p. 39). The papilla lacrimalis projects into the lacus lacrimalis and this apparatus serves to drain tears from the surface of the eyes.

From superficial to deep, the eyelid layers are as follows:

- skin
- subcutaneous tissue
- orbicularis oculi—striated muscle fibres
- orbital septum/tarsal plate
- conjunctiva.

The skin of the eyelids is extremely thin and folds easily. It becomes continuous with the conjunctiva at the site of the orifices of the meibomian glands. Eyelashes are present in double or triple rows on the lid margin from the lateral angle to the papilla lacrimalis. These are more numerous on the upper lid (150) than the lower lid (75), are commonly darker than scalp hair, and do not lose pigmentation with age. They do not possess erector pili muscles and the sebaceous glands of Zeiss and the ciliary glands of Moll open into the follicles of the lash or directly onto the lid margin.

The subcutaneous tissue is loose and mostly comprises elastic tissue and very little fat.

The orbicularis oculi muscle (Fig. 1.27) is a striated muscle that extends to the temporal region and cheek (orbital part) and onto the eyelids (palpebral part). The latter consists of thin bundles of fibres that arise from the medial palpebral ligament (MPL). These fibres are attached to the superficial and deep surface of the ligament but not to the inferior surface. They pass laterally and concentrically, and laterally they meet at the lateral palpebral raphe. The lacrimal part of the orbicularis oculi lies behind the lacrimal gland. Beneath the orbicularis oculi lies a layer of connective tissue that contains the blood vessels and nerves of the eyelids.

Medially, the attachments of orbicularis oculi are complex. The pretarsal muscles attach medially by a deep and superficial head. The superficial head forms the medial canthal tendon. The deep head has fibres which begin at the medial end of the tarsal plates and insert to the posterior lacrimal crest, behind the lacrimal sac. The preseptal orbicularis oculi muscles also insert medially by a superficial and deep head. The superficial head joins with the medial canthal tendon. The deep head inserts into the fascia overlying the lacrimal sac as well as into the medial orbital wall above and below the deep head of the pretarsal attachment of the orbicularis oculi.

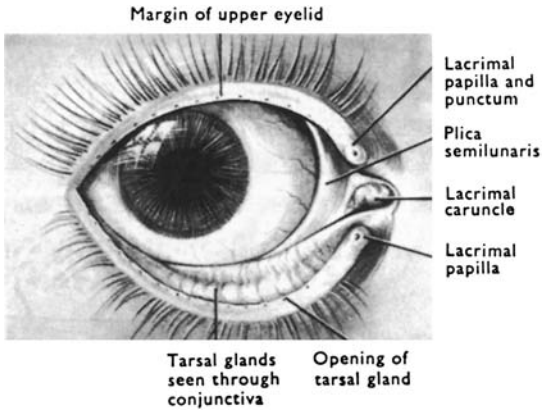


Fig. 1.26 Right eye with lids everted to show conjunctival sac, lacus lacrimalis, and the caruncula lacrimalis.

Reproduced from G.J. Romanes, *Cunningham's Manual of Practical Anatomy, Head and Neck and Brain*, Volume Three, Fifteenth Edition, Figure 15, Page 13, 1986 with permission from Oxford University Press.

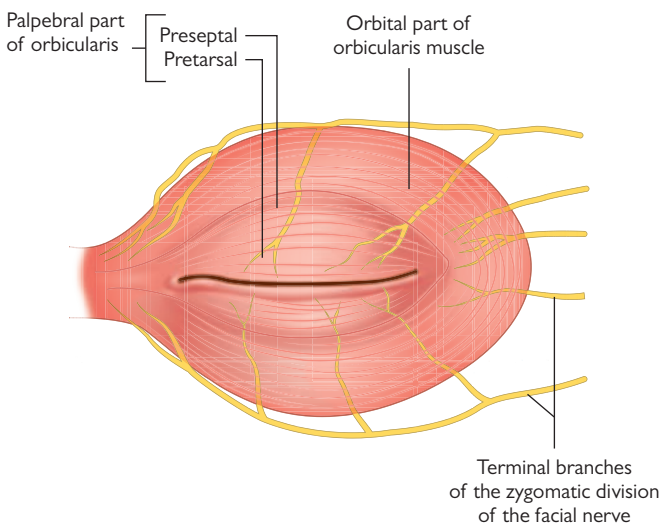


Fig. 1.27 Orbicularis oculi muscle and the terminal branches of the facial nerve.

This figure was published in *Colour Atlas of Ophthalmic Plastic Surgery*, Third Edition, AG Tyers and JRO Collin, Figure 1.6, p. 8, Copyright Elsevier 2007.

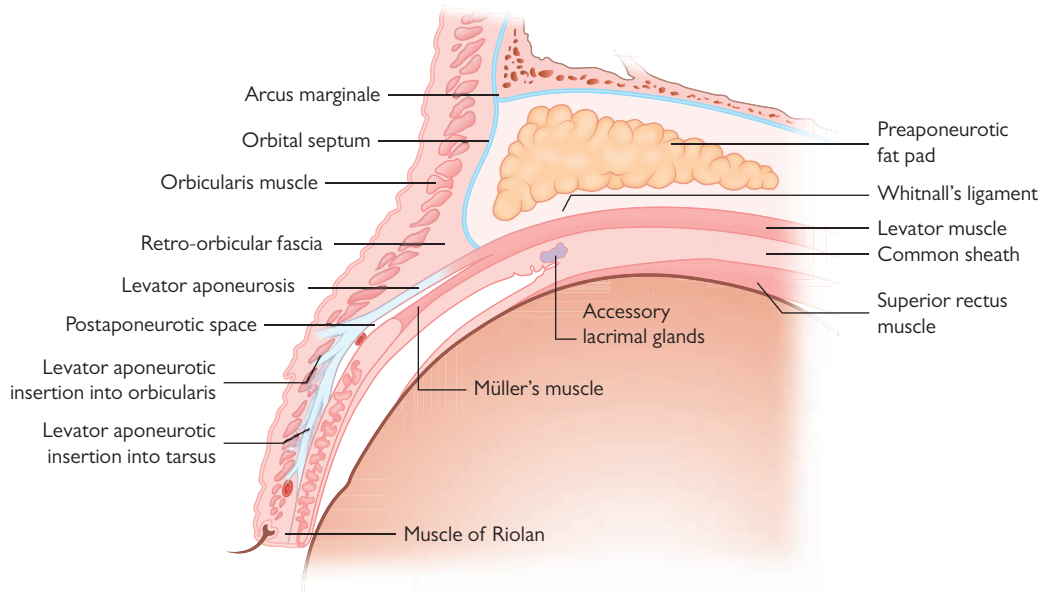


Fig. 1.28 Cross-section of the upper eyelid.

This figure was published in *Colour Atlas of Ophthalmic Plastic Surgery*, Third Edition, AG Tyers and JRO Collin, Figure 1.16, p. 18, Copyright Elsevier 2007.

The lateral palpebral ligament attaches the lateral ends of the tarsi to the marginal (Whitnall's) tubercle on the orbital margin, as formed by the zygomatic bone. It lies under the lateral palpebral raphe.

The orbicularis oculi is supplied by temporal and zygomatic branches of the facial nerve, which enter the deep surface of the nerve from laterally. The orbital part muscle is mostly under voluntary control and pulls on the surrounding tissues like a purse string to draw the lids towards the medial angle of the orbit. The palpebral part is under both voluntary and involuntary control. It draws up the lower lid and draws down the upper lid. The blink reflex (involuntary) is initiated by drying of the cornea. The lacrimal portion draws the eyelids medially and contributes to the pumping mechanism of tear drainage as well as positioning the puncta lacrimalis.

The orbital septum is a membranous sheet (composed of seven layers) that is attached to the orbital margin, where it is continuous with the periosteum. It acts to separate the eyelids from the contents of the orbit.

The tarsal plates consist of fibrous tissue that acts as a skeleton for the eyelids. The tarsal plate of the upper lid is around 10 mm in height in the centre. Attached at its upper end is the orbital septum and the smooth muscle fibres of the levator palpebrae superioris. The tarsal plate of the lower lid measures around 5 mm in height centrally. The orbital septum is attached at its lower end. The tarsal plate is actually 70% composed of glandular structure because of the presence of meibomian glands within them. There are 20–25 meibomian glands arranged in a single row in each upper and lower lid. These are modified sebaceous glands, made of a long central canal surrounded by 10–15 acini. They secrete an oily substance that prevents overflow of tears, reduces evaporation of tears, and allows the closure of the eyelids to be airtight.

The superior tarsal muscle (Müller's muscle) is continuous superiorly with the levator palpebrae superioris and inferiorly with the tarsal plate of the upper lid. This muscle assists the striated muscle of the upper lid in elevating the upper lid. The inferior tarsal muscle is attached to the lower margin of the tarsal plate of the lower eyelid and is continuous with the fascial sheath of the inferior rectus muscle. Both muscles have a sympathetic nerve supply.

The conjunctiva is described on page 42. Importantly with relation to the adnexa, the fornices of the conjunctivae (the points at which it is reflected back on itself) have attachments to the aponeurosis of the levator palpebrae superioris (superior fornix) and the aponeurosis of the inferior rectus muscle (inferior fornix).

The normal position of the upper lid is maintained by the levator palpebrae superioris. This is a striated muscle that is only present in the upper lid. It has an aponeurotic tendon that begins posterior to the orbital septum.

The orbital septum is attached to the anterior surface of the aponeurotic fibres as a thickened band about 8 mm below Whitnall's ligament and 3–4 mm above the tarsus. This inserts medially into the trochlea and laterally into the

capsule of the lacrimal gland and orbital wall. It acts as a fulcrum for the levator palpebrae.

The preaponeurotic fat pad, an important surgical landmark, lies between the posterior surface of the septum and the levator aponeurosis.

The levator aponeurosis passes anteriorly and inferiorly and has four attachments (Fig. 1.28):

1. anterior surface of the upper lid tarsal plate posteriorly
2. muscular fibres of orbicularis oculi anteriorly
3. upper lid skin to form the skin crease
4. superior conjunctival fornix.

The nerve supply to the levator palpebrae superioris is by the superior branch of the oculomotor nerve.

The medial and lateral extents of the levator are in the form of two 'horns' into the region of the canthal tendons. The lacrimal gland is closely associated with the posterior edge of the lateral horn.

Müller's muscle is a smooth muscle with sympathetic innervations that arises from the underside of the levator muscle. It is 15–20 mm wide and descends between the levator aponeurosis and conjunctiva for 15–20 mm to insert into the superior border of the tarsal plate.

Arterial supply of the eyelids comes from the lateral and medial palpebral arteries:

- The lateral palpebral artery branches from the lacrimal artery, itself a branch of the ophthalmic artery.
- The medial palpebral arteries (superior and inferior) arise from the ophthalmic artery below the trochlea of the superior oblique muscle, pass behind the lacrimal sac, and enter the eyelids.
- Each medial palpebral artery divides into two branches that pass laterally, forming two arches in each upper and lower lid. The arches anastomose with the lateral palpebral arteries as well as with branches of the superficial temporal, transverse facial, and infraorbital arteries.

The veins of the eyelids drain:

- medially into the ophthalmic and angular veins
- laterally into the superficial temporal vein.

Lymphatic vessels from the lateral two-thirds of the upper and lower lids drain to the superficial parotid nodes. Those from the medial angle drain to the submandibular nodes.

Sensation to the upper lids is from the ophthalmic division of the trigeminal nerve. The following branches are involved: infratrochlear, supratrochlear, supraorbital, and lacrimal nerves.

Sensation to the lower lid is supplied by the infratrochlear branch of the ophthalmic division of the trigeminal at the medial angle. The rest of the lower lid is supplied by branches of the infraorbital nerve, the terminal portion of the maxillary branch of the trigeminal nerve.

The lacrimal system

See Chapter 2.

Tears are produced by the lacrimal gland and the accessory lacrimal glands to form the tear lake. This lake drains via the superior (20%) and inferior (80%) punctae in the lids. These punctae are connected via canaliculi to the lacrimal sac, which then drains via the nasolacrimal ducts into the inferior nasal meatus (Fig. 1.29).

The lacrimal gland

Structure

- The lacrimal gland is a bilobed tubulo-acinar exocrine gland that resides in the anterior supero-temporal aspect of the orbit and lies entirely behind the septum (Fig. 1.30).
- It is, in essence, C-shaped. The lacrimal gland wraps around the lateral border of the aponeurosis of levator palpebrae. Each lacrimal gland is divided into two lobes. The palpebral lobe sits below the aponeurosis and extends to the upper lid. The larger orbital lobe sits above the aponeurosis and lies within the lacrimal gland fossa. The lobes are continuous with each other (Fig. 1.30).
- Despite being grossly divided into two lobes the lacrimal gland is largely an amorphous structure and possesses no strict capsule, although periorbita is believed to surround it.
- Ducts arising from the orbital portion of the gland travel through the palpebral portion to drain into the conjunctival sac. There are approximately 12 of these ducts.
- Ducts may also arise independently from the palpebral portion directly into the superior fornix.

In addition to the main lacrimal gland there exist small accessory lacrimal glands, which reside in the conjunctiva concentrating in the fornices. The sheer number of these additional glands means that it is possible to maintain a satisfactory tear volume despite the removal of the main gland.

Histology

As previously mentioned the lacrimal gland is a bilobed, lobulated, tubulo-acinar exocrine gland. Each lobule is separated by connective tissue.

Acini are small round masses of columnar epithelial secretory cells with a central lumen. In order to produce the relatively large quantities of aqueous tears, each acinus is rich in rough endoplasmic reticulum, Golgi apparatus, and mitochondria.

Lacrimal acini drain into lobular ducts that coalesce and then go on to drain into interlobular ducts. These ducts comprise a columnar or cuboidal epithelial lining with myoepithelial cells in the periphery.

Innervation

The lacrimal gland has three modes of innervation (Fig. 1.31):

1. sympathetic
2. parasympathetic
3. sensory.

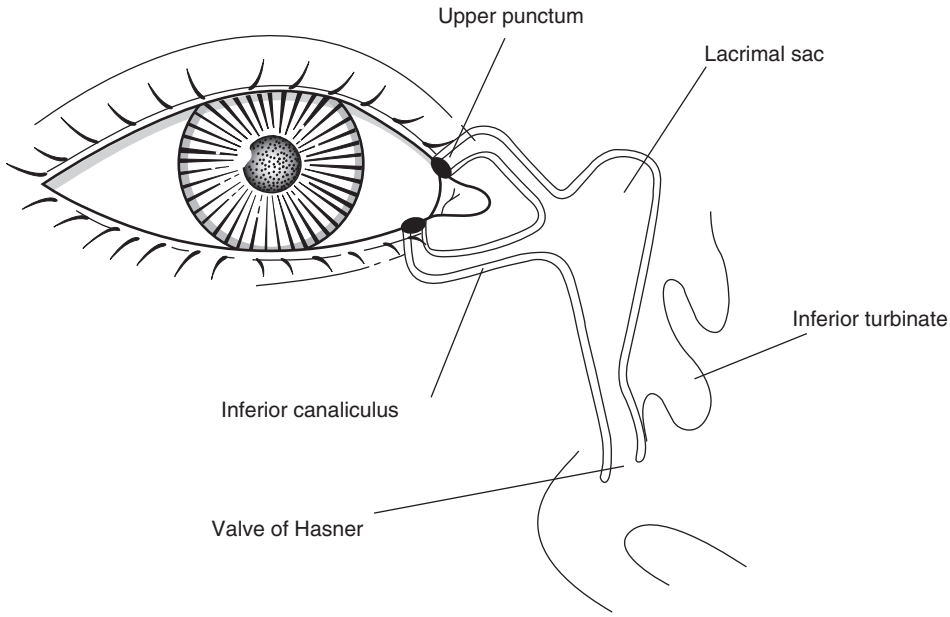


Fig. 1.29 The nasolacrimal system.

Reproduced from Michael Yen, *Surgery of the Eyelids, Lacrimal System, and Orbit*, Figure 14.1, Page 180, 2012, published by Oxford University Press, with permission from Michael Tze-Chien Yen.

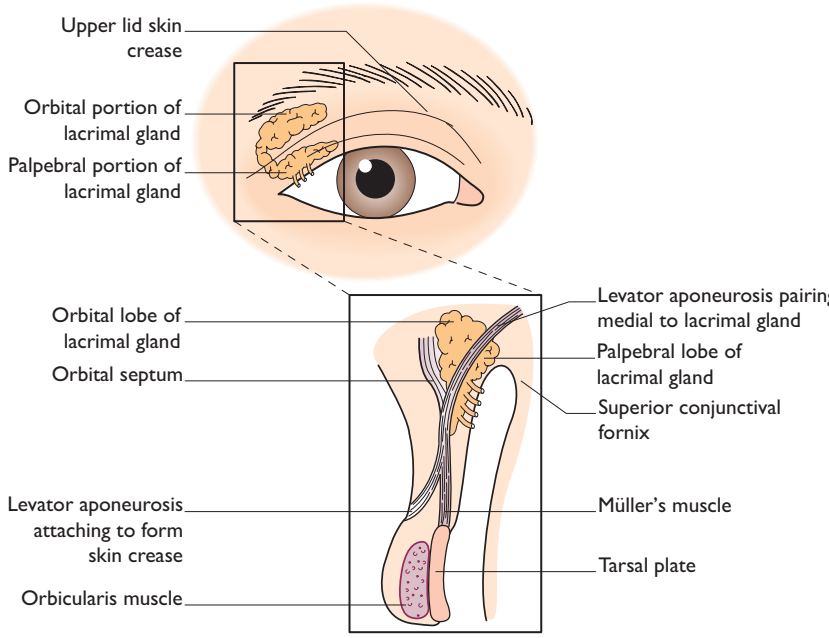


Fig. 1.30 Saggital section of the orbital and palpebral portions of the lacrimal gland. Note the relationship of the gland to the aponeurosis of the levator palpebrae superioris muscle and to the superior fornix of the conjunctiva.

With permission from Neil Modi.

Sympathetic

- Sympathetic innervation arises with postganglionic fibres from the superior cervical sympathetic ganglion.
- These move through the nerve plexus surrounding the internal carotid artery. They join the deep petrosal nerve, then the nerve of the pterygoid canal, then the maxillary nerve, the zygomatic nerve, the zygomaticotemporal nerve, and finally the lacrimal nerve.

Parasympathetic

- Parasympathetic supply arises from the lacrimal or lacrimatory nucleus (a portion of the superior salivatory nucleus) of the facial nerve. This is a visceromotor cranial nerve nucleus located in the pontine tegmentum.
- Preganglionic fibres are destined for synapse at the pterygopalatine ganglion. These fibres move from the lacrimal nucleus through the nervus intermedius and its great petrosal branch, then through the nerve of the pterygoid canal to the pterygopalatine ganglion. Postganglionic fibres leave to join the maxillary nerve. These fibres move along the zygomatic branch of the maxillary, then the zygomaticotemporal nerve, then the lacrimal nerve to reach the lacrimal gland.

Sensory

The sensory fibres are branches of the ophthalmic division of the trigeminal nerve. These reach the lacrimal gland, as with all lacrimal gland terminations, via the lacrimal nerve.

Lacrimal drainage

The lacrimal canaliculi appear as a pincer-like shape extending from the lacrimal sac (Fig. 1.29).

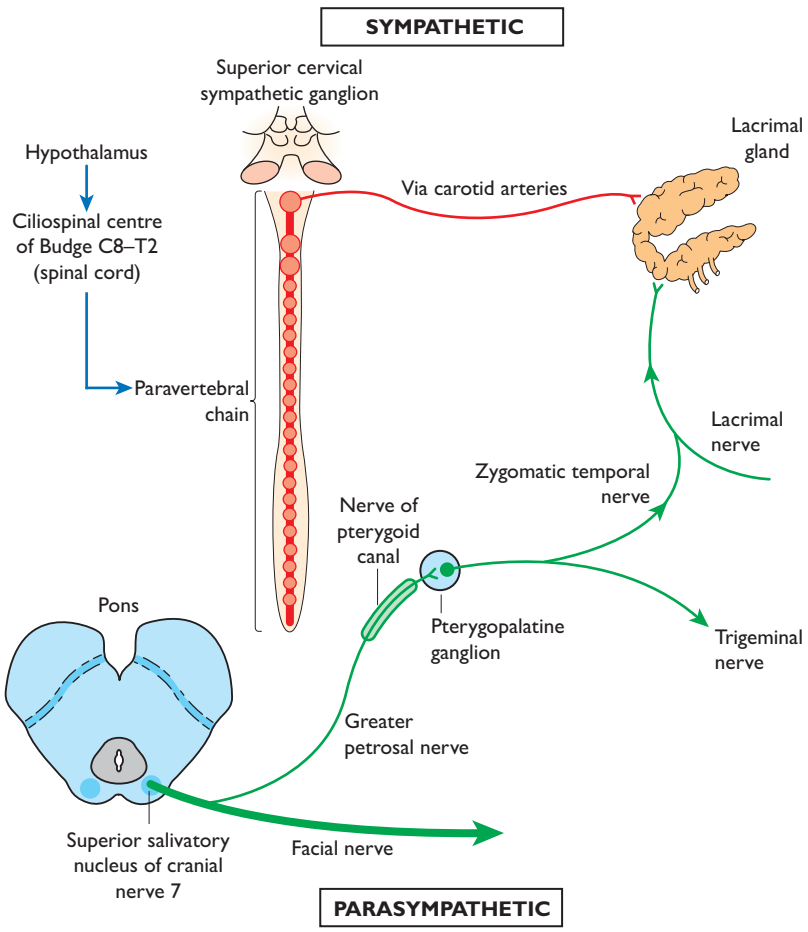


Fig. 1.31 The autonomic innervations of the lacrimal gland.

With permission from Neil Modi.

Starting from the punctum, both canaliculi start with a 2 mm vertical path which then turns medially to become more horizontal (10 mm). At this point the canaliculi are still surrounded by fibres of the palpebral portion of the orbicularis oculi.

Both canaliculi then move posteriorly and combine, in most cases, to form a common canaliculus that lies behind the MPL. The common canaliculus then pierces the lacrimal sac 2.5 mm below its apex—see *The lacrimal sac* section.

From the sac, the nasolacrimal duct continues in an osseous canal (consisting solely of maxillary bone) that leads it in an inferior, posterior, lateral direction eventually exiting in the inferior meatus.

On blinking, the deep heads of the pretarsal muscles pull the lids medially, shortening the canaliculi. At the same time, the lacrimal fascia and sac wall are pulled laterally by the contraction of the deep heads of the preseptal muscle. As a result, the punctae close and the tears in the canaliculi are pulled medially and sucked into the sac. After the blink, the lacrimal fascia and sac move medially again and the lids move laterally. The puncta reopen and the canaliculi refill with tears. In this way, there is a lacrimal 'pump' that drains tears from the ocular surface to the lacrimal sac.

CLINICAL TIP

Important relations to the lacrimal sac when performing a dacrocystorhinostomy (DCR):

- Upper portion of the lacrimal sac is bounded anteriorly by the MPL.
- Posteriorly the upper sac is bound by the lacrimal part of the orbicularis oculi.
- Medially the upper part of the lacrimal sac is related to the anterior ethmoid air cells.
- The sac is covered laterally first by a venous plexus and then by the lacrimal fascia (formed from but distinct from periorbita). This occurs only between the anterior and posterior lacrimal crests.
- The lower part of the lacrimal sac is medially related to the anterior middle meatus of the nose.

NB The angular vein is an important structure to avoid when performing a DCR. It crosses the anterior surface of the MPL, lying around 8 mm medial to the medial canthus.

The lacrimal sac

The lacrimal sac is around 12 mm long and is housed in the lacrimal groove or fossa (as opposed to the lacrimal gland fossa). This groove consists of the lacrimal bone and frontal process of maxilla.

The sac itself consists of fibroelastic tissue and is lined by two layers of columnar cells.

Drainage from the lacrimal sac to the nasolacrimal duct is the result of gravity.

Ocular anatomy

The conjunctiva

Structure

The conjunctiva is a mucous membrane that covers both the globe and the internal lid surface—these are termed bulbar and palpebral conjunctiva, respectively. The palpebral conjunctiva attaches firmly to the interior lid surface.

Bulbar conjunctiva attaches firmly 1 mm posterior to the corneo-scleral junction. The rest of the bulbar conjunctiva lies loosely over the sclera, serving as a protective 'sleeve' and aiding globe movement.

Bulbar and palpebral conjunctiva meet in pockets known as the fornices (Table 1.6). There are superior, inferior, and lateral fornices. All forniceal conjunctiva is related directly to the extensions of the extraocular muscle sheaths (Tenon's capsule), thus allowing the conjunctiva to be drawn towards the direction of eye movement without sagging.

Histology

Conjunctival histology varies throughout its structure (Table 1.7). On the whole, conjunctiva may be summarized into two components: an underlying lamina propria and an accompanying epithelium (see Table 1.7 for the differences).

Goblet cells feature heavily in conjunctival histology, reflecting its role in contributing mucus to the tear film. The goblet

Table 1.6 Dimensions of the conjunctival fornices

Fornix	Superior	Inferior	Lateral
Distance from limbus (mm)	10	8	14 (this is the deepest fornix of the three and extends posterior to the equator)

Table 1.7 Variations in histological appearance in the conjunctiva

Conjunctival site	Palpebral	Bulbar	Limbus
Histological differences	Stratified squamous epithelium + lamina propia	Stratified columnar epithelium + lamina propria	Stratified squamous epithelium + limbal pupillae formed from lamina propria

cells are found most frequently in the inferonasal bulbar conjunctiva.

Conjunctival stem cells

Conjunctival epithelium has been shown to have self-renewing properties due to its population of stem cells. Conjunctival epithelial stem cells also play a significant role in tissue homeostasis but are intrinsically distinct from the limbal stem cells. In fact, the two have been shown to have distinctly separate lineages and, unlike corneal stem cells, conjunctival stem cells have been shown to be bipotent, producing both goblet and non-goblet epithelial cells.

The location of conjunctival stem cells remains under investigation but is thought to be predominantly in forniceal conjunctiva, which has been demonstrated in the murine model.

Blood supply

Palpebral conjunctiva and peripheral bulbar conjunctiva are supplied by the same supply as the eyelids: the medial and lateral palpebral arteries.

The anterior ciliary arteries supply the bulbar conjunctiva more proximal to the limbus.

All conjunctival veins ultimately drain into the superior and inferior ophthalmic veins, which in turn drain into the cavernous sinus. Thrombosis of the cavernous sinus will therefore produce engorgement of conjunctival veins.

Conjunctival-associated lymphoid tissue

The conjunctiva is a mucosal tissue rich in lymphatic drainage. It contains focal aggregations of lymphoid tissue known as conjunctival-associated lymphoid tissue (CALT), which is part of the mucosa-associated lymphoid tissue (MALT).

MALT refers to specialized lymphoid tissue that exists in most mucosal surfaces and differs from primary lymphoid tissues (spleen and lymph nodes) in its structure, location, and cellular composition. The highly specialized MALT epithelium has developed a significant role in antigen uptake and processing. For example, MALT contains specialized epithelial cells with unique apical microfolds, known as M cells. These cells are able to engulf and deliver antigens to neighbouring antigen-presenting cells. The absence of goblet cells within MALT also permits a reduction in the mucin layer, allowing MALT epithelium to gain better access to antigen.

Ocular MALT components include the conjunctiva, lacrimal gland, tear film, and cornea.

CALT plays a key role in local active immunity and in developing immune tolerance. Aside from the conjunctiva, the lacrimal gland and the tear film are other components of CALT.

Lymphatic drainage

The conjunctival lymph vessels are an anastomotic network divided into superficial and deep plexuses that lie in the submucosa.

The small lymph vessels of the bulbar conjunctiva arise 1 mm peripheral to the limbus and join to form larger vessels in the substantia propria, which then drain towards the palpebral commissures. There, they join the lymphatic drainage of the palpebral conjunctiva and the eyelid. The lymph vessels of the lateral half of the lid drain into the preauricular (intraparotid) node system, whereas the medial half drains into the submandibular lymph nodes.

Innervation

- Bulbar conjunctiva: long ciliary nerves (these are branches of the nasociliary nerves—part of the ophthalmic division of the trigeminal, V1).
- Superior palpebral conjunctiva and superior fornix: frontal and lacrimal branches of V1.
- Lateral inferior palpebral conjunctiva and lateral fornix: lacrimal branch of V1.
- Nasal inferior palpebral conjunctiva and nasal fornix: infraorbital nerve of V2.

Summary of secretion types within the adnexae

Table 1.8 shows the secretion types matching the layers of the tear film.

The caruncle

The medial angle of the lid aperture typically lies around 6 mm medial to the globe as opposed to the lateral angle, which is in direct contact with the globe.

At the medial angle there lies a small pink triangle of tissue that contains the 'lake of tears' or lacus lacrimalis, i.e. the pool of tears that sits in the inner canthus. The medial fleshy elevation of the lacus is known as the caruncle or caruncula lacrimalis.

The flat lateral fold of the lacus lacrimalis is known as the plica semilunaris due to its half-moon shape (Latin *plica*, meaning fold or corrugation).

Table 1.8 Secretory glands within the adnexae

Secretory gland/cell	Secretory mechanism	Secretion	Location
Goblet cells	Holocrine	Mucin	Conjunctiva, caruncle, and plica
Accessory lacrimal (Kraus, Wolfring)	Exocrine	Aqueous	Kraus + Wolfring: sub-palpebral conjunctiva Other accessory lacrimal tissue: plica, caruncle
Lacrimal	Exocrine	Aqueous	Orbital and palpebral parts of the lacrimal gland Anterior, superolateral orbit
Moll	Eccrine	Oil	Lid
Zeis	Holocrine	Oil	Hair follicles in lids and caruncle
Meibomian	Holocrine	Oil	Within tarsal plate

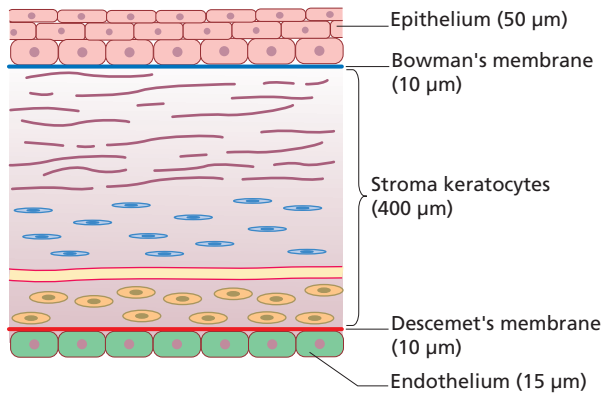


Fig. 1.32 Cross-sectional structural diagram of the cornea.
With permission from Louise Bye.

Histologically the caruncle is of modified skin containing sebaceous glands whilst its surface is of non-keratinized stratified squamous epithelium.

The cornea

See Chapter 2.

The cornea is the transparent layer of the anterior segment of the eye. The cornea, and in particular the interface between air and its anterior surface, has the greatest refractive index of the ocular optical system and, using its own structural rigidity, aids the formation of the anterior chamber.

Dimensions

The cornea itself is shorter vertically than it is horizontally. On average, the vertical measurement is 10.6 mm whereas the horizontal central diameter averages 11.7 mm. However, despite the difference of approximately 1.1 mm, the circumference of the cornea is circular when viewed posteriorly.

The cornea is also thinner centrally (0.52 mm) than it is peripherally (up to 1.1 mm at the limbus). Optically the cornea is described as aspheric because its radius of curvature varies from apex to limbus. Also, there is a difference between the anterior radius of curvature and the posterior radius of curvature, which are 7.8 and 6.5 mm, respectively. Quite often the vertical meridian is more convex than the horizontal.

Structure

The cornea itself comprises five major layers (beginning anteriorly) (Fig. 1.32):

1. the corneal epithelium
2. Bowman's layer (anterior limiting lamina)
3. corneal stroma (substantia propria)
4. Descemet's membrane (posterior limiting lamina)
5. corneal endothelium.

Epithelium

The corneal epithelium is the outermost layer and forms the air–tear film interface. It is formed of stratified squamous and columnar epithelial cells that regenerate quickly at the limbus. The squamous epithelial strata exists more anterior to the columnar cells.

Centrally, the epithelium is composed of five to six cell layers that are held together laterally by desmosomes and to

the underlying foundation that is the basal lamina by hemidesmosomes. Peripherally, the epithelium is of roughly 10 cell layers.

The thickness of the corneal epithelium typically lies between 50 and 60 μm.

Unlike the corneal endothelium, the epithelium is able to regenerate rapidly. Mitotic activity generates new cells from the basal lamina at the limbus. New cells are said to exhibit amoeboid sliding movements, which displace existing cells superficially and centripetally towards the centre of the cornea.

The basal epithelia lie on a distinct basal lamina that resides between the corneal epithelium and Bowman's layer. This can be further divided into two classes: lamina lucida 25 nm and lamina densa 50 nm. The corneal epithelium is anchored to Bowman's layer through a complex mesh of connecting tissues via the basal lamina, known as anchoring fibrils (collagen type VII) and anchoring plaques (collagen type VI). These interact with the lamina densa and the collagen fibrils of Bowman's layer to secure the epithelium to Bowman's layer.

There are also several groups of non-epithelial cells that appear within the corneal epithelium. These include lymphocytes, macrophages, pigmented melanocytes, and wandering histocytes.

Antigen-presenting Langerhans' cells are derived from bone marrow macrophage precursors and are located in the limbus and the peripheral cornea, but are usually absent in the central third of the cornea. Langerhans' cells migrate centripetally in response to trauma and infection but also as part of the ageing process.

Bowman's layer

Bowman's layer is an acellular region of connective tissue that is randomly spaced and consists of collagen fibrils of types I, III, V, and VI. It ranges in diameter from 20 to 30 nm.

Stroma (substantia propria)

By contrast to Bowman's layer, the stroma consists of very regular connective tissue. The substantia propria provides the majority of the cornea (approximately 90%). Despite its thickness, the corneal stroma is transparent for two reasons:

1. The regular arrangement of the collagen fibres in a lattice allows for destructive interference of light. Therefore light scattered by one lattice is cancelled out by similar scattering from an adjacent lattice.
2. The spacing between each lattice has a critical limit by which transparency can occur (< 200 nm).

Anatomically, the corneal stroma consists of 2-μm-thick flattened collagen fibres (collagenous lamellae), which are made up of 200–250 layers and run parallel to the corneal surface. Between the layers of lamellae lie modified stellate fibroblasts, which are keratocytes by classification and are the source of the lamellae.

There are around 2.4 million keratocytes within the cornea and they occupy between 3 and 5% of the stromal volume. They are most dense in the anterior cornea and are bound together by gap junctions. Histologically, keratocytes have abundant mitochondria, rough endoplasmic reticulum, and Golgi apparatus. Most of the substance of the corneal stroma is composed of a glycosaminoglycan-keratan matrix.

Descemet's membrane

Descemet's membrane is a basement membrane of the corneal endothelium and is generally rich in type IV collagen. This membrane is roughly 8–10 μm thick and its structure is based on two parts, both of which stain periodic acid-Schiff (PAS) positive and are rich in type IV collagen:

1. the anterior third
2. the posterior two-thirds.

The anterior third is a banded region that is also said to include type VIII collagen, and types V and VI collagen at the stromal–Descemet's interface. This may have a stabilizing role, providing adherence between the layers. The anterior third develops *in utero*.

The posterior two-thirds are the non-banded region and are laid down by the corneal endothelial cells throughout life.

Peripherally, Descemet's membrane runs continuously with the cortical zone of the trabecular meshwork, ending at Schwalbe's line.

Endothelium

The endothelium is a single-layered structure consisting of flattened polygonal cells and measuring around 15 μm thick. The free surface of these cells contains microvilli.

The cells cover the posterior face of Descemet's membrane and are continuous with the endothelial lining of the trabecular meshwork and the anterior cells of the iris.

One of the most important functions of the endothelium is to maintain a relative state of corneal dehydration by means of an active transport mechanism. This mechanism is supported by the abundance of mitochondria, rough endoplasmic reticulum, and Golgi apparatus in the endothelial cytoplasm. The endothelium also prevents aqueous from entering the stroma by acting as a barrier aided by its tight cell–cell junctions.

Should the endothelium become damaged, the cells do not regenerate but simply further flatten themselves to cover a larger surface area. This unfortunately provides a critical limit to which they operate. In Fuch's endothelial dystrophy there is a loss of endothelial cells, which can lead to stromal oedema and epithelial bullae.

Blood supply

The cornea is a rare body structure in that it is devoid of vasculature and lymphatic supply. The capillaries descending from the anterior ciliary arteries of the conjunctiva and sclera terminate as they reach the cornea.

The cornea receives its nourishment from the capillaries at its boundaries and, when the eye is open, indirectly from the oxygen dissolved in the tear film. When the eye is closed, it receives oxygen by diffusion from the lid vasculature and posteriorly from the aqueous.

Innervation

The cornea is supplied mainly by the long ciliary nerves, which originate from the ophthalmic division of the trigeminal nerve. As these myelinated nerves travel anteriorly through the perichoroidal space, they pierce the sclera just posterior to the limbus to form the annular plexus. Fibres then branch off the plexus in a radial pattern to enter the corneal stroma. These fibres further divide to form a subepithelial plexus. Finally, fibres from the subepithelial plexus branch off to Bowman's layer to terminate within the epithelium as an intraepithelial plexus of unmyelinated fibres.

The limbus

Although not an exact structure, the limbus is regarded as the transition zone at which corneal tissue becomes continuous with scleral tissue. It measures 1–2 mm wide and this is termed the corneoscleral limbus or junction. Pathologists, anatomists, and clinicians regard the limbus differently.

Limbal tissue comprises stratified-squamous, non-keratinizing epithelium (which bridges from the corneal epithelium to the conjunctival epithelium) and the subjacent vascularized loose connective tissue (Fig. 1.33).

Limbal epithelium

The limbal epithelium comprises seven to ten layers and is similar in structure to the corneal epithelium. However, unlike the cornea, Langerhans' cells are common and are found interspersed between the limbal epithelial cells.

The apical cells of the limbal epithelium display microvilli on their apical membrane, and tight junctions are present on their lateral borders.

The basal cells are smaller and less columnar than their corneal counterparts. Furthermore, the basal surface of these cells undulates and contains a far smaller number of hemidesmosomes as compared to the corneal epithelium. The basal cells are of particular importance as current opinion states that a small population of these cells comprises stem cells that replenish the corneal epithelium.

These stem cells are primitive and are able to divide symmetrically to self renew. They are also able to divide asymmetrically to give rise to transit amplifying cells (TAC), which migrate centripetally to populate the basal layer of the corneal epithelium.

The TAC divide and migrate superficially, becoming more differentiated. These cells eventually develop into terminally differentiated cells.

Limbal loose connective tissue

The connective tissue beneath the limbal epithelium is loosely arranged in radial undulations or folds known as the palisades of Vogt. It is hypothesized that the epithelial crypts that are formed by the palisades of Vogt help to provide deep housing for the basal cells and thus protect the stem cell population.

The collagen, proteoglycan, and glycoprotein constitution of the limbal connective tissue corresponds closely with

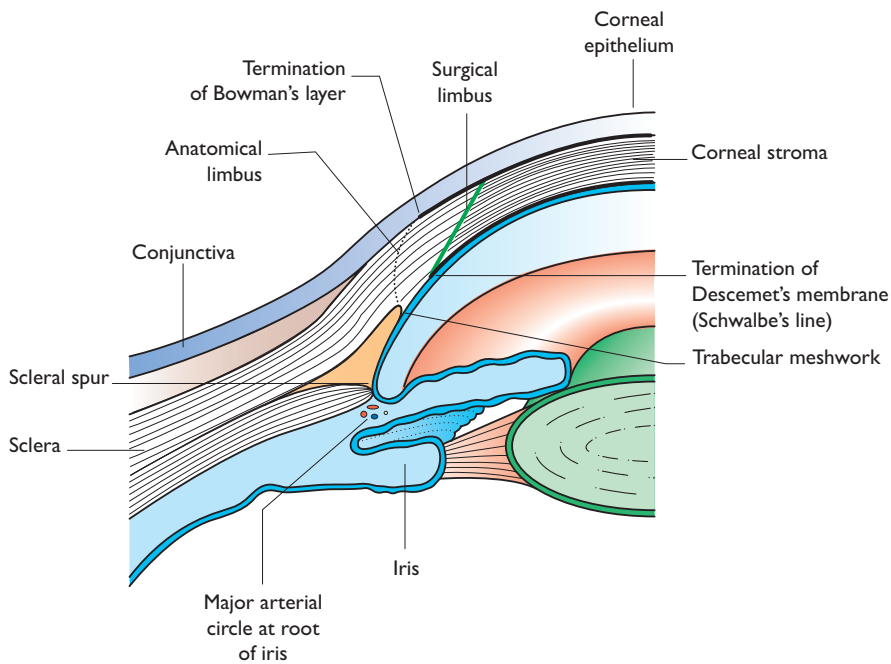


Fig. 1.33 The anatomical, surgical, and pathological limbus.
With permission from Neil Modi.

that of the cornea, with the exception of keratan, which is only found in the cornea.

Another feature of the limbal connective tissue is the peg-like outpockets of stroma known as rete. These lie between the ridges of the palisades and extend from the peripheral cornea through the limbus to the conjunctiva. It is thought that the rete contribute to the adhesion strength of the epithelial cells.

The anatomical limbus

The anatomical limbus takes up an arc as it traverses the tissues in an anterior to posterior manner and begins roughly 1 mm posterior to the conjunctival limbus. Schwalbe's line marks the posterior limit to the anatomical limbus.

Two grooves belie the anterior and posterior extremes of the limbal arc. The external scleral sulcus lies at the anterior aspect of the limbal arc.

The internal scleral sulcus lies at the posterior aspect of the limbal arc. It is this sulcus that contains the canal of Schlemm and the trabecular meshwork.

The surgical limbus

The surgical limbus lies slightly anterior to the anatomical limbus.

It is considered in two parts:

1. a blue/grey line that overlies clear cornea and extends from Bowman's layer to Schwalbe's line
2. a posterior white zone that includes the trabecular meshwork and extending from Schwalbe's line to the scleral spur.

The pathological limbus

The pathological limbus is more of a section of corneoscleral tissue. By joining an imaginary line between the peripheral terminations of Bowman's layer and Descemet's membrane, one creates the anterior wall of the pathological limbus. The posterior wall is formed by a line 1.5 mm posterior to the termination of Bowman's layer, which is perpendicular to the eye's surface.

The sclera

See Chapter 2.

The sclera is a tough, relatively avascular casing that encompasses the globe. It begins anteriorly as a continuation of the cornea, and runs posteriorly (internal to the conjunctiva) to finish by fusing with the dural sheath of the optic nerve.

The thickest point of the sclera is at the posterior pole, where it measures 1 mm. The sclera then thins to little more than 0.6 mm at the equator. However, the thinnest regions of the sclera lie directly behind the insertion of the extraocular muscles, where the sclera measures around 0.3 mm in thickness.

The sclera is penetrated by four types of structure:

1. the optic nerve and fibres
2. anterior ciliary arteries
3. ciliary nerves
4. vortex veins.

The optic nerve attaches to the sclera approximately 3 mm medial and 1 mm superior to the anatomical posterior pole. As the optic nerve head abuts the globe, the sclera becomes perforated to accommodate the penetration of the nerve fibres. This area of perforated sclera is known as the lamina cribrosa.

The anterior ciliary arteries pierce the sclera at the insertion of each of the rectus muscles. There are two arteries for every rectus muscle, except the lateral rectus, which has only one. The oblique muscles do not contribute to the anterior arterial circle, which is why it is possible to recess an oblique muscle whilst operating on two other rectus muscles without inducing anterior ocular ischaemic syndrome.

The long and short ciliary nerves pierce the sclera next to the optic nerve head and travel anteriorly within the perichoroidal space.

The vortex veins pierce sclera 4 mm posterior to the equator. There are two vortex veins in the superior hemisphere as well as another two within the inferior hemisphere.

Layers of the sclera

The sclera comprises three layers:

1. episclera
2. stroma
3. lamina fusca.

The episclera is the external layer of sclera lying just internal to, and in certain sites connected to, the tenon's capsule. It comprises a vascular network supplied by the anterior ciliary arteries and loose connective tissue.

The scleral stroma is formed by an arrangement of assorted fibrous collagen fibrils and elastic fibres. The collagen tends to be of types I and III and varies in diameter (28–280 μm). Unlike the cornea, it is the randomness of the interweaving of these collagen fibres that is largely responsible for the opacity of the sclera.

The lamina fusca (Latin for brown) is the deepest layer of the sclera. It is so named because of the occurrence of melanocytes along its body. The lamina fusca lies just exterior to the choroid and has numerous weak collagenous attachments between the two.

CLINICAL TIP

Differentiating episcleritis from scleritis

The episcleral vessels are typically constricted by topical phenylephrine but not by topical adrenaline. This may be used to distinguish between episcleritis and scleritis as deep inflamed scleral vessels will not constrict with topical phenylephrine alone.

The lens

See Chapter 2, Lens and accommodation.

Dimensions and growth

The lens is an ever-growing structure within the eye, starting with an equatorial diameter of 6.5 mm and a thickness of around 3 mm at birth. The typical adult human lens measures approximately 10 mm in diameter and 4 mm in thickness in its relaxed state.

The lens is defined as a 'positive' lens by its biconvex shape. It accounts for around 30% of the eye's refractive power (roughly 15 dioptres). The posterior surface of the lens tends to be more convex than the anterior surface, and the circumference where the two convex surfaces meet determines the equator. The apices of each convex surface are known as the poles (anterior and posterior).

The growth of the lens throughout life reduces the anterior radius of curvature, thus increasing its power to converge light. Ordinarily, this would lead to the development of myopia if not for the changes in the lens material leading to a compensatory change in refractive power over time.

Cross-section of the lens

In cross-section, the most exterior layer of the lens is the capsule (Fig. 1.34).

The capsule consists of an elastic basement membrane, which is rich in type IV collagen and is produced by the lens epithelium. Fine filaments arranged in lamellae form the capsule by placing themselves parallel to the surface, adding to its strength. Laminin has also been discovered within the anterior capsule; this is absent in the posterior capsule.

To account for the ever-increasing size of the lens, the synthesis of the anterior capsule also continues throughout life whereas the posterior capsule remains essentially constant. This may explain the increase in anterior convexity later in life.

The capsule becomes relatively thickened either side of the equator, sparing the poles and the equator itself. Conversely, the relative thinning of the capsule is demonstrated clearly in cataract surgery, where the risk of rupturing the posterior capsule is always present.

Lens epithelium

The lens epithelium forms a sheet a few cells thick and resides behind the anterior capsule only. The apices of the lens epithelia face towards the nucleus and their basal aspect lies directly on the capsule. Laterally, adjacent cells interdigitate with each other, forming multiple 'tongue and groove' connections. These connections obliterate any gaps between epithelial cells.

Other important qualities of the lens epithelial cells include large nuclei, but relatively few cytoplasmic organelles.

Regional distinctions between lens epithelial cells

Depending on the location, lens epithelium exhibits different characteristics. The following are the most obvious changes:

- Morphology—centrally, lens epithelial cells are typically cuboidal in shape at the anterior pole but become more columnar on reaching the equator.
- Mitotic activity—cells that lie in the central zone typically show a stable population with a very slow rate of decline. Just outside this area lies an intermediate zone, which contains smaller epithelial cells that display occasional mitotic activity.
- At the pre-equatorial areas, the epithelial cells form a 'germinative' zone. It is at the equator that lens epithelial mitotic activity is at its maximum.

Lens fibres

The lens fibres make up the main mass of the lens. They are formed by multiplication and differentiation of the lens epithelial cells at the equator, which continues throughout life.

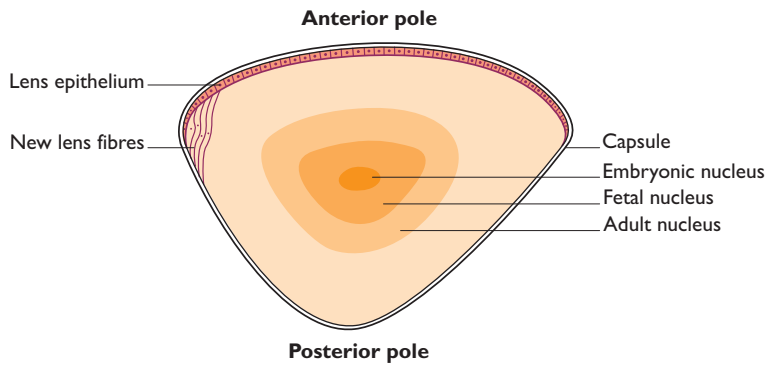


Fig. 1.34 The lens in equatorial view showing the different nuclei within the lens cortex.

With permission from Neil Modi.

The cells begin to elongate along the internal surface of the lens in a posterior direction. As the apical part of the lens epithelial cell elongates posteriorly, it slips beneath the internal surface of adjacent lens cells.

As this happens, the nuclei move anteriorly to form the classical convex nuclear pattern known as the 'lens bow'.

Each elongated lens cell is now called a lens fibre. This is a hexagonal prism in cross-section and is around 10 mm long. The fibres run meridionally from the posterior to the anterior lens surface and are U-shaped. The earliest formed fibres are those in the centre or nucleus of the lens. The outer fibres in the cortex are formed later.

As they bend around the nucleus to towards the anterior pole, the laterally opposing secondary fibres within the same layer connect to form the classical Y suture. The posteriorly elongating fibres curve around the nucleus and join just within the posterior pole to form an inverted Y or lambda (λ)-shaped suture (think L for last)

The lens fibres themselves have a few small vesicles, microfilaments, microtubules, and occasional mitochondrion in their cytoplasm. The fibres are packed tightly with very little intercellular space. There is interlocking of adjacent plasma membranes in the form of ball-and-socket type joints. These interactions are significantly less complicated in the superficial zones of the lens, and this may permit moulding of the lens shape during accommodation.

Lens fibres can exist at a distance from a circulation because of the high number of gap junctions they possess. This makes the lens a syncytium and this is the reason why the lens acts as a single cell.

Zonule (suspensory ligament)

The zonules are individual fibrous structures that attach to the external peri-equatorial areas of the lens in a radial pattern. They are derived from the basal laminae of the non-pigmented ciliary epithelium along both the pars plana and the pars plicata.

Each zonule consists of multiple tightly packed fibrillin filaments which integrate with the lens capsule.

Naturally, any weakening of the zonules may risk the subluxation of the lens. Such conditions include Marfan's syndrome and homocysteinuria.

When viewing distant objects the zonules are under maximal tension to flatten the lens. Conversely, when viewing near objects, the ciliary muscle contraction causes forward and inward movement of the ciliary zonule attachments. This relaxes the zonule tension to allow the lens to resume its natural, more convex shape.

The iris

Gross structure and cross-section

The iris is a contractile, tympanic-like structure with a central aperture—the pupil. The iris forms a wall between the cornea and the lens, separating the aqueous into two compartments: the anterior chamber (between cornea and iris) and the posterior chamber (between iris and lens).

In cross-section, the iris consists of:

- the anterior stroma
- a posterior bilayer of epithelium.

Anterior stroma

The stroma consists of the muscle layer, connective tissue (containing a vascular network and collagen), and nerve fibres.

Looking directly at the iris anteriorly, it is divided into two zones (Fig. 1.35):

1. the pupillary zone.
2. the ciliary zone.

The pupillary zone lies between the ruff and the collarette. The ruff is the absolute inner limit of the pupil and frequently it is possible to see the termination of the posterior pigmented layer of the iris slightly protruding from behind the pupil (a marked protrusion of the pigmented layer is termed ectropion uvea).

The collarette is the thickest structure of the iris and lies around 2 mm peripheral to the pupil margin.

Within the pupillary zone lies the sphincter pupillae. These concentrically arranged autonomic smooth muscle fibres cause pupil constriction. The muscle cells are connected by gap junctions and arranged in bundles that are separated by connective tissue carrying a blood and nerve supply. They are supplied by parasympathetic fibres from the short ciliary nerves derived from the oculomotor nerve.

The ciliary zone continues from the collarette until it reaches the iris root. Within this region lies the dilator pupillae muscle, which, as the name suggests, acts as the antagonist to the sphincter by dilating the pupil. The dilator is composed of myoepithelial cells that, again, are connected to each other by gap junctions. These cells form a thin flat

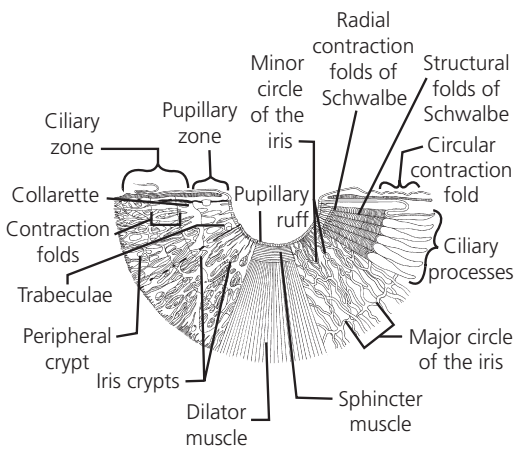


Fig. 1.35 The iris zones, musculature, and blood supply.

Reproduced from Daniel H. Gold and Richard A. Lewis, *Clinical Eye Atlas*, Second Edition, Figure 6.4, Page 396, 2011, with permission from Oxford University Press.

sheet that arises, embryologically, from the basal processes of anterior iris epithelium, posteriorly. The postganglionic sympathetic fibres from the superior cervical sympathetic ganglion make up the efferent supply to the dilator pupillae. These fibres reach the dilator through the long ciliary nerves and implant upon the myoepithelial plasma membrane as unmyelinated fibres.

Posterior bilayer of epithelium

The epithelium is divided into a posterior pigmented layer and an anterior non-pigmented layer of cuboidal cells.

The two layers of iris epithelium lie apex-to-apex and are an extension of the ciliary epithelium and both draw similarities to each other. The pigmented ciliary epithelium is continuous with the anterior non-pigmented iris epithelium.

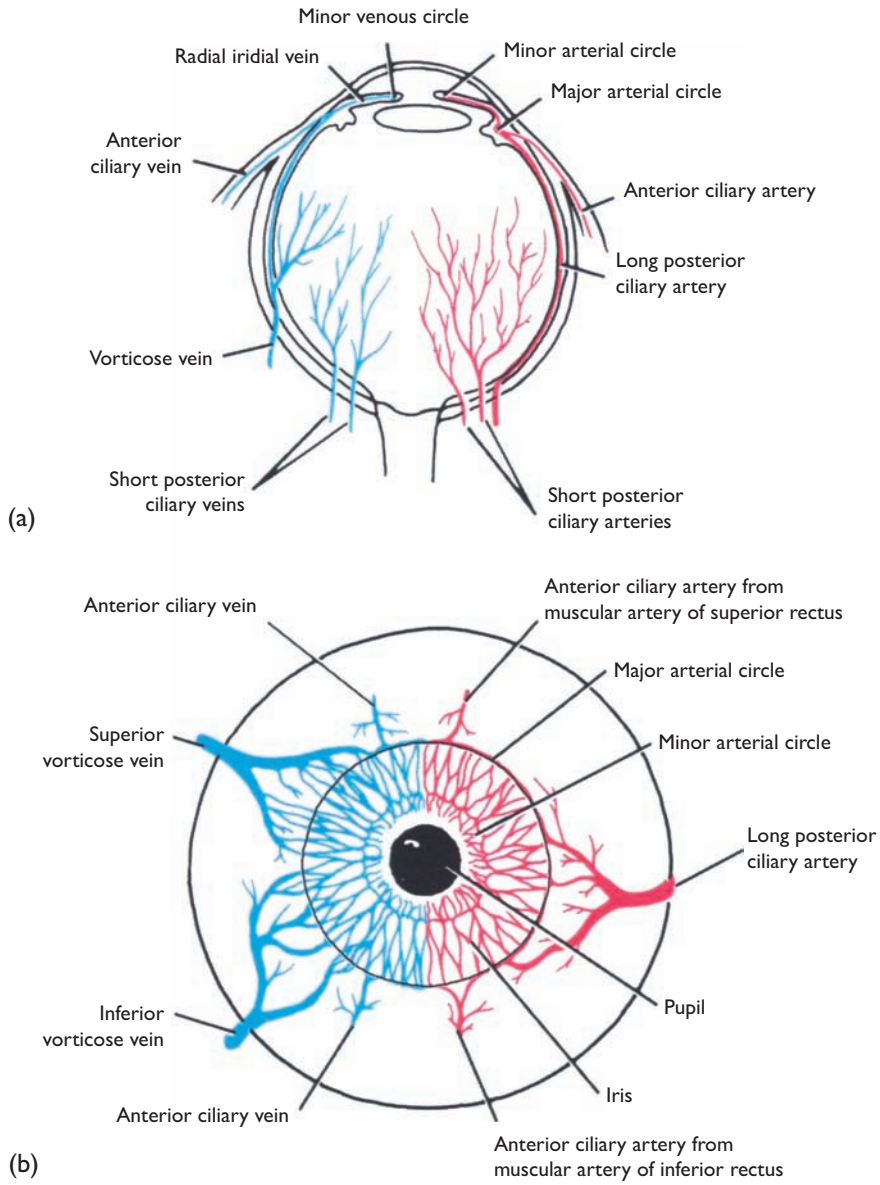


Fig. 1.36 (a) The arterial and venous supply of the eyeball. (b) Summary of the arterial and venous drainage of the iris.

Likewise, the non-pigmented ciliary epithelium is continuous with the posterior pigmented iris epithelium.

Blood supply

The major blood supply to both the iris and the ciliary body is from the major arterial arcade (Figs 1.35 and 1.36). Alongside the two long posterior ciliary arteries, the seven anterior ciliary arteries supply the major arterial arcade—one artery from the lateral rectus and two from the remaining recti. The two long posterior ciliary arteries travel anteriorly through the choroid to anastomose with the major arterial arcade.

Smaller arteries branch off the major arterial arcade in a radial pattern towards the pupil margin. On reaching the pupillary zone, these smaller arteries form an incomplete minor arterial arcade.

Similarly, the venous structure also creates a minor venous arcade. However, instead of forming a major venous arcade, the veins coalesce to eventually flow out of the vorticose (or vortex) veins.

All vessels within the iris are non-fenestrated and do contain tight junctions between the vascular endothelial cells. However, this is a pseudo-blood–ocular barrier as in times of inflammation the vessels will leak protein and other white cells. White cells may diapedese out of these vessels.

The ciliary body (the 'hairy body')

See Chapter 2, Lens and accommodation.

Structure

The ciliary body is a continuation of the uveal tract, the other components of which are the iris and choroid. It performs three main functions:

- aqueous formation *(see Chapter 2, Lens and accommodation.)*
- lens accommodation
- aqueous drainage via the uveoscleral outflow and trabecular system *(see Chapter 2, Lens and accommodation.)*

Macroscopically, the ciliary body is a torus in shape and broadly triangular in cross-section, measuring 6–7 mm in width. The iris projects from the ciliary body base whereas the apex of the ciliary body continues towards the ora serrata. In addition, the base of the ciliary body marks its only anchor to the sclera, where the ciliary body's longitudinal muscle strands insert directly into the scleral spur.

There are three main components to the ciliary body:

1. the stroma
2. the muscle
3. the epithelium.

The ciliary body may also be divided into two areas:

1. pars plana
2. pars plicata.

The pars plicata is highly vascularized in nature. Its rough appearance is created by the ciliary processes (hence the name ciliary). Here, the ciliary epithelium is thrown into multiple folds, maximizing its surface area. The troughs of the ciliary processes are the primary site of attachment for the lens zonules, although some also insert into the pars plana.

The pars plana measures approximately 4 mm and appears as a relatively avascular, pigmented region that lies between the ora serrata and the ciliary processes of the pars plicata. A trans-pars plana approach is the safest point of entry into the vitreous cavity and is estimated to be around 3–4 mm behind the limbus.

Ciliary body epithelium

The ciliary epithelium consists of a bilayer of cuboidal epithelial cells—a pigmented layer and a non-pigmented layer. The apices of both cell layers connect with each other and both play a principal role in aqueous humour formation.

Non-pigmented ciliary body epithelium

The non-pigmented epithelium represents the innermost half of the two layers. Its basal lamina lies in direct contact with the aqueous humour and is continuous with the internal limiting membrane of the retina and the posterior pigmented layer of the iris epithelium.

The morphology of the non-pigmented cells changes depending on location. The cells that are at the pars plana region are predominantly cuboidal, whereas those at the pars plicata tend to be columnar.

The non-pigmented epithelium is abundant in Golgi apparatus and rough endoplasmic reticulum to support its function as a producer of aqueous humour. They also have far less melanosomes compared with the pigmented counterpart. However, both pigmented and non-pigmented cells contain basal infoldings, mitochondria, and large nuclei.

Between the lateral border of adjacent non-epithelial cell apices are numerous tight junctions (zonulae occludentes). These tight junctions constitute one portion of the blood aqueous barrier.

There are differences between the non-pigmented epithelial cells at the tips of the ciliary processes and those between these. The cells at the tips are designed for fluid secretion and those in the valleys are designed for tetherinig zonular fibres.

Pigmented ciliary body epithelium

The basal lamina of the pigmented epithelium is thicker and more homogenous compared to that of the non-pigmented layer and faces the stroma. The basal lamina here is also continuous with the retinal pigment epithelium.

The cells of this layer tend to be cuboidal throughout the pars plana and pars plicata, and have abundant melanosomes.

Ciliary body stroma

The ciliary stroma is a collection of loose connective tissue where the main vascular supply is found. The stroma also houses the embedded ciliary muscle. This connective tissue

runs through each ciliary process to form a connective tissue core. A capillary plexus also runs in each ciliary process, the vascular tone of which plays a role in determining regional blood flow and hydrostatic capillary pressure. This can influence the surface area of the ciliary epithelium and therefore how much area may be available for fluid exchange.

Ciliary body muscle

The origin of the ciliary muscle anteriorly is at the sclera spur by Schlemm's canal, which acts as a fixed anchor against which ciliary muscle contracts. The posterior attachment of the ciliary muscle is to the stroma of the choroid. The anterior and inner surfaces of the ciliary muscle are bounded anteriorly by the stroma of the pars plicata and posteriorly by the pars plana of the ciliary body. Hence contraction of the ciliary muscle causes forward and inward movement of the ciliary body. As the ciliary body muscles contract, the zonules loosen, allowing the lens to retain a more convex shape (which is of a higher refractive order than when it is under tension) and thus accommodation occurs.

The ciliary muscle is within the ciliary body and is composed of three muscle fibre groups. These are orientated longitudinally, radially, and circularly. The largest group is the longitudinal muscle fibres (Brücke's muscle), which are attached between the sclera spur and the choroid. Inward to these are the radial fibres, which are attached anteriorly to the sclera spur and then to the insertion of the iris and the elastic tendons of the choroid. Beneath these, the radial (circular) fibres are positioned more anteriorly in the ciliary body and closest to the lens. The three muscle fibre groups are closely related and contract together to bring about accommodation.

The ciliary muscle receives its innervation from the postganglionic parasympathetic fibres of the oculomotor (CN 3) nerve. These pass through the short ciliary nerves to reach the ciliary body.

Blood supply

The major arterial circle is derived from the long posterior ciliary arteries. The vasculature of the ciliary body is fenestrated (unlike that of the iris).

The choroid

The choroid is the posterior part of the uveal tract.

The choroid lies in between the retinal pigment epithelium (RPE) in the retina and the sclera, and serves to provide the outer third of the retina with nutrients. It is attached to the sclera at two sites: the site of exit of the vortex veins and the optic nerve head, where the choroid becomes continuous with the pia and arachnoid. The choroid is smooth on the inner surface and rough externally.

The choroidal capillaries are most dense at the macula and fovea. The foveal avascular zone defines a region in which there is no retinal arterial supply. This area is supplied exclusively by the choroid.

The choroid may be broken down into two components: Bruch's membrane and a vascular layer.

CLINICAL TIP

In degenerative conditions such as age-related macula degeneration, Bruch's membrane may be breached and new vessels may pass through under the retina from the choroid.

Bruch's membrane

Bruch's membrane is a broadly homogenous structure comprising five layers and measuring 2–4 μm thick. The most internal layer is the basement membrane of the RPE. The most external layer consists of the basement membrane to the choriocapillaris. The middle three layers comprise elastic fibres.

Although the exact role for Bruch's membrane is not clear, it is thought that it may play a part in fluid transport from choroid to retina.

The vascular layer

The vascular layer is subdivided into three layers graded by successive vascular calibre:

- a capillary layer
- a medium vessel layer
- a large vessel layer.

The capillary layer is known as the choriocapillaris and occupies the innermost vascular layer. The choriocapillaris is a specialized continuous network of large (40–60 μm in diameter) capillaries lying in the plane directly external to Bruch's membrane. The distinctive feature of these vessels is their walls, which are thin and have numerous fenestrations, especially on the side that faces the RPE. The vessel walls do exhibit pericytes on their external surface, which helps to contribute to wall integrity. Each capillary displays a wide-bore lumen, which is widest at the macula.

External to the choriocapillaris is the middle layer of small to medium vessels and finally outside this there is a large vessel layer. These medium and large vessels have similarities with small vessels in the rest of the body in that both contain an internal elastic lamina and a smooth muscle media.

Blood supply

The suprachoroidal space occupies the plane between choroid and sclera. It is in this plane that the choroid receives its blood and nerve supply. The two long posterior ciliary arteries, the short posterior ciliary arteries, and the perforating anterior ciliary arteries perfuse the choroid.

The blood leaves the choriocapillaris via complementary choroidal veins, which drain into the vortex veins. This venous blood eventually flows into the ophthalmic vein.

Nerve supply

- The long ciliary nerves carry sensory and sympathetic fibres and are branches of the nasociliary nerve, which are derived from a branch of the ophthalmic division of the trigeminal nerve.

- The short ciliary nerves carry both sympathetic and parasympathetic information and project from the ciliary ganglion.

Both long and short ciliary nerves penetrate the sclera around the optic nerve head and travel anteriorly through the suprachoroidal space.

Summary of functions of the choroid

- Nourishment—as mentioned earlier, the choroid gives nutrition to the posterior third of the retina and carries a blood supply forward to the anterior of the globe.
- Light trap and heat sink—owing to the high melanocyte content, the pigmentation of the choroid is profound and serves as a 'light sink', preventing unwanted rays being reflected back on to the retina.

Blood flow in the choroid is particularly fast, so much so that only around 2–3% of arterial oxygen leaves the arteries. It is thought that this level of flow may help in heat extraction from the globe, thus counteracting the heat generated from absorbing light.

The retina

See Chapter 2, Lens and accommodation.

The retina forms the internal layer of the eyeball. This is the site of photochemical transduction, where nerve impulses are created and transmitted along visual pathways to the brain for higher cortical processing.

The thickness of the retina varies from 0.56 mm near the optic disc to 0.1 mm at the ora serrata. Its thinnest point is at the fovea.

The outer surface of the retina is in contact with Bruch's membrane of the choroid and the inner surface is in contact with the vitreous body. An approximate landmark for the ora serrata (anterior edge/termination of the retina) is the insertion of the medial rectus muscle medially and the lateral rectus muscle laterally.

Anteriorly, the retina becomes continuous with the pigmented and non-pigmented columnar cell layers of the ciliary body.

In cross-section, the retina forms an inner neurosensory layer and an outer pigmented layer (RPE) which is derived from neuroectoderm. There are ten layers to the retina: nine layers of the neurosensory retina coupled with the single RPE layer. The layers are (starting from inside out) (Fig. 1.37):

1. inner limiting membrane (ILM)—formed by the amalgamation of the inner foot processes of Müller cells
2. nerve fibre layer
3. ganglion cell layer
4. inner plexiform layer
5. inner nuclear layer
6. outer plexiform layer
7. outer nuclear layer
8. external limiting membrane (ELM) formed by the amalgamation of the outer foot processes of Müller cells and adherens junctions with the photoreceptor outer segments
9. photoreceptor layer
10. retinal pigment epithelium layer.

Inner limiting membrane

The ILM and the ELM are both layers that are formed by the amalgamation of the inner and outer foot processes of Müller cells, respectively.

The Müller cells themselves are the main neuroglial cell in the retina, are analogous to the oligodendrocytes in the central nervous system, and are derived from neuroectoderm.

Müller cells are arranged radially in the retina and, as alluded to previously, they extend from the ILM as far as the ELM. Their processes as well as Müller cell basement membranes contribute to the ILM.

There are blood vessels that pass within the layers of the neurosensory retina and neither the ELM nor the ILM take any part of the blood ocular barrier.

Other neuroglial cells in the retina include astrocytes, which are less common and reside mostly between the ganglion cell layer and the inner nuclear layer. Both these cells are responsible for creating the fibroglial scars seen when there is an insult to the retina. Gliosis is the mechanism by which the retina heals.

The ganglion cell layer

The ganglion cell layer comprises the bodies of ganglion cells. These ganglion cells serve as the last retinal integrator of information before leaving via the nerve fibres in the nerve fibre layer.

- Parasol ganglion cells or M-cells (taken from the magnocellular layer they project into the lateral geniculate nucleus) integrate information from a large receptive field, namely from several photoreceptors.
- Midget ganglion cells or P-cells (taken from the parvocellular layer, they project into the lateral geniculate nucleus) make connections with only one amacrine cell and one midget bipolar cell. This means that the information sent through this system is often from a single cone, allowing for more detailed information to be sent.
- Midget ganglion cells and midget bipolar cells are more common near the fovea.
- On exit, nerve fibres are ensheathed by glial cells and myelinated only after exiting the globe (hence the conical shape of the optic nerve head). They will next synapse at the lateral geniculate nucleus in the thalamus.

Inner and outer plexiform layers

The characteristic feature of the plexiform layers is the absence or distinct reduction of cell bodies. The major components of any plexiform layer in the retina are either:

- nerve axons or
- cell processes.

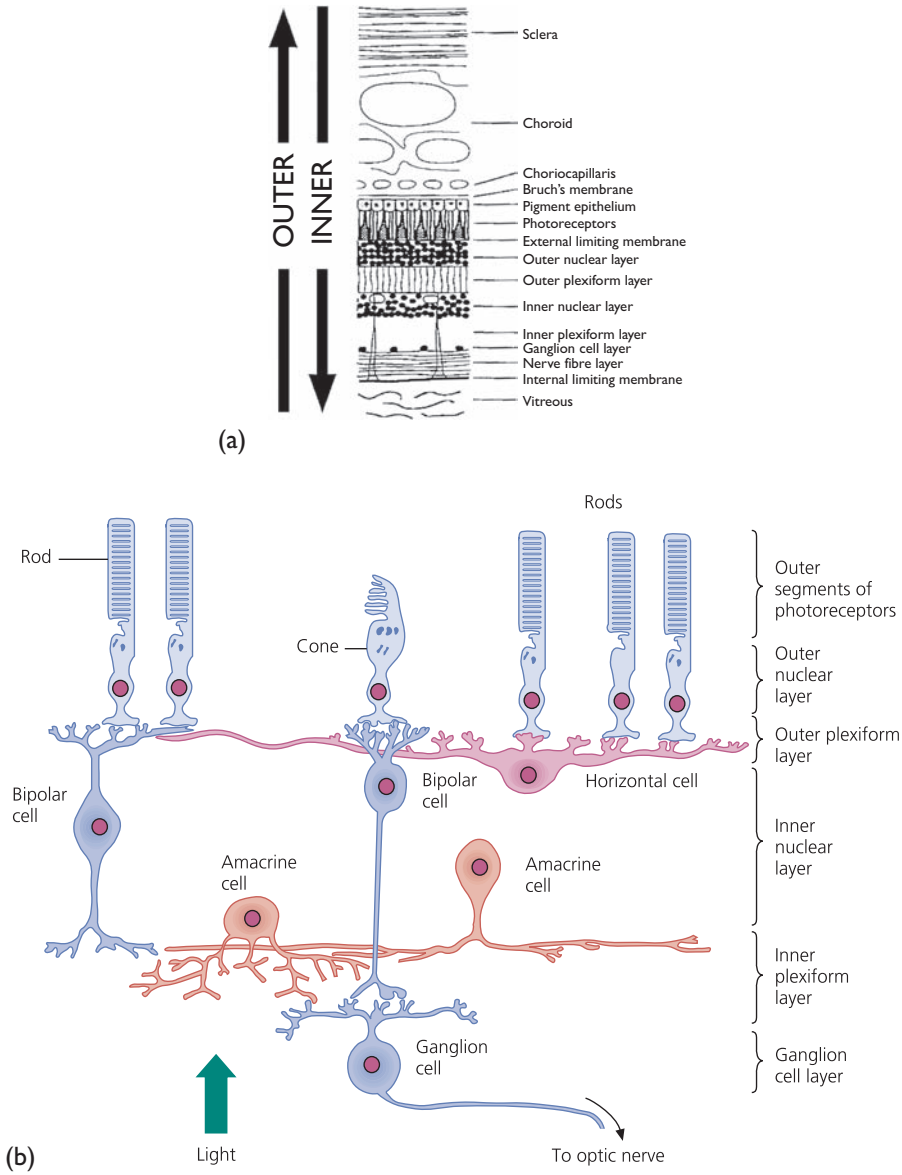


Fig. 1.37 (a) The terms inner and outer are based on a three-dimensional globe with a central inner spot, external to which all points are described as outer. The layers of the retina are inner or superficial (toward the vitreous) and outer or deep (toward the choroid). In cross-section, the 10 retinal layers are internal limiting membrane (ILM); nerve fibre layer (NFL); ganglion cell layer (G); inner plexiform layer (IPL); inner nuclear layer (INL); outer plexiform layer (OPL); outer nuclear layer (ONL); external limiting membrane (ELM); rods and cones (RC); and pigment epithelium (PE). (b) Schematic functional cross-section of neurosensory retina.

(a) Reproduced from Daniel H. Gold and Richard A. Lewis, *Clinical Eye Atlas*, Second Edition, Figure 9.1, Page 636, 2011, with permission from Oxford University Press.
(b) Reproduced from Gillian Pocock and Christopher D. Richards, *The Human Body*, Figure 14.12, Page 240, 2009, published by Oxford University Press, with permission from Gillian Pocock and Christopher D. Richards.

The inner plexiform layer comprises the processes of bipolar, amacrine, and ganglion cells. Naturally, this layer serves to communicate information between the inner nuclear layer and the ganglion cell layer.

Similarly, the outer plexiform layer contains photoreceptor and bipolar axons, and the processes of horizontal cells. These serve as a communication between the outer and inner nuclear layers.

The inner nuclear layer

This layer contains the cell bodies of the bipolar, amacrine, horizontal, and Müller cells.

- Bipolar cells connect in a radial fashion, usually from photoreceptor to synapse with ganglion cells. One or more dendrites from the bipolar cells pass outward to synapse with the photoreceptor cell terminals. The single axon is directed inward to synapse with ganglion and amacrine cells. There are several types of bipolar cell:
  - rod bipolar cells—connect several rod cells to one to four ganglion cells
  - flat or diffuse bipolar cells—connect many cone cells to many ganglion cells
  - midget bipolar cells—connect a single cone cell with a single midget ganglion cell.
- Amacrine cells connect amacrine to ganglion and ganglion to ganglion cells. This ensures that laterally placed ganglion cells are excited. They also modulate photoreceptor signals by their neurotransmitter content.
- Horizontal cells are situated close to the terminal expansions of rods and cones. They are multipolar cells with one long (up to 1 mm in length) and several short processes, which run both horizontally and parallel with the retinal surface. They connect in a horizontal fashion, mostly integrating information between the bipolar–photoreceptor synapses. The horizontal cells associated with cones have short processes that synapse with around seven cone pedicles. The horizontal cells associated with rods have longer processes that synapse with 10–12 rod spherules. These then make contact with bipolar cells some distance away. The horizontal cells release gamma-aminobutyric acid (GABA) in response to stimulation by the photoreceptor. This inhibits the activity of bipolar cells, thus increasing contrast and spatial resolution.
- Müller cells are similar to neuroglial cells. They are long, narrow, and pale-staining. They have long processes that run almost the entire thickness of the neural retina. Branches extend out horizontally and surround and support the neighbouring nerve cells. As well as forming an integral part of the ILM and ELM, these cells make extensive contacts with blood vessels in the retina, taking part in the structure of the blood–ocular barrier.

The outer nuclear layer

The outer nuclear layer contains the nuclei of the photoreceptors. These nuclei are separated from their respective photoreceptor bodies by the ELM.

Photoreceptors

There are two types of photoreceptors: rods and cones (Fig. 1.38). The rods are responsible for vision in dim light and give images in black and white. Cones are adapted for bright light and resolve fine details and colour vision.

There are approximately 120 million rods in the entire retina. There are no rods at the fovea and the number increases until the maximum density occurs in the juxtafoveal zone, about 160,000 cells/mm^2. This density reduces gradually to 30,000 cells/mm^2 in the peripheral retina.

There are around 6–7 million cones in the retina and their density is concentrated in the fovea, which has approximately 160,000 cells/mm^2. This figure declines rapidly to 5000 cells/mm^2 at around 10°s from the fovea.

Rod cells are slender and long (100–120 μm long). It is the outer segment that is the true photoreceptor of the cell and contains the photosensitive pigment, rhodopsin. Each outer segment contains 600 to 1000 transversely arranged membrane-bound 'discs', which are around 2 μm wide, 14 nm thick, and stacked one above another like a pile of coins. Rhodopsin is arranged within these membranes.

The connecting stalk that joins the outer segment to the cell body contains a modified cilium, which consists of about nine doublet microtubules but no central pair. This originates in the basal body found in the inner segment.

The inner segment consists of two areas:

- ellipsoid—next to the connecting stalk
- myoid—towards the vitreous.

The ellipsoid contains the basal body, with numerous mitochondria, and the myoid contains the granular and agranular endoplasmic reticulum, free ribosomes, and a Golgi apparatus.

The rod outer fibre joins the inner segment to the cell body and the inner fibre joins the cell body to the pear-shaped spherule.

Cone cells are also long and slender (65–75 μm long) and have a structure similar to the rods. The differences are that the outer segment is conical and wider than the rod at its base. The membranes of the discs are continuous with the outer cell plasma membrane and so the outer segment in rods is continuous with the extracellular space. The tips of the cone cells are not phagocytosed by the pigment cells. The body of the cone cell is connected at its inner end by the inner fibre to the expanded end, called the cone pedicle.

Macula

The macula lutea (Latin for yellow) is an oval area at the posterior pole of the eye, roughly 3 mm lateral to the optic nerve head. It is responsible for central vision and measures 4.5 mm in diameter. Histologically, it is defined as the area in which the ganglion cell layer is more than one cell thick. Directly within the centre of the macula lies a pit in the contour of the retina, known as the fovea centralis. The fovea measures 1.5 mm in diameter and contains the most dense area of cone cells. Histologically, the fovea is defined as an

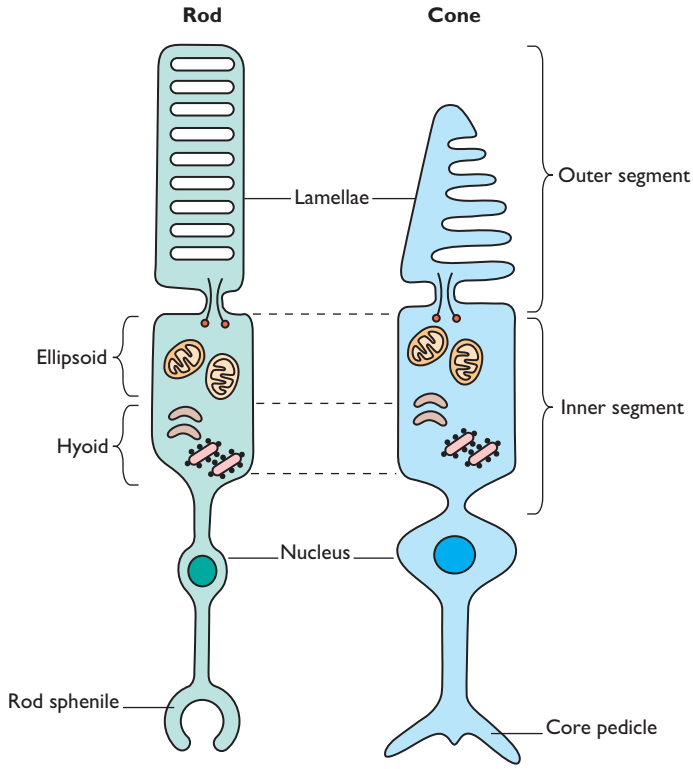


Fig. 1.38 Detailed structure of a rod and cone.
With permission from Louise Bye.

area with an absent ganglion cell layer and a receptor layer composed entirely of cones. In cross-section the central trough of the pit displays only photoreceptor cells, where bipolar cells and their axons are swept aside in a radial fashion; this is known as the foveola.

The fovea contains an avascular zone which measures 500 µm across. This avascular zone can be displayed on fluorescein angiography, and demarcates an area that must be avoided by laser treatment in conditions such as diabetic maculopathy.

The macula is not fully developed at birth, which explains, in part, the imperfect fixation at this stage.

Retinal pigment epithelium

See Chapter 2, Lens and accommodation.

This consists of a simple layer of hexagonal cells that extends from the margin of the optic nerve to the ora serrata anteriorly.

The basal end of each cell is infolded and rests on Bruch's membrane of the choroid.

The apical ends possess microvilli from 5 to 7 µm long, which project between and surround the outer segments of rods and cones, and there are no specialized attachments between them.

The microvilli are surrounded by a matrix of glucosaminoglycans, which act as a 'glue', binding the microvilli to the neural layer.

The basal regions are joined by zonula adherens and the apical cells are joined by zonula occludens, both of which surround the cells and join them tightly together. This is important in isolating the retina from the systemic circulation.

The apical microvilli continuously erode the tips of the photoreceptors and phagocytose the debris. This occurs in a cyclical fashion over a 24-hour period, with most of the sloughing of cells occurring in the morning. Lysosomes play an important role in breaking down the contents of the phagosomes and lipofuscin granules are the end result of this process.

Blood supply

The central retinal artery is the first branch of the ophthalmic artery (0.3 mm diameter). It runs forward adherent to the dural sheath of the optic nerve and enters inferior and medial to the optic nerve, 12 mm posterior to the eyeball. It first pierces the dura and arachnoid, from which it obtains coverings. It then turns 90° to enter the pia mater, acquiring another sheath. It then runs forward again to pierce the eyeball. Here the posterior ciliary arteries form an anastomotic

circle in the sclera around the optic nerve. A number of small anastomoses occur between branches of the posterior ciliary arteries and the central retinal artery. Occasionally a larger connection, known as the cilioretinal artery, exists between the two arterial systems.

The retinal arteries are distributed within the nerve fibre layer close to the internal limiting membrane. These arteries lose their internal elastic lamina after they bifurcate at the optic disc. They function as end arteries supplying all four quadrants of the retina and there are no arteriovenous anastomoses.

The arterioles travel in different layers of the retina, out as far as the inner nuclear layer. They do not anastamose but create capillaries. Non-fenestrated endothelial cells line the capillary walls. Whilst these capillaries are naturally absent in the foveal avascular zone, they are concentrated within the macula.

Those cells external to the outer plexiform layer are supplied by the choroidal circulation. Those cells that are deep to the outer plexiform layer (i.e. the inner two-thirds) are supplied directly by the central retinal artery and its tributaries.

The blood–retinal barrier is maintained because of the zonula occludentes between the non-fenestrated retinal blood vessels and the zonula occludentes between the pigment epithelial cells of the retina.

The central retinal vein comprises tributaries that accompany arteries. Its diameter is a third to a quarter greater than that of the arteries. Arteries travel superficially to the veins and the veins eventually leave the eyeball through the lamina cribrosa accompanied by the central retinal artery. This then drains to the cavernous sinus or superior ophthalmic vein.

There is no lymphatic drainage from the retina.

CHAPTER 2

Ocular physiology

Tear production and lacrimal drainage

Tears and ocular surface

The tear film covers the ocular surface, which includes the cornea and conjunctiva. It is responsible for maintaining the smooth optical properties of the corneal surface, for providing oxygen to the avascular cornea, for lubricating the interface between the lids and the cornea, for the removal of foreign bodies, debris, and cells, and finally for providing antibacterial properties to the cornea.

There are three layers to the tear film:

- deep mucous layer (30–40%)
- middle aqueous layer (60%)
- surface oily layer.

The corneal surface is hydrophobic, and without the layer of mucus it would be unable to spread and maintain the aqueous layer over it for a suitable time. The mucus is not wiped off by blinking, but only thinned in this process.

The mucus layer is made from epithelial cell glycocalyx and a layer of tear mucins (glycoproteins) produced by the conjunctival goblet cells. The goblet cells are mostly under neuroendocrine control. Goblet cell secretion is stimulated by activation by the sensory nerves in the conjunctiva and cornea. These in turn stimulate parasympathetic and sympathetic nerves around the goblet cells. The goblet cells are particularly aggregated in the tarsal conjunctival crypts (Henle's crypts) and on the bulbar conjunctiva nasal to the limbus (Manz's glands). Mucins are also produced by stratified squamous cells of the corneal and conjunctival epithelium. The mucus gives the tear film its viscosity. This property is the result of the structure of mucin with a central protein containing many threonine and serine residues and attached are many O-glycosylated short oligosaccharide chains. This gives mucus its ability to bind with hydrophilic or hydrophobic groups. MUC5AC is the main tear mucin, which is produced along with the trefoil proteins TFF1 and 3.

The aqueous component of the tear film (7–10 μm thick) is produced by the lacrimal gland and its accessory glands. It is 98% water with a solution of electrolytes (particularly K^+ and Cl^-) and protein, including immunoglobin A (IgA), lactoferrin, G protein, tear-specific prealbumin, and lysozyme. Traces of plasma protein are also found. The fluid electrolyte concentration varies with flow rate. At low flow rates, the fluid is hypertonic whereas at higher flow rates it becomes isotonic. The aqueous has antibacterial, antiadhesive, and lubricant properties.

The lacrimal gland is a tubuloacinar exocrine gland, of which the most common cell is the acinar cell, which secretes protein, electrolytes, and water. Basal tear secretion is 1.2 μl/min but can be increased by a number of mechanical and psychological stimuli. Lacrimal gland tear secretion is under autonomic neural control. The majority of the nerves to the lacrimal gland are parasympathetic (acetylcholine and vasointestinal protein (VIP), but there are also sympathetic (noradrenaline) and sensory nerves. The parasympathetic system exerts its effects directly on the acinar cells or the myoepithelial cells surrounding the acinar cells. Nerves and peptide hormones stimulate secretion of electrolytes, water, and regulated proteins. Steroid hormones stimulate secretion of constitutive proteins. Although hormonal control is recognized it is not well understood. Tear secretion decreases after menopause and testosterone increases the secretion of certain components of the tear film, such as IgA.

The accessory lacrimal glands contribute to a lesser degree to the aqueous component of the tear film. Little is known about their electrolyte and water secretion. It has previously been thought they contribute to basal tear secretion but they have good parasympathetic innervations, which does not support this. The corneal epithelium is also a minor contributor to the aqueous layer of the tear film. It secretes only electrolytes and water. Sympathetic nerves are responsible for regulation of this mechanism. Finally, the conjunctiva can secrete electrolytes and water to the aqueous part of the tear film. This is again under the control of the sympathetic system.

The lipid layer (0.2–1 μm thick) is formed from polar and neutral lipids by meibomian gland secretion (holocrine secretion) and is the thinnest layer (0.1 μm). The polar lipids face the aqueous component of the tear film and the non-polar lipids face the air. There are a number of different types of

lipid secreted. Blinking releases the stored material from the ducts. Androgen sex steroids regulate lipid synthesis and secretion, and neurotransmitters from nerves surrounding the acini can alter lipid synthesis or alveolar cell rupture. The function of the oily layer is to prevent the evaporation of tears, to prevent tears spilling over the lid margin, to prevent skin lipids from migrating onto the ocular surface, and to provide a clear ocular medium.

Lacrimal drainage

The lacrimal drainage system starts at the punctae (superior and inferior) at the medial aspects of the eyelids. These punctae are the openings of the canaliculi (10 mm diameter by 0.5–1.0 mm). The canaliculi are lined with stratified squamous epithelium and surrounded by orbicularis muscle. In 90% of individuals, the upper and lower canaliculi join to form the common canaliculus, which then enters the nasolacrimal sac and duct, which are both lined with non-ciliated columnar epithelium. The common canaliculus angles anteriorly, producing a valve-like effect, which prevents retrograde flow from the sac.

About 10–25% of secreted tear volume is normally lost to evaporation. The rest of the tears pass through the nasolacrimal drainage system to the nose. Some of the tears may be absorbed in this system.

The eyelids and medial canthal apparatus have a dynamic 'pumping' effect. During blinking, the deep heads of the pretarsal muscles (Horner's muscles) pull the medial ends medially, shortening the canaliculi. The lacrimal sac is pulled laterally by contraction of the deep heads of preseptal orbicularis muscle. The puncta close and the tears in the canaliculi are forced medially and sucked into the sac. As the deep insertions of the orbicularis muscle relax at the end of each blink the lacrimal fascia and sac wall move medially again, the medial ends of the lids move laterally, the puncta reopens, and the ampullae refill with tears. Drainage of tears for the lacrimal sac to the nasolacrimal duct is influenced by gravity.

Aqueous production and drainage

See Chapter 1.

Histologically the ciliary body (Fig. 2.1) is made of epithelium, stroma, and muscle. It has a role in:

- accommodation
- aqueous humour production
- production of lens zonules
- production of vitreal glycosaminoglycans and collagen
- non-conventional aqueous outflow.

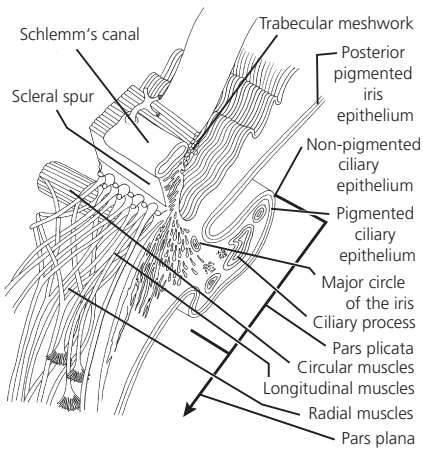


Fig. 2.1 Anatomy of the ciliary body.

Reproduced from Daniel H. Gold and Richard A. Lewis, *Clinical Eye Atlas*, Second Edition, Figure 6.9, Page 400, 2011, with permission from Oxford University Press.

Aqueous humour dynamics

Aqueous humour is specifically actively secreted by the non-pigmented ciliary epithelium in the ciliary processes of the anterior pars plicata. From here it circulates into the posterior chamber and through the pupil and towards the anterior chamber angle (defined as the area where the cornea and iris join). It then drains through the outflow apparatus into the episcleral veins (Fig. 2.2).

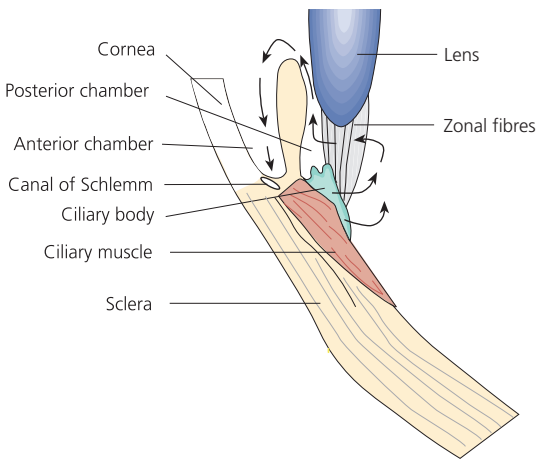


Fig. 2.2 Conventional aqueous outflow.

Reproduced from Gillian Pocock and Christopher D. Richards, *The Human Body*, Figure 14.14, Page 241, 2009, published by Oxford University Press, with permission from Gillian Pocock and Christopher D. Richards.

About 10% of the aqueous fluid also drains through the uveoscleral route. The flow of aqueous humour maintains intraocular pressure—as determined by production and drainage. This is affected by episcleral venous pressure, rate of flow of aqueous humour, circadian rhythm, and neural and hormonal influences.

Aqueous humour is produced at a rate of approximately 2–3 µl/min. Three physiologic processes contribute to the formation of aqueous humour:

- diffusion
- ultrafiltration (and the related dialysis)
- active secretion.

Diffusion and ultrafiltration are passive and hence require no active cellular participation. These two processes form a pool or reservoir of the plasma ultrafiltrate in the stroma, from which the posterior chamber aqueous humour is derived (by active secretion across the ciliary epithelium).

Active secretion (Fig. 2.3) requires energy, provided by the hydrolysis of adenosine triphosphate. Active transport of sodium across the non-pigmented epithelium (NPE) into the posterior chamber results in water movement from the stromal pool into the posterior chamber. Normally, 80–90% of total aqueous humour formation is by active secretion. This process is particularly mediated by two enzymes:

- sodium–potassium-activated adenosine triphosphate (Na^+–K^+-ATPase)
- carbonic anhydrase.

Na^+–K^+-ATPase is found predominantly bound to the plasma membrane of the basolateral infoldings of the NPE. The enzyme provides the energy for the metabolic pump, which transports sodium into the posterior chamber by converting ATP to ADP.

Carbonic anhydrase is responsible for production of HCO_3^- and H^+ from CO_2 and H_2O. Inhibition of HCO_3^- leads to inhibition of the active transport of Na^+ across the NPE into the first-formed aqueous humour, thus reducing aqueous humour formation. This is why carbonic anhydrase inhibitors have their effect on intraocular pressure.

As a result of active transport, aqueous humour has increased levels of ascorbate, some amino acids, and certain ions, such as Cl^-.

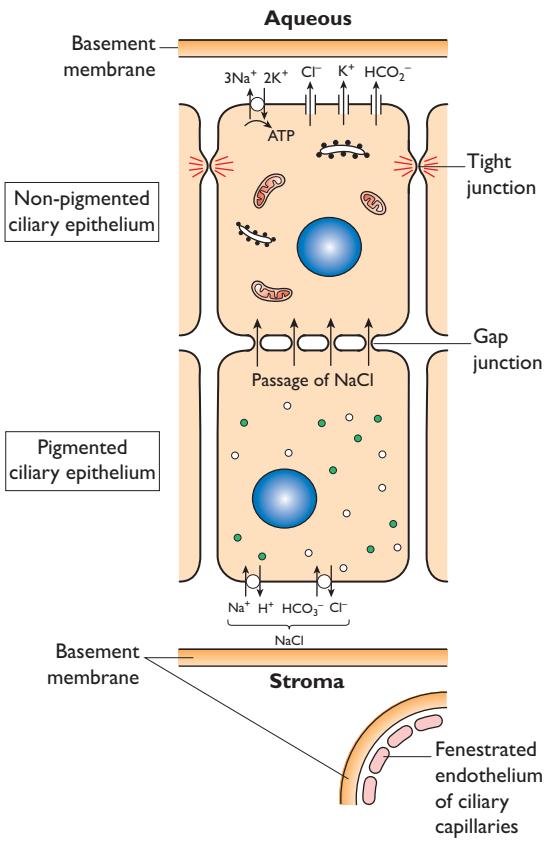


Fig. 2.3 Secretion proceeds in three steps: (1) uptake of NaCl from ciliary body stroma to pigmented epithelial cells by electroneutral transporters, (2) passage of NaCl from pigmented epithelial cells to non-pigmented epithelial cells through gap junctions, and (3) release of Na^+ and Cl^- through Na^+K^+-activated ATPase and Cl^- channels, respectively.

Aqueous humour constitution

Aqueous humour is derived from plasma but has a different composition of electrolytes, small molecules, and proteins. It has different properties to plasma because of the blood–aqueous barrier and because it is secreted by the ciliary epithelium.

The main differences result from the low protein and high ascorbate concentrations (200 times less and 20 times greater, respectively). The ascorbate is thought to reduce oxidative damage from UV radiation. Lactate levels are also elevated due to glycolysis in the cornea, lens, and other ocular structures. Cl^- and certain other amino acids are also in excess in the aqueous (Table 2.1).

Table 2.1 Composition of aqueous humour compared to plasma

Component	Aqueous	Plasma	Units
Na	142	130–145	MEq/L
K	4	3.5–5	
HCO_3	20	24–30	
Ca	1.2	2–2.6	
Cl	131	92–125	
Glucose	2.7–3.9	5.6–6.4	mmol
Lactate	4.5	0.5–0.8	
Ascorbate	1.1	0.04	
Albumin	5.5–6.5	3400	mg/dl
Transferrin	1.3–1.7		
Fibronectin	0.25	29	
IgG	3	1270	

See Chapter 1.

Cornea

The cornea is 80% hydrated (compared to sclera, which is 70% hydrated). Despite this, it readily takes up more water if available. This can be seen easily in corneal lacerations. The high concentration of glycosaminoglycans (GAGs) (lumican in particular) is responsible for this property, which is known as the swelling pressure. The swelling pressure produces an interfibrillar tension that helps to keep the corneal fibrils in their normal arrangement.

To maintain corneal transparency, corneal fibres only vary a little in diameter and the diameter is only a fraction of the wavelength of visible light. Also, the fibres are not arranged in a regular lattice, but in a quasi-random configuration. There is local ordering of fibrils to approximately 200 nm from individual fibrils, which is sufficiently uniform to account for corneal transparency.

The swelling pressure is counteracted by a metabolic pump in the endothelium, which actively removes water from the cornea. As well as this endothelial pump, there are ATP-driven ion pumps. H_2O and CO_2 are pulled into the endothelial cell, where they reversibly form H^+ and HCO_3^-. Na^+ and HCO_3^- are transported into the aqueous by the Na^+/HCO_3^- symport and the HCO_3^-/Cl^- antiport. The H^+ is driven into the stroma by the H^+/Na^+ antiport. In this way, H_2O (by conversion to HCO_3^- and H^+) is removed from the stroma. The net osmotic gradient of Na^+ drives water into the aqueous from the stroma.

The endothelial Na^+–K^+ pump is located on the basolateral membrane of the endothelial cell and is vital for maintaining normal corneal hydration. The basolateral membrane also contains an Na^+–H^+ exchanger, which moves sodium into the cell and hydrogen out of the cell. This makes extracellular fluid more acidic and drives more CO_2 into the cell. Bicarbonate is also important and removal of this from the solution irrigating the endothelium causes flux of bicarbonate from the stroma to the aqueous and corneal swelling. Intracellular bicarbonate is produced by the action of carbonic anhydrase inside the cell and is responsible for the –500 μV charge across the endothelium (aqueous humour negative to stroma) (Fig. 2.4).

Intraocular irrigating solutions therefore need to contain a number of substances to maintain normal function of the various ocular structures. An energy source (e.g. glucose), buffer (e.g. bicarbonate), and a substrate (e.g. calcium and glutathione) maintain junctional stability and the blood–aqueous barrier. The pH should be between 6.7 and 8.1, with osmolarity between 270 and 350 mOsm/kg. Adrenaline is also added to maintain mydriasis.

Corneal cell metabolism

The epithelium derives most of its glucose from the stroma and most of the metabolic demands for glucose, amino acids, vitamins, and other nutrients are supplied to the cornea ultimately by the aqueous humour (via the ciliary body). Additionally, the corneal epithelium contains stores of glycogen. This is converted to glucose-6-phosphate, and 85% of this is metabolized via the glycolytic pathway to pyruvate. The majority of this is converted to lactic acid, but some enters the citric acid cycle to produce ATP. The pentose phosphate pathway accounts for the rest of the glucose utilization, and this helps in dealing with free radical production.

The epithelium obtains oxygen from the atmosphere and tear film at 3.5–4.0 μl/cm^2 per hour. Minor amounts are supplied by the aqueous and limbal vasculature. Tears contain more oxygen than the aqueous. During sleep, oxygen is supplied to the cornea by the palpebral conjunctiva.

Keratocytes are not highly metabolically active and generate energy to support the stroma.

The endothelium, on the other hand, is five times as active as the epithelium. It mainly uses anaerobic glycolysis, but also uses the citric acid cycle and the pentose phosphate pathway. Because of the high activity of these cells, the endothelium uses the citric acid cycle more than epithelial cells do. Both the endothelium and deep stroma obtain their oxygen from the aqueous humour.

During contact lens wear, and resulting hypoxia and acidosis, oxygen consumption increases in the endothelial cells as a result of the activation of pH regulatory mechanisms such as the Na^+–H^+ exchange, which in turn stimulates Na^+–K^+-ATPase activity.

Corneal cell turnover and wound healing

The epithelium constantly turns over by mitotic activity in the basal layer of cells. Most mitotic activity occurs at the limbus, in the stem cells, which can differentiate into basal cells. The mitotic rate is 10–15% per day.

After epithelial damage, the cells slide horizontally to fill the defect. Hemidesmosomes and intercellular contacts reform and eventually the six-layer structure is reformed. Epithelial cells migrate due to the action and redistribution of actin-myosin fibrils, which cause the shape of the cell to change.

Attachment of epithelial cells to basement membrane and Bowman's layer is normally achieved by hemidesmosomes initially and later by the lamina densa and the anchoring type VII collagen fibrils.

During healing:

- protein synthesis by epithelial cells increases
- the focal adhesion component, vinculin, is synthesized *de novo*
- the cell surface glycoprotein CD44 is increased
- glycogen levels fall in migrating cells because cell migration is an energy-dependent process.

The main process used for the generation of energy during healing is anaerobic glycolysis. To provide the necessary energy,

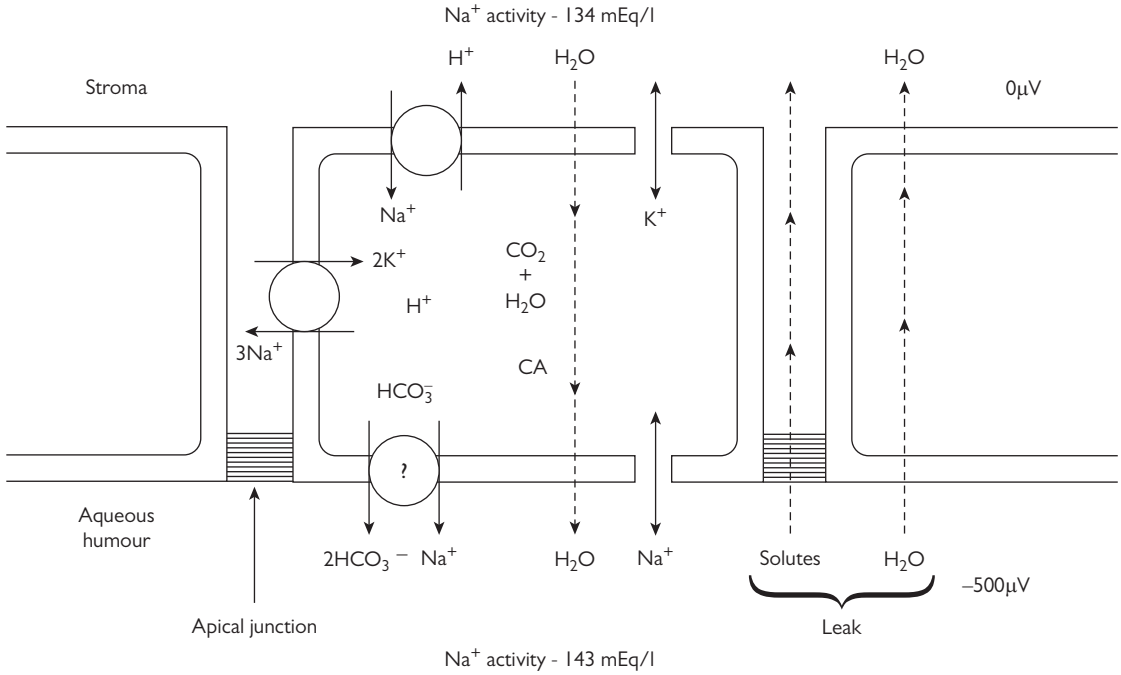


Fig. 2.4 Model of ion and water movements across the corneal endothelium. A metabolic pump sets up an osmotic gradient so that fluid moves from the stroma to the aqueous humour to balance a continuous leak in the other direction.

This figure was published in *Adlers Physiology*, Kaufman PL and Alm A, Figure 4.39 Model of ion and water movements across the corneal endothelium, p. 88, Copyright Elsevier, 2003.

within 2 hours of injury there is increased expression of GLUT1 messenger (mRNA). The resulting protein transporter levels increase within 4 hours, providing an adequate supply of glucose for wound healing.

The control of migration depends on autocrine and paracrine control by peptide growth factors. Synthesis of these factors and their receptors increases after corneal injury.

After injury, the corneal stroma becomes opaque as a result of oedema and disorganization of the regular fibrillar arrangement. Water is attracted by the GAGs.

Healing involves resynthesis and crosslinking of collagen, alteration in proteoglycan synthesis, and gradual wound remodelling, resulting in return of tensile strength.

The wound is closed by deposition of fibrin within the wound and activation of fibroblasts to generate collagen and GAGs. There is also rapid epithelialization of the wound.

Initially, the keratocytes act like unspecialized fibroblasts to lay down collagen and GAGs with an irregular size and arrangement. This results in corneal opacity, particularly in larger wounds. In smaller corneal wounds, attempts are made to restore normal corneal matrix components.

The corneal endothelium does not undergo mitosis, and with age there is a reduction in the number of corneal endothelial cells. The cells' size and morphology changes and they move to fill the gaps left in the endothelium, normally with no clinical consequence. Endothelial cells can move 80–100 μm/day in the initial stages of healing, and during this process they do not lose contact with each other. This is followed by remodelling to their normal hexagonal shape and the pump and barrier function is restored when there is a confluent monolayer of cells.

If sufficient numbers of cells are lost either after surgery or pathologically, the cornea can decompensate due to increased leak where cells cannot fill the gaps and due to decreased ion and water transport out of the cells. The cornea becomes opaque as a result.

Sclera

Most bulk transport of fluid occurs through the anterior chamber drainage angle and the uveoscleral meshwork. There is also some flow of fluid across the retina, most of which drains through the choroidal plexus. Some drains across the sclera, which is absorbed by the sclera proteoglycans. The sclera therefore performs its role by keeping proteoglycans with a low water-binding capacity.

Lens

See Chapter 1.

Lens fibres are joined together at apical (anterior capsule) and basal (posterior capsule) ends. During differentiation fibre cells:

- withdraw from the cell cycle
- elongate
- degrade all membrane-bound organelles
- express proteins (crystallins)
- acquire specializations of their plasma membranes.

Up to 40% of the wet weight of the lens is made up of crystallin protein (three times that of an ordinary cell). Crystallins are members of the heat shock protein family. An important role of these proteins is to stabilize proteins that are partially unfolded and prevent them from aggregating (chaperone activity). This is important as the proteins in lens fibres must persist for life. Excess aggregation could lead to light scattering and cataract formation.

Microtubules stabilize the fibre cell membrane. They may be important in transporting vesicles. There is also an abundant network of actin-containing microfilaments. These associate with the adhesive junctions between cells and the spectrin network within the cell. There is also a network of intermediate filaments, including those composed of vimentin (usually found in cells of mesodermal origin).

During elongation, the lateral fibre membranes are smooth. These become more interdigitated, forming ball and socket junctions that stabilize the lateral membranes of the fibre cells. The cell membranes of the fibre cells have the highest proportion of cholesterol of any plasma membrane, and this increases with age. There is also a high proportion of sphingomyelin. Both cholesterol and sphingomyelin makes the cell membrane more rigid. Lens fibre cell membranes contain some unique proteins, including major intrinsic polypeptide (MIP). This accounts for 50% of the total protein of the lens fibre cell. It is from the aquaporin family of proteins and its role is not well understood, although it is thought to play an important role in the model of fluid and ion flow in the lens.

Gap junctions of the lens are assembled from several subunits (connexins). The lens has the highest concentration of gap junction plaques of any body tissue as a result of the distance of its cells from a source of nutrient.

As well as the ball and socket joints, lens fibres are attached to each other along their length by N-cadherin. This is attached to the actin-containing cytoskeleton. The close attachment of neighbouring cells minimizes extracellular space and therefore light scattering.

The lens grows by adding new fibre layers to its outer surface. These fibres elongate, extending their apical and basal ends towards the sutures. At the sutures, the cells meet a fibre cell coming from the opposite side of the lens. The two fibres join and detach from the posterior capsule. They are gradually buried deeper within the lens by the production of further lens fibres. Soon after detaching from the posterior capsule the fibres lose their organelles, including the mitochondria, endoplasmic reticulum, and nuclei.

The human lens grows rapidly in the womb and first postnatal year but slows between years 1 to 10. It then continues at a much slower, linear rate throughout life.

The role of the epithelium is poorly understood. However, if it is compromised, the viability of the underlying lens fibres is compromised.

Deeper lens fibres lack mitochondria and enzyme systems are less active. There is a fine balance between oxidative damage and diffusion from the more superficial cells. There is a low oxygen tension around the lens, which contributes to protecting the lens from oxidative damage. This does not cause problems as glycolysis is the main source of energy. However, lactic acid is produced, which lowers the pH. This progressively decreases deeper in the lens and as a result the aquaporins do not work as well. Some energy arises from oxidative phosphorylation. Hydrogen peroxide from mitochondria and oxidation of ascorbic acid from the aqueous humour contributes to oxidative stress. Transferrin in the aqueous humour binds iron and prevents it from reacting with hydrogen peroxide and resulting in the release of free radicals. Solar radiation causes harmful damage to the lens fibre cells. Ultraviolet light is the most harmful and energetic, although this is mostly absorbed by the cornea. Interaction between UV light and cellular components results in free radical production. To protect against this free radical damage the tripeptide glutathione, which is present in high concentration in the lens, has a sulfhydryl group that is readily oxidized. Glutathione is produced by the lens fibres and epithelium and is imported from the aqueous humour. Diffusion is the only process by which glutathione reaches the deeper lens fibres, and this diminishes with age. Ascorbic acid can also protect against oxidative stress. It is transported by active transport to the aqueous, where it is present in much higher concentration than in blood. Oxidation of ascorbate by free radicals prevents these molecules from causing damage elsewhere. Catalase is present in the lens, which breaks down hydrogen peroxide to water and oxygen. Glutathione peroxidase is also present, which couples the reduction of hydrogen peroxide to the oxidation of glutathione.

Glycolysis is the main source of energy production in the deeper lens fibres because of the reduced rate of diffusion. The more superficial lens fibres and lens epithelium use glycolysis and oxidative pathways as they contain more mitochondria.

From superficial to deep layers there is an increasing protein concentration. However, water does not follow this protein concentration gradient, possibly because of

the reducing protein osmotic activity deeper in the lens. Electrolytes are thought to flow around the lens. The Na^+–K^+-ATPase is of importance as it draws Na^+ into the cells, leaving an electrochemical gradient across the membrane and causing the positive current at the sutures. The positive current flowing out of the lens at the equator is thought to be due to K^+ ions. Water follows the flow of ions in one model and this is the source of the internal circulation of the lens.

Calcium is actively pumped out of lens fibre cells and their cytoplasmic calcium is lower than that of other cells. Higher levels of calcium lead to degradation of the cytoskeleton, uncontrolled proteolysis, cell swelling, and opacification.

Lens transparency depends on the organization of the cells of the lens and the distribution of proteins within. The high protein concentration causes the lens to have a higher refractive index than the fluid around it. There is a gradient of refractive index as the surface of the lens has a lower refractive index than deeper fibre cells. This partly corrects for spherical aberration. The lens absorbs the shortest wavelengths as a result of the accumulation of chromophores. At birth, the lens is pale yellow and the yellow pigmentation increases with age. Eventually this can lead to so-called brunescent cataracts.

With age, biochemical changes occur in the lens which may result in cataract formation. When the fibres degrade their organelles, protein synthesis stops. Many of the soluble crystallins are increasingly truncated by proteolysis with age. Proteins are found in higher molecular aggregates and are less soluble deeper in the lens, where fibres are older. Importantly α-crystallin becomes less soluble towards the centre of the lens and functioning α-crystallin binds to hydrophobic regions of proteins that have undergone these changes, leaving less α-crystallin available for chaperone duties. Proteolysis and insolubilization of the components of the cytoskeleton result in disassembling, especially of vimentin intermediate filaments.

Vitreous

See Chapter 1.

This is mainly (99%) composed of water. The gel structure of the vitreous is a result of long unbranching collagen fibres (mostly type II) and highly hydrated hyaluronic acid. The gel structure acts as a barrier against the movement of solutes. These solutes move by either diffusion or bulk flow.

Retinal physiology, including phototransduction

See Chapter 1.

Phototransduction

The visual cycle relies on photon absorption by visual pigments, which lie in the photoreceptor outer segments. Rods and the three types of cone photoreceptor cells each express a specific visual pigment with a different spectral sensitivity (Fig. 2.5).

Each visual pigment is formed from a cell-specific apoprotein (opsin) and a chromophore (11-*cis*-retinaldehyde, 11-Ral) derived from vitamin A. The rod visual pigment, rhodopsin (Fig. 2.6), is the most abundant visual pigment and has an absorption spectrum that closely resembles that of the dark-adapted retina (peak at 500 nm).

Rhodopsin (R) is embedded in the flattened disc membranes of rod cells. The opsin protein is composed of 349 amino acids and spans the plasma membrane seven times (Fig. 2.7).

Activation

- Photon absorption triggers isomerization of 11-Ral to the all-*trans* form.
- As a consequence of this change in the chromophore, the whole rhodopsin complex changes to the catalytically active form metarhodopsin II (R*), which results in transducin activation. This process takes 10 ms following photon absorption. The active molecule stores 27 kcal/mol of the energy carried by the photon.
- Each metarhodopsin (R*) can freely move in the lipid bilayer and activate several hundred transducin molecules.
- R* activates G-protein and transducin ($T_{\alpha\beta\gamma}$) by promoting a guanosine diphosphate to guanosine triphosphate (GDP-to-GTP) exchange.
- After this occurs, the active unit of transducin (T_α) dissociates from rhodopsin and the transducin βγ-subunit dimer ($T_{\beta\gamma}$).
- This can then bind to cyclic guanosine monophosphate (cGMP) phosphodiesterase (PDE). The activated T_α binds to the PDE molecule and activates it. Two T_α molecules are required to fully activate PDE, which is not an amplification step.

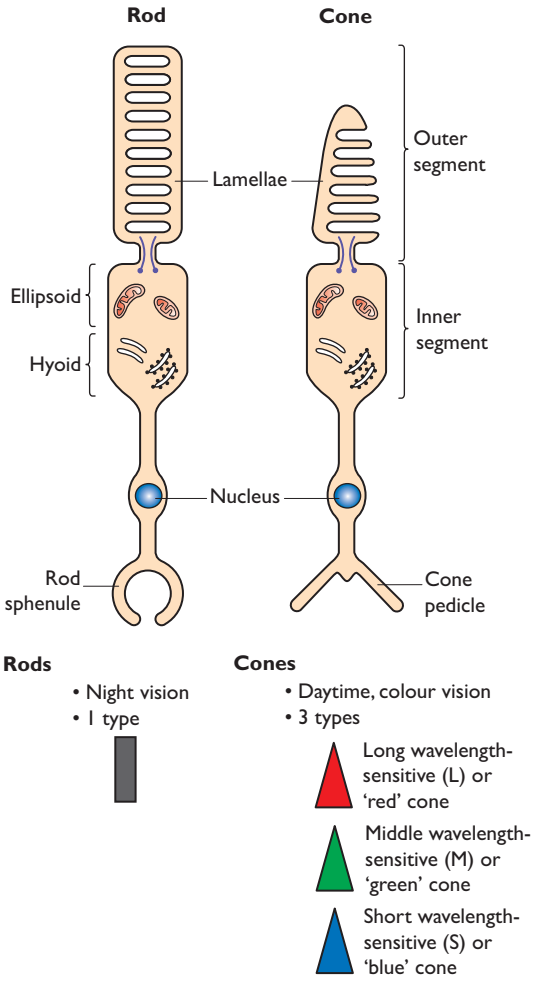


Fig. 2.5 Human photoreceptors.

Adapted from RW Rodieck, *First Steps in Seeing*, with permission from Sinauer. Copyright 1998.

- The active PDE molecule can then hydrolyse cGMP to GMP, which acts as another amplification step (as well as activation of transducin). The resulting decrease in cGMP results in closure of cGMP-dependent cationic (sodium) channels in rod outer segments.
- In the dark, these channels are constantly open, thus inducing a dark current that produces continuous photoreceptor depolarization. After light activation of the phototransduction cascade, cGMP hydrolysis by PDE triggers channel closure. This results in a reduction in dark current and then photoreceptor hyperpolarization. This is a graded response that depends on stimulus light intensity.
- Hyperpolarization causes voltage gated calcium channels to close.
- As calcium levels in the cell drop, the amount of glutamate released by the cell drops—calcium is required for glutamate.

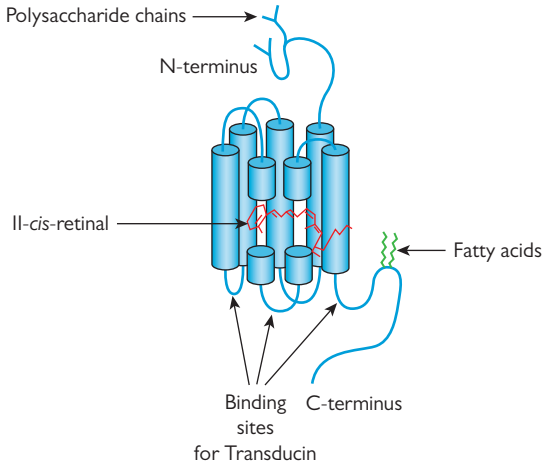


Fig. 2.6 Structure of rhodopsin.

- Decrease in glutamate release causes depolarization of on-center bipolar cells and hyper-polarization of cone off-centre bipolar cells.
- The recovery of the photoresponse requires an increase in cGMP concentration. This depends on the balance between hydrolysis of cGMP by PDE and production of cGMP by retinal guanylate cyclases. Fig. 2.8 demonstrates both activation and deactivation of phototransduction.

Range extension

As a result of the high level of amplification, the system would saturate if there was no modulation in the response to light. Reduction in calcium levels is important as it causes calmodulin to dissociate from cGMP channels, allowing less to escape and increasing the rate of synthesis of cGMP. The visual response time is also faster, which speeds up the integration time of the system.

Deactivation

To prevent ongoing transducin activity, the α subunit of transducin has intrinsic GTPase activity, which cleaves its bound GTP to GDP. This halts the action of transducin. Also, the activated rhodopsin molecule is inactivated through phosphorylation by rhodopsin kinase which is produced as a result of falling calcium levels. Under high free calcium levels, soluble rhodopsin kinase associates with membrane-bound recoverin and loses its activity, thus protecting rhodopsin from deactivation. Recoverin and high calcium levels thus prolong the lifespan of activated rhodopsin by preventing its phosphorylation.

Phosphorylation of rhodopsin greatly increases the affinity of rhodopsin for arrestin. Thus arrestin competes with transducin for phosphorylated inactivated rhodopsin. Binding with arrestin speeds up the recovery back to inactivated rhodopsin.

In high Ca^{2+} (dark adaptation), phosducin phosphorylation increases. This also facilitates transducin regeneration. Free intracellular calcium falls on illumination, which is important for light adaptation and for recovery from a light stimulus.

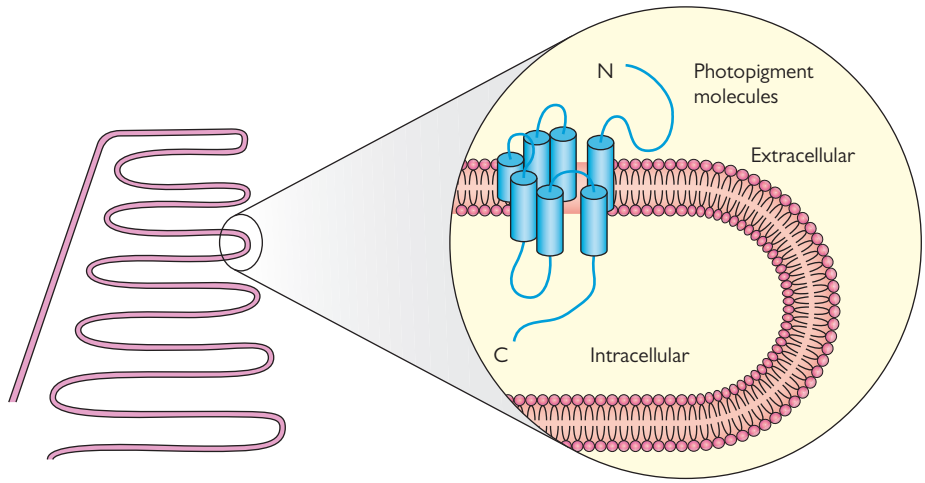


Fig. 2.7 Structure of rhodopsin.

Figure 1.2 from Lindsay T. Sharpe, Andrew Stockman, Herbert Jägle and Jeremy Nathans, 'Opsin genes, cone photopigments, colour vision and colourblindness' in Karl Gergenfurtner and Lindsay T. Sharpe (eds), *Colour Vision* (2000), with permission from Cambridge University Press.

Finally, all-*trans*-retinal is transported back to pigment epithelial cells to all-*trans*-retinol, the precursor to 11-*cis*-retinal. This is then transported back to the rods.

The chromophore cycle

A chromophore is the moiety that causes a conformational change of a molecule when hit by light. When light is absorbed in a photoreceptor, the visual pigment (rhodopsin—composed of an opsin and a chromophore) is degraded to form activated rhodopsin, as described above. In mammals, the chromophore of visual pigments is 11-*cis*-retinaldehyde. This chromophore is converted to all-*trans*-retinal (*t*-Ral), which separates from the opsin and is shuttled across from the luminal side to the cytoplasmic side by a member of the ATP-binding cassette (ABCR) family. *t*-Ral is reduced to all-*trans*-retinol (*t*-Rol) by all-*trans*-retinol dehydrogenase (RDH). This step is thought to be the limiting step in the visual pigment cycle because *t*-Ral is the only intermediate to accumulate following 40% rhodopsin bleaching in mice.

t-Rol can then diffuse to the subretinal space, where it may bind to the retinoid-binding sites of the interphotoreceptor retinoid-binding protein (IRBP).

t-Rol then enters retinal pigment epithelium (RPE) cells at their apical membrane. *t*-Rol (also known as vitamin A) can also enter RPE cells through their basolateral membrane from the bloodstream. *t*-Rol circulates in blood bound to retinol-binding protein (RBP), which is complexed to the larger transthyretin (TTR).

Once inside the RPE cell, however, the retinol becomes bound to a small carrier protein, the cellular retinol-binding protein (CRBP). The *t*-Rol is esterified by lecithin/retinol acetyltransferase (LRAT) to all-*trans*-retinyl ester (*t*-RE). *t*-RE can be stored or immediately converted by an isomerohydrolase to 11-*cis*-retinol (11-Rol), a free fatty acid. The RPE-65 protein, which is involved in a variety of retinal diseases, is thought to be a component of isomerohydrolase. The 11-Rol may be re-esterified and stored (by LRAT) or oxidized to 11-*cis*-retinal (by 11-*cis*-retinol dehydrogenase). This can then travel to the retinal membrane of the RPE cell, where it is released. This is then transported back to the photoreceptor outer segment.

The visual cycle is shown in Fig. 2.9.

Cone visual pigments

The red and green opsins have close absorption spectra and high homology. However, this is not shared with the blue opsin. Most vertebrates have only blue and green cone opsins. The red cone opsin is thought to have originated from a more recent duplication. These cone opsins are different from rhodopsin, but nevertheless have a similar structure with seven transmembrane helices.

The biochemical events of the visual pigment cycle appear identical in cones to those previously described in rods.

In cone cells, the photopigment is not stored in membrane discs but in invaginations of the cytoplasmic membrane. In cones, although the chromophore is 11-*cis*-retinal, the opsins differ such that the photopigments have maximal sensitivity to blue, green, and red light.

The light absorbed by the photoreceptor outer segments changes them such that every morning the rod outer segments have to shed their tips, which are then degraded by RPE phagocytosis. Replacement protein and other components are spread to the outer segments by the cilium and new discs are formed at the invaginations at the base of the outer segments.

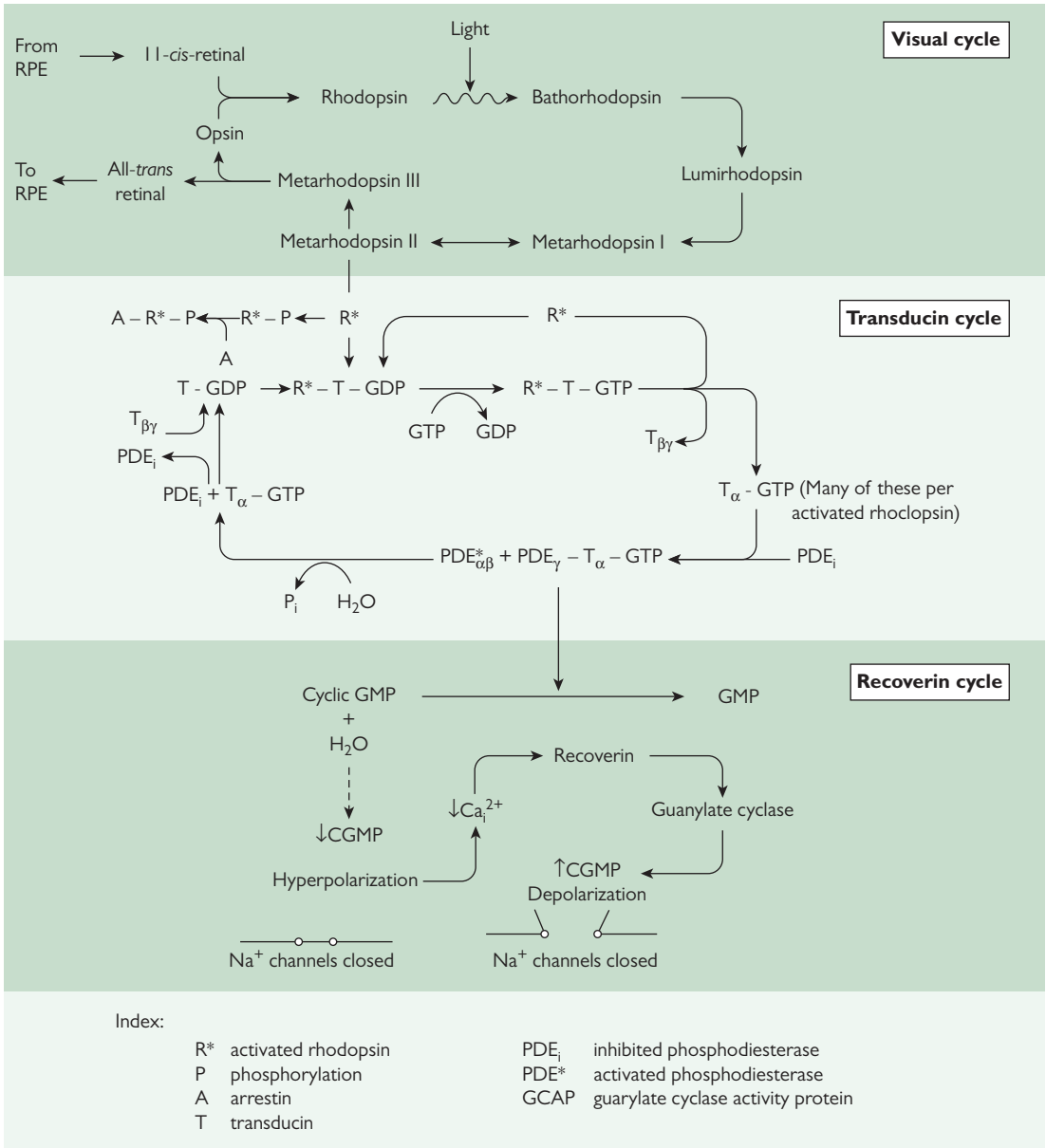


Fig. 2.8 Activation and in activation steps of the phototransduction cascade.

Reprinted from *Current Opinion in Neurobiology*, 9, 4, E.N. Pugh et al., 'Molecular mechanisms of vertebrate photoreceptor light adaptation', p. 411, Copyright (1999), with permission from Elsevier.

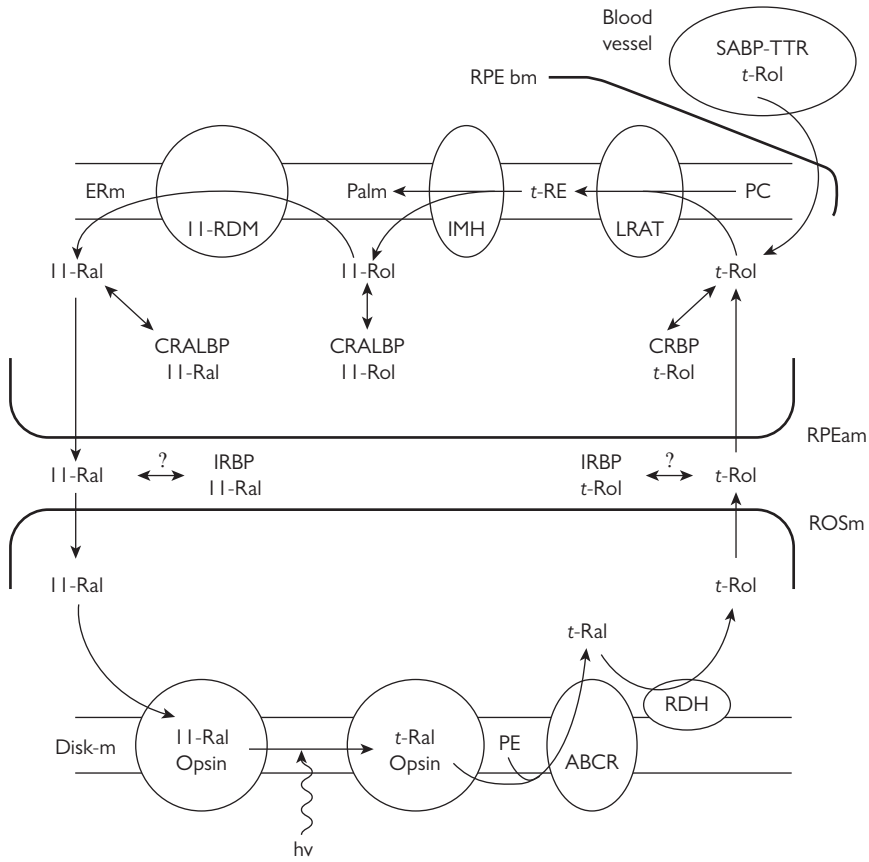


Fig. 2.9 Chromophore cycle flow diagram.

This figure was published in *Adlers Physiology*, Kaufman PL and Alm A, Figure 13.5 Visual cycle flow diagram, p. 386, Copyright Elsevier, 2003.

HELPFUL HINT

Cones vs rods—a comparison

Cones have different isoforms of visual pigment: transducin, arrestin, phosphodiesterase (PDE6), cGMP channel, and recoverin.

Quantitative differences:

- R* four times faster than rods
- R* decays 10–50 times faster (less amplification)
- GTPase activating protein expressed at much higher levels—shorter activated transducin time, faster recovery
- clearance of Ca^{2+} faster than in rods

Functional differences:

- cones are 25–100 time less sensitive to a single photon
- cones catch fewer photons
- cones have faster kinetics
- cones have greater ability to adapt to background light
- cones do not saturate at normal environmental light levels.

Information processing

After phototransduction by photoreceptors, information such as motion and contrast are elicited. These are coded at ganglion cell level in rates of action potential. For a particular stimulation, different ganglion cells produce different responses. This depends partly on the level of dendritic arborization in the inner plexiform layer (IPL), which in turn depends on the various configurations of bipolar and horizontal cells. As a result, the graded hyperpolarization of the photoreceptor is converted to a phasic ganglion cell activity.

Photoreceptors and bipolar cells mainly produce graded potential responses in the form of glutamate. Photoreceptors and OFF bipolar cells continuously release glutamate. Light stimuli are mediated by a decrease in glutamate concentration in the synaptic cleft. Glutamate release is mediated by exocytosis or vesicular release from specialized synapses called ribbon synapses. These exhibit an electron-dense bar that extends into the cytoplasm perpendicular to the plasma membrane. They also have a higher rate of neurotransmitter release than at a conventional synapse. Endocytosis occurs at a similar rate to regenerate synaptic vesicles. Both endocytosis and exocytosis are regulated by Ca^{2+} concentration. Increase in Ca^{2+} causes exocytosis, whereas a decrease causes more endocytosis.

Glutamate ionotropic receptors are all gating cation channels and result in a sign-preserving synapse.

By contrast, ON bipolar cells receive their information at a glutaminergic sign-inversing synapse, which relies on a metabotropic glutamate receptor, whose activation causes the closing of a cation channel (by comparison to its opening in ionotropic receptors). The ON bipolar cells hyperpolarize in the dark when photoreceptors tonically release glutamate.

Horizontal cells and most amacrine cells are thought to release GABA or glycine, which act as inhibitory transmitters.

Retinal metabolism

Glucose metabolism is the primary source of energy in the retina, as in the rest of the central nervous system. This starts with glycolysis, which provides two ATP molecules and two pyruvate molecules, and reduces two oxidized forms of nicotinamideadenine dinucleotide (NAD^+) molecules to the reduced form of NAD (NADH). Regeneration of NAD^+ can be achieved by conversion to lactate (in anaerobic conditions) or by entering the aerobic mitochondrial respiratory chain (in aerobic conditions).

Under aerobic conditions, the pyruvate molecules enter the tricarboxylic acid (TCA) cycle, which requires mitochondrial respiratory change and oxygen consumption. This generates an additional 36 molecules of ATP for each glucose molecule.

The retina can swap between glycolysis and oxidative metabolism according to its ATP needs and oxygen levels. Under aerobic conditions, it generates 85% of its oxygen needs by oxidative metabolism. Under anaerobic conditions, the retina can produce 75% of its normal ATP by glycolysis.

In the outer retina, beside the choroid, 80% of glucose results in lactate formation. Light decreases glucose consumption here, by reducing lactate. In the inner retina, 70% of the glucose is oxidized and the metabolic pathways are not affected by light.

Glucose can also follow the pentose phosphate pathway, which is particularly active in rod outer segments (ROS) for rhodopsin regeneration and ROS protection from oxidative damage.

Glucose enters the retina through the RPE and retinal blood vessels. Facilitated diffusion occurs through the RPE lying over the fenestrated vessels of the underlying choroid and in the endothelial cells of retinal blood vessels. This is due to the GLUT1 transporter.

After entering retinal cells, glucose can be stored as glycogen in Müller glial cells and also some retinal neurons, including rod bipolar cells, amacrine cells, and cone photoreceptors.

Lactate metabolism is another important source of energy. Retinal neurons prefer lactate as a source of energy, rather than glucose. Müller cells use the process of glycolysis to provide most of their energy and as a result they produce lactate, which may be used by the retinal neurons.

In the rod outer segment, a high energy demand is required to maintain the phototransduction cascade. The processes of importance include:

- rhodopsin phosphorylation
- transducin GTPase
- cGMP synthesis
- ABCR retinal transport
- Na^+ extrusion by Na^+-K^+ pump (Na^+ enters through the cGMP-dependent channels together with Ca^{2+}).

Mitochondria provide most of the energy by aerobic glucose metabolism. They are absent from ROS but are packed densely in the inner segments of the photoreceptor. The diffusion of ATP from the inner to the outer segments is slowed at the connecting cilium, which is a problem given the speed at which phototransduction occurs. A phosphocreatine shuttle facilitates the transfer of energy from the inner to the outer segment. This involves a creatine kinase in the ROS that transfers high-energy phosphate groups from creatine phosphate (the concentration of which is light dependent) to ATP.

Photoreceptors can obtain most of their energy by oxidative metabolism, but can shift to glycolysis under aerobic conditions. Glucose enters the ROS directly by the glucose transporter GlUT1. Here, all the enzymes of the glycolytic pathway are present. Glycolysis can provide for basal cGMP synthesis and Na^+ extrusion. However, to recover from illumination (where cGMP synthesis increases), the excess energy is provided by oxidative metabolism.

Rhodopsin regeneration occurs after *t*-Ral is converted to *t*-Rol. This is achieved by retinol dehydrogenase found in ROS, using redox potential from the reduced form of NAD phosphate (NADPH). The pentose phosphate pathway is thought to provide a source of NADPH in the ROS.

Oxygen consumption of the retina is higher per gram than in the brain. Photoreceptors are the main energy-consuming cells. In vascularized retina, the highest oxygen tension is close to the choroid vasculature. There is also a peak in the oxygen tension around the inner limiting membrane due to the retinal blood vessels. Oxygen is mainly consumed in the photoreceptor inner segments, the outer plexiform layer (OPL) and the outer part of the IPL, where OFF bipolar cells have their terminal endings. These sites in the IPL and OPL represent the locations of synaptic terminals from neurons depolarized in the dark. The energy demand is a result of the maintained Ca^{2+} influx (resulting in activation of Ca^{2+} exchanger) in the depolarized terminals and the continued glutamate release that triggers sustained activation of postsynaptic glutamate receptors. These all generate Na^{+} influx, which must be balanced by a pump, which requires energy. The high flow rate of the choroid is well suited to providing oxygen to the inner segments.

Functional organization of the retina

The visual system conveys information to the brain through a series of neurons. Photons of light are converted to a membrane-resistant change. The simplest chain of cells involves a bipolar cell followed by a ganglion cell. These take information to the lateral geniculate body in the brain. The optic tracts carry this information to the brain along the optic radiations. The signals are modified at each synapse. At the first synapse of the retina (OPL), static information is processed. However, phasic information (movement) is processed at the second synapse (IPL). In the rod pathway, an amacrine cell is found between the bipolar and ganglion cells.

There are separate channels for detecting the turning on and off of light. These are called the ON and OFF channels.

All photoreceptor cells hyperpolarize in response to light. This hyperpolarization is passed along unchanged in the OFF bipolar cells (sign-conserving synapse) or as a depolarization in the ON bipolar cells (sign-inverting synapse). In cone pathways, these are apparent at the OPL, where cone cells connect to two types of bipolar cells that have either 'flat' (OFF) or 'invaginating' (ON) junctions. These cells have endings in either sublamina a (OFF) or b (ON), respectively.

Rods are only connected to depolarizing bipolar cells (ON). In the scotopic visual pathway, the ON/OFF division appears when the signal is passed on to the ON and OFF cone channels via amacrine cells. These amacrine cells receive impulses from the rod bipolar cells and transmit via a chemical synapse to OFF bipolar cells and via an electrical synapse (gap junction) to an ON bipolar cell.

Other important characteristics include the resolution of fine detail and detection of general background illumination. This is achieved by the P (parasol) and M (midget) cells. The P system codes for colour and the M system responds to a range of other stimuli. Both systems have ON and OFF channels. For colour opponency, ON and OFF channels have been identified for red and green cones, but for blue-sensitive cone cells there has not been any OFF ganglion cell channel identified. Other parallel signalling systems include direction sensitivity.

Interactions between these various parallel channels in the retina result in centre-surround organization of the receptive fields of bipolar cells and many of the ganglion cells. The central area of the receptive field of bipolar cells is surrounded by an annular area, which has an opposing response. Central responses are usually dominating. In this mechanism, the central responses in the P cells area are generated by the direct pathway of photoreceptor cells, bipolar cells, and ganglion cells. The antagonistic effect of surround illumination is generated in the OPL by the inhibitory feedback of horizontal cells on photoreceptor cells.

Rod photoreceptor pathways

The rod photoreceptor pathway is concerned with vision in dim light (scotopic). Rods outnumber cones by a factor of 10 to 20. There is a high degree of convergence, but the rod pathway neurons outnumber cone pathway neurons everywhere except at the fovea. This convergence increases the sensitivity of the system at the expense of resolution. Approximately 75,000 rod photoreceptors drive 5000 rod bipolar cells and then 250 amacrine cells before converging to a single large ganglion cell.

Only ON-centre depolarizing-type bipolar cells make connections with rod photoreceptors. This is mediated by a metabotropic glutamate receptor. Each bipolar cell reaches 15 to 80 rod spherules. They send their processes to the IPL, where they terminate in the sublayers closest to the ganglion cell bodies. Here they make ribbon synapses that include two amacrine cells:

- all amacrine cells (ON-centre) and
- indoleamine-accumulating amacrine cell type.

The system diverges here, but several bipolar cells can synapse with one amacrine cell so there is also some convergence. Usually there is one of each of the two types of amacrine cell.

First, the A II-amacrine cell is a small field type that makes gap junctions with cone bipolar cells and other amacrine cells in sublamina b (30% of its input). Most of its conventional chemical synapses are to centre-hyperpolarizing (OFF-centre) ganglion cells whose dendrites lie in sublamina a. Hence the ON input from rod bipolar cells drives OFF-centre ganglion cells directly. The ON input from rod bipolar cells can drive the ON-centre ganglion cells through their gap junctions with cone bipolar cells that contact ON-centre ganglion cells.

The indoleamine-accumulating amacrine cell is a wide-field amacrine cell that usually makes synapses with the rod bipolar cell terminal in sublamina b. It contains up to 1000 rod bipolar cell terminals and therefore is sensitive at low light intensities.

Cone photoreceptor pathways

Cone photoreceptor cells contact bipolar cells, which contact ganglion cells. This is a three-neuron chain through the retina. There are three types of cones, sensitive to blue, green, and red light. The pathways for blue cones are different to those for red or green.

Cone cells can induce hyperpolarization or depolarization. They therefore come in two types depending on how they respond: OFF (centre hyperpolarizing) or ON (centre hyperpolarizing). Different types of glutamate receptors (ionotropic and metabotropic) are responsible for the two response types. The OFF receptors are ionotropic and sign conserving (OFF/centre-hyperpolarizing). The OFF-centre bipolar cells have their terminals in sublamina a, where OFF-centre ganglion cell processes occur.

The ON receptors are metabotropic and occur in bipolar cells that contact photoreceptors in triad synapses in rod and cone cells. This is a sign-inverting synapse, and the bipolar cells are of sign-inverting (ON-centre depolarizing) types. These ON-centre bipolar cells have their terminals in sublamina b, where ON-ganglion cell processes occur.

To obtain a high resolution, foveal cone cells connect to a system of small bipolar and ganglion cells, forming the midget cell system. There is minimal/absent convergence in the centre of the fovea. This small degree of convergence is maintained by midget ganglion cells, each of which projects to an individual parvocellular layer cell of the lateral geniculate body.

Each foveal cone cell joins to an ON and an OFF bipolar cell and hence to an ON and an OFF midget ganglion cell. Each foveal midget path carries information for red or green colour.

In monkey retinas, colour opponency chromatic centre-surround organization occurs where, if the centre is a red-sensitive ON response, there is an inhibitory surround from green-sensitive cone cells. There may be a similar colour opponency with blue and yellow.

Retinal pigment epithelium

See Chapter 1.

The functions of the RPE are repeated in this section as they are the intended result of the biochemical processes:

- photoreceptor renewal
- retinal attachment
- interphotoreceptor matrix (IPM) production
- transport of water and metabolites
- retinoid metabolism
- blood–retinal barrier
- immunoregulation
- free radical scavenging.

The RPE is a monolayer of hexagonal cuboidal cells. The nucleus and mitochondria are located in the basal part of the cell. Pigment granules in the apical cytoplasm give the epithelium a black appearance. There is no mitosis in these cells.

The RPE cells are bound together by junctional complexes with prominent tight junctions. The tight junctions limit diffusion of water-soluble substances between the subretinal space and the intracellular space of the choriocapillaris. The subretinal space is bound by the tight junctions of the RPE and the external limiting membrane of the neurosensory retina.

The RPE has a role in transport as it is positioned between the choriocapillaris and the avascular outer retina. There are tight junctions between the RPE cells so transport occurs through the RPE cells.

The retinal membrane of the RPE incorporates an Na^+-K^+ pump, which pumps Na^+ out of and K^+ into the cell using ATP. There is a gradient from outside to inside of Na^+. This is used to drive in other ions through secondary active transport. There are three such transport systems:

1. Na^+-K^+-$2Cl^-$ co-transport
2. Na^+-H^+ exchange mechanism
3. Na^+-$2HCO_3^-$—this accumulates Cl^- and HCO_3^- intracellularly from the retina and these can then exit through the choroidal membrane.

The RPE has been shown to absorb water, which is consistent with the finding that fluid under a rhegmatogenous retinal detachment is absorbed once the holes in the neurosensory retina have been closed. The rate of fluid reabsorption is up to 2 ml per 24 hours or more than 50% aqueous secretion in that period. This property may have a role in maintaining normal retinal adhesion by exerting suction on the neurosensory retina.

The IPM is composed of glycoproteins and proteoglycans (traditional extracellular matrix proteins may be absent, e.g. collagen). IRBP is the most abundant glycoprotein. Other important RPE proteins include RPE65, CRALBP (RLBP1), TIMP3, bestrophin (Best's disease), mertk, myosin VIIa (Usher's syndrome), and OA1(ocular albinism). Light evokes changes in the IPM molecular distribution.

RPE is important in the uptake, storage, and metabolism of vitamin A and related compounds (retinoids), which are important for the visual cycle as described above.

Exposure to light affects the RPE because of the necessity to break down photoreceptors, and as a result the composition of the subretinal space is affected.

In dark-adapted retinas there is a high rate of Na^+-K^+ pumping in the inner segments of these cells. Light onset causes reduction in the dark current and a decrease in the

rate of Na^+-K^+ pumping and metabolism. In the subretinal space, light onset results in a decrease in K^+ concentration, an increase in pH, and a decrease in lactate concentration. The decrease in extracellular K^+ concentration causes a hyperpolarization of membrane potential across the predominantly K^+ permeable retinal membrane of the RPE.

Müller cells

Müller cells, the main glial cells of the retina, extend through the whole thickness of the retina. They have many fine processes, which cover most surfaces of the neuron cell bodies in the nuclear layers. In the plexiform layers, Müller cells cover the dendritic processes of the neurons to the synaptic clefts, thus insulating them both electrically and chemically. In the nerve fibre layer, Müller cells cover most ganglion cell axons. Also, blood vessels within the retina are covered by Müller cells. The blood vessels on the vitreous side are covered by astrocytes, a second class of glial cells.

As well as structural support, Müller cells regulate the extracellular environment of the retina by buffering the light-evoked variations of K^+ concentrations, in particular in the extracellular space. They also take up glutamate from the extracellular space. Müller cells perform duties in the retina analogous to the oligodendroglia and astrocytes in the brain, for example they store glycogen and provide glucose to retinal neurons.

Uveal tract

See Chapter 1.

This includes the iris, ciliary body, and choroid, and is a continuous structure.

The iris

See The pupil, p. 74.

The ciliary body

The ciliary body provides the neurovascular supply to the anterior segment. It also constitutes a part of the blood–aqueous barrier and is responsible for accommodation.

Blood to the cililary body arrives from the long posterior ciliary arteries and the major iris circle. Seven per cent of total ocular blood flow travels through the ciliary body. However, the blood vessels do have fenestrations and they leak most of their contents into the stroma. The barrier between the ciliary body stroma and the anterior segment of the eye is formed by tight junctions between the non-pigmented ciliary epithelial cells. This is in contrast to the iris, where the blood–aqueous barrier function is maintained by tight junctions between the endothelial calls and lack of fenestrations.

There are numerous gap junctions between the pigmented and non-pigmented cells. These allow the two layers to act as a metabolic and transport synctium.

Both tight junctions in the iris and ciliary body are composed of the same proteins, occludin and cingulin.

The choroid

This is the lymphovascular supply of the posterior segment. It is designed to nourish the overlying retina. It is composed mostly of blood vessels embedded in a loose connective tissue matrix with a high concentration of type III collagen. As a result, 85% of ocular blood flow passes through the choroid via the anterior and posterior ciliary arteries.

The choriocapillaris is made of highly fenestrated, leaky blood vessels. These are arranged in a lobular fashion and collect in vessels of increasing size. Eventually, all vessels drain into four vortex veins, one in each quadrant of the globe. Blood passing into these veins is not desaturated and has lost only 5–10% of its oxygen. This is because of the high blood flow of the choroid.

There are no recognized lymphatics, but there are a number of immune cells, including mast cells, macrophages, and dendritic cells.

Blood–ocular barrier

Lipid-soluble substances, such as oxygen and carbon dioxide, pass readily through the endothelial cells. Water can also pass readily through the vessel wall, probably between and through endothelial cells. However, permeability to water-soluble substances depends on the structure of the capillary endothelium. This may be continuous, fenestrated, or discontinuous. All three types have variable networks of tight junctions (zonula occludens) between endothelial cells, which determine their permeability. The eye contains both continuous and fenestrated types of endothelium. The continuous type is the least permeable of the three. Even continuous membranes contain pores and hence act as porous membranes. The permeability depends on the size and number of these pores.

The retina has a similar arrangement to the blood–brain barrier, including tight junctions between the endothelial cells of the retinal vessels and the cells of the RPE.

There is a defect in the blood–retinal barrier at the optic disc, where water-soluble substances enter the anterior optic nerve by the process of diffusion directly from the extravascular space of the choroid.

The anterior segment of the eye has a corresponding blood–aqueous barrier. This consists of tight junctions between the endothelial cells of the iris capillaries and the non-pigmented epithelium of the ciliary body.

Because blood–ocular barriers are largely impermeable to even small water-soluble substances, such as glucose and amino acids, these substances must be transported by carrier-mediated systems. These are found in the blood–aqueous and blood–ocular barriers.

Various stimuli can increase permeability rapidly, e.g. histamine, bradykinin, and serotonin. The mechanism is thought to be a contraction in the endothelial cells, causing a widening of the clefts between them.

The technique of fluorescein angiography is clinically used to assess the permeability of the blood–retinal and blood–aqueous barriers and this topic is therefore of relevance to the clinician.

The retinal capillaries have a diameter of 5–6 μm. A continuous layer of endothelial cells is surrounded by a basement membrane, which contains a discontinuous layer of intramural pericytes. These are pluripotent cells that are small and run along the length of the vessels, giving off shorter processes that encircle the capillary walls. They are in intimate contact with the endothelial cells by peg and socket contacts, adhesion plaques, and gap junctions. They may have a role in permeability, modulation of endothelial cell growth, and angiogenesis. They may also act as progenitor cells for smooth muscle cells and phagocytes. They have contractile properties and thus it is also thought that they are involved in regulation of the microcirculation.

The capillaries of the choroid and ciliary body are fenestrated, resulting in greater permeability to low-molecular-weight substances and even some larger polypeptides/proteins such as albumin. For the choriocapillaris, this is necessary to keep a high glucose concentration at the RPE and to supply the amino acids necessary for production of vitamin A to the retina.

Accommodation

The act of accommodation causes three physiologic responses (the accommodative triad/near reflex):

- pupil constriction
- convergence
- accommodation.

Accommodation is a dynamic, optical change in the dioptric power of the eye. It allows the eye to focus from distant to near objects. It is achieved because of the ability of the ciliary muscle in conjunction with the lens zonules to alter the refractive power of the lens.

These three actions (accommodative triad) are coupled by preganglionic parasympathetic innervations. The intraocular muscles are innervated by postganglionic innervations. The extraocular muscles are innervated by cranial nerves III, IV, and VI, the axons of which originate from motor nerve nuclei in the brainstem, which receive impulses from the Edinger–Westphal nucleus.

Accommodation and convergence are coupled in both eyes, so that a monocular stimulus for accommodation results in a binocular accommodation and convergence response. Both accommodation and convergence occur in response to blur (for example if a myopic lens is placed in front of the eye) or convergence (for example if a base-out prism is placed in front of the eye) stimuli.

During accommodation, when the ciliary muscle contracts there is an anterior and inward shift of the ciliary body which serves to release tension on the zonules. This slackening causes the oblong lens to revert to its most relaxed state, which is spherical, thereby increasing the refractive power of the lens through an increase in both anterior and posterior lens surface curvatures.

Accommodation is measured in dioptres. A dioptre is a reciprocal metre and is the measure of the vergence of light. When the eye is focused at infinity, there is no accommodation. If the eye focuses an object from infinity to 1 m, it uses 1 diopter of accommodation. If the eye accommodates from infinity to 0.5 m, it uses 2 diopters of accommodation. At rest, eyes have resting accommodation of around 1.5 dioptres (tonic accommodation).

Depth of focus is the range over which an object can be moved towards or away from the eye without a noticeable change in blur or focus of the image on the retina. In practice this is measured subjectively by moving a near-reading target towards the eye. For any given accommodative state there is a range over which an object is perceived to be in focus. Visual acuity is important as it affects the subject's ability to perceive that an object is in focus. This is dependent on pupil size—a small pupil results in a larger depth of focus and vice versa. The size of the pupil is dependent on illumination and hence this too affects accommodation. Both accommodation and ageing cause pupillary constriction. The depth of focus results in an overestimate of true accommodative amplitude and even in older patients about 1 dioptre of 'accommodation' is still seen.

The accommodative apparatus includes the lens, elastic lens capsule, ciliary body and muscle, zonular fibres, and choroid. The ciliary muscle is within the ciliary body and is composed of three muscle fibre groups. These are orientated longitudinally, radially, and circularly. The ciliary muscle is a smooth muscle, so contraction is brought about by its parasympathetic innervations (M_3 muscarinic receptors). Relaxation is brought about by its sympathetic innervations

(β_2-adrenergic receptors). Ciliary muscle is different from other smooth muscle on account of its large motor neurons, the longer distance between muscle and motor neurons, and its ultrastructure (which resembles skeletal muscle).

As a result of these innervations, muscarinic agonists such as pilocarpine cause ciliary muscle contraction and involuntary monocular accommodation. Pharmacological blockage of accommodation is termed 'cycloplegia', and can be brought about by using a muscarinic antagonist such as atropine, cyclopentolate, or tropicamide. These competitively bind the same muscarinic receptors but do not activate them, hence blocking accommodation.

The zonular fibres are elastin-based elastic fibres that arise from a posterior insertion at the posterior pars plana near the ora serrata. There they enter the valleys between the ciliary processes and insert here to the ciliary epithelium. From here the zonular fibre forks and extends to the anterior and posterior lens surface with a superficial attachment (few fibres penetrating into the capsule).

When the eye is at rest, the zonules exert an outward-directed tension on the lens equator through the lens capsule. The lens is thus held in a flattened and unaccommodated state. On contraction, as described above, the inner aspect of the ciliary body moves inwards and forwards. This releases the resting tension on the zonular fibres, so the lens assumes a more spherical, accommodated form. This increases the optical power of the crystalline lens. Also, this decreases the anterior chamber depth as a result of the forward movement. The capsule provides the major force for the lens to form its more spherical shape.

Accommodative amplitude declines until about 50 years of age and is termed presbyopia (Greek *presbys* means aged person and *opsis* means vision) (see Chapter 7).

The pupil

Control of pupil size depends on the actions of the muscles of the iris. A balance between the dilator and sphincter pupillae exists. The pupil is able to change diameter from 2 to 9 mm (87% increase). It is usually located inferonasally under the cornea and has a normal physiologic tremor, hippus. In 25% of the population there is simple anisocoria (unequal pupils) where there is no pathology responsible for a difference in pupil size between the two eyes. Factors that affect the pupil size include (predominantly) ambient light, age, emotional state, state of arousal, and intraocular pressure.

Dark-adapted eyes are more sensitive to light. With increasing light stimulus intensity up to 9 log units above the stimulus threshold, amplitude of constriction increases and duration of constriction increases. Above 9 log units, the pupillary response plateaus off. Between the start of the light stimulus and the effect on the pupil there is a time lag (latent period) of 0.2–0.5 s. Because of this latency period, the pupil cannot respond to stimuli quicker than 5 Hz. The time lag for the pupillary light reflex is much quicker than for accommodation.

Pupil constriction occurs in response to light of any colour, but the response is related to the apparent brightness (Purkinje shift). As well as constricting to light, the pupil constricts to gratings and other recognition stimuli, e.g. the Snellen chart—this response is small but can be used to measure visual acuity in small children. Finally, another important phenomenon is that there is a summative effect on the pupils from binocular stimulation (binocular summation).

Constriction (miosis) of the pupil gives greater depth of focus and decreased chromatic aberration but also an increase in diffraction of light, whereas the converse is true of dilation (mydriasis). With increased dilation, however, light from the periphery can more easily enter the eye but because of the oblique angle at which this light enters, it is less effective at stimulating photoreceptors (Stiles–Crawford effect).

Dilation (mydriasis)

- This comprises a three-neuron adrenergic pathway.
- The first-order neuron descends caudally from the hypothalamus to the synapse in the cervical spinal cord (C8-T2)—also known as the ciliospinal centre of Budge.
- A second-order neuron travels from the sympathetic trunk, through the brachial plexus, over the surface of the lung, and then ascends to the superior cervical ganglion located near the angle of the mandible and the bifurcation of the common carotid artery.
- The third-order neuron ascends within the adventitia of the internal carotid artery, through the cavernous sinus (where it lies in close proximity to the sixth cranial nerve). This then joins the ophthalmic division of the trigeminal/fifth nerve (VI). This goes on to innervate the iris dilator as well as Müller's muscle (smooth muscle in the upper eyelids, causing a proportion of eyelid elevation).

Constriction (miosis) to light and near stimuli

- Light information from retinal ganglion cells travels through the optic nerves, optic chiasm (nasal fibres decussate), and optic tracts. This first synapses in the pretectal nuclei of the dorsal midbrain (each pretectal nucleus receives input from both eyes).

- Axons travel from the pretectal nuclei to the Edinger–Westphal nuclei. (Because of the duality of input, both pupils constrict equally to light input to either eye, giving the anatomic basis for both the direct and consensual light reflexes.)
- Parasympathetic (cholinergic) fibres for pupillary constriction travel along the third cranial nerve to the ipsilateral ciliary ganglion within the orbit.
- The final target for these postganglionic parasympathetic fibres is the pupillary sphincter muscle (and ciliary muscle for lens accommodation), which they innervate.

Pharmacological agents

Mydriasis

Pharmacological agents may cause dilation of the pupil because of the noradrenergic sympathetic stimulation of the dilator pupillae (Fig. 2.10).

See Chapter 9 for details of therapeutic agents.

Miosis

Pharmacological agents may cause miosis by cholinergic stimulation of the sphincter pupillae (Fig. 2.11).

See Chapter 9 for details of therapeutic agents.

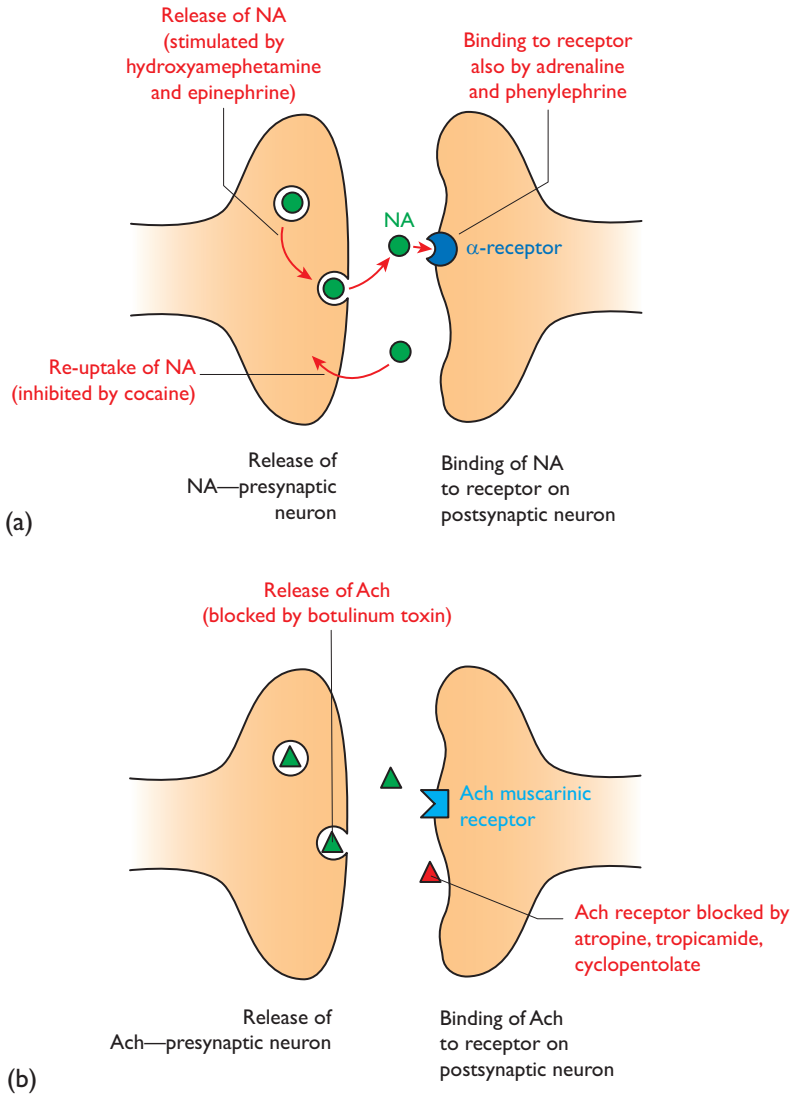


Fig. 2.10 Mydriasis: (a) the adrenergic synapse and action of sympathomimetics; (b) the cholinergic synapse and actions of cholinergic blockade. NA, noradrenaline; Ach, acetylcholine.

With permission from Neil Modi.

Near reflex—accommodation, convergence, miosis

For near stimuli, the term 'near vision complex/triad' is used to describe the three major components: accommodation, convergence, and pupil miosis (cyclorotation can be considered a minor component). Afferent input is first processed in the striate cortex (area 17), but the functionally specific area for the near-vision response is the pre-occipital cortex (areas 19 and 22), which receives an input from the entire cerebral cortex. The near response is therefore holistic and affected by the overall state of the individual. The nerve pathways for the three components merge in the midbrain at the Edinger–Westphal nuclei. Here the reflexes come under more primitive reflex, postural, and cerebellar influences. This pathway bypasses the pretectal nuclei of the dorsal midbrain. This results in a dichotomy between the light and near reflexes that can be elicited in some pathologies, i.e. pupils that react to near stimuli but not to light. The

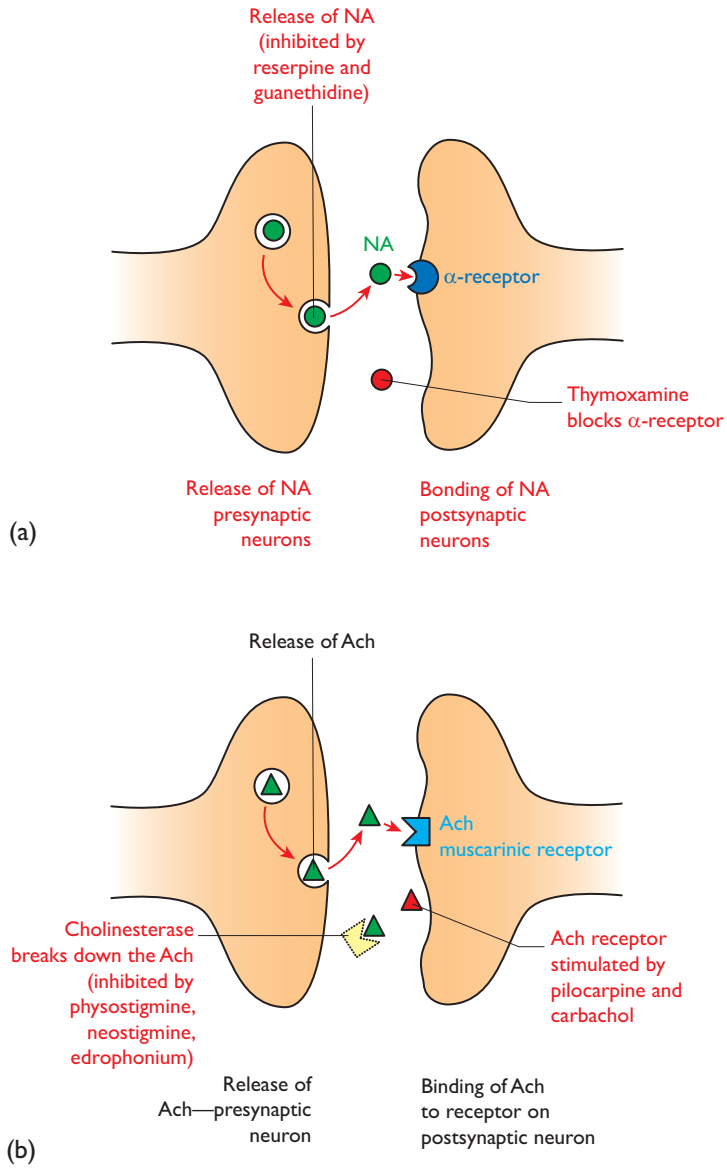


Fig. 2.11 Miosis: (a) the adrenergic synapse and actions of sympathetic antagonists; (b) the cholinergic synapse and action of cholinomimetics. NA, noradrenaline; Ach, acetylcholine.

With permission from Neil Modi.

CLINICAL TIP

Pupillary light reflexes

To examine the light reflexes, lights must be dimmed. In particular, Horner's syndrome is overlooked by a failure to dim the lights. A bright light such as an indirect ophthalmoscope must be used to examine the patient, who must be fixing on a distant object. If the subject fixes on the examiner's light the pupil may constrict due to accommodation.

The magnitude, speed, and symmetry of the direct and consensual reflexes should be observed. The pupil should be observed when the light is shone on it but also when it is removed. Dilation lag is the asymmetry in redilation between the two eyes when the light source is moved away from the eye. This suggests an abnormality of sympathetic innervations causing slower dilation of the affected pupil (Horner's syndrome), but this can also be a sign of a tonic (or Adie's) pupil.

Horner's syndrome is caused by an interruption of the sympathetic chain along its path or central (brainstem) lesions, and preganglioic and postganglionic lesions. Anisocoria is greater in dim light as the normal pupil will dilate, whilst the Horner's pupil will not. The other features that characterize this condition are facial sweating and ptosis. The postganglionic sympathetic nerves that control facial sweating travel along the external carotid artery (as opposed to the internal carotid artery) and thus are not affected in postganglionic Horner's syndrome.

Preganglionic lesions may be caused by a Pancoast tumour of the lung. Pharmacological testing with cocaine may be useful—cocaine prevents reuptake of noradrenaline at the adrenergic synapse. A Horner's pupil will not dilate with cocaine but a normal pupil will. Hydroxamphetamine will dilate a preganglionic or central Horner's if the postganglionic pathway is intact, but will have no effect on a postganglionic Horner's. Pharmacological testing, however, cannot differentiate between a preganglionic and central Horner's.

Relative afferent pupillary defect (RAPD) describes the paradoxical dilation when the light is swung from the healthy eye to the eye with a lesion in the anterior visual pathway. Although there may be no abnormality of the direct or indirect reflexes, the afferent pathway on the side with the lesion leads to less input/stimulus to the pretectal nucleus and as a result of this perceived reduction in input the level of constriction is lessened, causing an apparent dilation. This can be elicited by the swinging flashlight test. Cataracts cannot cause an RAPD. It is most commonly caused by retinal or optic nerve pathology, e.g. glaucoma.

The near response is tested when there is abnormality of the light reflex and can be assessed using an accommodative target such as reading the near-acuity card at 30 cm.

Light-near dissociation is impaired when the light reaction is impaired but accommodation is preserved. Causes include bilateral anterior afferent visual pathway disease, dorsal midbrain lesions, Argyll–Robertson pupils, diabetes (autonomic neuropathy), and tonic pupils (e.g. Adie's pupil, systemic neuropathic disorders, and local orbital disease).

Adie's pupil is a relatively common phenomenon. The light reaction is sluggish but the near response is fast and tonic. Super-sensitivity to cholinergic agonists is a postdenervation phenomenon, as this problem is caused by a ciliary ganglion lesion.

final common pathway to the eye is the third cranial nerve, from which there are branches to medial recti, ciliary muscle, and sphincter muscle (the latter two synapse in the ciliary ganglion).

Visual acuity and contrast sensitivity

Visual acuity

Visual acuity is a measure of the ability of the eye to distinguish two stimuli separated in space. Within this general definition there are three main subdivisions of acuity:

1. identifying presence of a single feature or 'minimum visible'
2. identifying features within a visible target or 'minimum resolvable'. This is what is measured on a standard chart
3. identifying the relative location of visible features or hyperacuity.

Clinically, visual acuity is determined by discriminating letters on a chart, which requires discrimination at a retinal level accompanied by an ability to interpret the form and shape of the letters at a higher level. In a situation where higher centres are not yet fully developed or are abnormal, contrast sensitivity can be used to give a prediction of visual acuity.

Assessment of visual acuity using letter charts

Measurement of visual acuity in clinical practice assumes that the cone photoreceptor has the ability to discriminate

two objects in space subtended by an angle of 1 minute of arc at the nodal point of the eye (the minimum angle of resolution, MAR). In fact, in a normal observer the resolution is somewhere between 30 seconds and 1 minute of arc. Cones have the highest discriminatory capacity. Level of acuity falls off rapidly the greater the distance from the fovea (5° from the central fovea visual acuity is a quarter of foveal acuity). The essential requirements for discriminating two objects are:

1. an underlying retinal image pattern with two peaks separated by a trough
2. a retinal illuminance difference between the peaks and the trough
3. separate localization of the differentially stimulated regions.

Types of chart used in clinical practice

Snellen and Landolt's C

The standard normal visual acuity equates to the vision of 6/6 or 20/20 when viewing a predetermined standard target at 6 m (UK) or 20 ft (USA). Test conditions describe ambient illumination and the illumination of the letters (optotypes) on the chart to provide contrast is standardized. The Snellen chart is based on the concept that the smallest spatial target that can be resolved subtends 1 minute of arc at the nodal point of the eye.

LOGMAR

The logarithm of the MAR has definitive sizing and spacing of the letters, and can provide a more quantitative evaluation of visual acuity. It therefore tends to be the standard in clinical trials.

Vernier acuity or hyperacuity

Vernier or hyperacuity is not present in infancy and is optimal at around age 14 years. It is used in situations such as measuring objects with a ruler. It requires the observer to make spatial judgements at thresholds of 2–10 seconds of arc and allows the observer to determine the relative location of two or more visible features with respect to each other. It is not yet understood how hyperacuity is achieved. It is thought to be absent in strabismic amblyopia.

Factors influencing visual acuity

Many factors influence visual acuity. Well-documented factors include those shown in the *Clinical tip* box.

Contrast sensitivity

Visual acuity is affected by contrast. Contrast sensitivity can be thought of as the ability to discriminate a thin white line against a uniform background illumination and is thought to be less dependent on optics of the eye than visual acuity measurements. Contrast sensitivity measurements may be useful in the diagnosis of certain ophthalmic disorders.

Contrast sensitivity is measured using a sinusoidal grating (Fig. 2.12). The basic element of the test is a grating target of given spatial frequency, expressed in cycles per degree of visual angle. Its contrast is changed by keeping the mean luminance the same and varying the difference between the luminance at the peaks (I_{max}) and troughs (I_{min}) until the observer's threshold is reached. The modulation is described in terms of the maximum (I_{max}) and minimum (I_{min}) and is brought together in the equation: $M = (I_{max} - I_{min})/(I_{max} + I_{min})$.

Modulation sensitivity can vary from 0.5 cycles/° (covering 2° of visual angle) to 20 or 30 cycles/° (covering 3 or 2 minutes of arc).

Both low and high spatial frequencies require higher contrast to be seen. Modulation sensitivity is a reciprocal measure of the grating contrast needed for threshold. Contrast sensitivity is higher in the intermediate range and falls off for both higher and lower spatial frequencies.

Like visual acuity, many factors affect contrast sensitivity. Major factors include:

- the direction of the grating lines (most sensitive in vertical and horizontal directions)
- luminance
- bar width and length
- grating motion.

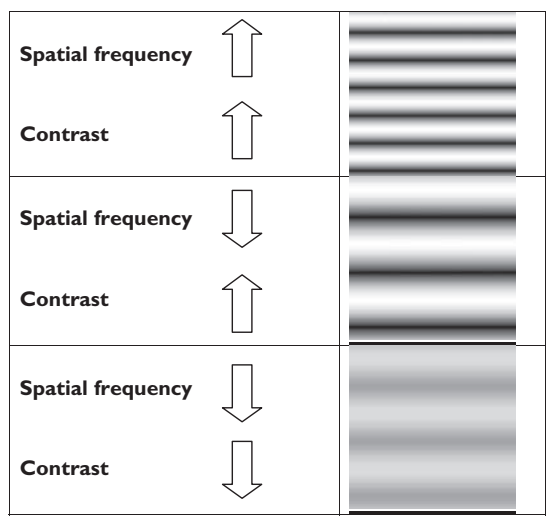


Fig. 2.12 Spatial frequency and contrast of gratings.

CLINICAL TIP

Factors influencing visual acuity

Factor	Explanation
Refractive error	As optics defocus, the point spread function widens and objects need to be farther apart to be interpreted as separate.
Retinal eccentricity	The centre of the fovea has maximum acuity. Cones are farther apart as the distance from the fovea increases, causing an increase in retinal summation areas. Visual acuity and hyperacuity fall off at different rates with eccentricity but both exhibit better values in the temporal rather than the nasal visual field.
Luminance	Visual acuity remains constant over a wide range of photopic luminances. High luminance causes an unexplained reduction in acuity.
Contrast	As contrast reduces, resolution is reduced. Stereoscopic acuity suffers when contrast is reduced. Vernier acuity is less severely affected.
Pupil size	Visual acuity remains relatively constant between 2.5 and 6 mm. Outside this there is a reciprocal relationship between minimal angle of resolution and pupil size.
Exposure duration	Visual acuity reduces with decreased exposure duration in the millisecond range. Vernier acuity is not so affected by a reduction in exposure but this is not the case for stereoscopic acuity.
Target and eye movements	Significant movement of a retinal image or saccadic eye movements cause a reduction in visual acuity. Pursuit eye movements and small movements of an image do not significantly affect acuity.
Meridional variation	Horizontal and vertical meridians are favoured generally.
Interaction effects	Visual acuity suffers when targets are too close together, a phenomenon known as crowding. This effect is eliminated by use of a LOGMAR chart as optotypes are evenly spaced.
Development	It takes several months after birth for the full development of pursuit eye movement. It is therefore not easy to differentiate between inability to resolve a pattern and inability to execute movement. Preferential looking methods use the frequency of intersaccadic fixation in regions containing a pattern as compared with those not containing a pattern. This yields consistent results.
Ageing	As the eye and visual pathways age, visual acuity decreases. Increased intraocular scatter of light is associated with an ageing eye.

Light detection and dark adaptation

Light detection

It is thought that between 50 and 150 quanta of light are required to strike the cornea for a signal to be detected. Approximately 10–15 rods must be stimulated. Signal detection is dependent on both photoreceptor and neural pathway function. Responses must be collected together and summed at the bipolar or ganglion cell level to induce visual sensation. This process is known as summation.

Factors affecting light detection (think of a visual field test) include:

- background illumination
- spatial frequency or temporal summation (Bloch's law)
- the intensity of the threshold stimulus (inversely proportional to the duration of the stimulus)
- spatial summation or stimulus size (Ricco's law) (the threshold intensity of a stimulus is inversely proportional to the area of the stimulus)
- wavelength
- the optics of the eye
- dark adaptation.

Light adaptation

Light adaptation works quickly, increases as background light increases, and depends on calcium flux. The light-adapted eye is most sensitive at a wavelength of 555 nm.

The visual system must detect light of varying intensity over a range of at least 12 log units. The pupil area varies around 16-fold (1.3 log units). Cone function (intensity required for detection, ΔI) is proportional to the background intensity (I) over a range of 3 log units. This relation, where $\Delta I/I$ is a constant, is known as the Weber–Fechner relation, the constant is called the Weber fraction. At higher light levels saturation occurs and the relationship breaks down (Fig. 2.13).

Dark adaptation

As discussed previously, the minimum visual stimulus varies depending on whether the stimulus is viewed in the dark or under normal/bright light conditions. In the dark, the eye becomes more sensitive to light, reaching a minimum threshold at around 30 minutes. Dark adaptation is therefore much slower than light adaptation. The dark-adapted eye is maximally sensitive at about 505 nm.

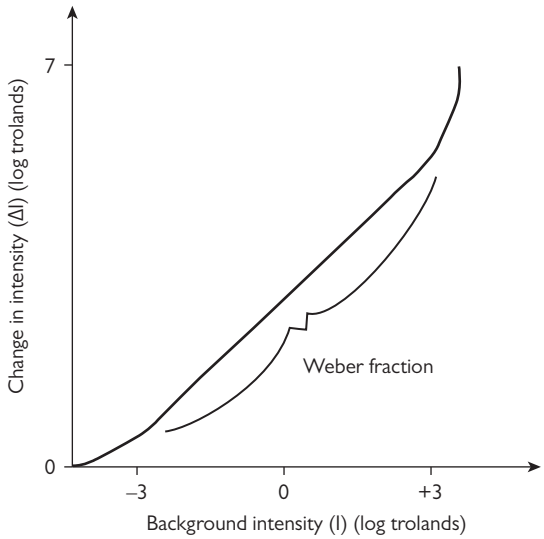


Fig. 2.13 The light adaptation curve.
With permission from Louise Bye.

The dark adaptation curve

The dark adaptation curve demonstrates the adaptation of rods and cones to diminishing light levels (Fig. 2.14). It has two main components: an early rapid response resulting from increases in the cone sensitivity and a second slower response produced by increased rod sensitivity.

The normal dark adaptation curve varies if the conditions are varied: with a small central white target, rods are not stimulated at all and the curve flattens out. If the cones are first light adapted by weakly stimulating them to maximum sensitivity or by adapting subjects to red light before placing them in the dark, the cone component can be lost.

Dark adaptation on a molecular level

There is an approximate relationship between the amount of rhodopsin bleached and the state of dark adaptation. This is delineated in the Dowling–Rushton equation: $\log(z)/A = aB$, where A is the threshold in complete dark adaptation, B is the fraction of bleached rhodopsin, and a is a constant of proportionality.

Dark adaptation and regeneration of rhodopsin are dependent on the local concentration of 11-*cis* retinal. The limiting factor for recovery after a bleach is the rate at which it is delivered to opsin in the bleached photoreceptors. A healthy RPE is essential to this process and the age-related decline in dark adaptation may be related to a failing RPE.

Melatonin and circadian rhythms

The light–dark cycle is generated in the suprachiasmatic nuclei and drives production of melatonin from the pineal gland.

Melatonin is synthesized from tryptophan via serotonin. It is secreted mainly at night and its secretion is suppressed in light. It is therefore an important component of the

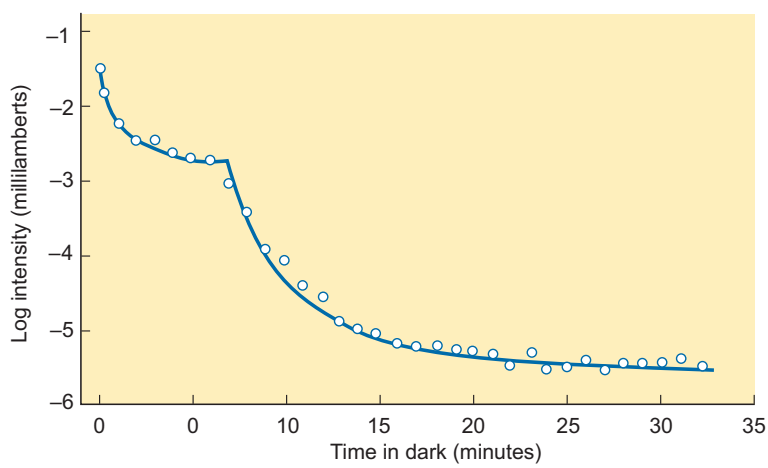


Fig. 2.14 The dark adaptation curve, outlining dark-adaptation responses: mixed rod and cone response of dark adaptation.
Reproduced from Gillian Pocock and Christopher D. Richards, *Human Physiology, The Basis of Medicine*, Third edition, Figure 8.28, Page 122, 2006, with permission from Oxford University Press.

sleep/wake cycle and in stabilizing circadian rhythm. It can act to influence immune defence, glucose regulation, haemostasis, and ocular functions such as control of aqueous secretion. In part this occurs via photosensitive retinal ganglion cells, which are distinct from those involved in image forming. These cells make up 2% of retinal ganglion cells and produce melanopsin. Melanopsin has a peak sensitivity of 434 nm (blue light). This allows for control of circadian rhythm and neural and endocrine signals, which regulate behavioural and physiological circadian rhythms.

Alterations in melatonin production are thought to predispose to certain conditions, e.g. evidence that night-shift workers have a higher incidence of diabetes may be related to melatonin secretion.

Colour vision

The concept of colour

Colour is a perceptual phenomenon and not just due to the physical property of an object. Colour (C) is specified in terms of the three primary colours by the equation:

$$C = (r \times R) \times (g \times G) + (b \times B)$$

It is measured with colour-matching instruments such as a flicker photometer or with spectrophotometers fitted with arrays of wavelength-selective photodiode detectors.

Many factors influence the colour perceived.

Spectral composition of light from the object

Most of the light we detect is light reflected from objects. Colour depends on their surface properties. In spite of the wavelength of the reflected light varying in different light conditions the 'perceived' colour remains the same. This is a function of higher visual processing by the cortex and is known as colour constancy.

Spectral composition of light from the surroundings (Purkinje shift)

In the dark-adapted state wavelengths of 500 nm (blue–green) appear brightest and in photopic conditions wavelengths of 555 nm (yellow–green) appear brightest. This phenomenon of short wavelengths becoming brighter compared with long wavelengths as luminance is reduced is known as the Purkinje shift.

State of light adaptation of the eye

Interestingly, by choosing suitable conditions of light adaptation (by adapting with blue light to desensitize the rods) it is possible to measure cone thresholds with different wavelengths of light. This can produce cone-specific spectral curves.

Colour detection

Wavelength discrimination requires a panel of photoreceptors with peak responsiveness at specific wavelengths independent of their responses to changes in luminosity.

A specific colour or hue is detected by a summation of responses from a mixture of receptors, and the contribution from each of the three primary photoreceptor types can be deduced from spectral mixing curves. This is the trichromatic theory of colour detection (Young, Helmholz, and Maxwell). Hue discrimination requires higher-order colour perceptual mechanisms.

In the primate retina there are three classes of cones (Fig. 2.15):

- long wavelength sensitive (L) or red cones with peak sensitivity 570–590 nm
- medium wavelength sensitive (M) or green cones with peak sensitivity 535–550 nm
- short wavelength sensitive (S) or blue cones with peak sensitivity 440–450 nm (blue cones are by far the most infrequent in the human fovea).

Mixing the three primary colours will produce any secondary colour or hue. A specific quantity of the primary colour is required to produce each hue and this amount is determined by spectral mixing curves.

The three cone opsins differ in their transmembrane regions of the proteins. This allows them to respond to different wavelengths of light. Cone opsins are less sensitive than rhodopsin, but their responses are several folds faster. The parvocellular pathway deals with spatial and colour vision. P cells have small receptive fields and a single cone cell may have sole input to a single bipolar cell. There are thought to be different P ganglion cells for chromatic function as there are for spatial function. There is no convergence at the retinal level for colour vision, although there is convergence in the cortex for colour.

The bipolar cell that is fired cannot distinguish whether it has been stimulated by a red or green cone. This creates some confusion. It is thought that this confusion is sorted out at the level of the ganglion cell through organization of receptive fields. This is the colour opponent theory (of Hering).

Colour opponent theory

In the colour opponent scheme there are three colour opponent arrangements:

- ON-centre red/OFF-centre green—the bipolar cell receives stimulatory or inhibitory input from a single red or green cone
- ON-centre blue/OFF-centre yellow
- ON-centre white/OFF-centre black—the bipolar cell receives input from all three cones and colour mixing occurs.

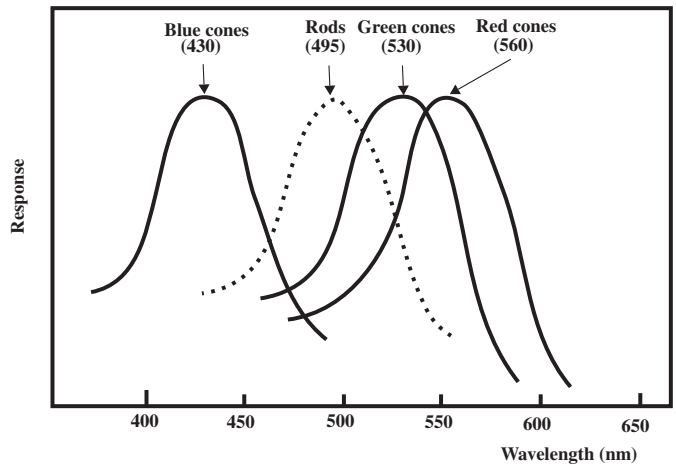


Fig. 2.15 Spectral mixing curves. Mixing the three primary colours will produce any secondary colour or hue. A specific quantity of the primary colour is required to produce each hue and this amount is determined by spectral mixing curves.

This refines the spectral sensitivity of the ganglion cell responses and allows the perception of hues and unsaturated colours.

Colour blindness

Clinically, colour vision is tested using hue-discrimination techniques (e.g. the Farnsworth–Munsel 100 hue test). Peak ability is around age 19 years.

Some colour vision defects are explained by loss of one or other specific type of receptor. In practice the situation is more complex, involving a combination of genetic defects that reduce the responsiveness of cone opsins at various levels.

The main types of colour defect include the following:

- Monochromatism
  - Rod monochromatism affects 1 in 30,000 of the population—these individuals have true achromatic vision, low visual acuity, photophobia, nystagmus, and some have macular dystrophy.
  - Cone monochromatism affects 1 in 100,000—these individuals have normal visual acuity but cannot discriminate coloured lights of equal luminosity. Cone monochromats possess all three types of cone and the defect occurs in cortical processing.
- Dichromatism—the affected individual matches all colours with mixtures of two primaries. This means the range of secondary colours is restricted:
  - protanopes are missing red
  - deuteranopes are missing green
  - tritanopes are missing blue.
- Anomalous trichromatism—these individuals use different proportions of the three primaries to match colour: protans use more red, deutans use more green, and tritans use more blue. This means the individual will not accept those matches that the normal trichromat agrees on. This is the most common form of colour blindness and occurs in 6% of the male population.
- Achromatopsia—colour blindness can be the result of defects in cortical processing (area V4). Congenital or acquired lesions in the lingual or fusiform gyrus are associated with cerebral achromatopsia.

Electrophysiology of the visual system

Conversion of light energy to an electrical response

Transmission of nerve impulses in retinal receptors and neurons is mediated by changes in the electrical potential across the cell membrane, leading to an electrical discharge. The action potential in ganglion cells is considered an all or nothing event. However, bipolar, horizontal, amacrine, and photoreceptor cells have a graded response and the direction may be positive or negative.

Electrophysiology of the photoreceptor

This is the reverse of what we have learned previously regarding neurons and is dependent on cell permeability to Na^+ and Ca^{2+}.

In resting conditions (in the dark):

- The outer segment is maintained in a depolarized state through open Na^+ channels.
- Sodium ions influx from the extracellular fluid into the photoreceptor.

In the active state (in the light):

- Light stimulates the outer segment, the sodium channels are closed. This stops the influx of sodium and leads to a reduced level of depolarization known as a relative hyperpolarization.
- Hyperpolarization is transmitted by an influx of calcium ions through the cGMP-gated Ca^{2+} channel and the Na^+-Ca^{2+}-K^+ exchanger located in the plasma membrane.
- Hyperpolarization is transmitted along the length of the photoreceptor to the synapse with the bipolar cell.

Electrophysiology of the bipolar cell

The Ca^{2+} wave induces the bipolar cell to release its neurotransmitter glutamate. Bipolar cells will produce one of two responses:

- an ON response, which is a hyperpolarized state
- an OFF response, which is a depolarization response.

Dark currents

Dark currents are created from the resting potential difference across the photoreceptor cell membrane and the difference in potential difference between the depolarized outer segment and the relatively hyperpolarized synaptic region of the cell at its interaction with the bipolar cell.

Electrophysiology of retinal neurons

The concept of summation

There are 150×10^6 photoreceptors but only 1×10^6 optic nerve fibres. This means that many receptors must converge signals to feed into a single neuron. These will be a combination of stimulatory and inhibitory signals, hence a 'sum' of these signals will produce one final outcome. This is known as summation.

The concept of receptive fields

A receptive field is an area of retina where summed responses converge (Fig. 2.16). It is thought that within a given area of retina while the rate of discharge in certain fibres increases (ON response) when light is presented to the eye, in other fibres it decreases (OFF response). This ON/OFF response in optic nerve fibres correlates with a centre/surround organization seen in receptive fields and is caused by interneuronal interactions producing inhibition in surround cells.

ON/OFF receptive fields

- Apply to cone–bipolar–ganglion cell circuitry. Rod cells synapse indirectly into cone circuits at the bipolar cell level. Rod bipolar cells connect with cone bipolar cells through an amacrine cell known as the AII cell.
- Apply to several different types of signal, including light–dark, blue–yellow, and red–green
- Are large when dealing with big ganglion cells such as magnocells or M cells (large receptive field = high degree of convergence from many amacrine or bipolar cells)
- Are small when dealing with small ganglion cells such as parvocells or P cells (small receptive fields = less convergence)
- Are created by midget ganglion cells at the fovea.

The role of amacrine and horizontal cells in receptive fields

Amacrine and horizontal cells modify the ON/OFF microcircuitry at the bipolar–ganglion and receptor–bipolar cell interfaces, respectively. This is done by a combination of inhibitory and stimulatory signals.

Amacrine cells are more numerous than horizontal cells and have a wider range of functions, including more fine control of signals via a larger range of neurotransmitters (dopamine, acetylcholine, glutamate, GABA, glycine). They can affect orientation selectivity, light–dark, colour perception, and centre/surround organization.

The ON/OFF response in the ganglion cell can only be graded with regard to frequency or rate of discharge but cannot be graded as seen in the hyperpolarization of the photoreceptors or horizontal, amacrine, and bipolar cells.

Electrophysiological tests

The eye possesses a resting potential, which is generated at the interface between the RPE and the photoreceptors (the resting retinal potential), and is about 60 mV.

At a cellular level:

- the electro-oculogram (EOG) represents the trans-pigment epithelial potential
- the electroretinogram (ERG) represents the response from all retinal layers
- the pattern ERG represents the response after processing in the bipolar cell layer
- the visual evoked potential (VEP) records electrical activity from the occipital cortex following presentation of a light stimulus to the retina.

The electro-oculogram

The EOG is representative of the trans-retinal pigment epithelial potential differences generated by separation of ionic gradients across the RPE and maintained by tight junctions (Fig. 2.17).

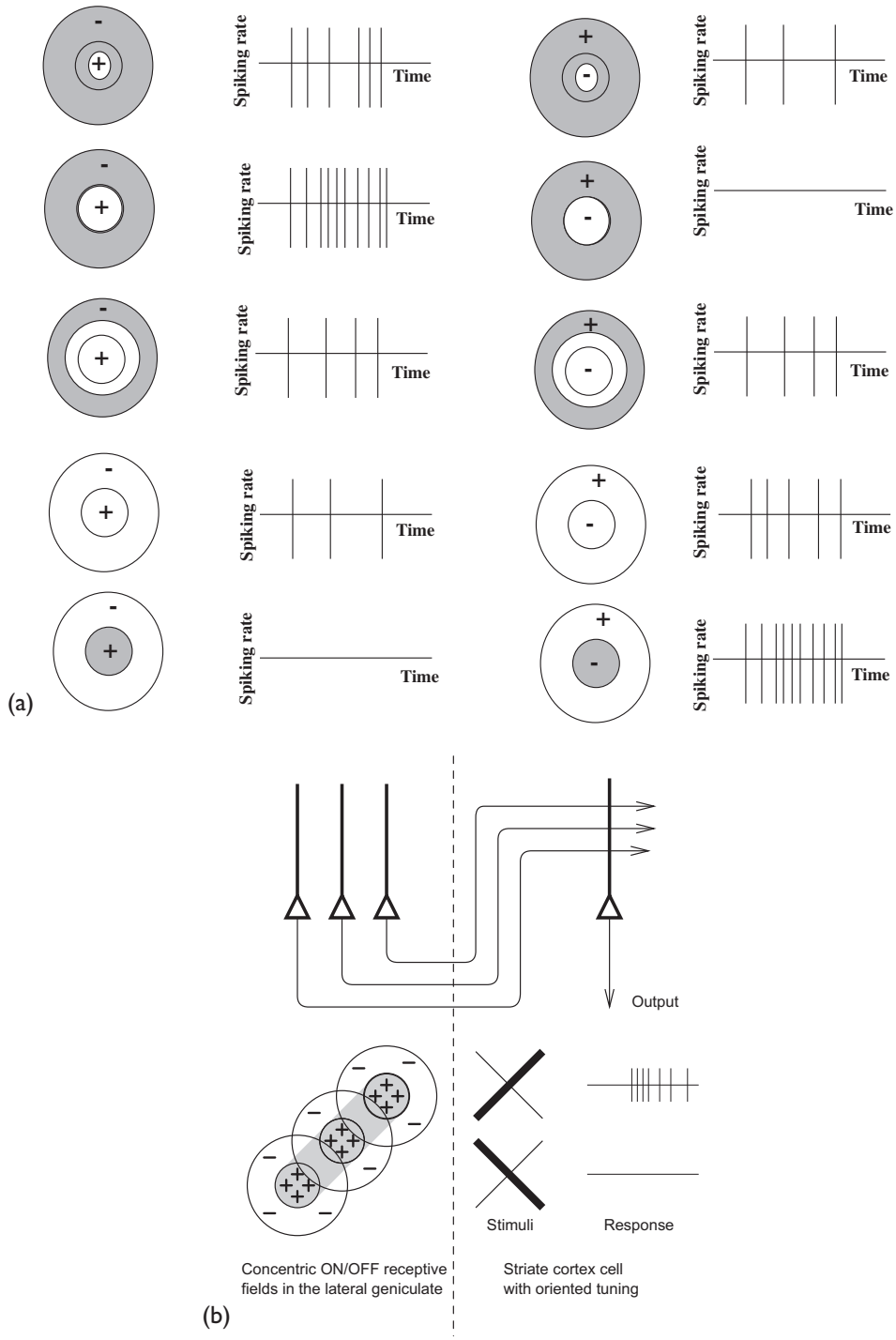


Fig. 2.16 Receptive fields. (a) Illustration of a receptive field. (b) The size of the receptive field varies between 200 and 600 μm in diameter. Receptive fields may have an ON centre and an OFF surround, an OFF centre and an ON surround. In addition, each of these occurs for light–dark, red–green, and yellow–blue detection.

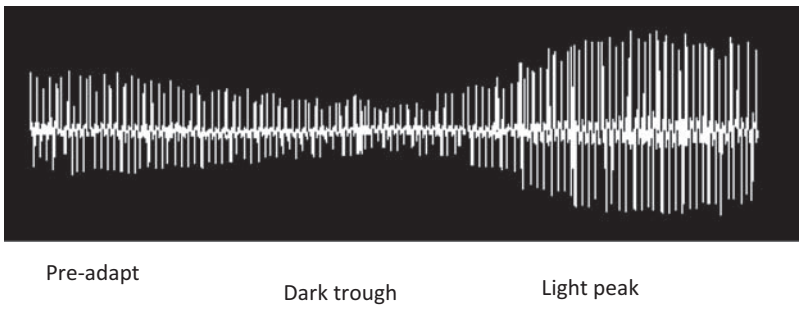


Fig. 2.17 The electro-oculogram.
With permission from Dr Robson.

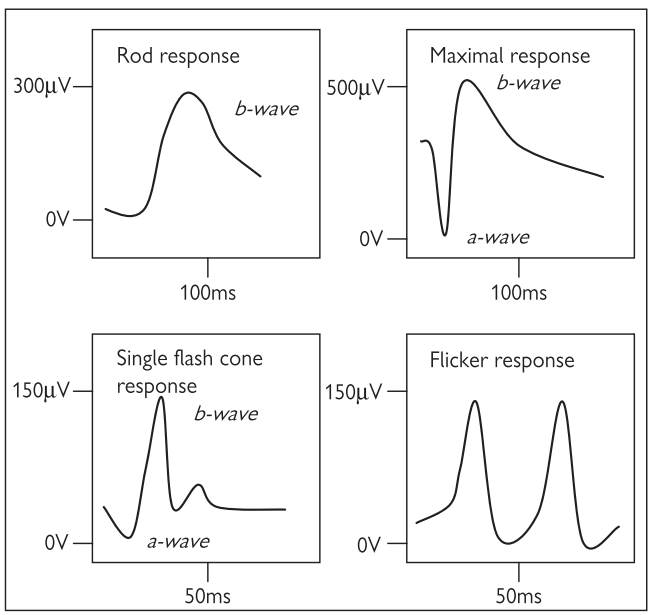


Fig. 2.18 Dark-adapted response to a flash retinogram.
Reproduced from Alastair K.O. Denniston and Philip I. Murray, *Oxford Handbook of Ophthalmology, Rapid Reference for all Ophthalmologists*, Figure 2.15, Page 79, 2009, with permission from Oxford University Press.

The EOG is a recording of the electrical dipole occurring between the cornea and retina, and reverses in direction when the eye moves from side to side. Measurements are taken between electrodes at the outer and inner canthi in the dark (20 minutes) then in the light. The amplitude increases in conditions of bright illumination. The critical value is the Arden ratio, which is light peak/dark trough and should be over 180% in a normal eye.

The light rise is the result of an extracellular flow of current caused by changes in the potassium concentration in the interphotoreceptor matrix. The EOG is lost in conditions that disrupt the RPE–photoreceptor relationship, such as retinal detachment.

The electroretinogram

The ERG is superimposed on the EOG and is the cumulative electrical response to a light stimulus from all retinal elements. It is affected by the intensity, duration, wavelength, and pattern of the stimulus and the level of light–dark adaptation of the retina. The ERG allows an assessment of the photoreceptor and bipolar cell function in photopic and scotopic conditions.

Major components of the ERG include (Fig. 2.18):

- negative a wave—this is generated by hyperpolarization in the photoreceptors' inner segments (Granit's PIII component), a^1 comes from cones and a^2 from the rods.

- positive b wave (Granit's PII component)—this comes from bipolar cells directly or indirectly spread to the Müller cells: b^1 is cone-dominated bipolar cells and b^2 is rod-dominated bipolar cells (the b wave is lost in certain retinal vascular conditions such as central retinal vein occlusion).

Measurements include implicit time or latency of the components from onset of stimulus to the trough (a) or peak (b) of the component. The a wave amplitude is measured from onset and the b wave amplitude is measured from the trough of the a wave.

Minor components of the ERG include:

- slow positive c wave—superimposed on the b wave this is a series of fast waves which have limited clinical value and is a positive component generated in the retinal pigment epithelium, which are lost in patients with diabetes
- the d wave—this is an OFF response following cessation of illumination.

In order to separate out the function of rods and cones ERG settings can be varied:

- photopic conditions—with a bright white light superimposed upon a rod-saturating background or with a 30-Hz flicker stimulus a cone response can be measured
- scotopic conditions—when a very dim white light or blue flash is used in the dark the response is generated purely by the rods.

Examples of ERG changes in clinical practice include:

- retinitis pigmentosa—primarily affects the rods and will therefore affect the scotopic ERG
- retinoschisis and congenital night blindness—the rod bipolar cell component of the b wave is missing, producing a negative ERG
- early cone abnormalities—an increase of the implicit time is seen
- in choroidal disease both a and b waves are affected but the a wave is spared in central retinal artery and vein occlusion.

The pattern electroretinogram

The pattern ERG can be used to distinguish retinopathy from maculopathy (Fig. 2.19). The stimulus is a reversing checkerboard (two to six reversals per second).

It has two main components:

- P_{50} (positive at 50 ms) is affected by damage to the inner retina and ganglion cells
- N_{95} (negative at 95 ms) is affected by optic nerve damage.

The amplitude/unit area of the pattern ERG varies with cell density and is less dominated by the fovea than the VEP. It is abnormal in local macular conditions and early retinal degenerations.

The visual evoked potential

The VEP is a measure of the response of the occipital cortex to visual stimulation (Fig. 2.20). It is extracted using scalp electrodes. Usual stimuli are a reversing checkerboard or a diffuse flash. Pupillary dilation is contraindicated for pattern VEPs and correction of refractive error is necessary for recording pattern VEP. Flash VEPs can be used in uncooperative patients such as infants or preverbal infants, or patients who are unconscious or in a coma. Assessment of optic nerve and chiasmal function is possible.

Components of the VEP include:

- P_{100} (positive at 100 ms). The latency of this is frequently measured. Optic nerve demyelination produces a delay and can cause a reduction in amplitude, which can correlate with visual acuity
- N_{70} and N_{135} (negative at 70 and 135 ms)
- an amblyopic eye will give a normal flash VEP and an abnormal pattern VEP. In albinism the pattern VEP and flash VEPs are abnormal due to the reduced number of uncrossed nerve fibres at the chiasm.

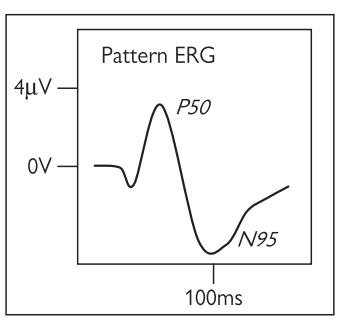


Fig. 2.19 The pattern electroretinogram.

Reproduced from Alastair K.O. Denniston and Philip I. Murray, *Oxford Handbook of Ophthalmology, Rapid Reference for all Ophthalmologists*, Figure 2.16, Page 81, 2009, with permission from Oxford University Press.

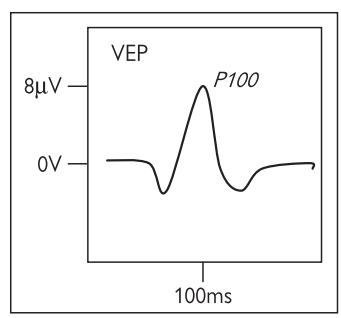


Fig. 2.20 The visual evoked potential in a normal subject.

Reproduced from Alastair K.O. Denniston and Philip I. Murray, *Oxford Handbook of Ophthalmology, Rapid Reference for all Ophthalmologists*, Figure 2.16, Page 81, 2009, with permission from Oxford University Press.

Visual fields

The visual field represents the area of retina sensitive to visual stimuli. In the UK, a minimum field of binocular single vision is required to qualify for a driver's vehicle licence and is defined as a field of binocular single vision (BSV) 20° above and below the horizontal meridian and 60° to either side of the vertical meridian (Table 2.2).

Retinal sensitivity to light varies according to the distance from the fovea in degrees of fixation. This concept has been described as the hill of vision.

Visual field testing

Visual field (VF) testing can allow for early detection of ocular and neurological disorders by looking for sensitivity losses throughout the VF. It can also be used to monitor the progress of the VF over time to determine any deterioration or improvement of the condition. It is important to be able to interpret both static and kinetic VFs as both are used in clinical practice and frequently come up in exam questions. This section will concentrate on understanding VF testing methods but will not cover typical field patterns seen in various pathologies. These can be found in clinical texts.

Measurement of VFs can be performed by many techniques:

- confrontation
- Amsler—for central visual field, e.g. macular lesions
- static (target is stationary), e.g. Humphrey
- kinetic (target is moving), e.g. Goldmann
- flicker-based fields.

For the static and kinetic tests the basic methodology for all patients is similar:

- Patients need to detect a stimulus of a given size and intensity (e.g. a small light) on a uniform background.
- Patients keep one eye fixed on a target directly forward, while the other eye is covered.
- A test object is presented to the test eye at various positions.
- Patients signal when they see these objects.
- Areas of perception deficits are indicative of VF defects: scotoma.

Table 2.2 The monocular and binocular visual field

Field type	Width	Height
Monocular	100° temporally, 60° nasally	60° superiorly, 75° inferiorly
Binocular	120°	135°
Total	200°	135°

Goldmann visual field

This is a form of manual kinetic perimetry. A target of fixed size and luminance is moved from the periphery to the centre.

The Goldmann test is good for detecting and monitoring neurological deficits as it covers a wider VF than static tests. It is also good for patients with poor fixation and for patients in whom malingering is suspected—this produces a classic spiral defect.

The disadvantages are that staff members need to be trained specifically in order to perform a Goldmann test, it is less useful for producing reproducible fields for central defects or scotomas, and as it is a manual test it takes longer than a static automated test.

Stimulus is varied in intensity (b) and size (c). Retinal sensitivity is measured in decibels (dB) and light intensity in apostilbs (asb) (see Chapter 8 for a full explanation of these units). dB and asb are inversely related: 40 dB = 4 log units less than 10,000 asb = 10,000/10 (4) = 1 asb. 1 asb is considered to be the minimum perceivable stimulus. 40 dB is therefore the maximum sensitivity. Goldmann maximum intensity is 1000 asb, whereas Humphrey maximum intensity is 10,000 asb.

The centre of the VF is represented by the intersection of the lines on the grid. The areas in which these targets can be

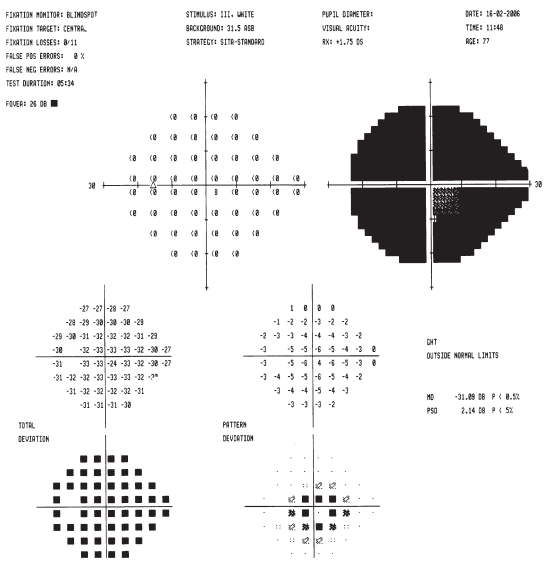


Fig. 2.21 The Humphrey visual field.

Reproduced from Venki Sundaram, Allon Barsam, Amar Alwitry, and Peng Khaw, *Training in Ophthalmology*, 2009, Figure 7.17, p. 289, with permission from Oxford University Press.

seen are encircled at equal size and brightness (isopters). In the normal eye, larger, brighter targets can be seen farther into the periphery than the smaller, dimmer targets.

Humphrey visual field

This is a form of automated static perimetry. A static stimulus Goldmann III at 25 dB is presented for 0.2 seconds at random locations across the central (~30°) VF. If the stimulus is seen the luminance is reduced in discrete steps until it is no longer seen. Background illumination of the bowl is 31.5 asb.

The advantages of the Humphrey test are:

1. Testing conditions are standard and highly reproducible.
2. It does not require a highly trained experienced examiner.
3. A standard analysis package allows easy assessment and comparison.
4. Results are reported in statistical terms as compared to the visual response of normal age-matched sighted people.

The main disadvantage is that only the central visual 30° field is examined. In addition initial field testing may be unreliable as the patient learns the technique.

The Humphrey perimeter has different threshold strategies which can be used according to patient need:

- Original standard. Stimulus is increased by 4 dB until the threshold is crossed, then reversed in 2 dB steps.
- FASTPAC. Stimulus advances by 3 dB. Once the threshold is reached there is no reversal. Seventy per cent quicker than the original standard but lower resolution and higher intratest variability.
- Swedish Interactive Threshold Algorithm (SITA). Available in 'standard' and 'fast'. A sophisticated algorithm that reduces the number of times a stimulus is presented to arrive at threshold. Complex statistical analysis allows the machine to perform the test twice as fast with similar reliability. This method is used most frequently in the diagnosis and monitoring of glaucoma patients (Fig. 2.21).

Eye movements and stereopsis

Extraocular muscles and eye movements

Primary, secondary, and tertiary actions

The fixed point of the centre of rotation of the eye is 13.5 mm posterior to the corneal apex and 1.6 mm to the nasal side of the geometric centre. The primary action of a muscle is the movement that occurs when the muscle contracts with the eye in the primary position.

- Medial rectus is an adductor, lateral rectus is an abductor. They have no secondary actions.
- Superior oblique is primarily an incyclotortor, but when adducted to 54° its action is depression.
- Inferior oblique is primarily an excyclotortor and an elevator in adduction.
- Superior rectus causes elevation, which is maximal when the eye is abducted at 24° and incyclotorsion (maximal on adduction).
- Inferior rectus causes depression, which is maximal at 24° of abduction and excyclotorsion (maximal in adduction).

The eye can be said to move about Fick's axes (Fig. 2.22). This includes a horizontal *x*- and *y*-axis sitting at 90° to each other and a vertical *z*-axis. Secondary gaze positions are achieved by rotation around the *x*- or *z*-axes, whereas tertiary gaze positions are achieved by simultaneous rotation about the *x*- and *z*-axes.

The concept of yoked muscles, Hering's law (versions), and vergence

Yoked muscles (Fig. 2.23) contract to move both eyes in the same direction, for example the right lateral rectus and left medial rectus move the eyes to the right. The superior oblique on the right and inferior rectus on the left move the eyes down and to the left.

In versions the eyes move in the same direction.

Hering's law states that in all voluntary conjugate movements equal and simultaneous innervations flow from ocular motor centres to the muscles establishing the direction of gaze.

Vergence is movement of the eyes in opposite directions. It is slower than versions and is stimulated by retinal disparity.

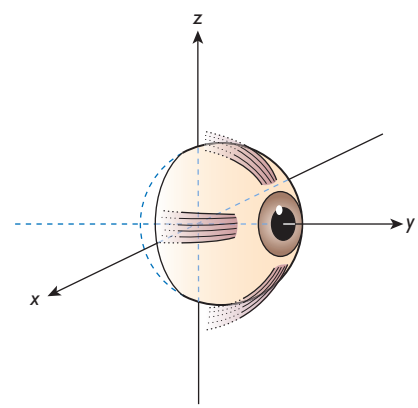


Fig. 2.22 Fick's axes.
With permission from Neil Modi.

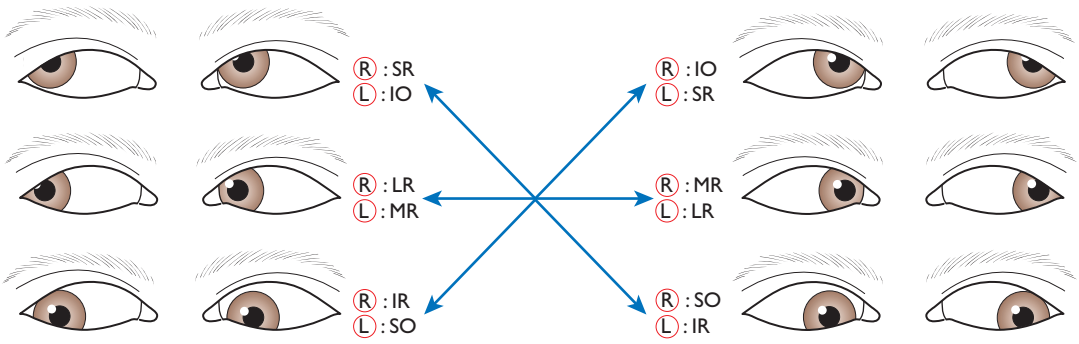


Fig. 2.23 Yoked muscles.
With permission from Neil Modi.

Control of eye movements, smooth pursuits, and saccades

Eye movements are controlled by various different inputs. The superior colliculus is retinotopically mapped and is connected to the extraocular muscle nuclei. The cerebellar vermis is involved in initiation of eye movements.

In addition, the semicircular canals detect rotator head movements and connect with the superior and medial vestibular nuclei. The utricle and saccule detect head tilt and connect with the inferior and medial nuclei. The floculonodular lobe of the cerebellum mediates vestibular inputs to the oculomotor system.

Smooth pursuit movements are designed to keep an object of interest on the fovea and are initiated by an area in occipitoparietal region. They have a velocity of up to 100° per second, a latency of 125 ms, and make up the slow component of optokinetic nystagmus. Lesions of the occipitoparietal region can lead to disruption of smooth pursuit to the ipsilateral side, leading to asymmetrical optokinetic nystagmus (Cogan's law).

Saccades are designed to place an object of interest in the peripheral visual field on to the fovea and are initiated by the frontal cortex (frontal eye fields). They are the fastest of all eye movements, with a velocity of up to 400° per second.

Fixation movements are small movements designed to move the retinal image at regular intervals to prevent image fade due to bleaching of photoreceptors. This is known as Troxler's phenomenon.

Optokinetic nystagmus is a biphasic movement that can be tested using a striped drum revolving at 30–100° per second. The smooth pursuit component is a compensatory movement and is followed by a quicker saccadic movement.

Optokinetic drums can be used to crudely assess visual acuity and in cases of hysterical visual loss. The presence of a central scotoma will increase the frequency at which optokinetic nystagmus can be elicited.

Vestibulo-ocular reflexes and caloric testing

When a patient's head is held backwards at 60° the horizontal canal predominates and caloric testing is carried out.

Warm water causes the endolymph to rise in the horizontal canal, stimulating the end organ. This would be equivalent to an ipsilateral head turn causing a smooth eye movement to the contralateral side and a fast saccadic corrective movement to the ipsilateral side.

The opposite is true if cold water is used. This is depicted in Floren's law, which states that stimulation of a semicircular canal leads to nystagmus in the plane of that canal. The direction refers to the direction of the fast saccadic component.

Fixation on an object and an optokinetic stimulus will reduce this reflex whereas scotopic conditions and visual blur will stimulate the response. Reduced visual acuity has no effect.

Binocular vision and stereopsis

The fusion of images to create binocular single vision and stereopsis (depth perception) requires certain sensory inputs:

- Corresponding points on both retinas (i.e. the same angular distance from the two foveal centres). This is known to have zero binocular disparity. These points are connected to the same areas of the visual cortex. The horopter (Fig. 2.24) is a line in space connecting a set of points whose binocular disparity is zero. Images found anywhere along the border of a horopter known as Panum's area will be perceived as single even if they fall on slightly non-corresponding retinal points. This requires normal functioning extraocular muscles. However, remember that a patient with a paretic muscle will still retain binocular single vision in certain fields of gaze (away from the action of the paretic muscle).
- A certain proportion of points that are non-corresponding allow integration of information from corresponding and disparate points. This allows for sensory fusion of images with slight binocular disparity but that still fall inside Panum's area.

Stereopsis is the ability to see an object in three dimensions. Binocular single vision is not a prerequisite for stereopsis. Depth perception can be possible even though retinal

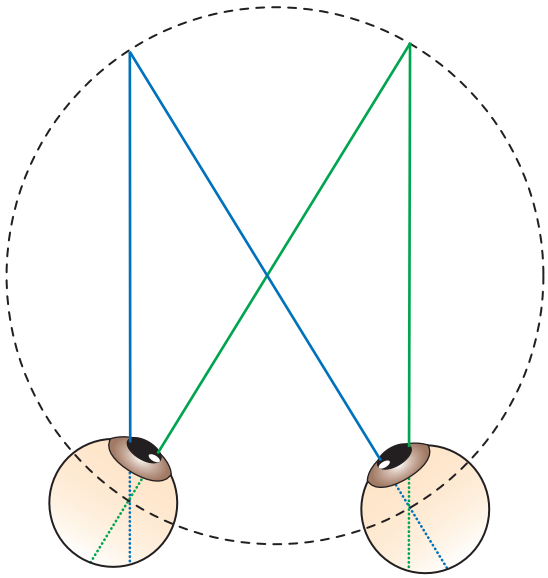


Fig. 2.24 The horopter field of vision.
With permission from Neil Modi.

images are sufficiently disparate to be seen as double. Resolution of stereopsis allows disparities of 10° of arc to be detected but is poor beyond 20° from the fovea and varies with object size (detecting disparity is easier with larger objects). Stereoacuity is a measure of stereopsis.

Clinically, measurements of stereopsis include the following:

- A stereoscope measures the angle of convergence required to fuse images from two slightly dissimilar objects using an instrument incorporating two base-out prisms. However, light and shade or monocular cues provide image disparity and can skew results.
- Random dot stereogram, the random dot E test, and the Frisby test measure stereoacuity. These remove the error from monocular cues and use a camouflaged object in which elements that are non-resolvable monocularly are presented in a random pattern at different disparities. Ability to perceive depth and form is assessed. The limits of stereoacuity are in the region of 4° of arc, which is equivalent to an image disparity less than the diameter of a cone.

To a point, stereoacuity improves the further away the object is from the observer, probably due to the presence of monocular cues. In addition an increase in duration of stimulus improves stereopsis.

Diplopia: terminology

Diplopia is the sensation produced by stimulating two points outside Panum's area. Diplopia can be physiological or non-physiological (heterotropia). A heterotropia is a manifest deviation. Heterophoria is a latent deviation. Concomitant means that the size of the deviation does not vary with the direction of gaze. Incomitant means size does vary with position of gaze. An uncrossed or homonymous diplopia is produced by the image of an object distant to the fixation site falling on the nasal retina. A crossed diplopia (seen in exotropia) is caused by stimulation of the temporal retina.

Retinal rivalry and ocular dominance

Retinal rivalry describes simultaneous perception by each eye individually without fusion of the images. Ocular dominance refers to the preferential use of one eye when performing monocular activities. These effects are thought to occur on a retinal and cortical level.

Visual perception

Visual perception is the end product of processing by the cortex of sensory responses made by the retina to visual stimuli. Visual perception has been covered in previous sections where appropriate, e.g. colour vision and stereopsis. In addition, spatial perception and measuring by eye are covered here.

Spatial perception

Images of objects are projected onto definite positions in space. In spite of observer movement an object will appear fixed in position.

It is thought that this image stabilization is achieved by compensatory events both at a retinal level through selective inhibition of ganglion cells and at a cortical level.

In addition, if object or observer is moving, the observer must integrate direction of movement. This is only accurately achieved by messages from proprioceptors from extraocular and head and neck muscles.

In oscillopsia image stabilization is lost and the patient experiences 'retinal slip'. Here the defect may lie in the vestibulo-ocular reflexes or cervico-ocular reflex.

Spatial perception also involves an element of memory. For example, if an object is illuminated and then viewed in the dark the eyes will adopt a position close to fixation.

Aubert's phenomenon and measuring by eye

The accuracy of measuring by eye is high: for detecting the orientation of horizontal lines the error rate is 0.2% and for vertical lines it is 1.0%. This is better than for oblique lines due to cortical processing. There is evidence that the visual cortex contains specific cells responsive to line orientation. These are essential for the analysis of form and shape.

Aubert's phenomenon is the ability to perceive a line in its true position despite a change in observer position (Fig. 2.25).

A vertical bright light viewed in a completely dark room will tilt to the left if the head is slowly tilted to the right (B). If the head is tilted suddenly or if the line is viewed in the light, the line appears upright in its normal position (C). Thus information on retinal position is fed via the semicircular canals to the object-positioning centre. (A) is the resting position.

The retina uses horizontal and vertical meridians to pinpoint objects in space. These can be displaced by eye or head movement, hence compensatory mechanisms are required to reinterpret the position and project the real position and direction of the object in space.

Pattern and directions are important in estimating dimensions. There are many optical illusions that are produced where objects appear to be different lengths or sizes if one of them has been altered by the addition of other visual cues.

Binocular viewing of objects permits depth perception. Depth perception can be aided from what are known as monocular cues such as comparing the size or colour of objects, overlapping edges, light/dark shades, and texture, and by movement of the observer's head.

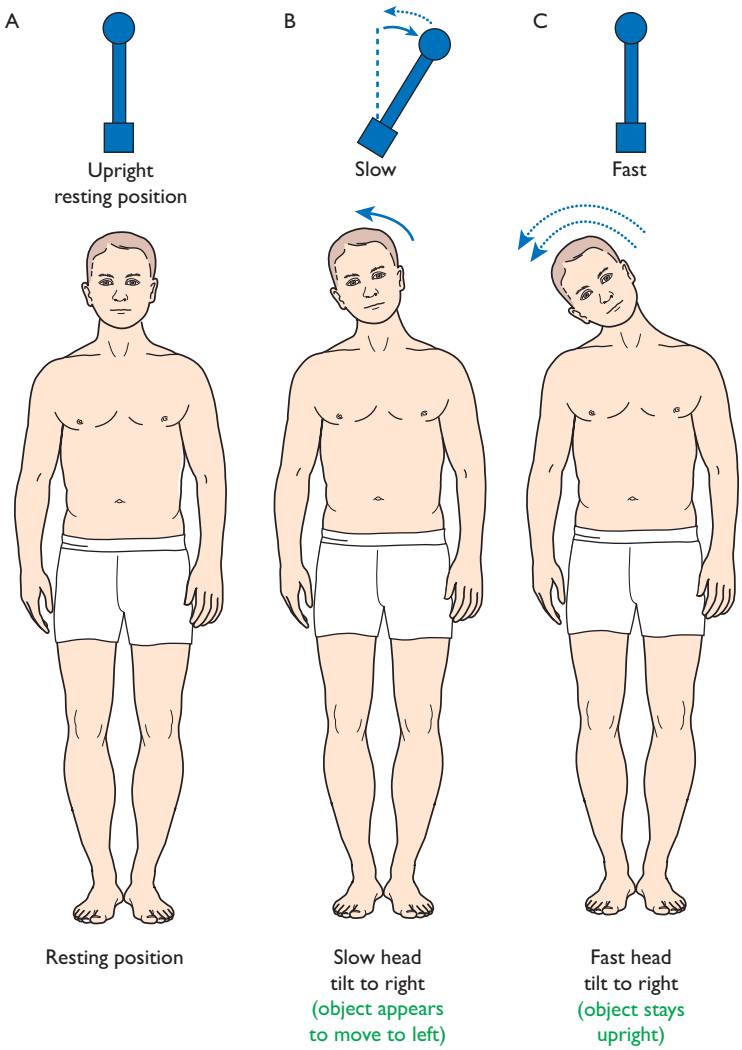


Fig. 2.25 Aubert's phenomenon.

With permission from Neil Modi.

Visual pathways, including magnocellular and parvocellular pathways

The lateral geniculate nuclei

Retinal ganglion cells and fibres from the visual cortex synapse in the lateral geniculate nuclei (LGN). These are paired thalamic nuclei that are retinotopically mapped with information from the superior and inferior retina relaying to the medial aspect. The LGN are made up mostly of X and Y cells, which exhibit a similar centre/surround organization as retinal ganglion cells.

Magnocellular ganglion cells are mostly linked to rod photoreceptors. They are large and fast, dealing with motion detection and luminance. They synapse with Y cells in layers 1 and 2 of the LGN.

Parvocellular or P ganglion cells are small and slow, and are linked to cones mainly at the fovea. They deal with spatial resolution and colour. They synapse with X cells in layers 3–6 of the LGN. Layers 3–6 fuse anteriorly, forming four ventral layers. Layers 2, 3, and 5 receive input from the ipsilateral eye.

The striate cortex

The striate cortex (otherwise known as V1 or Brodmann's area 17) is found in the occipital cortex. Cortical cells, like all other neurons in the visual pathway, exhibit a receptive field structure.

The striate cortex consists of six layers.

- Layer II cells synapse with the contralateral visual cortex via the corpus callosum.
- Macroscopically the striate cortex has one visible layer IV, also known as the stripe of Gennari. Layer IV is the thickest layer, receiving efferents from the LGN and containing mainly stellate cells.
- Layer V cells relay to the superior colliculus, whose connections with the medial longitudinal fasciculus enable ocular movements to be coordinated.
- Layer VI sends efferents to the LGN and contains mainly pyramidal cells. These efferents synapse with ganglion cells and interneurons in the synaptic glomeruli of the LGN.

There are thought to be four main cell types in the visual cortex:

- concentric cells with a centre surround receptive field
- simple cells found in layer IV and with receptive fields in parallel bands
- complex cells with binocular input, found in layers III and V, which respond to a stimulus of specific orientation moving in a specific direction
- hypercomplex cells, which have a further function of responding to a stimulus of specific length, orientation, and direction, known as 'end inhibition'.

Visual field

Clinically it is important to remember that the superior half of the visual field (inferior retina) corresponds to the area below the calcarine sulcus.

At the most anterior end of the sulcus there is an area of cortex that corresponds to the most peripheral nasal retina. A lesion here will produce a defect in the temporal crescent of the contralateral eye only.

The posterior end of the sulcus lies at the watershed between the middle and posterior cerebral arteries, and the area of the cortex devoted to the macula is large.

The prestriate cortex

The striate cortex is connected directly to the prestriate cortex (areas V3 to V8) and via V2. Each of these areas has one or more specific functions, for example:

- V3 cells are selective for form
- V4 cells respond to colour, line orientation, and shape/form detection
- V5 cells respond to motion and are directionally sensitive.

V2 is organized into areas of thin (colour detection) and thick (motion detection) stripes. Form-selective detectors are present in thick and thin stripes.

CHAPTER 3

Biochemistry and genetics

The cell

The cell is the smallest functional unit of living organisms. Cells are classified as:

- prokaryotic (absence of nucleus) or
- eukaryotic (nucleus present).

Viruses are prokaryotic whereas human cells are eukaryotic.

The cells of the eye contain multiple strands of deoxyribonucleic acid (DNA) in the membrane-bound nucleus and a number of other membrane-bound structures, called organelles. These organelles allow cellular processes to be compartmentalized, allowing greater cellular specialization and diversity.

The largest compartment is the cytosol or cytoplasm, which makes up around half the volume of the cell. It is composed of dissolved ions, metabolites, proteins, and ribonucleic acids, and is the site of many metabolic pathways.

Plasma membrane

Cell contents are compartmentalized by a plasma membrane. This is a selective two-way barrier to passive diffusion. It is composed of a phospholipid bilayer in which membrane proteins float.

- The bilayer forms because of the physicochemical properties of the polar phospholipids. The polar groups are external and the hydrophobic groups face each other on the inside of the bilayer.
- Transporters are proteins in the membrane which allow active transport.
- Receptors are found in the membrane. These proteins often form a three-part structure, with an extracellular component, a transmembrane component, and an intracellular component, which is coupled to the second messenger system.
- The cell surface is irregularly pitted with invaginations and evaginations. These include specialized structures for endocytosis and exocytosis.
- Other specializations occur depending on the role of the cell, for example the rod photoreceptor develops an evagination of the plasma membrane, which folds upon itself many times to form stacked membranous discs.

Mitochondria

- These are small (2 μm) oval organelles that have their own complement of DNA (but no histones), ribosomal ribonucleic acid (rRNA), and transfer RNA (tRNA), so they can generate their own mitochondria-specific proteins (Fig. 3.1).
- They contain the sites of oxidative phosphorylation and enzymes of the citric acid cycle and fatty acid β-oxidation pathways.
- They are bound by a phospholipid bilayer and are composed of two compartments.
- The outer membrane contains large channel-forming proteins (porin) that are permeable to molecules up to 5000 Daltons (Da). The membrane also contains enzymes involved in mitochondrial lipid synthesis and those that convert lipid into forms that are metabolized in the matrix.
- The intermembrane space contains the proteins responsible for the transport of metabolites between the two compartments and also between the cytosol and the outer compartment.
- The inner membrane is thrown into folds or cristae, which increases its surface area. It carries proteins with three functions: to carry out the reactions of the electron transport chain, the ATP synthase which makes ATP in the matrix, and transport proteins that allow the passage of metabolites into and out of the matrix. An H^+ gradient drives the ATP synthase, so the membrane is impermeable for this to be established.
- Mitochondria have an important role in metabolism. They are the site of ATP/GTP formation and are involved in ATP breakdown. For this, the inner compartment is composed of a concentrated mixture of hundreds of enzymes, including those responsible for oxidation of pyruvate and fatty acids, and for the citric acid cycle.
- Mitochondria also function as a calcium store (mainly as calcium phosphate) and they are involved in apoptosis.

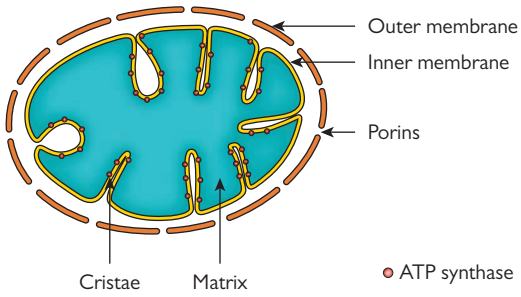


Fig. 3.1 Mitochondrion.

The nucleus

Every eukaryotic cell contains a nucleus (except red blood cells).

The nuclear matrix contains a complete copy of all genomic DNA. This is packaged with histone proteins into chromosomes.

The nucleolus is the site of ribosomal RNA synthesis. Nucleoli develop during the late stages of mitosis in association with specific regions on the chromosomes (nucleolar organizer centres). They represent the site of rRNA synthesis and assembly.

The nucleus is surrounded by a nuclear envelope, which consists of a double layer of inner and outer nuclear membrane, between which is the periplasmic space.

The outer nuclear membrane is continuous with the rough endoplasmic reticulum (ER), and the periplasm is continuous with the ER lumen.

The nuclear envelope contains nuclear pores, which allow the two-way passage of small molecules or proteins. They are found at points of contact between the inner and outer membranes. Molecules up to 60 kDa can pass freely through these pores, but transport of larger molecules is ATP dependent and requires receptor-mediated recognition of a nuclear targeting sequence by the pore complex.

Chromatin is the term for long strands of DNA, RNA, and associated nucleoproteins. This is dispersed throughout the nucleus, becoming more compact during mitosis or meiosis. Heterochromatin is electron dense and distributed around the periphery of the nucleus and in discrete masses within the nucleus. The DNA is closely associated with nucleoproteins but is not actively undergoing RNA synthesis. Euchromatin is electron translucent and is thought to be DNA that is actively involved in RNA synthesis.

Endoplasmic reticulum

- The cell is filled with a cytoplasmic gel, which is crossed by a network of bilayered phospholipid membranes, known as the ER (Fig. 3.2). They are constantly being formed and reformed and are similar to the plasma membrane in structure.

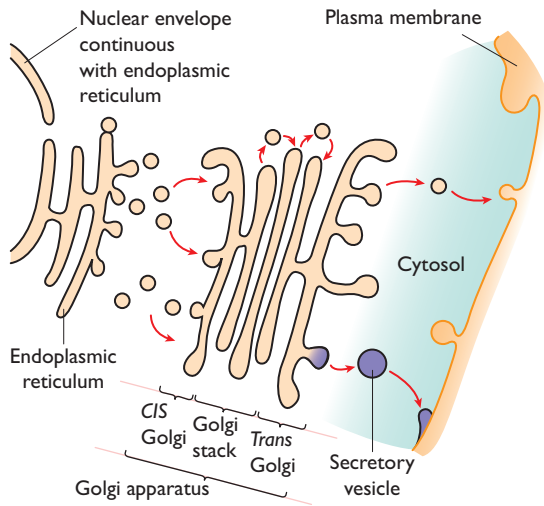


Fig. 3.2 Rough endoplasmic reticulum and Golgi apparatus.

- Rough ER has ribosomes (made of RNA and protein), responsible for secretory, lysosomal, and membrane protein assimilation on its surface, and is more highly developed in secretory cells such as the lacrimal gland acinar cell.
- In the rough ER, newly synthesized proteins come from the ribosome and enter the ER, where they are post-translationally folded. These are made ready for secretion via the Golgi apparatus by vacuolar budding and fusion with the plasma membrane for exocytosis.
- Smooth ER is the site of synthesis for molecules such as lipids, triglycerides, and steroids. It is highly developed in cells such as the RPE and meibomian gland cells.
- During telophase, when a series of flat vesicles surround the chromosomes and fuse at their edges, the ER forms the nuclear envelope.

Golgi apparatus

The Golgi apparatus consists of stacks of flattened smooth membrane sacs and vesicles, and is responsible for post-translational modification of secretory and membrane proteins by sequential glycosylation reactions (Fig. 3.2).

Lysosomes and peroxisomes

Lysosomes contain hydrolytic enzymes that break down carbohydrate, lipids, and proteins. These activities are sequestered into distinct regions in order to protect other regions from breakdown.

Lysosomes are formed by directly budding off from the Golgi complex. They then fuse with intracellular vesicles,

which contain material that has been taken up by phagocytosis and endocytosis. Material that has been taken into the cell is therefore broken down.

Peroxisomes are also small organelles, but they contain enzymes (e.g. catalase) that function in oxidative reactions which produce hydrogen peroxide and other organic peroxides.

Cytoskeleton

This is a dynamic structure, responsible for changes in cell shape and movement. It is important in the movement of intracellular organelles and other cytoplasmic components.

Cytoskeletons are linked to the extracellular matrix (ECM) by specialized junctional complexes. The connections are made by special transmembrane proteins. On the intracellular side, the cytoskeleton is attached to these proteins by accessory proteins such as α-actinin, talin, and vinculin.

Cytoskeletons are based on three basic filamentous structures:

- microfilaments (7 nm diameter), actin, dynamic, contractile
- microtubules (24 nm diameter), tubulin, mobile, dynamic, cellular highway
- intermediate filaments (10 nm diameter), robust, durable.

They form helical filaments with polarity that have several isoforms that are tissue specific. Cytoskeletons may be composed of one or more of these three filamentous elements. Accessory proteins are also involved, which are important in controlling the length of the filamentous structures and their association with other protein complexes, organelles, membranes, and the ECM. They also permit modification of the cytoskeleton in response to metabolic cellular changes and signals, such as changes in Ca^{2+}.

Microfilaments are formed from actin, a flexible protein. G-actin, the monomer, can form stable dimers or trimers. Each monomer has a polarity so they associate head-to-tail. Elongation can occur at either end. G-actin is bound to either ATP or ADP. The rate of binding of G-actin–ATP to filaments is much faster and this results in hydrolysis of ATP. Hydrolysis allows dissociation of G-actin–ADP at the other end of the filament, allowing movement of the cell or structure. Actin filaments are formed from two-stranded polymers of diameter 5–9 nm. Actin has a role in cell–cell interaction in adherens junctions (zona adherens), cell–substrate interactions in focal adhesions, and membrane attachment in striated muscle. It is therefore most concentrated below the cell membrane in the cortex of the cell.

Microtubules are cylindrical polymers of 13 tubulin subunits with a diameter of 25 nm. Each unit is an αβ-dimer which polymerizes bidirectionally. α and β subunits alternate, giving the structure a polarity. Treadmilling can occur, whereby one end of the filament can grow and the other become shorter. Hydrolysis of bound GTP allows treadmilling to occur and dissociation of the polymer, and thus one unit can dissociate from the other end, in much the same way as actin.

Microtubules are more rigid than actin, and therefore are long and straight. One end is attached to a microtubule organizing centre—a centrosome complex—which is necessary for polymerization. Centrosomes are composed of a pair of centrioles at right angles to each other, which consist of tubules further composed of nine microtubules. In cells, the slow-growing end is capped to stop dissociation. Polymerization is inhibited by raised Ca^{2+} and low temperature. Microtubules have a role as structural components within cells and are involved in many cellular processes, including mitosis, cytokinesis, and vesicular transport.

Intermediate filaments are of intermediate diameter (10 nm), hence the name. There is greater heterogeneity than in microfilaments and microtubules. A greater variety of proteins make up the intermediate filaments, e.g. nuclear laminin, vimentin, keratins, and neurofilaments.

Each intermediate filament protein has an α-helix core of 310 amino acids which forms a dimer with another intermediate filament protein. These dimers polymerize to form a rope-like intermediate filament. Intermediate filaments have a role in cell–cell interactions in desmosomes and cell–substrate interactions in hemidesmosomes.

Cellular interaction

Cells interact with one another via junctions and adhesion molecules. Junctions are found between cells and between cells and the ECM.

There are three groups of cell junction:

- tight (occluding) junction
- anchoring junction
- communication (gap) junction.

A junctional complex consists of a tight junction, an adherent junction, and a desmosome.

Tight (occluding) junctions are composed of the proteins occludin and claudin. They exist between neighbouring cells, sealing them together so that water-soluble molecules cannot pass between cells. This is an important feature of all epithelia.

Anchoring junctions are important in maintaining tissue integrity and are most common in cells under stress, such as cardiac muscle. They link the cellular cytoskeleton with that of the neighbouring cell or the ECM. They consist of an intracellular attachment protein, transmembrane-linked glycoproteins, and an ECM or transmembrane-linked glycoprotein on another cell. If joining actin cytoskeletons together, these are called adherens junctions. Desmosomes act as anchoring sites for intermediate filaments in cell–cell adhesion and provide tensile strength.

The cell adhesion proteins of the desmosome are desmoglein and desmocollin. These are members of the cadherin family of cell adhesion molecules. These transmembrane proteins bridge the space between adjacent epithelial cells by way of homophilic binding of their extracellular domains to other desmosomal cadherins on the adjacent cell. Since these are Ca^{2+} dependent, changing the concentration of Ca^{2+} alters their interactions. On the cytoplasmic side of the plasma membrane, there are two dense structures called the outer dense plaque (ODP) and the inner dense plaque (IDP). These are spanned by the desmoplakin protein. The ODP is where the cytoplasmic domains of the cadherins attach to desmoplakin via plakoglobin and plakophillin. The IDP is where desmoplakin attaches to the intermediate filaments of the cell. Hemidesmosomes attach cells to the ECM. Instead of using cadherins, hemidesmosomes use integrin cell adhesion proteins. Hemidesmosomes are asymmetrical and are found in epithelial cells, generally connecting the basal aspect of cells to the basement membrane.

Communicating (gap) junctions do not seal membranes together or restrict the passage of material between membranes. They allow inorganic ions carrying current and water-soluble molecules to pass from one cell to another, thus permitting electrical and metabolic cell coupling. A gap junction is composed of two connexons (hemi-channels) that cross the intercellular space. Each connexon is formed from six connexins. Molecules of up to 1000 Da can pass through a pore.

Cell adhesion molecules

Cadherins (calcium-dependent adhesion molecules) are glycoproteins that form part of a desmosome and are involved in cell adhesion. They take part in calcium-dependent homophilic interactions. Since these are Ca^{2+} dependent, changing the concentration of Ca^{2+} alters their interactions.

Integrins are receptors that mediate attachment between a cell and the tissues surrounding it, which may be other cells or the ECM. They are obligate heterodimers containing two distinct chains, called the α (alpha) and β (beta) subunits, which span the plasma membrane.

Their two main functions are attachment of the cell to the ECM and signal transduction from the ECM to the cell. They are also involved in immune responses, cell migration, and binding to the cell by certain viruses.

Integrins attach the ECM to the cytoskeleton (microfilaments) inside the cell. Inside the cell, they bind to ligands such as fibronectin, vitronectin, collagen, and laminin. An important role of surface integrins is their role in cell migration. Cells adhere to a surface through their integrins. During movement, the cell makes new attachments to the substrate at its front and concurrently releases those at its rear. When released from the substrate, integrin molecules are taken back into the cell by endocytosis; they are transported through the cell to its front by the endocytic cycle, where they are added back to the surface. In this way they are cycled for reuse, enabling the cell to make fresh attachments at its leading front.

Integrins also play an important role in cell signalling. Connection with ECM molecules can cause a signal to be relayed into the cell through protein kinases that are indirectly and temporarily connected with the intracellular end of the integrin molecule, following shape changes directly stimulated by ECM binding.

Selectins are a family of cell adhesion molecules. There are three types:

- L-selectin is expressed on leukocytes
- E-selectin is expressed on activated endothelial cells
- P-selectin is expressed on activated platelets and endothelial cells.

They are Ca^{2+} dependent and undergo heterophilic binding to carbohydrate ligands, initiating leukocyte–endothelial interactions. Expression is induced by local chemical mediators such as histamine, thrombin, cytokines, and tumour necrosis factor (TNF)-α. When expressed, the lectin recognizes specific oligosaccharides on the surface of neutrophils, allowing leukocytes to stick to the endothelial lining of blood vessels, resulting in the characteristic 'rolling' until integrins are activated.

Cell signalling

Intercellular signalling is classified by function into three groups:

- hormones: secreted by specialized cells, usually from endocrine glands, into the circulation, in which they travel to their target cells
- local mediators (paracrine signalling): released by many cells into the local extracellular fluid; usually short lived and not often found in the general circulation
- neurotransmitters: released from nerve terminals into synapses; they bind to receptors on post-synaptic membrane, resulting in targeted transfer of information.

For signalling to occur, a ligand binds specifically to a receptor site. This may activate a receptor, in which case the ligand is termed an agonist. If a ligand binds without resulting in activation, it is called an antagonist. If a ligand binds but does not produce the maximal effect possible for that receptor, it is called a partial agonist. If unbound, receptors remain functionally silent.

Hydrophilic signalling molecules cannot easily cross the lipid bilayer. Hence, the signalling molecule must be on the extracellular cell surface. Hydrophobic signalling molecules can cross the lipid bilayer by diffusion, but this nevertheless

requires an intracellular receptor, e.g. steroid and thyroid hormones.

Receptors are classified by the signalling molecule they recognize, e.g. acetylcholine receptors.

Receptor binding:

- is specific
- is usually reversible
- results in conformational change of the molecule
- does not result in modification of the ligand.

There are four main mechanisms by which receptors act when stimulated:

- membrane-bound receptors with integrated ion channels
- membrane-bound receptors with integral enzyme activity
- membrane-bound receptors that couple to effectors through transducing proteins
- intracellular receptors.

On binding to a membrane-bound receptor with an integrated ion channel, a conformational change occurs in the ion channel. This results in the opening of a gated channel, permitting the flow of ions down a concentration gradient. G-proteins may also act by opening ion channels and these types of receptors may also be present within cells.

Agonists binding to membrane-bound receptors with integral enzyme activity results in a conformational change in the receptor that activates an intrinsic enzyme activity within the cytoplasmic domain of the receptor. The result of activation of intrinsic enzyme activity is the formation of an intracellular chemical 'second messenger' (e.g. cyclic guanosine monophosphate, cGMP) or direct action on cytoplasmic proteins by phosphorylation or dephosphorylation.

Tyrosine-kinase-linked receptors are an example of receptors with integral enzyme activity. Hormones binding to these activate tyrosine-kinase, which autophosphorylates (transfers a phosphate from ATP onto itself) tyrosine residues on the cytoplasmic region of the receptor. These residues are recognized by transducing proteins or enzymes containing phosphotyrosine recognition sites. These in turn become activated allosterically or by tyrosine phosphorylation by the receptor kinase. Thus an extracellular signal produces an intracellular chemical effect.

Several transmembrane domain receptors that are not bound to ion channels and have no integral activity can couple to effector molecules via a transducing molecule, known as GTP-binding regulatory protein (G-protein). A range of signalling molecules use this type of receptor, e.g. muscarinic acetylcholine receptors, adrenoceptors, dopamine receptors, 5-hydroxytryptamine receptors, opioid receptors, peptide receptors, light receptors (rhodopsin), and smell and taste receptors. Receptor activation causes conformational change in the G-protein, which in turn causes the release of GDP. GTP takes its place and the α-GTP and βγ complex dissociate from the receptor and go on to interact with effector molecules, which may be enzymes or ion channels. The α-subunit has GTPase activity, which hydrolyses GTP and returns the α-subunit to its inactive configuration and the α-GDPβγ heterotrimer reforms. The rate of GTP hydrolysis therefore controls the duration of action of the G-protein.

Hydrophobic ligands which can penetrate the lipid bilayer can stimulate intracellular receptors. Examples are steroid hormones, cortisol, oestrogen, testosterone, and thyroid hormones. When inactive, the receptors are stabilized by association with heat shock or chaperone proteins. On activation, these receptors dissociate from their stabilizing proteins and translocate to the nucleus. Here, they control the transcription of certain genetic sequences. The effects of these receptors are therefore slower as they rely on transcription and translation to occur before they can exert their effect.

The signalling molecules are present usually in a low concentration, but their signal is amplified as a result of each receptor exerting effects on many substrate molecules. A cascade of such events can result in further amplification.

The effect of activation of many receptors is not direct, and cascades of intracellular cell signalling events are initiated via cell signalling pathways. Receptor activation can lead to activation of an enzyme effector which, in turn, produces an intracellular molecule (second messenger). For example, several receptors activate the integral plasma membrane enzyme adenylyl cyclase. Receptors that stimulate adenylyl cyclase include β-adrenoceptors, dopaminergic D_1, glucagon, and growth hormone receptors. Receptors that inhibit adenylyl cyclase include α_2-adrenoceptors and dopaminergic D_2 receptors. Adenylyl cyclase converts ATP to cyclic adenosine monophosphate (cAMP) and pyrophosphate. cAMP activates cAMP-dependent protein kinase, which phosphorylates target proteins to stimulate or inhibit these. cAMP is hydrolysed by cellular cAMP phosphodiesterase to restore basal levels once stimulation at the receptor has stopped.

Similarly, cGMP is produced from GTP by guanylyl cyclases. cGMP activates a specific protein kinase (protein kinase G). cGMP is important in the relaxation of smooth muscle cells in response to atrial natriuretic peptide (ANP) and NO (endothelium-derived relaxing factor). Reduction in cGMP signals detection of light in retinal cells. cGMP is hydrolysed by cGMP phosphodiesterase to guanosine-5′-monophosphate (5′-GMP).

Protein phosphorylation or dephosphorylation caused by protein kinases or phosphatases, respectively, causes a conformational change in a protein that activates or inhibits it, depending on the protein.

An example of target proteins that are phosphorylated by an activated protein kinase is the phospholipid, phosphatidylinositol 4,5-bisphosphate ($PtdInsP_2$). Activated $PtdInsP_2$ phosphodiesterase acts on $PtdInsP_2$ to give inositol 1,4,5-triphosphate ($InsP_3$) and diacylglycerol (DAG), which stays in the membrane.

$InsP_3$ is hydrophilic and diffuses to the ER, where it acts on $InsP_3$ receptors. These second messenger-gated receptors

act as Ca^+ channels, resulting in Ca^+ release from the ER lumen. The Ca^+ also acts as a second messenger. DAG and Ca^+ activate protein kinase C (PKC). This reduces the concentration-dependence on Ca^{2+}.

A number of G-protein-coupled receptors activate the above pathway, e.g. muscarinic M_1 and M_3 receptors, H_1 receptors, α_1-adrenoceptors, and 5-hydroxytryptamine ($5HT_2$)-receptors. They are important in vascular, gastrointestinal, and airways smooth muscle, platelet aggregation, and mast cell degranulation.

There is a low concentration of cytosolic Ca^{2+} compared to the extracellular fluid. Increase in cellular Ca^{2+} causes several processes, including contraction, secretion, and glycogenolysis. This increase could be mediated by ligand-gated *N*-methyl-D-aspartate (NMDA) receptors and $InsP_3$ receptors in the ER.

Cellular processes

Fig. 3.3a shows the three main ways in which food is broken down to yield ATP for cellular metabolism.

Glycolysis

Glycolysis is the metabolic pathway that converts glucose ($C_6H_{12}O_6$) into pyruvate ($C_3H_6O_3^-$) (Fig. 3.3b). The free energy released in this process is used to form the high-energy compounds ATP and NADH. It involves a sequence of 10 reactions with 10 intermediate compounds. Most monosaccharides, including fructose, glucose, and galactose, can be converted into one of these intermediates.

Glycolysis occurs with variations in most organisms, including both aerobic and anaerobic. The most common type of glycolysis is the Embden–Meyerhof pathway:

$$\text{D-glucose} + 2[NAD]^+ + 2[ADP] + 2[P]_i \rightarrow$$
$$2\ \text{pyruvate} + 2[NADH] + 2H^+ + 2[ATP] + 2H_2O$$

As a result of glycolysis, each molecule of glucose produces two molecules of ATP. NAD^+ is used as a result of glycolysis, and most cells carry out further processes to reform NAD^+. The final product is ethanol or lactic acid.

There are three regulation points in the glycolysis pathway. The three regulated enzymes are hexokinase, phosphofructokinase, and pyruvate kinase. The reactions catalysed by these enzymes are irreversible. Hexokinase converts glucose into glucose-6-phosphate. This intermediate, instead of continuing through glycolysis, can be converted into glucose storage molecules such as glycogen. This can also happen in reverse, with glycogen breaking down to glucose-6-phosphate and thus entering glycolysis. The second regulation step involves the enzyme phosphofructokinase converting fructose-6-phosphate to fructose-1,6-bisphosphate, which in turn can be converted to glyceraldehyde-3-phosphate and dihydroxyacetone phosphate. The latter can be removed from glycolysis by conversion to glycerol-3-phosphate and then go on to form triglycerides. The reverse can occur and triglycerides can be broken down to fatty acids and glycerol. The latter can be converted to dihydroxyacetone phosphate, which can enter glycolysis.

Passage through the glycolytic pathway is modulated by intra- and extracellular conditions. The rate of glycolysis in the liver is regulated to meet the need for ATP, the need for building blocks of biosynthetic reactions, and the need to lower blood glucose. When blood sugar falls, glycolysis is halted to allow the reverse process of gluconeogenesis.

Aerobic and anaerobic metabolism

Further metabolic processes may be aerobic or anaerobic. Aerobic respiration synthesizes much more ATP (34 additional molecules). This uses pyruvate and NADH + H^+ from glycolysis. The steps are as follows:

- Pyruvate is converted to acetyl-CoA and CO_2 (pyruvate decarboxylation).
- The acetyl-CoA enters the citric acid cycle, where it is fully oxidized to carbon dioxide and water, producing more NADH.
- The NADH is oxidized to NAD^+ by the electron transport chain, using oxygen as the final electron acceptor. This process creates a 'hydrogen ion gradient' across the inner membrane of the mitochondria.
- The proton gradient is used to produce a large amount of ATP in a process called oxidative phosphorylation.

In anaerobic respiration, pyruvate is converted to lactate (the conjugate base of lactic acid) by lactic acid fermentation:

$$\text{pyruvate} + NADH + H^+ \rightarrow \text{lactate} + NAD^+$$

Many single-celled organisms use glycolysis and anaerobic fermentation as their only energy source. They use a wide variety of compounds (e.g. nitrogenous compounds, sulphur compounds) as the terminal electron acceptors in cellular respiration.

The citric acid cycle (tricarboxylic acid cycle or Kreb's cycle)

The citric acid cycle (or Kreb's cycle) (Fig. 3.4) is a series of enzyme-catalysed chemical reactions that uses oxygen as part of cellular respiration. This occurs in the matrix of mitochondrion in eukaryotes. As well as resulting in the formation of ATP, CO_2, and H_2O from pyruvate, it provides precursors for various compounds, including amino acids.

The citric acid cycle starts with acetyl-CoA transferring its two-carbon acetyl group to the four-carbon acceptor

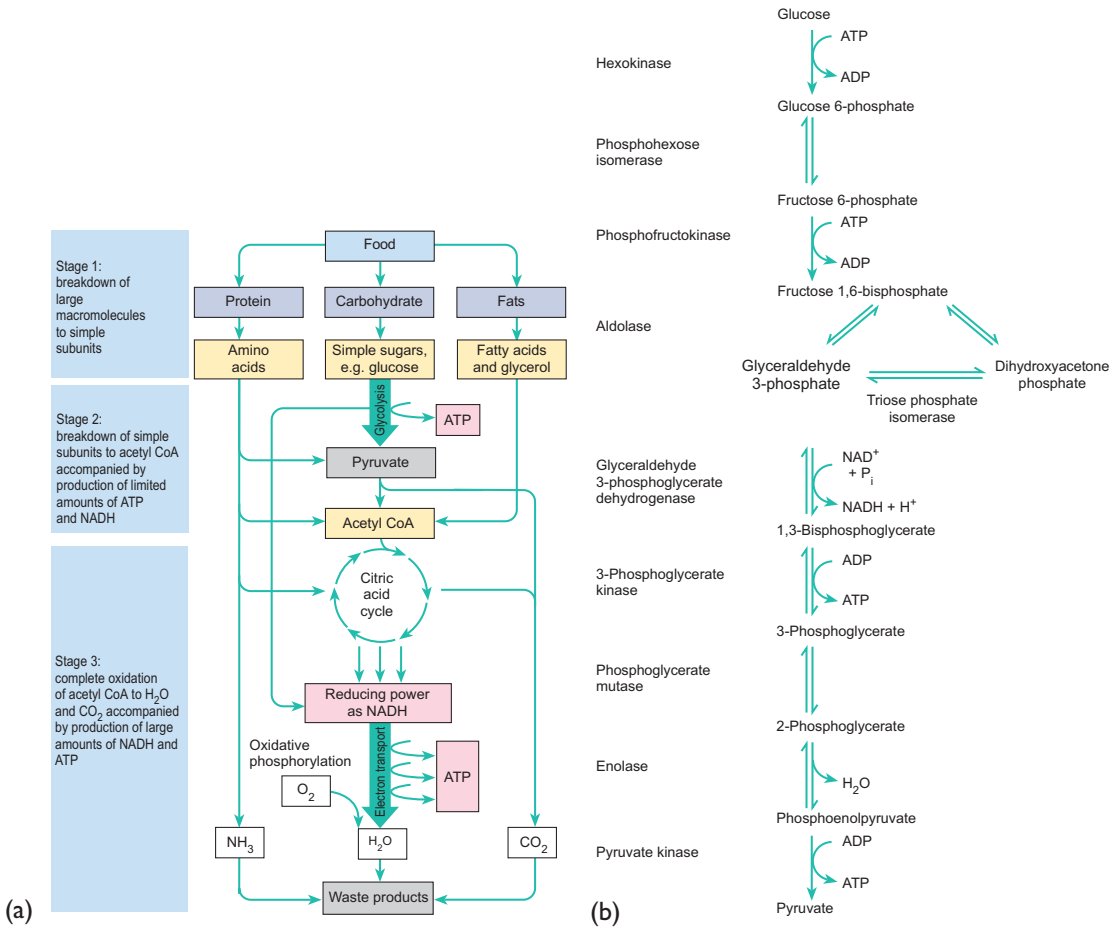


Fig. 3.3 Glycolysis. (a) The three main stages by which foodstuffs are broken down to yield ATP for cellular metabolism. (b) Principal steps in the glycolytic breakdown of glucose.

Reproduced from Gillian Pocock and Christopher D. Richards, *Human Physiology, The Basis of Medicine,* Third edition, Figure 3.7 and 3.8, Page 28, 2006, with permission from Oxford University Press.

compound (oxaloacetate) to form a six-carbon compound (citrate). The citrate goes through a series of transformations, losing two carboxyl groups as CO_2. These carbons originate from what was oxaloacetate. The carbons lost as CO_2 originate from what was oxaloacetate, not directly from acetyl-CoA. Loss of the acetyl-CoA-donated carbons as CO_2 requires several turns of the citric acid cycle. The energy made available by the oxidative steps of the cycle is transferred as energy-rich electrons to NAD^+, forming NADH. For each acetyl group that enters the citric acid cycle, three molecules of NADH are produced. Electrons are also transferred to the electron acceptor Q, forming QH_2. At the end of each cycle, the four-carbon oxaloacetate has been regenerated, and the cycle continues with the addition of a further acetyl-CoA. The regulation of the citric acid cycle is largely determined by substrate availability and product inhibition. Calcium acts as a regulator in this cycle.

The products of the first turn of the cycle are one GTP (or ATP), three NADH, one QH_2, and two CO_2. Each glucose molecule, however, forms two acetyl-CoA molecules so the products from each glucose molecule are two GTP (or ATP), six NADH, two QH_2, and four CO_2. The net result is 30 molecules of ATP per glucose molecule.

Several catabolic pathways converge on the citric acid cycle. It is the third step in carbohydrate catabolism (the breakdown of sugars). As described above, glucose is broken down to pyruvate and then acetyl-CoA. It may then enter the citric acid cycle. Proteins are broken down by proteases to their constituent amino acids. These can then become a source of energy by being converted to acetyl-CoA and then entering the citric acid cycle. Triglycerides are hydrolysed to

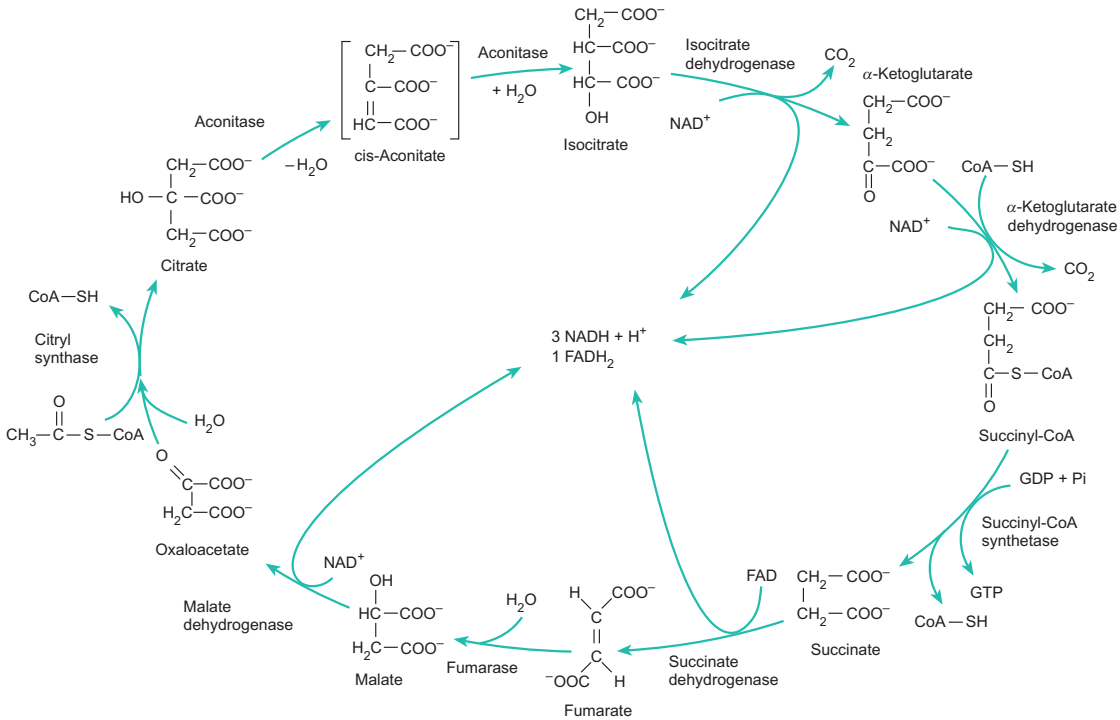


Fig. 3.4 The citric acid cycle.

Reproduced from Gillian Pocock and Christopher D. Richards, *Human Physiology, The Basis of Medicine*, Third edition, Figure 3.9, Page 29, 2006, with permission from Oxford University Press.

break them into fatty acids and glycerol. Glycerol is converted to glucose in the liver by gluconeogenesis. Fatty acids are broken down by β-oxidation, resulting in acetyl-CoA, which is fed into the citric acid cycle.

Oxidative phosphorylation

Catabolic biochemical processes such as glycolysis, the citric acid cycle, and β-oxidation produce the coenzyme NADH. This contains electrons that have a high transfer potential and release a large amount of energy on oxidation. Oxidative phosphorylation extracts the energy (as electrons) from NADH and QH_2, oxidizing them to NAD^+ and Q, respectively, thus allowing the citric acid cycle to continue. The citric acid cycle does not use oxygen, but the process of oxidative phosphorylation does.

Oxidative phosphorylation is a very efficient way of releasing ATP compared to alternative fermentation processes such as glycolysis. The process depends on electron transfer from donors to acceptors in redox reactions. This releases energy, which is used to form ATP. These redox reactions are carried out by a series of protein complexes (electron transport chains) within mitochondria in eukaryotes.

The energy released is harnessed to transport proteins across the inner mitochondrial membrane. This generates a pH gradient and an electrical potential across the membrane (Fig. 3.5). The flow of protons back across the membrane down this gradient, through an ATP synthase, generates an ATP from ADP in a phosphorylation reaction. Oxidative phosphorylation produces reactive oxygen species (ROS) such as superoxide and hydrogen peroxide, which leads to propagation of free radicals.

Free radicals

ROS are highly reactive ions or small molecules that possess unpaired electrons. They are produced in mitochondria as a byproduct of the normal metabolism of oxygen and have a role in cell signalling. In the mitochondria, electrons are passed through an electron-transport chain. The last destination for the electron is oxygen, which is normally reduced to produce water. Some of the oxygen is prematurely and incompletely reduced to form the superoxide radical, O_2^-. Alone, it is not particularly reactive, but can inactivate specific enzymes or initiate lipid peroxidation in its H_2O_2 form. The effects of ROS are minimized but not completely erased by the use of enzymes such as catalase

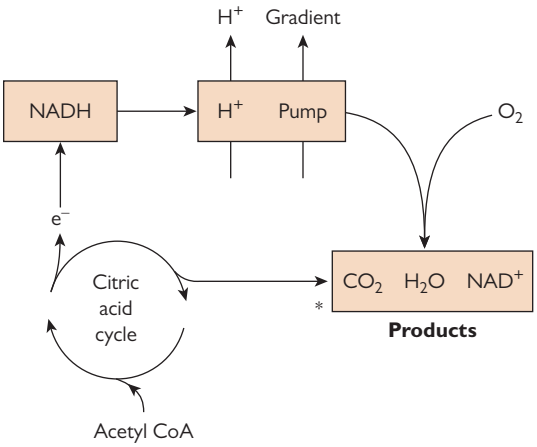


Fig. 3.5 Oxidative phosphorylation in the mitochondrion.

and superoxide dismutase. Small molecule antioxidants such as ascorbic acid (vitamin C), tocopherol (vitamin E), uric acid, and glutathione also play important roles as cellular antioxidants.

Generally, the harmful effects of ROS on the cell are most often:

- damage to DNA
- oxidation of polydesaturated fatty acids in lipids (lipid peroxidation)
- oxidation of amino acids in proteins
- inactivation of specific enzymes by oxidation of co-factors.

Too much cell damage triggers apoptosis. ROS are also linked to the immune system by inducing host defence genes and ion transport systems. Platelets involved in wound repair and blood homeostasis release ROS to recruit more platelets.

Gluconeogenesis

This is an anabolic pathway that generates glucose from non-carbohydrate substrates, including lactate, glycerol, and some amino acids. It helps to prevent the body from becoming hypoglycaemic.

The pathway can begin in the mitochondria or cytoplasm, depending on the substrate being used. Many of the reactions are the reversible steps found in glycolysis. The first step involves the formation of oxaloacetate through the carboxylation of pyruvate. After several steps, glucose-6-phosphate is hydrolysed to produce glucose in the lumen of the ER. Most steps are the reverse of steps in glycolysis. Lactate enters this pathway by returning to the liver, where it is converted to pyruvate by lactate dehydrogenase. All intermediates in the citric acid cycle may be used in gluconeogenesis. Transamination or deamination of amino acids allows their carbon skeleton to enter the cycle directly, as pyruvate or oxaloacetate. They can enter indirectly via the citric acid cycle.

There is reciprocal inhibition between glycolysis and gluconeogenesis. The rate is also controlled by the action of fructose-1,6-bisphosphate, which is regulated through phosphorylation of cAMP.

Pentose phosphate pathway

This is an alternative to glycolysis, involving the oxidation of glucose. It is an anabolic process that results in generation of NADPH and pentose sugars. It occurs exclusively in the cytoplasm.

The pathway is an important way of creating molecules with reducing power (NADPH). These are used to prevent oxidative stress by reducing glutathione via glutathione reductase, which in turn converts H_2O_2 to H_2O.

Transcription and translation

Transcription is the process in which RNA is synthesized according to a DNA template (Fig. 3.6a). An RNA copy of the sense strand of DNA is formed by the enzyme RNA polymerase. The anti-sense strand is read and a complement RNA molecule is created in the 5′-3′ direction, which has an identical base-pair sequence to the sense strand of DNA.

To initiate transcription, the polymerase enzyme must recognize and bind to a promoter sequence. This is located immediately upstream of a transcription start site. To achieve this, the polymerase attaches to other proteins or transcription factors to form an initiation complex. Basal transcription is therefore modulated by general transcription factors, but many genes rely on specific enhancers or repressors binding to *cis* or *trans* elements, thus affecting the rate at which RNA polymerase initiates transcription.

The 5′ end of a messenger RNA (mRNA) is capped with a methylated guanosine residue (by triphosphate bridge rather than a phosphodiester bond). This is necessary for translation by a ribosome and for prevention of degradation. The 3′ end is determined by site-specific cleavage rather than simply termination of transcription. After cleavage, large numbers of adenine residues are added to produce a poly-A tail. This prevents the 3′ end being degraded and is a signal for the mRNA to be exported from the nucleus.

After transcription in prokaryotes, genes are expressed in continuous sequences. In eukaryotes, the coding sequence after transcription is often punctuated with 'junk' sequences called introns. These are transcribed along with coding sequences or exons. The introns are removed by splicing, which results in the formation of a mature mRNA, which can then be translated.

Messages are localized to different parts of the cell either when the mature protein is formed, when only part of the protein is formed by translation and the mRNA is still attached to the ribosome, or when the mRNA itself is taken to a different part of the cell. This occurs due to polyadenylation signals carried in their 3′-untranslated region (UTR). mRNAs have different turnover rates, which corresponds to their

duration of action. Lack of a poly-A tail results in more rapid turnover.

Translation is the synthesis of polypeptides using a sequence of mRNA. The information carried on the mRNA codes for amino acids by groups of three bases, called codons. Since there are four bases, this equates to 4^3 (64) codons coding for 20 amino acids, meaning that there is some redundancy in this code. This includes 61 amino acid coding codons and three termination codons (UAA, UAG, and UGA), which stop the translation process.

mRNA is translated by ribosomes, which are complexes of RNA and protein. Each ribosome consists of two unequal subunits called the S (small) and L (large) subunits. The RNA molecules undergo extensive intramolecular base-pairing, which determines the ribosomal structure.

These, along with accessory proteins, identify the initiation codon (AUG in eukaryotic and most prokaryotic genes) and form a polypeptide chain in the sequence represented by the codons (three consequent bases) on the mRNA.

The amino acids are brought to the ribosomes by tRNAs, which carry a triplet sequence (anticodon) that is complementary to the codon on the mRNA (Fig. 3.6b).

tRNA:

- carries specific amino acids to the site of protein synthesis
- has two active sites
- is a linear molecule with an average 76 nucleotides
- has extensive intramolecular base-pairing, giving it a 'clover-leaf'-shaped secondary structure.

At the end of the mRNA message, there is one of three termination codons. These do not code for an amino acid, but signal that translation is complete.

After translation, there is extensive post-translational modification of proteins. This gives it functional activity and includes peptide cleavage and covalent modifications such as glycosylation, phosphorylation, carboxylation, and hydroxylation of specific residues. Proteins are taxied to their correct locations by amino acid sequences (e.g. signal peptide) or post-translational modifications.

The signal peptide sequence is a hydrophobic amino acid sequence of 18–30 amino acids that directs non-cytoplasmic polypeptides into the ER lumen, in which the polypeptide undergoes further modification. Proteins destined for secretion are moved to the Golgi apparatus in vesicles that bud off the ER.

Glycosylation of peptides occurs in the ER or Golgi apparatus. It involves the addition of oligosaccharides to specific amino acid residues. Glycosylation is important for correct functioning or compartmentalization. Phosphorylation of serine or tyrosine residues regulates enzyme or protein activity. Sulphation targets tyrosine and is important for compartmentalization and biological activity. Hydroxylation targets lysine and proline residues, and is important in collagen formation. Cleavage is a process that activates some enzymes and hormones.

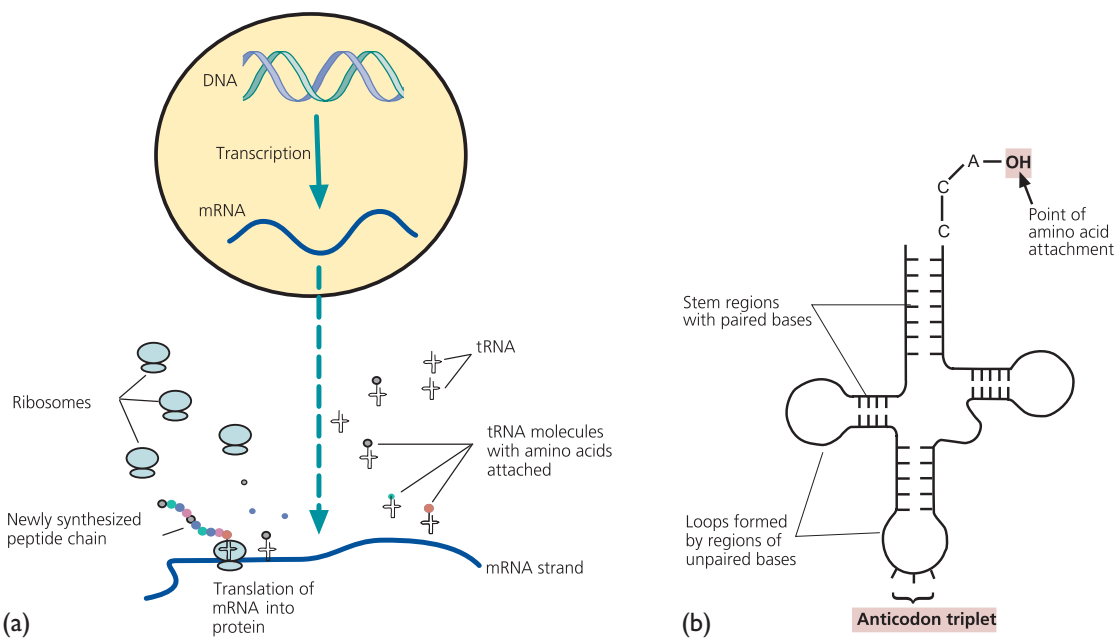


Fig. 3.6 (a) Transcription and (b) translation.

Reproduced from Gillian Pocock and Christopher D. Richards, *The Human Body*, Figure 30.6, and Figure 30.5, Page 653, 2009, published by Oxford University Press, with permission from Gillian Pocock and Christopher D. Richards.

Cell cycle

There are two mechanisms of cell division. Mitosis is the type of cell division that results in the production of two genetically identical daughter cells.

Meiosis occurs in gamete formation, producing two daughter cells containing half the genetic information of the parent cell (haploid). Crossing over ensures a re-assortment of genetic material between homologous or 'paired' chromosomes.

The cell cycle is a series of events that brings about cell growth and division of a cell to produce two daughter cells. There are four phases, each of which must be complete before the next stage can begin.

The G_1 phase is the gap period after mitosis of the cell from the previous cycle and before DNA synthesis that is required for the next cell cycle. The restriction point in the G_1 phase denotes the start of the next cell cycle.

The S phase is the DNA synthesis phase. Here, the cell's chromosomes are replicated in preparation for cell division.

The G_2 phase is the gap period between completion of DNA synthesis and the start of cell division.

These three phases, G_1, S, and G_2, are collectively known as the interphase.

The M phase (mitosis) is the cell division phase, which results in the production of two daughter cells.

Non-dividing or quiescent cells, such as corneal endothelial cells and neurons, remain in a resting, non-cycling state known as G_0.

Progression through the cell cycle is controlled by the activity of cyclins. These control the transition from one phase to another.

Cyclin-dependent kinases stimulate cell-cycle progression by phosphorylating specific proteins in the cell required for transition to the next stage. For example, chromosome condensation at the beginning of mitosis is initiated by the phosphorylation of H1 histone, a nuclear-associated protein.

Extracellular signalling pathways further modulate the activity of cyclin-dependent kinases, resulting in coordinated cell division. Mitogens induce mitosis. There are a number of factors (including growth factors, hormones, and cell–cell interactions) that activate the mitogen-activated protein kinase (MAPK) pathways.

There are also points along the cycle that are responsible for regulation of the cycle. During the restriction point during G_1, the cell becomes independent of external mitogens and once passed, the cell is committed to completing its cycle of cell division.

The G_1-S DNA integrity checkpoint ensures that the previous cycle of division has been completed and any resultant damage repaired. It is the main site of action of p53, a tumour suppressor gene that restricts the entry of cells with damaged DNA to the S phase. It therefore regulates the passage of the cell through the restriction point. p53 also has a role as a transcription activator, regulating certain genes involved in cell division, and a role in the initiation of apoptosis (programmed cell death).

The G_2-M DNA integrity checkpoint ensures that DNA synthesis and resultant damage are repaired before mitosis occurs.

The spindle assembly checkpoint ensures that the conditions are correct for spindle formation before commencing chromosome segregation.

Mitosis

There are six phases in mitosis (Fig. 3.7):

- Prophase: the cell's chromatin condenses and the chromosomes become visible as a pair of sister chromatids, joined by the centromere. The centrioles duplicate and migrate towards opposite poles of the cell and a spindle of microtubules is formed. The nucleoli disperse.
- Prometaphase: this involves the breakdown of the nuclear envelope and the formation of kinetochores (attachment points between the chromosomes and the spindle) at the centromeres of the chromosomes.
- Metaphase: the chromosomes attach to the spindle by the kinetochore and line up to form the equatorial plate.
- Anaphase: the centromeres separate and the chromatids are pulled to opposite poles by the spindle. By the end of anaphase, there is a clustering of a complete set of chromosomes at each pole of the cell. Each of the sister chromoatids is a double helix of DNA; thus after separation in anaphase each daughter cell receives a complete copy of the genome (i.e. the strands of the double helix are not separated in this phase).
- Telophase: the chromosomes begin to uncoil again, the nuclear membrane reforms, and nucleoli reappear.
- Cytokinesis: a cleavage furrow forms in the middle of the cell, dividing the cell into two. Each resulting cell has a complete diploid chromosome complement.

Meiosis

Meiosis involves two divisions: meiosis I and meiosis II. The first division results in two genetically different haploid cells. The second division results in duplication of each haploid cell. The end result is the formation of four cells (two identical pairs), each with a haploid number of chromosomes.

- Prophase 1: homologous chromosomes come together and exchange segments in homologous recombination:
  - Leptopene: spindle forms.
  - Zygotene: homologous chromosomes pair, shorten, and thicken. They arrange to form bivalents (pairs of homologous chromosomes).
  - Pachytene: chiasmata form, at which point crossing-over occurs.
  - Diplotene: exchange of genetic material between homologous chromosomes occurs at chiasmata and the nuclear membrane disappears.
  - Diakinesis: recombinant chromosomes are formed.

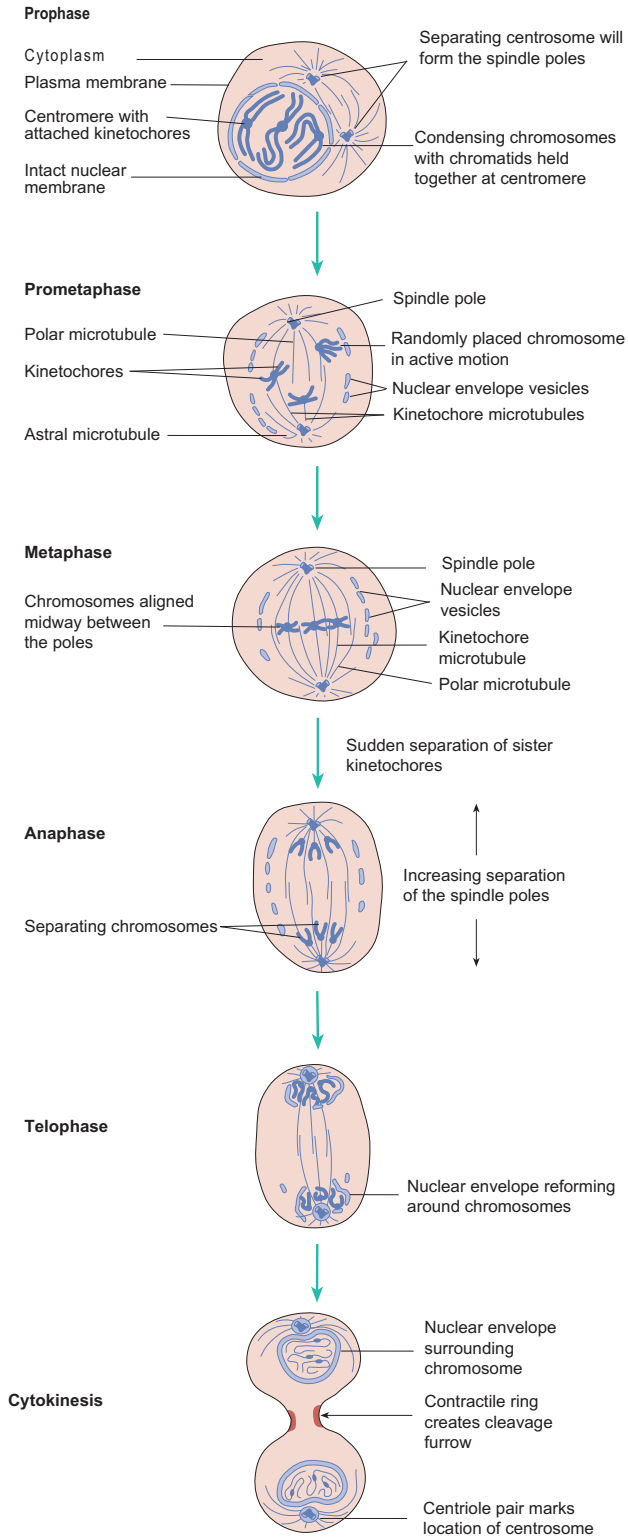


Fig. 3.7 Mitosis.

Reproduced from Gillian Pocock and Christopher D. Richards, *Human Physiology, The Basis of Medicine*, Third edition, Figure 3.3, Page 24, 2006, with permission from Oxford University Press.

- Metaphase 1: chromosomes attach to the spindle.
- Anaphase 1: instead of chromatids separating, whole chromosomes migrate to opposite poles.
- Telophase 1: two genetically different haploid cells are formed.

The second division is similar to mitosis, but only a haploid number of chromosomes are involved. In anaphase II, sister chromatids separate rather than chromosomes.

There are two processes that generate genetic diversity. First, chiasmata formation or crossing-over allows random exchange of genetic material between homologous chromosomes. Second, independent segregation of homologous chromosomes occurs during anaphase I. Since it is random which chromosome of a pair ends up in a particular gamete, there is a possibility of 2^{23} combinations in the gametes.

In oogenesis, each primordial germ cell results in only one definitive oocyte. After meiosis I, one secondary oocyte and one polar body are formed. If fertilization occurs, meiosis II can occur. This results in a definitive oocyte and another polar body.

Apoptosis

Cell death may be accidental (necrosis), which occurs after injury. It is characterized by the swelling of organelles and the breakdown of the plasma membrane, resulting in leakage of cellular contents and resulting inflammation. Programmed cell death (apoptosis) occurs after intra- or extracellular physiological triggers during development/remodelling, defence, homeostasis, and ageing. Cells undergoing apoptosis shrink and the cytoplasm shrinks, loosing microvilli and cell junctions. Organelles maintain their structure whereas the cell membrane forms irregular buds (blebs) and becomes convoluted. The cell breaks into smaller apoptotic bodies, which in turn are phagocytosed by macrophages (efferocytosis). There is no leakage of cellular constituents and therefore there is no inflammatory response. There is a characteristic condensation and fragmentation of chromatin, which condenses into patches against the nuclear envelope (pyknosis) as a result of endonuclease activity. The nuclear envelope becomes discontinuous and the DNA becomes fragmented (karyorrhexis).

Apoptosis may be important in disease processes by eliminating virally infected cells and removing those with damaged DNA that might otherwise contribute to the development of cancer. Failure of apoptosis of self-reactive T-cells may be important in autoimmune conditions, whereas inappropriate apoptosis might be important in degenerative conditions.

The mechanisms are varied. A cell initiates intracellular apoptotic signalling in response to a stress, which may bring about cell suicide. Commonly, raised Ca^{2+}, increased mRNA, and protein synthesis (including cell cycle regulatory proteins, proto-oncogenes, and tumour suppressor genes) are implicated. p53 is induced by DNA damage. This transcription factor results in arrest of the cell cycle in G_1 to allow DNA repair. Bcl-2 suppresses some apoptotic pathways, and reduction in levels of this molecule is associated with apoptosis. Telomeres are repetitive DNA sequences found at the end of chromosomes. They are synthesized by telomerase and are not replicated in the same way as other chromosomal DNA. With age, these telomeres shorten, allowing a maximum number of replications for any cell.

Regulatory proteins must, in response to these signals, initiate or halt the apoptosis pathway. There are two main methods of regulation: first, targeting of mitochondrial functionality; second, by directly transducing the signal via adaptor proteins to the apoptotic mechanisms.

Direct initiation of apoptosis may be a result of TNF (cytokine) production by macrophages. Another mechanism causing initiation of apoptosis involves the Fas receptor, which binds the Fas ligand (fasL), a transmembrane protein that is part of the TNF family. The interaction between Fas and FasL results in the formation of the death-inducing signalling complex (DISC), which contains the Fas-associated protein with Death Domain (FADD), caspase-8, and caspase-10. This sequence of events triggers the execution of apoptosis.

The process of apoptosis involves condensation of the nucleus (pyknosis) and cell shrinkage. Later the nucleus fragments (karyorrhexis) and the cells break up into apoptotic bodies, which are then engulfed by phagocytes.

Connective tissue, extracellular matrix, and cellular interactions

Connective tissue consists of cells and extracellular material. Connective tissue provides structural and metabolic support for other tissues and organs.

Fibroblasts synthesize and maintain the extracellular material. They are active in wound healing, where specialized contractile fibroblasts (myofibroblasts) bring about shrinkage of the scar tissue. Adipocytes store and maintain fat. They are found in clumps in loose connective tissue. Fat stored in these adipocytes forms a large droplet that occupies most of the cytoplasm. The adipose cells clump in loose connective tissue and form the main cell type in adipose tissue. Chondroblasts and chondrocytes produce and maintain cartilage. Osteoblasts, osteocytes, and osteoclasts are specialized cells that produce, maintain, and break down bone, respectively.

The extracellular material is a hydrated polysaccharide gel containing a meshwork of glycoproteins. It is secreted by the cells and determines the physical properties of the tissue. It is composed of fibrous adhesive proteins (laminin, fibronectin, and tenascin), fibres (collagen and elastin), and structural proteoglycans.

Laminins are a family of glycoproteins that are an important component of the basal lamina. They are large heterotrimeric molecules formed from an α-, β-, and γ-chain held together in a cross shape by disulphide bonds. They form networks with type IV collagen via perlecan (a proteoglycan) and with cell membranes (via integrin receptors). They are responsible for anchoring cell surfaces to basement membrane. Through these interactions, laminins critically contribute to cell attachment and differentiation, cell shape and movement, maintenance of tissue phenotype, and promotion of tissue survival. Defects can lead to, for example, junctional epidermolysis bullosa or nephrotic syndrome.

Fibronectin is a high-molecular weight (~440 kDa) 'adhesive' ECM glycoprotein. It binds to membrane-spanning receptor proteins called integrins and ECM components such as collagen, fibrin, and heparan sulphate proteoglycans (e.g. syndecans). Fibronectin exists as a dimer, consisting of two nearly identical monomers linked by a pair of disulphide bonds.

Fibronectin has roles in cell adhesion, growth, migration, and differentiation, and it is important for processes such as wound healing and embryonic development. Altered fibronectin expression, degradation, and organization have been associated with a number of pathologies, including cancer and fibrosis.

Proteoglycans form the hydrated polysaccharide gel that acts as a ground substance and allows diffusion of substances (e.g. nutrients and hormones) from the blood to the tissue and vice versa. They are composed of 95% protein (whereas glycoproteins are only 1–60% carbohydrate by weight) and consist of a core protein that is linked via a serine to one or more GAGs. GAGs consist of repeating disaccharide units forming unbranched polysaccharides (compared to glycoproteins, which have branched carbohydrate chains). These GAGs are formed in the Golgi apparatus. The main type of GAGs are:

- hyaluronic acid—a lubricant in synovial fluid
- chondroitin sulphate—in cartilage
- dermatan sulphate
- heparin sulphate
- heparin
- keratin sulphate—in skin.

These molecules are all very hydrophilic and have an extended coil structure that takes up extensive space. They have a net negative charge and attract cations such as Na^+. This brings water into the matrix, giving turgor pressure so that the tissue is able to withstand external forces.

Collagen is a fibrous protein with great tensile strength and resistance to stretching. It makes up 25% of the protein in mammals. It is rich in proline (ring structure) and glycine (the smallest amino acid, allowing the strands to fit together). Each individual collagen molecule forms a left-handed helix. Three molecules twist together to form a microfibril. Microfibrils bundle together to form fibrils, which in turn bundle together to form fibres. Synthesis of collagen is carried out in the ER and Golgi.

Most collagen is in the form of fibrils, which are 10–300 nm in diameter and a few millimetres in length. Type IV collagen is non-fibrillar, forming a sheet-like network, and is found in the basal lamina.

Elastin is a highly glycosylated hydrophobic protein that is rich in non-hydroxylated forms of prolene and glycine. One in seven amino acids is valine. The sheets are organized with the help of a microfibrillar glycoprotein, fibrillin. Fibrillin deficiency leads to Marfan's syndrome.

Biochemical and molecular biological techniques

The process of extracting nucleic acids from cells relies on their different solubility when compared to other cellular constituents or on their binding to synthetic resins.

Restriction enzymes are endonucleases that cleave DNA on recognizing a specific sequence. These sequences are between four and six base sequences long. Thus, a six base-pair cutter will produce many fragments with an average length of 4 kb. Since the genome is random, the fragment sizes will be variable about a mean.

These fragments can then be separated by gel electrophoresis. In this process, they are loaded into wells in a tray of agarose gel, across which an electric field is applied. DNA is negatively charged and therefore will move towards the positive charge, with small fragments migrating more quickly than large fragments. Ultraviolet light is used to visualize DNA after staining with ethidium bromide.

Polymerase chain reaction

Polymerase chain reaction (PCR) is a technique that is used to amplify short segments of DNA (2–3 kb) (Fig. 3.8). Conditions are optimized so that DNA polymerase repeatedly replicates a specific sequence of DNA. Five reagents are essential for PCR. Target DNA acts as a template for the first round of replication. Taq polymerase synthesizes DNA in a 5′ to 3′ direction and requires a primer to initiate synthesis. Deoxynucleotide triphosphates are the substrates used by Taq polymerase to synthesize new strands of DNA. 5′ and 3′ primers are used to start the process of replication by binding to complementary sequences on opposing DNA strands. It is this primer that gives PCR its ability to amplify targeted regions of DNA. Buffer maintains the optimum pH and chemical environment for the polymerase enzyme.

Once the reagents have been combined, they are placed in a thermocycling machine, which repeatedly heats and

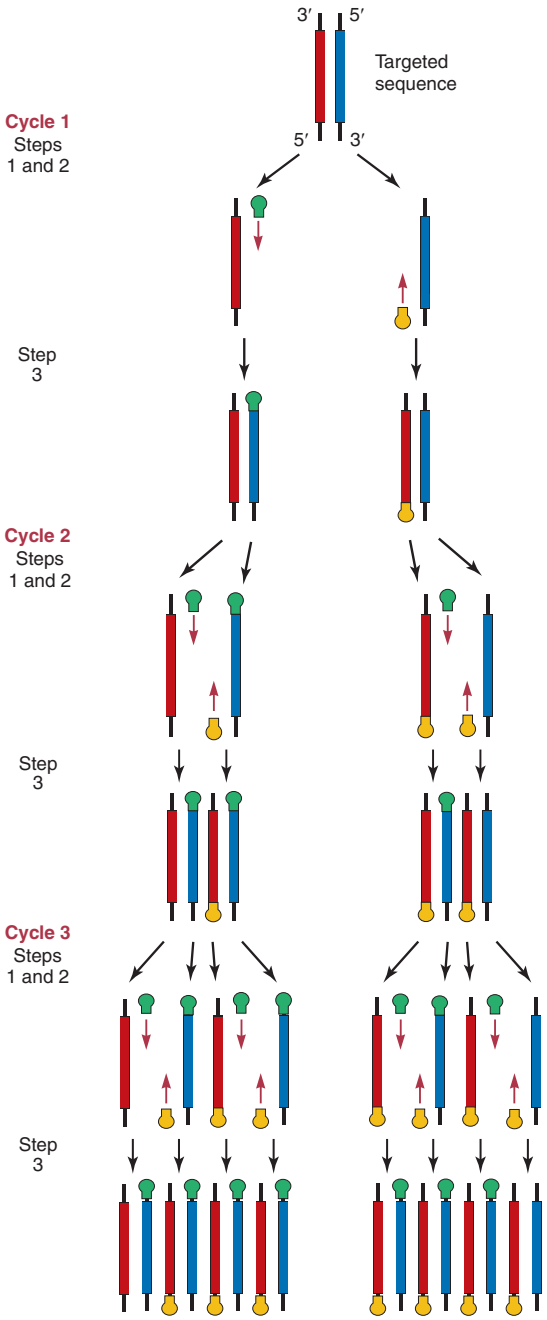


Fig. 3.8 Mechanism of the polymerase chain reaction.

Reproduced from Trudy McKee and James R. McKee, *Biochemistry, The Molecular Basis of Life*, International Fourth Edition, Figure 18e, Page 650, 2009, with permission from Oxford University Press.

cools the mixture in a cyclical manner. Each cycle theoretically doubles the amount of DNA. The products of PCR can be visualized directly on an agarose gel after staining with ethidium bromide and after spreading by gel electrophoresis. They appear as distinct bands, the size of which corresponds to the number of base pairs. Every experiment contains controls that can be used to identify contaminants.

RNA can be used, but the enzyme reverse transcriptase is required so that a strand of DNA can be produced from the RNA template. The products of this reaction are then used directly in a PCR reaction, as described above.

Compared to Southern blotting, this technique is quick, requires a small amount of DNA, and is easier to automate. However, it is more likely to produce spurious results and have contamination problems.

Reverse transcriptase (RT) PCR uses the viral enzyme reverse transcriptase. This produces a strand of DNA using RNA as a template. The RNA is incubated with this enzyme and an appropriate primer to allow the synthesis of an appropriate strand of cDNA. The products of this are used directly in a conventional PCR reaction.

Nucleic acid hybridization

In a mixture that contains many strands of DNA or RNA, the identification of a single sequence is possible. This relies on the property that two complementary strands of DNA can form a double helix. To achieve this, both the target and probe sequences must be denatured so they are in their single-stranded forms. The probe sequence is labelled and this is added to the mixture with the target sequence to form a hybrid. The probe sequence does not bind to non-complementary sequences. Hybrids can be detected by autoradiography (if the probe is radioactively labelled) or by chromogenic detection (using fluorescent probes).

DNA: Southern blotting

Southern blotting is used for detection of a specific DNA sequence in DNA samples.

This technique combines transfer of electrophoresis-separated DNA fragments on the agarose gel to a filter membrane (made of nylon or nitrocellulose) and subsequent fragment detection by probe hybridization.

RNA: Northern blotting

This is a technique used to study the expression of genes. The technique is analogous to Southern blotting, but instead of DNA, molecules of RNA (or isolated mRNA) are separated by electrophoresis and transferred to a nylon or nitrocellulose membrane. This membrane is then used for hybridization experiments.

Protein: Western blotting

This is the protein equivalent of the above techniques. Gel electrophoresis is used to separate the denatured proteins by mass. These are then transferred to a nitrocellulose membrane and probed for using a specific fluorescently labelled antibody, which binds only to the protein under investigation.

Biochemistry of ocular tissue

Cornea

See Chapter 1.

Three types of cells exist in the corneal epithelium. The innermost layer consists of basal cells, which are responsible for mitosis. This adheres to the basal membrane. After cell division, the cells move to the more superficial layers, forming up to three layers of wing cells. The most superficial cells are terminally differentiated squamous cells (three to four layers). Basal cells have more prominent metabolic and synthetic activity (and therefore more mitochondria, ER, and Golgi apparatus) as a result of their role in mitosis. They also have good stores of glycogen. The superficial corneal cells get sloughed off and this correlates to their lack of organelles and clumped chromatin in the nucleus. There are also reduced levels of RNA. The surface cells have microvilli, to which there is an adherent glycocalyx. The superficial cells contain the MUC 1 and MUC 4 genes.

The basal cells of the epithelium lie on the basal lamina. This is 40–60 μm thick and is similar in structure to the basement membranes of other squamous epithelia. It consists of type IV and VII collagen, laminin, fibronectin, fibrin, and the proteoglycan perlecan.

Bowman's layer lies below the basement membrane and is only approximately 12 μm thick. It is made of randomly arranged collagen fibres. It is acellular and can be considered to be a modified superficial layer of stroma.

Hemidesmosomes are responsible for the interaction between the epithelium and the basement membrane. These are integral membrane protein complexes in the epithelial basal cell plasma membrane. The intracellular domains of these hemidesmosomes are linked to keratin filaments.

In the basement membrane, the hemidesmosomes are bound to anchoring fibrils (type VII collagen). The anchoring fibrils in turn pass through the basement membrane, penetrating the stroma and branching to end in anchoring plaques. The anchoring plaques are composed of laminin.

The basal, wing, and superficial cells are linked by desmosomes. These are more numerous between superficial cells than between basal cells.

There are also numerous tight junctional (zonula occludens) complexes between the superficial cells of the epithelium. These represent an anastomosis of the lipid bilayer of two cells. Hence, superficial cells form a semipermeable membrane at the anterior surface of the eye.

The cells of all layers are also connected by gap junctions, more numerous in the basal layers than in the superficial layers. Basal cells are linked to their corresponding wing cells to form a functional syncitium, which is of importance in coordinating cell differentiation and migration. However, neighbouring wing cells are not joined by gap junctions.

The cornea is essentially made of collagen fibres running parallel to the corneal surface, with the individual collagen fibres separated by a matrix of proteoglycans.

The corneal stroma is 50–55% type 1 collagen. The regular arrangement of this is important in determining corneal transparency. The fibril diameter is less than 30 nm and the interfibrillar distance is 55 nm. There is also type V (10%) and type III collagen (1–2%) in the corneal stroma. The remainder is type VI collagen. Types I, III, and V collagen are fibrillar collagens whereas type VI collagen is a globular polypeptide. There are 200–250 lamellae of collagen fibres. Each lamella is formed of a bundle of collagen fibres, each 2.0 μm in thickness and 9–260 μm wide.

The extrafibrillar substance is mostly made up of proteoglycans. These contain a core protein with carbohydrate side chain. Previously the proteoglycans found in the cornea were named by their carbohydrate side chains as chondroitin sulphate/dermatan sulphate (CD/DS) proteoglycans and keratin sulphate (KS) proteoglycans. Now, the core protein has been identified and the proteoglycans found in the cornea are named decorin, lumican, keratocan, and mimecan.

Decorin is a CD/DS proteoglycan that binds to type IV collagen. It is more common in the anterior cornea. It is the only CD/DS proteoglycan in the cornea. Lumican, keratocan, and mimecan are related KS proteoglycans. Lumican is more prominent in the posterior stroma. Proteoglycans are found in other tissues, but elsewhere they either are nonsulphated or carry no KS side chains. Sulphation affects water retention and is therefore important in corneal transparency. At the limbus, there is an increasing concentration of hyaluronic acid.

There are specific binding sites between collagen fibrils and proteoglycans. These are important in the spacing of fibrils. The refractive index of collagen fibres is 1.411, compared to 1.365 for the extrafibrillar matrix. Despite the difference between these two, the regularity of the collagen fibres results in minimal light scattering.

Corneal fibroblasts (keratocytes) are the source of stromal collagens and proteoglycans. These are activated when the stroma is damaged. Some differentiate to form myofibroblasts. They contain numerous gap junctions and form a functional syncitium.

The cornea is 80% hydrated (compared to sclera, which is 70% hydrated). Despite the high concentration, it readily takes up more water if available. This can be seen easily in corneal lacerations. The high concentration of GAGs (lumican in particular) is responsible for this property, which is known as a swelling pressure. The swelling pressure produces an interfibrillar tension that helps keep the corneal fibrils in their normal arrangement.

To maintain corneal transparency, corneal fibres vary in diameter, but only within a small range. Their diameter is only a fraction of the wavelength of visible light. Also, the fibres are not arranged in a regular lattice, but in a quasi-random configuration. There is local ordering of fibrils to approximately 200 nm from individual fibrils, which is adequately uniform to account for corneal transparency.

The swelling pressure is counteracted by a metabolic pump in the endothelium, which actively removes water from the cornea. As well as this endothelial pump, there are ATP-driven ion pumps.

The corneal endothelium is formed by a monolayer of hexagonal cells. They contain a large nucleus that fills a large portion of the cell, causing the apical membrane to bulge outwards. There are numerous mitochondria, a prominent ER, and Golgi apparatus. This is usual in a cell that is highly metabolically active in transport, synthesis, and secretion.

Endothelial cells are interconnected by tight junctions. The apical tight junctions between cells are macula occludens (rather than zonula occludens); thus they do not completely encircle the cell and the endothelium leaks more than the epithelium.

Gap junctions are found on the lateral membranes of the cells. They only have a role in intercellular communication.

The endothelial cells have no special adhesion complexes, unlike the corneal epithelium. They lie on a basement membrane, Descemet's membrane, which is 1–15 μm thick. This is secreted by the endothelial cells, and its thickness increases throughout life. The main components are type IV collagen, laminin, and fibronectin. It is highly resistant to proteolytic enzymes and may remain intact even in severe corneal ulceration, forming a descemetocele.

Corneal cell metabolism

Epithelium derives most of its glucose from the stroma. This is converted to glucose-6-phosphate, and 85% of this is metabolized via the glycolytic pathway to pyruvate. The majority of this is converted to lactic acid, but some enters the citric acid cycle to produce ATP.

The pentose phosphate pathway accounts for the rest of the glucose utilization, and this helps in dealing with free radical production.

The epithelium obtains oxygen from the atmosphere and tear film at 3.5–4.0 μl/cm^2 per hour. Minor amounts are supplied by the aqueous and limbal vasculature. Tears contain more oxygen than the aqueous. During sleep, oxygen is supplied to the cornea by the palpebral conjunctiva. Most of the metabolic demands for glucose, amino acids, vitamins, and other nutrients are supplied to the cornea by the aqueous humour (via the ciliary body). Additionally, the corneal epithelium contains stores of glycogen. In hypoxia and normal conditions, some glucose is diverted to the hexose monophosphate shunt, which regulates levels of NADP. Additionally, glucose from the aqueous or epithelium is converted to pyruvate by the anaerobic Embden–Meyerhof pathway (glycolysis), which produces two molecules of ATP from each glucose.

Under normal aerobic conditions, pyruvate is then oxidized in the citric acid cycle to yield 36 molecules of ATP per glucose molecule.

Keratocytes are not highly metabolically active and generate energy to support the stroma.

The endothelium, on the other hand, is five times as active as the epithelium. It mainly uses anaerobic glycolysis. It also uses the citric acid cycle and the pentose phosphate pathway. Due to the high activity of these cells, the endothelium uses the citric acid cycle relatively more than the epithelial cells.

Both the endothelium and deep stroma obtain their oxygen from the aqueous humour.

During contact lens wear and resulting hypoxia and acidosis, oxygen consumption increases in the endothelial cells as a result of activation of pH regulatory mechanisms such as Na^+-H^+ exchange, which in turn stimulates Na^+-K^+-ATPase activity.

Corneal cell turnover and wound healing

The epithelium constantly turns over by mitotic activity in the basal layer of cells. Most mitotic activity occurs at the limbus, in the stem cells, which can differentiate into basal cells. The mitotic rate is 10–15% per day.

After epithelial damage, the cells slide horizontally to fill the defect. Hemidesmosomes and intercellular contacts reform and eventually the six-layer structure is reformed. Epithelial cells migrate due to the action and redistribution of actin-myosin fibrils, which cause the shape of the cell to change.

Attachment of epithelial cells to basement membrane and Bowman's layer is normally achieved by hemidesmosomes initially and later by the lamina densa and the anchoring type VII collagen fibrils.

During healing, protein synthesis by epithelial cells increases and the focal adhesion component vinculin is synthesized *de novo*, the cell surface glycoprotein CD44 is increased, and glycogen levels fall in migrating cells as cell migration is an energy-dependent process. The main process used for generation of energy is anaerobic glycolysis. To provide the necessary energy, within 2 hours of injury there is increased expression of GLUT1 mRNA. The resulting protein transporter levels increase within 4 hours, providing an adequate supply of glucose for wound healing.

The control of migration depends on autocrine and paracrine control by peptide growth factors. Synthesis of these factors and their receptors increases after corneal injury.

The corneal stroma, when injured, becomes opaque as a result of oedema and resulting disorganization of the regular fibrillar arrangement. Water is attracted by the GAGs.

Healing involves resynthesis and crosslinking of collagen, alteration in proteoglycan synthesis, and gradual wound remodelling resulting in return of tensile strength.

The wound is closed by deposition of fibrin within the wound and activation of fibroblasts to generate collagen and GAGs. There is also rapid epithelialization of the wound.

Initially, the keratocytes act like unspecialized fibroblasts to lay down collagen and GAGs with an irregular size and arrangement. This results in corneal opacity, particularly in larger wounds. In smaller corneal wounds attempts are made to restore normal corneal matrix components.

The corneal endothelium does not undergo mitosis and with age there is a reduction in the number of corneal endothelial cells. The cells' size and morphology changes and they move to fill the gaps left in the endothelium, and normally there is no clinical consequence. Endothelial cells can move 80–100 μm/day in the initial stages of healing, and during this process they do not lose contact with each other. This

CLINICAL TIP

If sufficient numbers of cells are lost either after surgery or pathologically, the cornea can decompensate due to increased leak where cells cannot fill the gaps and due to decreased ion and water transport out of the cells. The cornea becomes opaque as a result.

is followed by remodelling to their normal hexagonal shape and the pump and barrier function is restored when there is a confluent monolayer of cells.

Sclera

See Chapter 2

The sclera is almost acellular. It contains few fibroblasts and non-branching traversing blood vessels. The type I and III collagens are of varying diameter and distributed irregularly. There are fewer proteoglycans than in the cornea (so it is only 70% hydrated) and the main ones are proteodermatan and proteochondroitin sulphate. There is little proteokeratan sulphate compared to the cornea.

Most bulk transport occurs through the anterior chamber drainage angle and the uveoscleral meshwork. There is also some flow of fluid across the retina, most of which drains through the choroidal vessels. Some drains across the sclera, which is absorbed by the sclera proteoglycans. The sclera therefore performs its role by keeping proteoglycans with a low water-binding capacity.

Uveal tract

See Chapter 2

This includes the iris, ciliary body, and choroid, and is a continuous structure. These are discussed more fully in Chapter 1.

Lens

See Chapter 2

The refractive properties of the lens are due to the high proportion of lens crystallins (three times higher than in normal cells and about 40% wet weight of the cell).

The majority of crystallins are classical crystallins, which includes members of the α-Crystallin family and β/γ-crystallin family. α-crystallins are members of the heat-shock protein family and have a role in stabilizing partly unfolded proteins to prevent them from aggregating (chaperone function). This role is important as the lens fibres must last for the life of the individual, and without this property preventing aggregation of protein, cataract formation occurs. α-Crystallins also have serine–threonine enzyme activity, but the role of this is not clear.

Light travels without distortion through the lens due to the regular structure of lens fibres, the absence of membrane-bound organelles, and the small and uniform extracellular space between the fibre cells. Microtubules are important in stabilizing the fibre cell membrane and for transporting vesicles to the apical and basal ends of elongating fibre cells. As well as microtubules, there are also actin-containing microfilaments. These associate with the cytoplasmic surfaces of the adhesive junctions between lens fibres, and also interact with the spectrin-containing submembrane meshwork. Tropomyosin and tropomodulin proteins are present in the submembrane meshwork, and they can alter the structure of the microfilaments. Lens fibres also contain intermediate filaments, including those composed of vimentin.

There are other specializations in lens fibres. The plasma membranes have the highest cholesterol concentration of any plasma membrane in the body, and this increases as the cell matures. There is also a high proportion of sphingomyelin. The high levels of cholesterol and sphingomyelin cause the lens membranes to be rigid.

There are also several unique proteins in the lens fibre membrane. MIP accounts for up to 50% of the lens fibre membrane protein. It is thought to play an important role in fluid and ion transport in the lens.

The gap junctions of the lens are assembled from connexins, which are responsible for cell-to-cell transport of small molecules.

Lens fibres are linked to their neighbours along their lateral membranes by N-cadherin, a cell adhesion molecule. This is linked to the actin cytoskeleton by α- and β-catenin. Along with these, there are ball-and-socket joints that hold lens fibres together. Tight adhesion holds fibres together, thus reducing light scatter.

Lens fibres are also joined together at apical (anterior capsule) and basal (posterior capsule) ends. These complexes contain N-cadherin and vinculin.

Deeper lens fibres lack mitochondria and enzyme systems are less active. There is a fine balance between oxidative damage and diffusion from the more superficial cells.

There is a low oxygen tension around the lens, which contributes to protecting the lens from oxidative damage. This does not cause problems as glycolysis is the main source of energy. Some energy does arise from oxidative phosphorylation. Hydrogen peroxide from mitochondria and oxidation of ascorbic acid from the aqueous humour contribute to oxidative stress. Transferrin in the aqueous humour binds iron and prevents it from reacting with hydrogen peroxide and resulting in the release of free radicals. Solar radiation causes harmful damage to the lens fibre cells. Ultraviolet light is the most harmful and energetic, although this is mostly absorbed by the cornea. Interaction between UV light and cellular components results in free radical production.

Glutathione, a tripeptide, is present in high concentration in the lens and has a sulfhydryl group that is readily oxidized, thus protecting against free radical damage. Glutathione is produced by the lens fibres and epithelium, and is imported from the aqueous humour. Diffusion is the only process by which glutathione reaches the deeper lens fibres, and this diminishes with age. Ascorbic acid can also protect against oxidative stress. It is transported by active transport to the aqueous, where it is present in much higher concentration than in blood. Oxidation of ascorbate by free radicals prevents these molecules from causing damage elsewhere. Catalase is present in the lens, which breaks down hydrogen

peroxide to water and oxygen. Glutathione peroxidase is also present, which couples the reduction of hydrogen peroxide to the oxidation of glutathione.

Glycolysis is the main source of energy production in the deeper lens fibres due to the reduced rate of diffusion. The more superficial lens fibres use glycolysis and oxidative pathways as they contain more mitochondria.

From superficial to deep layers there is an increasing protein concentration. However, water does not follow this protein concentration gradient, possibly due to the reducing protein osmotic activity deeper in the lens. Electrolytes are thought to flow around the lens. The Na^+-K^+-ATPase is of importance as it draws Na^+ into the cells, leaving en electrochemical gradient across the membrane and causing a positive current at the sutures. The positive current flowing out of the lens at the equator is thought to be due to K^+ ions. Water follows the flow of ions in one model and this is the source of the internal circulation of the lens.

Calcium is actively pumped out of lens fibre cells and their cytoplasmic calcium is lower than in other cells. Higher levels of calcium lead to degradation of the cytoskeleton, uncontrolled proteolysis, cell swelling, and opacification.

With age, biochemical changes occur in the lens, which may result in cataract formation. When the fibres degrade their organelles, protein synthesis stops. Many of the soluble crystallins are increasingly truncated by proteolysis with age. Proteins are found in higher molecular aggregates and are less soluble deeper in the lens, where fibres are older. Importantly α-crystallin become less soluble towards the centre of the lens and functioning α-crystallin binds to hydrophobic regions of proteins that have undergone these changes, leaving less α-crystallin available for chaperone duties. Proteolysis and insolubilization of the components of the cytoskeleton result in disassembling, especially of vimentin intermediate filaments.

Vitreous

See Chapter 2

This is 99% water. The gel structure of the vitreous is a result of the long, unbranching collagen fibres (mostly type II) and highly hydrated hyaluronic acid. The gel structure acts as a barrier against movement of solutes. These solutes move either by diffusion or by bulk flow.

Neurosensory retina

See Chapter 2

The retina has the highest rate of aerobic glucose consumption in any tissue, most of which is generated by the photoreceptors.

However, a large proportion of this glucose is nevertheless converted to lactate. Other substrates, including glutamate, glutamic acid, malate, and succinate, can also be used. The glucose supply to the retina is insulin-independent. Hence, glucose can enter retinal cells by transport mechanisms regulated by extracellular glucose concentration directly. This facilitated diffusion relies on GLUT-1 and GLUT-3 transporters. Müller cells contain high levels of glycogen, which may be used by the photoreceptors.

The high concentration of lipid in the retina predisposes it to oxidative damage. The main lipids are phospholipids, phosphatidylcholine, phosphatidylethanolamine, and polyunsaturated fatty acids. Disc membranes have high levels of phosphatidylethanolamine whereas the plasma membrane has high levels of cholesterol (which inhibits the activity of rhodopsin).

Rhodopsin integration into plasma membrane has been well characterized. Lipids are replaced by membrane and molecular turnover. Protein composition is mainly rhodopsin (90%), whilst plasma membrane has more cell-specific proteins.

Other proteins, such as peripherin and the spectrin-like protein Rom-1, are structurally important in maintaining shape. The photoreceptor 'rim' proteins are important structurally but also have a transporter role.

The high metabolic rate and turnover of photoreceptors is reflected by their susceptibility to damage. Rod outer-segment renewal takes 9–10 days. Cone membranes and integral proteins are renewed less often.

Under conditions of stress, e.g. from illumination, levels of glutathione peroxidise and glutathione-S-transferase are increased. There are high levels of superoxide dismutase, but levels of catalase are low. The macula is particularly susceptible to damage from light. It contains the (carotenoid) pigments lutein and zeaxanthin. These are thought to reduce glare from short wavelength blue light and they have an antioxidant effect. Vitamin E is also important in reducing photoreceptor damage by inhibiting lipid peroxidation.

Iron can interact with hydrogen peroxide to produce hydroxyl radicals. Release of iron might be due to haemorrhage or intraocular foreign bodies.

Retinal pigment epithelium

See Chapter 2

The RPE (Fig. 3.9) is a monolayer of multifunctional pluripotent hexagonal cuboidal cells. The nucleus and mitochondriae are located in the basal part of the cell. Pigment granules in the apical cytoplasm give the epithelium a black appearance.

Photoreceptor outer segments lie in apposition to the RPE in the interphotoreceptor matrix, between the apical microvilli of the RPE cell. They contain several cytokeratins

HELPFUL HINT

The functions of the RPE are repeated in this section as they involve a number of biochemical processes described here:

- photoreceptor renewal
- IPM
- retinal attachment
- transport of water and metabolites
- retinoid metabolism
- blood–retinal barrier
- immunoregulation
- free radical scavenging.

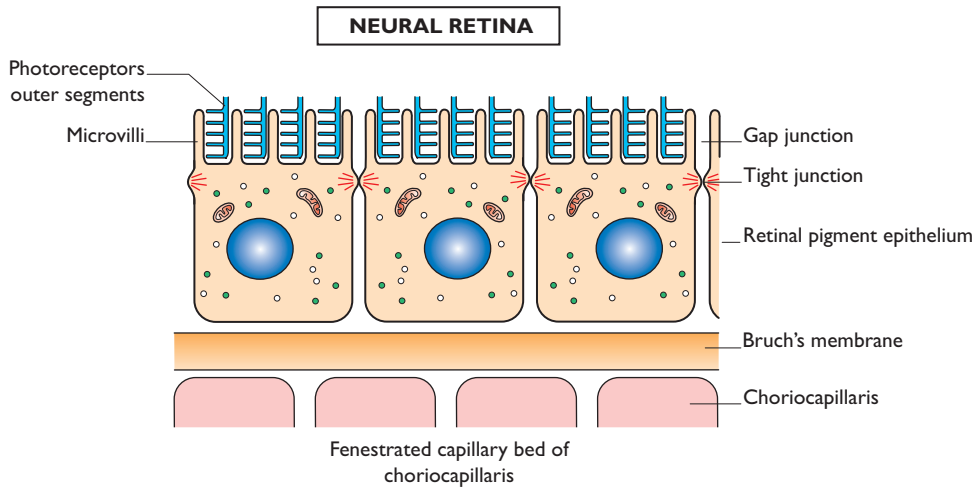


Fig. 3.9 Retinal pigment epithelium.
With permission from Neil Modi.

characteristic of other cell types as well as vimentin. Under conditions of stress, photoreceptor cells express proteins more typical of macrophages and other myeloid cells. These cells also contain high levels of phosphatidylcholine and phosphatidylinositol with high levels of saturated fatty acids. Turnover is slow or non-existent. RPE cells can synthesize and secrete growth factors such as fibroblast growth factor (FGF), insulin-like growth factor (IGF-1), and interleukin-1 (IL-1), which have a role in the normal physiology of the retina. These can, however, contribute to pathological conditions such as diabetic retinopathy and choroidal neovascularization.

The RPE cells are bound together by junctional complexes with prominent tight junctions. The tight junctions limit diffusion of water-soluble substances between the subretinal space and the intracellular space of the choriocapillaris. This, along with the tight junctions between endothelial cells in the retinal vessels, is the site otherwise known as the blood–retinal barrier.

Bidirectional transport of various metabolites occurs, but the main bulk flow (particularly for digested outer segment material) is from the retinal side to the choriocapillaris. The apical plasma membrane contains a Na^+-K^+-ATPase located in the apical plasma membrane of the RPE.

The cell is therefore a polarized epithelium, with its apical microvilli in apposition to the photoreceptor cell and its basal infoldings towards the choroid. The RPE sits on a basement membrane (Bruch's membrane), composed of five layers. This consists of proteoglycans, matrix proteins, hyaluronic acid, and chondroitin sulphate along with types I, III, VI, and VII collagen, and elastin.

The IPM is composed of glycoproteins and proteoglycans (traditional ECM proteins may be absent, e.g. collagen), including non-sulphated chondroitin and hyaluronic acid. IRBP is the most abundant glycoprotein (70% of IPM). This is important for photoreceptor function but is not required for the visual cycle.

Other important RPE proteins include RPE65, CRALBP (RLBP1), TIMP3, bestrophin (Best's disease), mertk, myosin VIIa (Usher's syndrome), and OA1 (ocular albinism). Light evokes changes in the IPM molecular distribution.

Melanin is an effective free radical scavenger, acting as a trap for free radicals from photoreceptors. Melanin granules in RPE cells are connected to lysosomal enzyme systems and loss of melanin is associated with age-related macular degeneration.

RPE is important in the uptake, storage, and metabolism of vitamin A and related compounds (retinoids). The chromophore is the moiety that causes a conformational change of the molecule when hit by light. The chromophore in mammals, 11-Ral, allows the visual cycle to take place.

Organization of the genome

The nucleus is the largest organelle in the eukaryotic cell. The nucleoplasm is in constant contact with the cytoplasm via the nuclear pores. The nucleus is involved in protein synthesis and the passage of genetic information from one generation to the next.

DNA is a polymer of four nucleotides. Each nucleotide is composed of a triphosphate, a deoxyribose sugar, and a nucleotide (adenine, cytosine, guanine, or thymine). The nucleotide can be either a purine (adenine or guanine) or a pyrimidine (cytosine, uracil, or thymine). These can be synthesized *de novo* in eukaryotes.

The sugar and phosphate groups are linked by phosphodiester. The order of the nucleotides determines the genetic code. The linkages are formed between C3 and C5 of

each pentose sugar, hence the notation 5′ to 3′ or 3′ to 5′ in a nucleic acid chain.

In 1953, Watson and Crick discovered that two complementary strands of DNA intertwine to form the double-helix structure of DNA. The strands run anti-parallel to each other and are annealed by specific base-pairing between adenine and thymine and between cytosine and guanine by hydrogen bonds. The two strands are therefore complementary in base sequence and are described as sense and antisense. The sense strand carries the coding sequence that determines the final protein.

There are progressive levels of packaging in the nucleus. First, DNA winds onto nucleosome spools. The nucleosome is formed by 146 base pairs of DNA wound twice around an octamer of histone protein, which contains a high proportion of positively charged amino-acid residues that can form ionic bonds with the negatively charged DNA. The nucleosome chain coils into a solenoid, which itself forms loops that attach to a central scaffold. The solenoid configuration corresponds to heterochromatin. The scaffold plus loops arrange themselves into a giant supercoil that is thought to represent transcriptional units. At metaphase, chromatin is maximally condensed, forming 14 nm fibres. Each metaphase chromosome is composed of two sister chromatids. These are connected at a central region called a centromere. This divides the chromosome into long and short arms (q and p, respectively). Centromeres are responsible for movement of chromosomes at cell division. A kinetochore is an organelle located at the centromere region which facilitates spindle formation by polymerization of tubulin dimers to form microtubules during mitosis.

Human cells contain 46 chromosomes (diploid): 22 identical pairs and a pair of sex chromosomes. The chromosomes are numbered 1 to 22 in order of decreasing size. Each pair of chromosomes (homologous chromosomes) carries matching genetic information and carries the same sequence of gene loci. At any locus, the two chromosomes may have slightly different forms, called alleles.

The ends of chromosomes are protected by telomeres, which are tandem repeats of a hexameric DNA sequence, ending in a 3′- single-stranded overhanging sequence of 50–400 nucleotides, which loops back on itself to form the T-loop. Telomeres prevent abnormal fusion between chromosomes, protect the ends of chromosomes from degradation, ensure complete DNA replication, and have a role in chromosome pairing during meiosis.

The DNA double helix is organized to form functional units or genes. Each one encodes a single protein. The bases that make up the coding sequence for the protein are read in triplets (codons), each of which codes for one amino acid.

In order for DNA to be converted into protein, it is first ‘transcribed’ into RNA. RNA is similar to DNA but contains ribose instead of deoxyribose, does not form a double helix, and uses the base uracil instead of thymine.

This mRNA leaves the nucleus. Translation is the process by which mRNA is used to form a protein molecule. Translation occurs in the cytoplasm. The flow of genetic information is one way, from DNA to RNA to protein. Only retroviruses can form DNA from RNA.

Each cell contains the whole genome of the organism. However, every cell is different and has a different complement of proteins at different times. Control of gene expression results in differential expression of proteins. Many genes (housekeeping) are expressed at relatively constant levels.

Mendelian and population genetics

Mendelian inheritance

Monogenic disorders show characteristic patterns of ‘Mendelian’ inheritance. The pattern of inheritance depends on the location of the gene, i.e. whether it is sex-linked or autosomal. There are five patterns: autosomal dominant (AD), autosomal recessive (AR), X-linked dominant, X-linked recessive, and Y-linked/holandric. A dominant gene is one that is expressed in heterozygotes.

An AD gene (Fig. 3.10a) is phenotypically expressed in homozygotes and heterozygotes for that gene. Affected parents have affected children and, conversely, unaffected members have unaffected parents and unaffected children. Both sexes are usually equally affected. Homozygotes are rare. Mutations may account for a proportion of cases, e.g. achondroplasia and familial adenomatous polyposis. AD genes show sex limitation, reduced penetrance (reduced proportion of individuals that carry a gene that actually express that trait), variable expressivity (variation in phenotype amongst individuals that carry the same gene), imprinting (a gene is only expressed depending on its parental origin, e.g. insulin-like growth factor 2 is only expressed when inherited from the allele inherited from the father), and anticipation (the symptoms of a genetic disorder become apparent at an earlier age, e.g. in Huntington’s disease an unstable gene can undergo a trinucleotide repeat that leads to a gain of function in the resultant protein—this increases with each passing generation).

In AD diseases there is usually a normal gene product from the second of the pair of chromosomes. The disorder may occur because of haploinsufficiency, e.g. familial hypercholesterolaemia, dominant negative effect, e.g. osteogenesis imperfecta, or simple gain of function from increased expression of the normal protein, e.g. Charcot–Marie–Tooth disease type 1a.

An AR gene (Fig. 3.10b) is only expressed in homozygotes with the abnormal gene. Usually, affected individuals have normal parents. If a carrier has an affected partner,

there is a 50% chance of the children being affected and 50% of their being carriers. Matings between heterozygotes have a 1 in 4 chance of producing a homozygote. Both sexes are equally affected. Consanguinity is implicated due to the sharing of genes in families.

In AR disorders, when the individual is a heterozygote, the normal allele provides sufficient functional protein to prevent disease. In homozygotes, however, there is no functional protein and this results in the associated disorder. Examples include sickle-cell anaemia, β-thalassaemia, phenylketonuria, and haemochromatosis.

X-linked disorders (XR; Fig. 3.10c) are transmitted by the female carrier. If the mother is a carrier and the father is normal, the sons have a 50% chance of being affected and the daughters have a 50% chance of being carriers. As a result, many more males demonstrate the recessive, affected phenotype. An affected male will usually have no affected children, but all of his daughters will be carriers. None of his sons will have the affected gene as they will only inherit his Y chromosome.

Males only have one X chromosome, so they are considered 'hemizygous'. They do not have any compensating normal or wild-type gene and hence will express any XR traits. In females, there is a double complement of X chromosomes. In the case of X and Y chromosomes, this is unnecessary and the X-linked genes from only one of these is expressed in any given cell (lyonization). The selection of inactivation of chromosomes in any cell is random, but once established the selection is transmitted to daughter cells. Women can therefore be affected if they are the daughter of an affected male and a carrier female, if there is skewed lyonization, if there is X chromosome-autosome translocation (this is a chromosomal abnormality caused by rearrangement of parts between an X-chromosome and an autosomal chromosome), or if XO (Turner's syndrome) is present. Examples of XR conditions are G6PDH deficiency, haemophilia A, and Duchenne's muscular dystrophy.

Y-linked inheritance is passed on to males from fathers. There is no interchromosomal genetic recombination.

Mendelian disorders demonstrate heterogeneity in their presentation. Genetic heterogeneity is the presentation of identical or similar phenotypes arising despite different genetic mechanisms such as different mutations of the same gene (allelic heterogeneity) or mutations in different loci (locus heterogeneity).

Non-Mendelian inheritance of single genes

Anticipation is the inheritance of a disease at a progressively earlier age. The mutation involves a trinucleotide repeat expansion that gets larger in successive generations (Huntington's disease and myotonic dystrophy).

Imprinting is a differential expression of genetic material depending on the parent from whom it has been inherited. It results from selective inactivation of genes in different patterns in the course of gametogenesis. It is important in Prader–Willi and Angelman syndromes, which both result from a microdeletion in the same chromosome, but the phenotype depends on whether the abnormal gene was paternally or maternally inherited.

Uniparental disomy is caused by duplication of a chromosome from one parent and loss of the homologue from the other parent.

Mitochondrial inheritance arises because mitochondrial DNA is maternally inherited. Sperm does not contain mitochondria, so males cannot transmit the disease to their children. An example is Leber's hereditary optic neuropathy.

Mosaicism is the presence of multiple cell lines within an individual from a single zygote. Germline mosaicism occurs when an abnormal cell line is confined to the gonads. This may result in unaffected parents producing more than one child with an AD condition. Somatic mosaicism occurs when either the maternal or the paternal chromosome is inactivated in each somatic cell. Each of that cell's progeny will have inactivation of the same chromosome. As a result, there may be unusually mild symptoms in AD conditions, for example Down's syndrome. Also, there may be expression of X-linked disease in female carriers if there is disproportionate inactivation of the normal gene.

In fact all women are genetic mosaics for the X chromosome by the process of lyonization described above.

Polygenic inheritance and multifactorial disease

Multifactorial diseases are more common than single-gene disorders. They do not show a Mendelian pattern of inheritance. Typically there is a continuous normal (Gaussian) distribution curve of the trait.

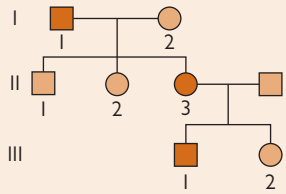
There may only be a few gene loci involved but these may then be affected by environmental factors. Heritability is the proportion of the total phenotypic variation that is genetic in origin in a given population. It is expressed as a percentage.

Gene frequency

The Hardy–Weinberg equilibrium is used when contemplating the genetic makeup of whole populations. This assumes a constant genetic constitution from one generation to the next for any given population where random mating occurs. In this equilibrium, p = proportion of normal alleles and q = proportion of abnormal alleles. Thus $p + q = 1$. The chances that both male and female gametes will carry the normal gene and abnormal gene is p^2 and q^2, respectively. The chance of the gametes producing a heterozygote is $(p \times q) + (p \times q) = 2pq$.

Since there is a pair of alleles for each individual at each locus, the relative proportions of normal homozygotes, heterozygotes, and abnormal homozygotes is therefore given by $p^2 + 2pq + q^2 = 1$. Hence, for an AR condition, the disease incidence = q^2, gene frequency = q, and heterozygote frequency (carrier) = 2pq.

(a)

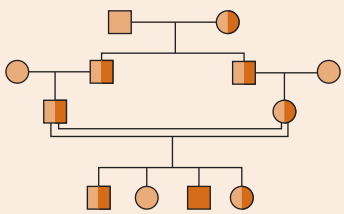


Mating of affected parent (Aa)

		Affected parent	
Gametes		A	a
Unaffected parent	a	Aa	aa
	a	Aa	aa

50% children affected on average (i.e. Aa as shown)

(b)



Mating of two carriers (Aa)

		Carrier	
Gametes		A	a
Carrier	A	AA	Aa
	a	Aa	aa

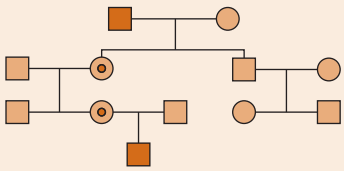
25% affected, 50% carriers, 25% normal

Mating of carrier and unaffected

		Carrier	
Gametes		A	a
Unaffected	A	AA	Aa
	A	AA	Aa

50% carriers, all phenotypically unaffected

(c)



Mating of affected male

		Affected male	
Gametes		X^D	Y
Unaffected female	X	X^DX	XY
	X	X^DX	XY

All daughters carriers, all sons normal

Mating of carrier female

		Unaffected male	
Gametes		X	Y
Carrier female	X^D	X^DX	X^DY
	X	XX	XY

Half of children inherit gene regardless of sex, 50% daughters carriers, 50% sons affected, 50% children normal

Fig. 3.10 Example pedigree and typical offspring of mating in (a) autosomal dominant inheritance (A, disease allele), (b) autosomal recessive inheritance (a, disease allele), and (c) X-linked recessive inheritance. (X0, disease allele).

For example, an AR disorder with a frequency of 1/1600 liveborn births has an incidence of $q^2 = 1/1600$, a gene frequency of $q = 1/40$ (the dominant allele has a gene frequency of 39/40), and a heterozygote frequency of $2pq = 2 \times 39/40 \times 1/40 = 1/20$.

For an AD disorder, nearly all those affected are heterozygotes, so $q^2 = 0$ (approximately), $p^2 = 1$ (approximately), and the disease frequency = $2pq = 2q$ (approximately).

A population is not in Hardy–Weinberg equilibrium if the genotype frequencies do not appear in the predicted proportions. This may be due to non-random mating, mutation, selection, genetic drift (one allele may be transmitted to a large proportion of offspring), migration, and founder effect (inbreeding within small populations).

Cytogenetics

Cytogenetics is the branch of genetics that is concerned with the study of the structure and function of the cell, especially the chromosomes.

Chromosomal disorders only occur in 0.6% of live births, although most result in spontaneous abortion. These disorders may be numerical or structural. The numerical disorders include extra chromosomes, missing single chromosomes, and extra haploid sets. As a result of these abnormalities there may be loss or gain of function. Structural disorders include translocation, inversion, isochromosome, duplication and deletion, and ring chromosome. Disease may be a result of dosage effects or misexpression of critical genes.

Polyploidy can arise as a result of fertilization by two sperm, a diploid sperm from failure in meiosis, or a diploid ovum due to failure in meiosis. Trisomy may result from a failure of separation (non-disjunction) of homologous chromosomes at meiosis I or from failure of separation of chromatids at meiosis II. Monosomy may result from non-disjunction or 'anaphase lag'. The latter occurs when there is delay in the movement of one chromosome from the metaphase plate during anaphase.

There are only three autosomal trisomies that are compatible with survival.

Down's syndrome (trisomy 21) occurs in 1/700 live births and the incidence increases with maternal age due to increased incidence of maternal non-disjunction of chromosomes during meiosis. Five per cent occur as a result of a Robertsonian translocation. This is a non-reciprocal translocation involving two homologous (paired chromosomes) or non-homologous chromosomes which have a long and a short arm. During the translocation, the participating chromosomes break at their centromeres and the long arms fuse to form a single chromosome with a single centromere. The short arms also fuse but these contain non-essential genes and are lost within a few cell divisions. Clinically, children with Down's syndrome present with hypotonia, short small middle phalynx of the fifth finger (clinodactyly), congenital heart defects, learning difficulty and low IQ, increased risk of leukaemia, and increased risk of cataract formation and Alzheimer's disease.

Patau's syndrome (trisomy 13) has an incidence of 1/10,000 live births. Twenty per cent arise as a result of a Robertsonian translocation. Clinically, this presents as dysmorphic features, scalp skin defects, incomplete cleavage of embryonic forebrain, and congenital heart disease.

Edwards' syndrome (trisomy 18) has an incidence of 1/8000 live births. Ninety-five per cent die *in utero* and of those born, fewer than 10% survive past 1 year of age. There is a high incidence of cardiac and renal malformations, and background apnoea.

Sex chromosome abnormalities include Klinefelter's syndrome (47, XXY), Turner's syndrome (45, XO), and trisomy X syndrome (47, XXX).

Mechanisms of chromosomal disorder

Translocation describes the exchange of chromosome segments that generally involves dissimilar chromosomes. Translocations may be reciprocal (balanced/non-Robertsonian) or Robertsonian (non-balanced). In the former, there is no loss of genetic material, so the individual is phenotypically normal. Gametes, however, depending on chromosomal separation during meiosis, may not contain a complete copy of the genome. The latter occurs when the long arms of two chromosomes fuse at a centromere and the short arms are lost. The short arms do not code for essential genes and these individuals will have 45 chromosomes but appear phenotypically normal.

Inversion occurs when there have been two breaks in a chromosome and the DNA between the breaks rotates by 180°. These are usually balanced. However, there may be difficulty in homologous chromosomes lining up in meiosis if the loop formation corresponds. There is therefore an increased chance of duplication and deletion of chromosomal regions, and thus of unbalanced chromosomal rearrangements in the offspring.

Isochromosome occurs where there is duplication of one arm but lack of the other. It is the result of breakage of a chromatid in the transverse direction. Usually this is lethal.

Duplication involves an extra copy of a chromosome region. This can be a result of inherited balanced structural disorders and *de novo* duplication from unequal crossing-over in meiosis, translocation, or inversion.

Deletion and insertion (frame shifts) of one or two nucleotide bases results in incorrect reading during translation and hence an incorrect amino acid sequence and premature termination of the polypeptide chain. Codon deletion or insertion (repeat expansions) are found in myotonic dystrophy, the fragile X syndrome, and Huntington's disease.

Mutations can range from a single base pair to deletion of large gene segments involving millions of base pairs. Point mutation (nucleotide substitution) can occur as a:

- missense mutation—single nucleotide change leading to a different amino acid substitution in the polypeptide chain
- nonsense mutation—single nucleotide change results in a stop codon; the polypeptide chain is shortened or vice versa, resulting in a lengthened polypeptide chain
- splice-site mutation—single nucleotide base change alters a critical splice junction and leads to abnormal RNA processing or a reduction in the normal gene product, or prevents the addition of a poly-A chain for effective transcription.

The genetic basis of eye conditions

Retinitis pigmentosa (RP) is a heterogeneous group of rod–cone dystrophies with a variety of clinical appearances. Inheritance can be: AD (20–25%), AR (20%), or X-linked recessive (8–25%). Defects in a large number of genes may cause identical phenotypic manifestations. By contrast, defects in one gene may cause a wide range of different clinical entities.

Autosomal dominant RP (adRP) is mostly caused by photoreceptor-specific genes. RHO was the first gene shown to cause RP (30–40% of adRP). Mutations in RDS/peripherin and ROM1 also result in adRP.

Autosomal recessive RP (arRP) may be caused by a large number of genes, suggesting that there is a final common pathway of photoreceptor and RPE dysfunction, apoptosis, and retinal atrophy. A number of these arRP genes are expressed in the photoreceptor. In particular, the phototransduction cascade and its recovery phase are implicated. The majority of mutations result in loss of function and, in almost all cases, one normal allele is compatible with normal function.

X-linked retinitis pigmentosa (XLRP) is a severe form of RP that affects males in their first decade and progresses to blindness by the third or fourth decades. Female carriers manifest milder symptoms than males. The most common responsible gene mutation is in RPGR, which modulates intracellular vesicular transport, critical to endocytic pathways.

Leber's congenital amaurosis is an AR retinal dystrophy that is a common cause of visual impairment. There are a number of gene mutations that have been identified (RPE65, GUCY2D, CRX, etc.). The RPE65 gene is responsible for retinal vitamin A metabolism and is thought to be essential for isomerization of 11-RoI. Recently, trials have shown some success after subretinal injection of RPE65 gene (using viral vectors) in returning some function to patients with RP.

Albinism may cause reduced visual acuity and nystagmus in children. It can be 'oculocutaneous' or 'ocular' albinism. Oculocutaneous may be 'tyrosine producing' or 'tyrosine non-producing'.

This is an AR condition. The responsible genes are tyrosinase (catalyses the first two steps in melanin synthesis), pink-eye gene (transmembrane protein found within melanosomes), and tyrosine-related protein 1 (regulates the formation of insoluble melanin and acts late in the melanin production pathway).

Retinoblastoma is a tumour of primitive photoreceptor cells and is the most common ocular malignancy of childhood. Unilateral tumour is nearly always sporadic. Forty per cent of cases are inherited. Those with a germline mutation (AD) carry a predisposing mutation in all cells and have an increased risk of non-ocular malignancy such as osteosarcomas, soft tissue sarcomas, and melanomas. Deletion of 13q14 may be responsible in 4% of patients. Knudson proposed a 'two-hit hypothesis', suggesting that there are two mutational events in inherited retinoblastoma. The first is present in the germinal cell and therefore every cell. The second somatic mutation must then occur to induce tumour growth in cells with the initial mutation by releasing suppression of the retinoblast. Two somatic mutations must occur in each retinoblast for a sporadic mutation; thus this type of mutation is more likely to be solitary and unilateral. The responsible gene is RB-1 (a tumour suppressor gene), which expresses a nuclear protein involved in cell cycle regulation and transition from G1S phase.

Using screening techniques, 80–90% of germline mutations can be detected.

Aniridia may occur sporadically (due to a deletion adjacent to the WT1 gene, which underlies Wilm's tumour) or due to AD inheritance (or less commonly AR). Around 140 mutations have been described for the 'paired box gene 6' (PAX6), most of which lead to termination of the PAX 6 protein due to haploinsufficiency.

Autosomal dominant conditions

Disease	Phenotype	Genotype
Cornea	Intraepithelial microcysts, symmetrical uniform intraepithelial opacities, recurrent erosions by first two decades	Cornea-specific keratin 3 and 2
Meesmans	Geographic asymmetric subepithelial opacities, recurrent erosions by fifth decade	Beta-Ig-H3/transforming growth factor
Reiss–Bücklers	Type 1: fine, grey, linear, anterior stromal opacities	
Lattice	Intervening cornea clear but becomes progressively hazy	
Granular	Type 2: corneal lattice + neuropathy (classified as an inherited systemic amyloidoses)	BIGH-3 gene mutations BIGH-3 gene
Lens	Progressive discrete grey-white opacities in central anterior stroma	

(*continued*)

Disease	Phenotype	Genotype
Marfan's	Multisystem connective tissue disorder with skeletal, cardiac, and ophthalmic manifestations	Fibrillin 1
	Main ocular manifestation is upward lens dislocation	
Aniridia	Total absence of iris tissue	Paired box gene 6 (PAX 6)
	Peripheral corneal vascularization causes stem cell failure and keratitis	
Congenital cataract	Increased risk of glaucoma	Numerous genes
	Cataract (particularly anterior polar)	
	Foveal and optic nerve hypoplasia	
Vitreoretinal		
Stickler's syndrome	Early onset myopia, lens opacities, degenerate vitreous gel, increased risk of rhegmatogenous retinal detachment	Type II and XI collagen gene
	Giant retinal tears	
Retinitis pigmentosa (AD = 20–25% cases)	Night blindness, constriction of peripheral field and eventually loss of central vision	Mostly photoreceptor-specific genes
	Poor dark adaptation and photophobia	
Best's vitelliform dystrophy	Bright yellow cyst under the macula	VMD2
	Late deterioration in central vision	

Autosomal recessive conditions

Disease	Phenotype	Genotype
Cornea	Macular dystrophy	
Lens		
Homocystinuria	Hypopigmentation, marfanoid habitus, chest deformities, stiff joints	Cystathionine beta synthase
	Ocular complication: progressive ectopia lentis	
Anterior polar cataracts	Often asymptomatic	Unknown
Congenital cataract		
Vitreoretinal		
Goldman–Favre disease	Degenerative change around vascular arcades	Nuclear receptor subfamily 2
Retinitis pigmentosa	Macular changes, including CMO	
(arRP = 15–30% of cases)	As above	
Stargardt's disease (fundus flavimaculatus)	Linear or pisciform macular lesions, macular degeneration, loss of central vision	ABCA4 gene
Gyrate atrophy		

arRP, autosomal recessive retinitis pigmentosa; CMO, cystoid macular oedema.

Lysosomal storage disorders

Disease	Phenotype	Genotype
Niemann–Pick	Cherry red spot	NPC 1
Metachromic leukodystrophy	Optic atrophy, cherry red spot, dystonia, mental retardation	
Tay–Sachs	Cherry red spot	
Sphingolipidoses		
Mucopolysaccharidoses	Optic atrophy, corneal clouding, mental retardation, physical deformity	
Refsum's syndrome	Polyneuropathy, ataxia, retinitis pigmentosa, cataracts	
Batten's disease	Pigmentary retinopathy, optic atrophy, psychomotor retardation	Ceroid lipofuscinosis, neuronal type 3 (CLN3)

X-linked recessive disorders

Disease	Phenotype	Genotype
Norrie's disease	Congenital visual disability Retrolental yellow vascularized mass Hearing loss and developmental delay	Norrie Disease Protein (NDP)
Incontinentia pigmenti	Cutaneous bullous eruptions, dysplastic retina, optic atrophy, cataract formation	
Retinitis pigmentosa	Severe form of retinitis pigmentosa	
Oculocutaneous albinism	Hypopigmentation, delayed visual maturation, nystagmus, refractive error, strabismus, iris transillumination, albinotic retina	Tyrosinase, pink-eye gene, tyrosine-related protein 1

Genetic techniques

DNA sequencing

This looks at the order of bases in the DNA sequence. The methods are chemical cleavage, chain termination, and thermal cycle sequencing. These will not be described here.

Cytogenetics

This involves the study of chromosomes and their abnormalities. Techniques such as G-banding or fluorescence *in situ* hybridization (FISH) are used to identify chromosomes or specific sequences within them.

G-banding

Chromosomes undergo digestion with trypsin and are stained with Giemsa. As a result, the chromosomes become stained with dark and light bands specific to each chromosome. This is used to diagnose monosomies and trisomies, translocations, and large deletions or insertions. FISH is superceding G-banding.

Fluorescence in situ *hybridization*

This is a method of visualizing specific regions of metaphase or interphase chromosomes. DNA is attached to a microscope slide and denatured to its single-stranded form. It is then hybridized with fluorescently labelled probe DNA that binds to a complementary sequence. Where hybridization has occurred, this region can be viewed under a fluorescent microscope. This technique can be used in gene mapping or diagnostically to identify a variety of chromosomal abnormalities, including monosomies and trisomies, translocations, microdeletions, and insertions.

Genetic linkage and linkage analysis

The gene responsible for a particular disease is identified without necessarily knowing anything about the protein product. Two genes may be transmitted together more frequently than could be expected by independent assortment. Gene loci are said to be linked when their alleles do not show independent segregation during meiosis.

At meiosis, homologous chromosomes exchange segments as a result of the formation of chiasmata (crossovers) before separating into two daughter cells.

As a result, children may inherit a different combination of alleles at two different loci than those of the parent. Linked markers segregate together more often than expected by chance because they lie closer together on the same chromosome. If markers are far apart, separation is more likely to occur between them. The recombination fraction (RF) is 0.5 for unlinked loci. For linked loci it lies between 0 and 0.5.

Genes can be mapped by techniques such as FISH. However, genetic linkage may be demonstrated by family studies, which can confirm the genetic background to a disease. Heterozygous family members can be studied at each of the gene loci. First, it must be determined which allele at locus 1 to be studied is on the same chromosome as a particular allele at locus 2 (marker locus). The ideal marker locus is highly polymorphic and thus heterozygous in most of the population.

DNA polymorphisms are inherited differences in the DNA that are present in all individuals, whether normal or otherwise. They often arise in non-coding regions of the genome. Examples include restriction fragment length polymorphisms (RFLPs), microsatellites, and single nucleotide polymorphisms (SNPs).

RFLPs occur as a result of point mutations in the human genome. Recognition sites for the restriction endonucleases are created or abolished, and these enzymes will create variable lengths of DNA fragments, giving characteristic patterns on Southern blotting. If an individual is heterozygous, there will be two bands in the electrophoresis gel. Hence, the transmission of a single chromosome can be tracked through the family. Many thousands of microsatellite markers exist, the locations of which are known. These are extremely polymorphic and can be detected by PCR.

It is important that sufficient families with the disease are identified and that they have the same disease and that all members of the family are identified. All family members are genotyped with respect to around 400 markers, which are usually microsatellites with known genetic and physical maps. Once linked to a particular chromosome, more closely packed markers are used. Once the relevant genetic region has been identified, the human genome project sequence maps can be accessed directly and candidate genes identified.

Alternatively, to identify a disease-causing gene, the candidate gene approach can be used. Here, a previously isolated gene is used as it is known to have a role in the physiology of the diseased tissue. This approach was used to identify rhodopsin mutations associated with RP. Finally, functional cloning is the process by which the gene responsible for a disease can be found based on an understanding of the underlying molecular defect. If the protein is known, the mRNA can be isolated and used as a probe for the gene. This technique was used to find the genes involved in sickle-cell anaemia.

The above techniques are not of great use in identifying the genetic basis of multifactorial disease or mutations.

Cloning and vectors

The human genome is extremely large. Cloning vectors can be used to break up the genome into small sections. These sections are easier to characterize by techniques such as restriction mapping and sequencing.

Cloning involves the transfer of one such section of the genome into a single microorganism. This is then replicated along with the microorganism's own DNA. Specific modifications can then be made to the cloned sequence.

The most simple form of cloning is where the same restriction enzyme is used to cut the insert and vector DNA. The DNA molecules therefore have complementary 'stick ends'. The vector and insert are mixed and incubated with DNA ligase, which catalyses the formation of phosphodiester bonds between molecules of double-stranded DNA, joining DNA fragments together.

The products of the ligation reaction are incubated with host cells, some of which take up DNA. These are then plated onto a selective media, where vector-containing cells only can grow. Techniques such as 'blue/white selection' are used to identify clones containing vector sequences that include insert DNA.

PCR can be used to generate insert DNA. Primers are designed to include the appropriate restriction enzyme sites. After amplification, the product is cut by the restriction enzyme and added to the ligation reaction above. This process has been used to produce therapeutically useful proteins such as insulin.

The management of genetic disease

Genetic disease can potentially be avoided or ameliorated. Screening techniques are advancing, especially with advances in knowledge from the Human Genome Project.

The mainstay of treatment presently is:

- conventional supportive treatment, e.g. contact lenses and glasses for Marfan's syndrome
- surgical correction of gross phenotype, e.g. for pyloric stenosis
- substrate limitation, e.g. avoidance of phenylanine in patients with phenylketonuria (they lack phenylalanine hydroxylase)
- environmental modification, e.g. avoidance of tobacco smoke in patients with α_1-antitrypsin deficiency
- enzyme replacement, e.g. type 1 diabetes is treated by insulin replacement.

Gene therapy has already been attempted with some success, for example in Leber's congenital amaurosis. Inhalation of the CFTR gene in a disabled adenovirus proved temporarily successful in cystic fibrosis patients. The adenosine deaminase gene has successfully been inserted into the bone-marrow cells of children with subacute combined immunodeficiency.

Prerequisites are that the gene concerned and its control elements must be fully characterized and cloned. Accessible target cells must be identified that have a reasonable and productive life span. Finally the chosen vector must be efficient and safe.

Currently, two main methods of gene sequence introduction are used. First, viruses have been manipulated to use their inherent ability to insert genetic material into their host cells. Second, non-viral delivery methods include the introduction of naked DNA directly into the cell and the use of liposome-mediated DNA transfer.

More recently, stem cells utilize their pluripotency and multipotency to replace defective organ cells. These are already commonplace for some haematological conditions such as leukaemia. The hope is that soon structures in which the cells do not multiply, such as the RPE, might be replaced using stem cells.

CHAPTER 4

Pathology

Inflammation

Inflammation is the response of tissue to a noxious stimulus. The response can be localized or generalized, and the stimulus infectious or non-infectious.

Non-infectious stimuli can be broken down into exogenous or endogenous. Exogenous causes originate outside the eye, e.g. penetrating trauma, alkali chemical injury, or external allergens.

Endogenous causes originate in the eye and body such as uveitis secondary to leaked lens matter (phacoantigenic uveitis), spread from adjacent structures (the sinuses in orbital cellulitis), and haematogenous spread.

Infectious stimuli may be viral, bacterial, fungal, or parasitic.

Phases of inflammation

Acute phase

The acute phase is recognized by five cardinal signs:

1. redness (rubor), caused by increased blood flow
2. heat (calor), caused by increased blood flow
3. mass (tumour)—oedema caused by leakage of fluid and cells
4. pain (dolor)
5. loss of function.

The triple response

The triple response of acute inflammation refers to the behaviour of blood vessels in damaged tissue. For example, if you were unfortunate enough to get hit by a golf ball on your leg, initially the impact would cause blood vessels to constrict and a white mark would appear; capillaries would then dilate, leaving a red mark (flush). Arteriolar dilation follows, leaving a larger red mark (flare) followed by fluid leakage from capillaries and local tissue swelling (wheal).

Cellular mechanisms and chemical mediators

During the triple response, leukocytes (neutrophils and monocytes) migrate to the site of injury. Neutrophils are the main inflammatory cells in the acute phase of inflammation. They are thought to follow a chemotactic stimulus released by vascular endothelial cells, leukocytes, mediators of complement, or even bacteria themselves at the site of injury (Fig. 4.1).

Chemical mediators such as histamine cause a local increase in vascular permeability by causing the tight junctions between vascular endothelial cells to open. This allows the leukocytes to invade the tissue and, coupled with complement, the process of phagocytosis of the offending organism can occur.

There are many chemical mediators involved in acute inflammation. In addition to this, those present in plasma are interrelated cascade systems, including the clotting cascade, fibrinolysis, complement, and bradykinin systems.

The outcome of acute inflammation is dependent on many factors, for example in infectious cases the organism, host response, and extent of necrosis.

Acute inflammation may resolve, repair with scarring, or progress to chronic inflammation.

Chronic phase

Chronic inflammation may start as chronic or result from acute inflammation.

It is a proliferative inflammation characterized by a cellular infiltrate of lymphocytes and plasma cells (sometimes polymorphonuclear neutrophils (PMNs) or eosinophils).

The two main types of chronic inflammation are granulomatous and non-granulomatous.

Granulomatous

The characteristic cell type in granulomatous inflammation is the epithelioid or giant cell. Classic examples include tuberculosis and sarcoid.

Epithelioid cells are derived from monocytes or macrophages (Fig. 4.2). They have abundant eosinophilic cytoplasm and tend to blend together in palisades around areas

HELPFUL HINT

Mediators of acute inflammation

Chemical mediator	Action	Origin
Histamine	Increase vascular permeability	Inflammatory cells, e.g. histamine and serotonin from mast cells
Serotonin	Vasodilatation	
Prostaglandins	Increase vascular permeability	
Leukotrienes	Macrophage activation	
Cytokines		

of necrosis. They can interact with T cells and phagocytose, and bind complement and immunoglobulin.

Inflammatory giant cells are likely to be formed by fusion of macrophages and come in three forms:

- Langhans' giant cell—typically found in tuberculosis and shows a homogenous, eosinophilic central cytoplasm and peripheral rim of nuclei
- foreign body giant cell—containing foreign material
- Touton giant cell—has a rim of foamy cytoplasm peripheral to a rim of nuclei and is seen in lipid disorders such as juvenile xanthogranuloma.

Patterns of chronic granulomatous inflammation

There are three patterns found in granulomatous inflammation:

- Diffuse type: sympathetic uveitis, juvenile xanthogranuloma, Vogt–Koyanagi–Harada syndrome, toxoplasmosis. Epithelioid cells are distributed randomly against a background of lymphocytes and plasma cells.
- Discrete type: sarcoidosis, tuberculoid leprosy, miliary tuberculosis. Nodules or tubercles form due to accumulation of epithelioid or giant cells surrounded by a narrow rim of lymphocytes and plasma cells.
- Zonal type: caseous necrosis of tuberculosis, chalazion, ruptured dermoid cyst, reaction to suture material, rheumatoid scleritis, toxocara. A central area of necrosis surrounded by a palisade of epithelioid cells. In addition, PMNs, Langhan's giant cells, and macrophages are in turn surrounded by lymphocytes and plasma cells.

Non-granulomatous

Examples include the many forms of anterior and posterior uveitis, Behçet's disease, multiple sclerosis, retinal vasculitis, and endocrine exophthalmos.

Cell types may include T and B lymphocytes, plasmacytoid cells (a variation of the plasma cell), and plasma cells with a Russell body. A Russell body is an inclusion in a plasma cell whose cytoplasm is filled and enlarged with eosinophilic structures. The nucleus is eccentric or absent. These are seen in B cell lymphomas.

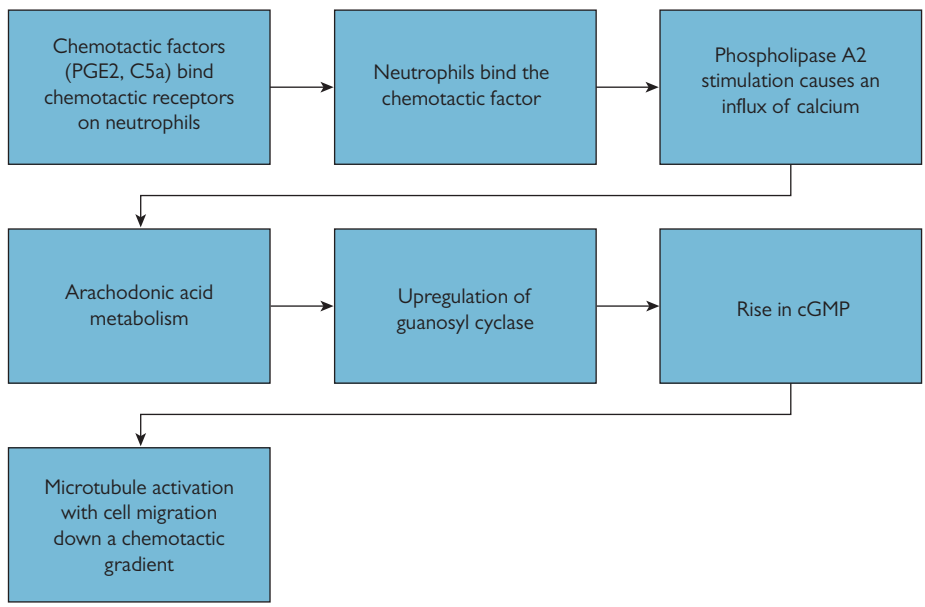


Fig. 4.1 The mechanism for chemotaxis.

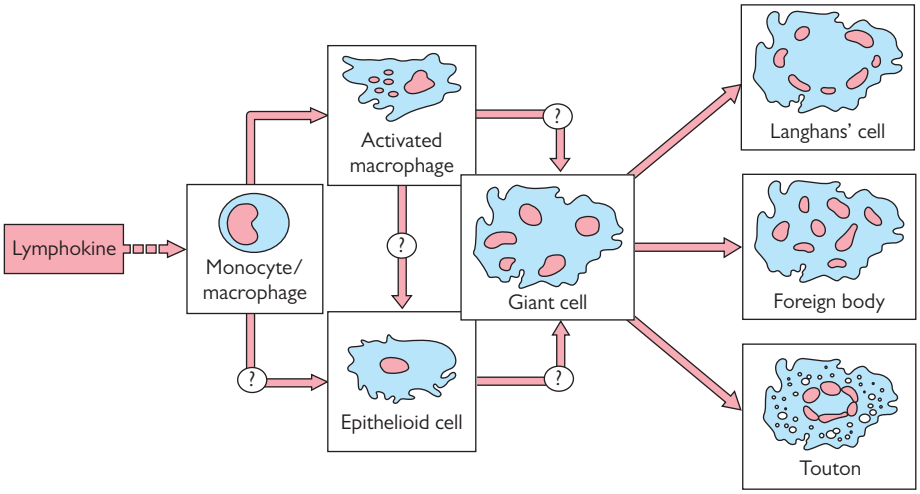


Fig. 4.2 Macrophages and granulomatous inflammation.

This figure was published in *Ocular Pathology*, Yanoff M and Sassani JW, Figure 1.15 Macrophages and granulomatous inflammation, p.12, Copyright Elsevier, 2009, with permission of Mosby. Yanoff Ocular Pathology.

Wound healing

It is worth reviewing the stages of healing in a full thickness corneal laceration due to its relevance in both surgery and penetrating trauma.

Stages of corneal healing in a full thickness laceration include the following:

- Immediate phase: retraction of Descemet's membrane and stromal collagen, anterior and posterior wound gaping of the wound, fibrin plug formation from aqueous fibrinogen, and stromal oedema.
- Leukocytic phase: at around 30 minutes, polymorphonuclear leukocytes from the conjunctival vessels and from the aqueous invade the wound. Limbal wounds have an invasion of mononuclear cells from limbal vessels. These can transform to fibroblasts after 12–24 hours.
- Epithelial phase: at 1 hour full thickness ingrowth is inhibited by healthy endothelium.
- Fibroblastic phase: central corneal wound fibroblasts are derived from keratocytes. They produce collagen and mucopolysaccharides to form an avascular matrix.
- Endothelial phase: at 24 hours endothelial sliding allows for coverage of the posterior wound.

Hypersensitivity, transplants, and graft rejection

There are four types of hypersensitivity reaction.

Type I hypersensitivity

This is an 'allergic' reaction that immediately follows contact with an antigen, which would normally not cause a marked immune response (an allergen). Examples of this are seasonal allergic conjunctivitis and perennial allergic conjunctivitis.

The mechanism is as follows:

- Mast cells bind IgE via their Fc receptors.
- On encountering antigen the IgE becomes cross-linked.
- This leads to degranulation and release of mediators such as histamine, serotonin, platelet-activating factors, and eosinophil chemotactic factors.
- Histamine then acts as a mediator of negative feedback to inhibit mast cell degranulation (Fig. 4.3).

Type II (antibody-dependent cytotoxicity) hypersensitivity

This arises from antibody directed against antigens expressed on an individual's own cells. Examples of this include incompatible blood transfusions, rhesus incompatibility of the newborn, hyper acute graft rejections, and myasthenia gravis (Fig. 4.4).

HELPFUL HINT

Tissue response to trauma

Tissue type	Description of response
Skin	*Epidermal* (i) Cell migration, (ii) proliferation, (iii) differentiation *Dermal* Invasion of fibrin clot by buds of endothelial cells from intact vessels Form blood vessels within 1 week Macrophages and fibroblasts invade wound Macrophages clear clots and fibroblasts produce collagen and glycosaminoglycans Myofibroblasts allow wound contraction by around 1 week
Conjunctiva	Can form granulation tissue Epithelium heals by sliding and proliferation similar to skin or cornea
Cornea	*Epithelium* Regenerates at the limbus and spreads rapidly across cornea *Bowman's layer* Does not regenerate *Stroma* Keratocytes form fibroblasts to heal stromal wounds *Descemet's membrane* Does not regenerate Is elastic and can recoil at the edge of a deficit *Corneal endothelium* Fills in defects by sliding and therefore deposits secondary layers in Descemet's
Iris	Fibrinolysins in the aqueous inhibit fibrin clot formation, hence the patency of iris defects If the edges of a wound appose, reactive proliferation of the iris pigment epithelium can occur
Lens	Epithelium responds to trauma by fibrous metaplasia
Retina	Glial cells replace damaged nerve cells, which are derived from perivascular astrocytes and Müller cells RPE can become metaplastic and proliferate and form fibrous tissue, for example preretinal membranes
Choroid and ciliary body	Melanocytes do not proliferate after trauma Granulation followed by scar tissue forms from fibroblasts
Sclera	Scars from episcleral and uveal fibroblasts
Optic nerve	Axonal loss and demyelination with reactive proliferation of glial cells and connective tissue cells

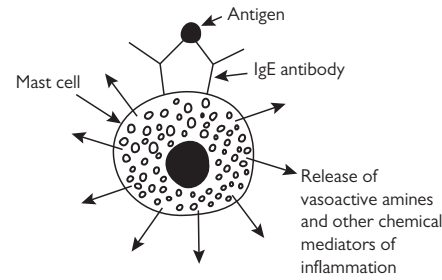


Fig. 4.3 Type I hypersensitivity.

Reproduced from R. Wilkins et al., *Oxford Handbook of Medical Sciences, The Essential Guide to the Sciences that Underpin Medicine*, Figure 12.7, Page 857, 2011 with permission from Oxford University Press.

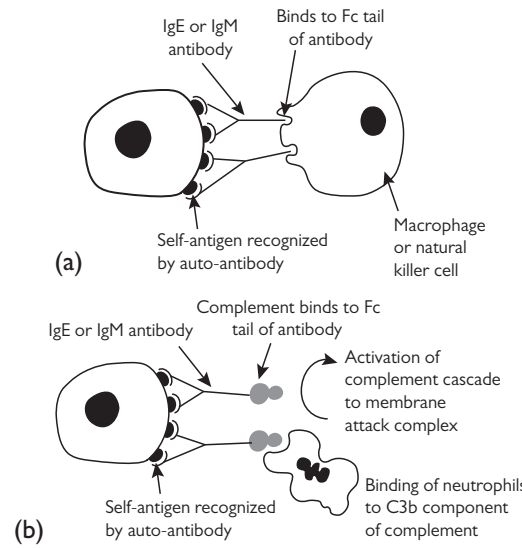


Fig. 4.4 Type II hypersensitivity.

Reproduced from R. Wilkins et al., *Oxford Handbook of Medical Sciences, The Essential Guide to the Sciences that Underpin Medicine*, Figures 12.8 and 12.9, Page 857, 2011, with permission from Oxford University Press.

Type III hypersensitivity

Immune complexes are deposited in the tissue. Complement is activated and polymorphs are attracted to the site of deposition, causing acute inflammation. This is seen in persistent infections (viral hepatitis) and some autoimmune diseases, such as rheumatoid arthritis and systemic lupus erythematosus. It also occurs in the Arthus reaction.

The mechanism is dependent on (i) turbulent blood flow allowing for deposition of immune complexes, e.g. kidney, (ii) increased vascular permeability due to histamine

release, and (iii) specific antigen–antibody complexes to a single organ.

The Arthus reaction occurs after injection of antigen intradermally in individuals who have previously been exposed (e.g. in immunization) and therefore have high antibody levels. The antigen–antibody complex is deposited, causing an acute inflammatory reaction lasting between 4 and 10 hours.

Type IV (cell-mediated) hypersensitivity

In this type, antigen-sensitized T cells release cytokines following a second contact with the same antigen. These cytokines induce inflammatory reactions and attract and activate macrophages to release mediators.

Examples include the following:

- Contact hypersensitivity by an epidermal reaction via Langerhans' cells. This peaks at 48 hours.
- Tuberculin type hypersensitivity caused by a subdermal injection of tuberculin producing a reaction in the dermis that peaks at 48–72 hours (the Mantoux test).
- Cell-mediated hypersensitivity results in a granulomatous reaction and is usually caused by persistent antigen in macrophages (Tb). Reaction peaks at approximately 4 weeks.
- Giant papillary conjunctivitis, vernal keratoconjunctivitis, and atopic keratoconjunctivitis, all of which are part type I and type IV reactions.

Transplants and graft rejection

See Chapter 6.

Ageing, degeneration, and dystrophy

Many ocular diseases are due to an accumulation of the effects of age on the eye.

Ageing involves:

- a decline in tissue cellularity
- a reduction in blood flow due to vascular disease
- atrophy of tissue
- tissue replacement with an acellular collagenous matrix.

Degeneration is a secondary phenomenon resulting from previous disease (Table 4.1). It occurs in tissue that has reached its full growth and can come in many forms. It commonly involves connective tissue components such as collagen, elastin, and proteoglycans.

Dystrophy is a primary, inherited disorder that can occur at any age. Dystrophies may involve a single matrix component. They are discussed in more detail in the section on corneal dystrophies on page 134.

Cellular and tissue reactions

Cells respond in various ways to stress:

- Hypertrophy: an increase in size of cells, fibres, or tissues without an increase in number (e.g. RPE hypertrophy).
- Atrophy: a decrease in size of cells, fibres, or tissues.
- Hyperplasia: an increase in the number of individual cells in a tissue; their size may or may not increase. Growth will reach equilibrium and is not indefinite (e.g. RPE hyperplasia secondary to trauma).
- Hypoplasia: arrested development of a tissue during embryonic life (e.g. aniridia).
- Aplasia: lack of development of a tissue in embryonic life (e.g. aplasia of the optic nerve).
- Metaplasia: the transformation of one type of tissue into another type (e.g. in anterior subcapsular cataract fibrous metaplasia of the lens epithelium). This can arise due to chronic irritation and will usually involve columnar or cuboidal epithelium changing to squamous epithelium.
- Dysplasia: abnormal growth of tissue with increased mitoses and reduced differentiation (e.g. retinal dysplasia). Dysplastic tissue is not invasive and will not pass through the basement membrane.

Neoplasia and preneoplastic conditions

Neoplasia is an increase in the number of cells in a tissue; growth exceeds and is uncoordinated with that of the normal tissues. This results in a neoplasm, which persists in an excessive manner after cessation of the stimuli that evoked the change. Neoplasia differs from hyperplasia as its growth never attains equilibrium.

Neoplasia is caused by an upregulation of proliferation (excessive or inappropriate oncogene action) or a failure

Table 4.1 Types of degeneration

Type of degeneration	Histopathological mechanism	Clinical examples
Hyaline	Replacement of normal cells with an acellular, amorphous, eosinophilic material	Walls of arteriolosclerotic small vessels of the eye in ageing, benign hypertension, and diabetes
Elastotic	Defective fibroblast function leads to an altered elastic matrix and reduced elasticity	Skin in ageing individuals, pterygia and pseudoxanthoma elasticum in which ruptures in Bruch's membrane expose the choroid (angioid streaks)
Calcification	Calcium is deposited as hydroxyapatite crystals, which can be metastatic in hypercalcaemic states or dystrophic in normocalcaemic states	Band keratopathy involves calcification of Bowman's layer and the superficial stroma This also occurs in hyperparathyroidism, hypervitaminosis D, and sarcoidosis Cataracta ossea is calcification in the fibrous and degenerative cortex of the lens Bruch's membrane can be calcified in Paget's disease In phthisis bulbi, ossification of the metaplastic fibrous tissue derives from proliferation of the RPE in a hypotonic eye Woven and lamellar bone is located on the inner surface of Bruch's membrane Ossification can extend into the vitreous and choroid
Amyloid	Insoluble protein deposited in tissue around blood vessels and basement membranes In H&E stains amyloid has a homogeneous pink appearance, staining with Congo red followed by examination with a polarized light, giving an apple green birefringence appearance Amyloid deposition can be localized or systemic	Localized: • solitary nodule in eyelid, orbit, conjunctiva—in the cornea it can be seen in lattice dystrophy and gelatinous drop-like dystrophy • amyloid from polypeptide hormones in endocrine tumours • amyloid from prealbumin leads to cerebral deposits in Alzheimer's diease Systemic: • pseudoexfoliation syndrome—an amorphous, eosinophilic substance is deposited on the anterior capsule of the lens, ciliary processes, iris surface, and trabecular meshwork, leading to secondary glaucoma; it is also deposited in the skin and viscera • Waldenström's macrogloblinaemia—amyloid is light-chain derived from immunoglobulin • rheumatoid arthritis and familial Mediterranean fever amyloid is derived from serum protein, an acute phase reactant in inflammation
Hydropic	Reversible change—cells are enlarged, containing cytoplasmic vacuoles	Infection, intoxication, anaemia, or circulatory disturbance
Cloudy swelling	Reversible change—cells are enlarged and filled with granules or fluid, representing intracellular oedema	Mild infection, intoxication, anaemia, or circulatory disturbance
Fatty change	Fat accumulates in cells for unknown reasons or after damage by a variety of agents	Arcus senilis of the cornea—fatty infiltration of the peripheral corneal stroma Xanthelasma—lipid within clumps of macrophages in the dermis of the eyelid seen in ageing and hypercholesterolaemia Deposition of lipid and cholesterol in the intima of arteries, leading to atheroma, can lead to thrombosis
Glycogen infiltration	Glycogen infiltration into tissue, leading to structural change	Diabetes mellitus—lacy vacuolation of iris pigment epithelium Long-standing neural retinal detachment due to lack of nutrition and in proliferating RPE cells

H&E, haematoxylin and eosin.

of mechanisms that lead to cell death (tumour suppressor genes).

A neoplasm may be benign or malignant.

Carcinogenesis

Carcinogenesis is non-lethal genetic damage that contributes to a cell undergoing neoplastic change. Genetic damage may be DNA deletion, mutation, amplification, translocation, or insertion which leads to loss or gain in function.

Environmental carcinogenesis can be divided into chemical carcinogenesis, physical carcinogenesis, and microbial carcinogenesis.

In chemical carcinogenesis (benzopyrenes, polycyclic hydrocarbons, 2-naphthylamine after liver hydroxylation, nitrosamines in gastric carcinoma, cyclophosphamide, aflatoxins, and arsenic) initiation involves a short exposure with mutation. Promotion is a long-term exposure which encourages cellular proliferation and is usually not mutagenic. Some substances can initiate and promote. These are complete carcinogens.

HELPFUL HINT

Benign versus malignant neoplasms

Benign	Malignant
Non-invasive (some benign tumours can invade, e.g. meningioma and pleomorphic adenoma)	Invasive (infiltrates and actively destroys surrounding tissue)
Non-metastatic	Metastatic (develops secondary deposits of neoplastic growth distant from the primary site via local, lymphatic, haematogenous, intraepithelial, or disseminative spread)
	Tumour cells spread via secretion of lytic enzymes to breach basement membranes, loss of contact inhibition, and increased cell movement
Reasonable differentiation	Anaplastic
May maintain some normal polarity/function	Variation from the normal structure or behaviour, including heterogeneity, nuclear pleomorphism, high mitoses with loss of differentiation, and polarity
Sustained angiogenesis, e.g. VEGF	Sustained angiogenesis, e.g. VEGF
Limitless replicative potential	Limitless replicative potential

VEGF, vascular endothelial growth factor.

In physical carcinogenesis (ionizing radiation and UV radiation) ionizing radiation directly damages DNA, especially in cell proliferation, and can result in a range of changes from single-gene mutations to major chromosome deletions. UV mainly affects the skin, forming pyridimine dimers that are usually excised by DNA repair mechanisms.

In microbial carcinogenesis (viral and bacterial) oncogenic viruses may cause conjunctival papillomas (type 16) or lacrimal papillomas (type 11). Epstein–Barr virus can cause orbital Burkitt's lymphoma and intraocular large B-cell lymphoma in the immunosuppressed.

Gene control in neoplasia, including retinoblastoma

Proto-oncogenes and tumour-suppressor genes act normally to balance cell growth, regeneration, and repair. This balance is lost in neoplasia.

In addition a loss of the ability to control apoptosis or repair DNA can lead to neoplasia.

Proto-oncogenes

Proto-oncogenes code for proteins involved in cell proliferation, including growth factors and their receptors, signal transducers, and nuclear regulating proteins. In neoplasia, proto-oncogenes become oncogenes through structural change, chromosomal translocations, or amplification (Table 4.2).

Tumour-suppressor genes, including the retinoblastoma gene

Tumour-suppressor genes switch off cell proliferation. Loss of both copies of a tumour-suppressor gene is required for neoplasia to develop. The gene in neurofibromatosis type 1 is located on the long arm of chromosome 17 and acts as a tumour-suppressor gene. In addition to this the retinoblastoma (Rb) and p53 gene are good clinical examples of mutated tumour-suppressor genes that lead to neoplasia.

Table 4.2 Mechanisms of oncogene production

Mechanism	Example
Structural change	A single point mutation in the ras gene leads to an inability of the protein to hydrolyse GTP Signalling activity is therefore permanently switched on This produces an oncogenic effect
Translocation	Burkitt's lymphoma—*myc* proto-oncogene is translocated from chromosome 8 to 24, leading to permanent expression Chronic myeloid leukaemia—*abl* translocation from chromosome 9 to 22
Amplification	Overexpression of DNA segments—*myb* oncogene is overexpressed in some colorectal and gastric carcinomas

The Knudson 'two hit hypothesis' describes a theory in neoplasia such as retinoblastoma that can be inherited or sporadic. In the inherited form one gene is already defective in the germ line (the first 'hit'). The second 'hit' is due to a mutation in the second allele. Sporadic mutations involve two 'hits' in somatic cells.

Retinoblastoma and control of the cell cycle

Control of the cell cycle is regulated by the retinoblastoma protein (pRB) and E2F proteins (Fig. 4.5):

- E2F activates transcription of genes involved with DNA synthesis and production of cell cycle regulators.
- pRB binds to E2F and inhibits the activation of transcription by E2F.
- pRB/E2F complex binds E2F promoters and prevents unbound E2F initiating transcription.
- pRB can be inactivated by phosphorylation, mutation, or viral oncogene binding.
- Cyclin D1 and cdk4 mediate the phosphorylation of pRB.
- Cyclin D1/cdk4 complex is most active in the G1 phase of the cycle, causing phosphorylation of pRB and release of E2F. This allows for the G1 phase to enter the S phase.
- P16 is a cyclin-dependent kinase inhibitor that indirectly prevents phosphorylation of pRB.

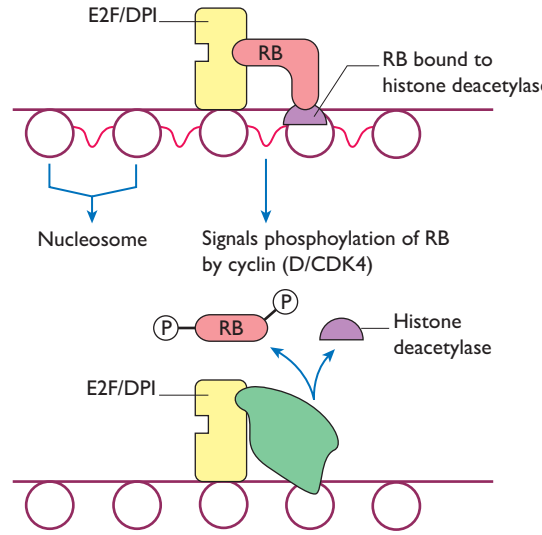


Fig. 4.5 Control of the cell cycle by pRB and E2F.
With permission from Louise Bye.

p53 and the cell cycle

The p53 gene is located on chromosome 17p13.1. It is the most common target for genetic alteration in human tumours. The major functions of p53 in response to DNA damage are cell cycle arrest and initiation of apoptosis.

Loss of control of apoptosis

Cell survival is regulated by genes that promote or inhibit apoptosis. The BCL/BAX family of genes is an example. BCL-2 is expressed in high levels in follicular B cell lymphoma. A translocation t (14:18) produces a fusion between the bcl-2 gene and the heavy chain gene. This leads to overexpression of bcl-2 protein, enhanced B cell survival, and neoplasia. Bcl-2 protects the cell from apoptosis through the mitochondrial pathway. The apoptosis repressor effects are counteracted by the BAX gene family, which induces apoptosis. BCL-2 and BAX can form homodimers and heterodimers. The ratio of homodimers to heterodimers will determine whether apoptosis occurs or not (Fig. 4.6).

Defects in DNA repair

In addition to possible DNA damage from environmental agents, the DNA of normal dividing cells is susceptible to alterations resulting from errors that occur spontaneously during DNA replication.

Genomic instability occurs when both copies of these genes are lost; thus they resemble tumour-suppressor genes.

Defects can occur in three types of DNA repair systems:

- mismatch repair
- nucleotide excision repair
- recombination repair.

Preneoplastic conditions

Hamartomas

A hamartoma is a non-neoplastic malformation that consists of a mixture of tissue normally found at a particular site. Two main types exist:

- Haemangiomas are a proliferation of vascular channels with a lobulated growth pattern. These may be capillary haemangiomas or cavernous haemangiomas, the latter of which involve large thick-walled calibre vessels. They can occur in the eyelid, orbit, or choroid. Capillary haemangiomas may regress during childhood whereas cavernous haemangiomas do not. Extensive haemangiomas occur as part of Sturge–Weber syndrome.
- Naevi are melanocytes that migrate through the dermis to reach epithelial cells. A naevus occurs due to abnormal migration, proliferation, and maturation. Naevi can occur in the conjunctiva. In the iris and choroid, naevi are static flat brown or black areas. Naevi at any site can progress to melanoma.

Choristomas

A choristoma is a non-neoplastic malformation consisting of a mixture of tissues not normally present at a particular site. For example:

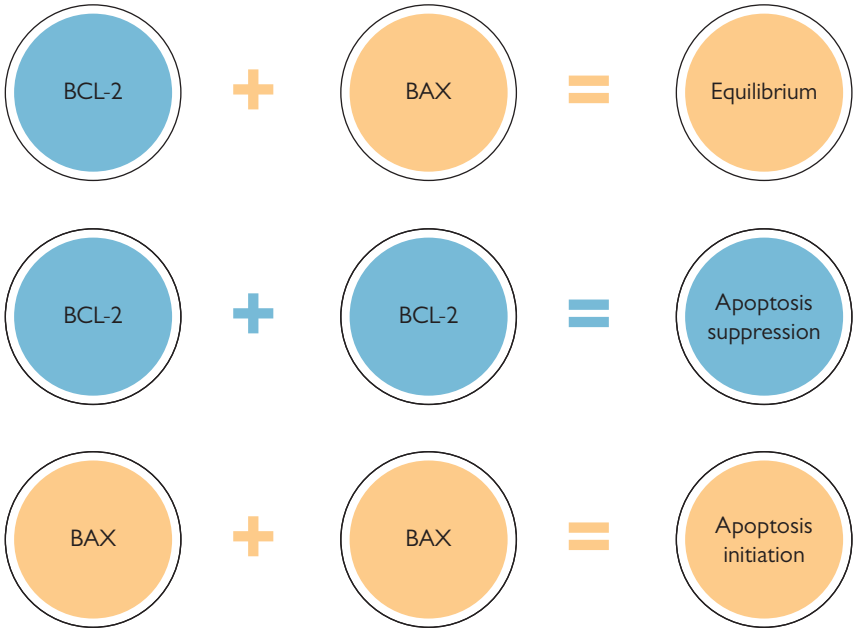


Fig. 4.6 Bcl-2 and BAX in apoptosis.

- Epibulbar dermoids present as a smooth white nodule with or without hair on the bulbar conjunctiva or at the outer angle of the bony orbit on the skin. They can be made up of fibrous tissue, fat, hair, and sweat glands.
- Phakomatous choristoma presents as a nodule in the eyelid. It consists of epithelial and basement membrane cells resembling a lens capsule in a fibrous stroma.

Thrombosis, emboli, and atheroma

Vascular disorders can be inflammatory (giant cell arteritis) or degenerative (diabetic retinopathy). In order to understand the pathology it is important to revisit the basic mechanisms of thrombosis, emboli, and atheroma formation.

Thrombosis

Before considering thrombus formation it is important to understand platelet structure.

Platelets are made up of four zones:

- peripheral
  - rich in glycoproteins; needed for platelet adhesion and aggregation
  - contains platelet factor 3, which promotes clotting during aggregation
- sol-gel
  - contains microtubules and microfilaments
- organelle
  - contains α granules, which contain factor VIII, factor V, fibrinogen, fibronectin, platelet-derived growth factor, and chemotactic factors
  - contains dense bodies that contain ADP, calcium, and 5HT
- membrane
  - contains the dense tubular system responsible for platelet contractile functions and prostaglandin synthesis.

Once a thrombus has formed (Fig. 4.7), it may:

1. detach from the vessel wall forming an embolus
2. be lysed by plasmin
3. persist at the vessel wall to form an occlusive thrombus—recanalization can occur through an occlusive thrombus
4. form a mural thrombus covered by smooth muscle cells which then becomes vascularized by blood vessels from the main lumen.

The homeostasis of clotting is much more involved than the relatively basic explanation above and is maintained by a combination of prothrombotic and anti-thrombotic factors (Fig. 4.8):

- Protein C is a vitamin-K-dependent serine protease that potently inhibits factors Va and VIIIa. Protein C deficiency

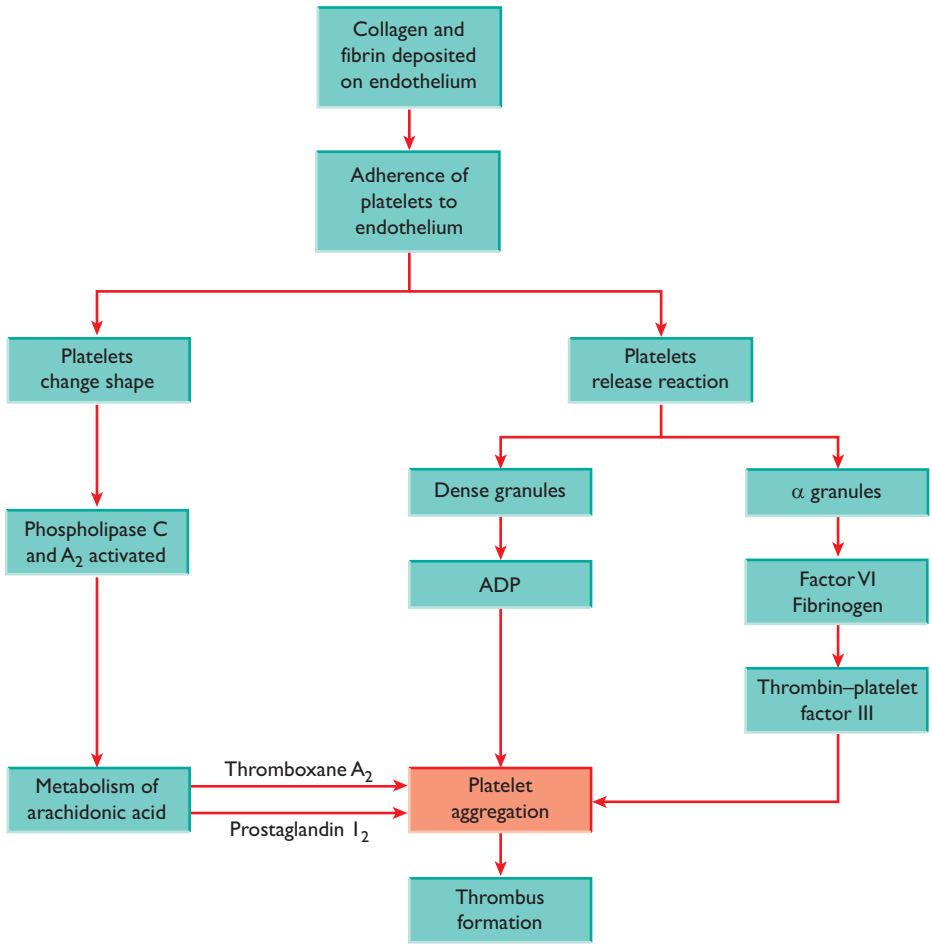


Fig. 4.7 The process of thrombus formation.
With permission from Louise Bye.

is inherited as an autosomal dominant trait and clinically affected individuals are heterozygous, with a protein C concentration of 50%. Prevalence is 6–8% of young patients with venous thrombosis.

- Factor V Leiden mutation leads to resistance of factor V to breakdown by protein C. Five per cent of the white population are carriers.
- Protein S is a cofactor in the inactivation of factor Va and VIIIa by activated protein C. Protein S deficiency inheritance is autosomal dominant and prevalence in young patients is 5–8%.
- Prevalence of anti-thrombin III deficiency in young patients with venous thrombosis is 0–5%.

Risk factors for thrombosis

When considering the risk factors it is important to consider Virchow's triad of conditions that predispose to thrombus formation:

1. A change in blood flow, e.g. venous stasis, arrhythmia, valvular disease.
2. A change in the vessel wall, e.g. atherosclerosis, trauma, inflammation, or neoplastic change.
3. A change in blood constituents, e.g. an increase in the number of platelets or altered platelet function.

Emboli

An embolus is an abnormal mass of matter carried in the bloodstream that is large enough to occlude a vessel.

Atheroma

Atheroma form following vascular endothelial cell damage. Platelets adhere to the endothelium and promote the smooth muscle cells within the vessel intima to proliferate. The main stimulation factor here is platelet-derived growth factor. Intracellular and extracellular lipid accumulates at the site due to a breakdown in the endothelial cell barrier function. This forms what is

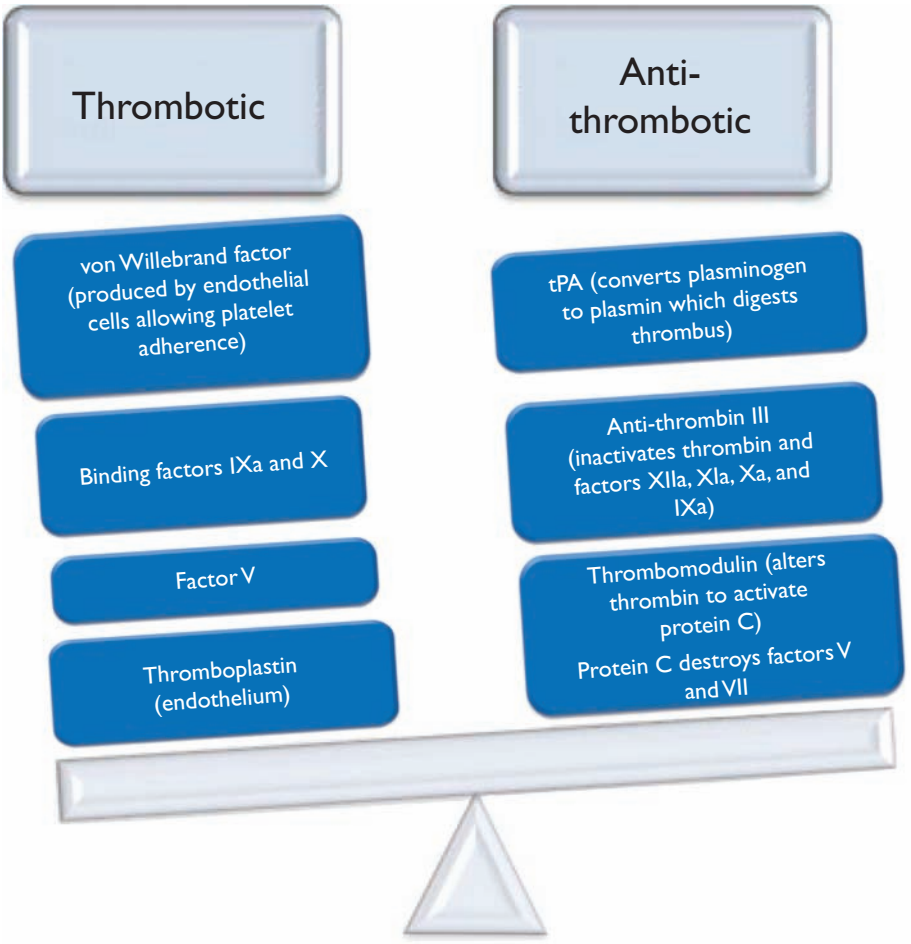


Fig. 4.8 The homeostasis of clotting: a delicate balance.

known as a fibrolipid or atherosclerotic plaque. One school of thought believes that the plaques are derived from fatty streaks, which can be seen as early as 10 months of age (Fig. 4.9).

Atherosclerotic plaques can lead to complications:

- Growth of the plaque can lead to necrosis and softening of the plaque base, which in turn leads to aneurysm formation due to thinning of the adventitia media.
- Small plaque fissures can lead to microthrombi and larger fissures can lead to emboli.

Risk factors for atherosclerosis are either reversible or non-reversible. Reversible risk factors include cigarette smoking, hypertension, diabetes, and hyperlipidaemia. Irreversible include age, male sex, and race. High levels of high-density lipoprotein are protective.

Additional examples of basic ocular pathology

This section covers some specific examples of basic ocular pathology not mentioned in detail in the previous sections.

Granulomatous inflammation

Sympathetic uveitis or ophthalmia

This is a bilateral diffuse granulomatous inflammation. It is a T-cell-mediated panuveitis that occurs anytime from 5 days to many years following a penetrating eye injury. The injury is usually associated with traumatic uveal incarceration or prolapse.

Presenting symptoms are blurred vision and photophobia in the non-injured eye. Granulomatous uveitis develops with the appearance of mutton fat keratic precipitates. These are collections of epithelioid cells plus lymphocytes, macrophages, multinucleated giant cells, or pigment on the endothelium of the cornea.

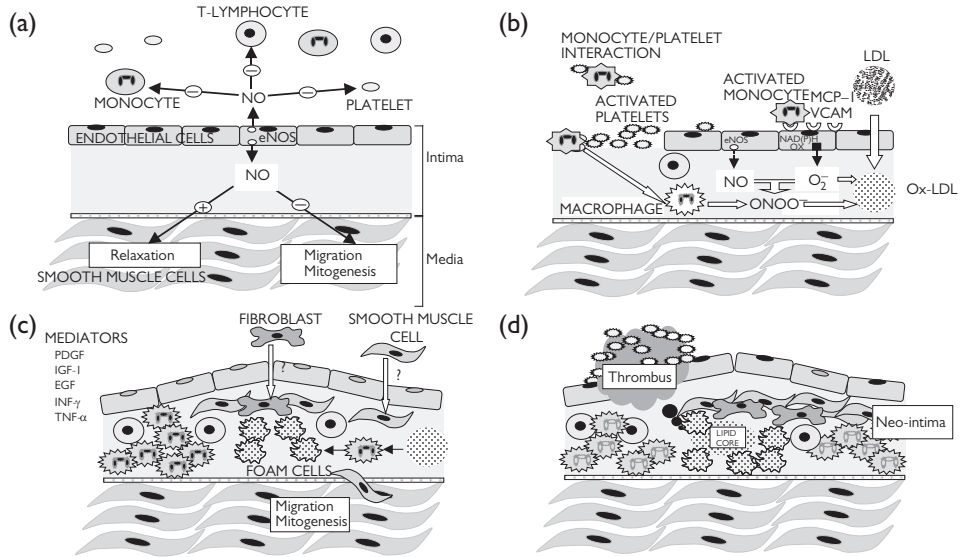


Fig. 4.9 The process of plaque formation and the structure of an atherosclerotic plaque.

Reproduced from R. Wilkins et al., *Oxford Handbook of Medical Sciences, The Essential Guide to the Sciences that Underpin Medicine*, Figure 6.43, Page 461, 2011, with permission from Oxford University Press.

HELPFUL HINT

Different types of emboli and their origins

Type of embolus	Origin
Thrombotic	Leg and pelvic thrombi are the cause of pulmonary embolism and can lead to postoperative death In the retina an embolus can originate from valvular thrombi, atheromatous plaques in the aorta, or carotid arteries or a mural thrombus from the left ventricle
Air	Can occur during thyroid surgery
Tumour	Can occur in the retinal or choroidal circulation
Fat	Following severe trauma and multiple fractures
Atheromatous	Cholesterol/calcified tissue/fibrin platelets can migrate into the retinal circulation
Septic	Subacute bacterial endocarditis, leading to Roth's spots
Amniotic fluid	Complication of delivery, particularly if the foetus is manipulated

Inflammation also involves the retinal pigment epithelium with accumulation of macrophages, here called Dalen–Fuchs' nodules.

The cause is thought to be related to an uveal reaction to antigens localized on the RPE or uveal melanocytes. The reaction is delayed type hypersensitivity. Certain human leukocyte antigen (HLA) types are associated with its development, including HLA DRB1*04, DQA1*03, and DQB1*04.

Phacoanaphylactic endophthalmitis

This is an autoimmune, zonal granulomotous inflammation caused by rupture of the lens capsule and a consequent reaction to the lens material. This may result from the breakdown of tolerance at the T-cell level and consequently the formation of an antibody–antigen reaction.

Macrophages and lymphocytes enter the anterior chamber from dilated blood vessels in the iris and ciliary body. The macrophages engulf the lens matter and can block the anterior chamber angle, leading to phacolytic glaucoma.

Non-granulomatous inflammation and autoimmune disease

Endophthalmitis is inflammation of one or more coats of the eye and adjacent cavities. Panophthalmitis is inflammation of all three coats of the eye and can spread to orbital structures.

Suppurative endophthalmitis

The term 'suppurative' describes a tissue necrosis and is characterized by the presence of polymorphonuclear leukocytic infiltration into the involved tissues.

The source of inflammation may be:

- exogenous: sources originate outside the eye and body, e.g. surgical trauma, penetrating trauma, radiation, and chemical injury

or

- endogenous: sources originate inside the eye, e.g. inflammation due to cellular immunity such as Behçet's disease, spread from contiguous structures, or haematogenous spread.

Non-suppurative uveitis and endophthalmitis

The inflammation is generally termed anterior, intermediate, or posterior, although all three may be affected in a panophthalmitis.

The source of inflammation may be:

- exogenous: traumatic anterior uveitis in blunt trauma, penetrating injury inducing a sterile inflammation secondary to foreign body exposure

or

- endogenous: idiopathic (the most common form) inflammation associated with viral or bacterial infection or local ocular disease such as pars planitis and Posner–Schlossman's syndrome or inflammation associated with systemic disease such as Reiter's syndrome, Crohn's disease, or ulcerative colitis.

Sjögren's syndrome

This can be a primary entity or occur related to a secondary systemic disease such as rheumatoid arthritis. It is a disorder of the acinar glands of the conjunctiva, lacrimal gland, and oral mucosa. The secretory acinar tissue of the lacrimal gland is destroyed by a lymphocytic infiltrate. In addition there is loss of the conjunctival goblet cells, leading to impaired tear secretion and dry eyes.

Rheumatoid eye disease

This is an immune complex and T-cell-mediated mechanism. The effect on the eye covers a wide spectrum from mild to severe. This includes:

- necrotizing scleritis and peripheral corneal ulceration due to immune complex deposition, which leads to complement activation, PMN infiltration, collagenase production, and corneal melt
- spontaneous central—peripheral is more common—corneal ulceration with or without inflammatory cell infiltrate, otherwise known as corneal melt
- scleromalacia perforans due to thinning of the sclera and exposure of the underlying uveal tract, although perforation is rare (this is usually due to arteriolar infarction)
- brawny scleritis due to inflammation and a reactive fibrosis
- posterior scleritis, which can lead to macular oedema.

Thyroid eye disease

This can occur in hyperthyroidism, hypothyroidism, and euthyroid patients. The orbits become congested, swollen, and painful. Symptoms range from mild dry eye to sight-threatening disease. It may be bilateral or more rarely unilateral. Features include dry eye due to corneal exposure, conjunctival chemosis, proptosis, eyelid retraction and lid lag, superior limbal keratopathy, diplopia, and compressive optic neuropathy.

Histology reveals a perivascular lymphocytic infiltration with mast cells and glycosaminoglycan accumulation within and around the extraocular muscles and fat. In the later stages this progresses into fibrosis.

Physical trauma and chemical injuries

Physical trauma may be mechanical (surgical or accidental or assault), radiation, or thermal damage.

Mechanical

In blunt trauma the globe remains intact. There is an anterior–posterior deformation, which has a variety of effects:

- separation of attachments, for example separation of the ciliary muscle attachment to the scleral spur leads to trabecular meshwork collapse and angle recession glaucoma
- retinal oedema or commotio retinae thought to be due to either (i) retinal vessel spasm producing ischaemia and endothelial cell damage with leakage into tissue or (ii) transient interruption of axoplasmic flow in the ganglion cell processes
- shearing of photoreceptors leading to a reactive RPE proliferation and a pseudoretinitis pigmentosa
- perforating and penetrating injuries leading to disruption of the normal ocular anatomy and thermal damage, which can be used to induce scarring in the peripheral retina during treatment of detachments, e.g. cryotherapy.

Radiation

Light in the visible wavelengths at normal intensity does not have any acute effects on the cornea, lens, or retinal photoreceptors as the cornea and lens absorb UV and blue wavelengths. Production of free radicals by photons and antioxidant mechanisms are in fine balance, which has the potential to be disturbed, e.g. snow blindness in which the corneal epithelium is damaged.

LASER (light amplification by stimulated emission of radiation)

The damage caused depends on wavelength, site of absorption, and quantity of energy released.

Panretinal photocoagulation is performed for high-risk proliferative diabetic retinopathy. Here the outer third of

the retina is destroyed with underlying RPE cells. Reactive proliferation of RPE at the edge of the burn leads to pigmentation around a white scar formed by glial cells.

Ionizing radiation

Ionizing radiation may be charged, uncharged, or electromagnetic. The unit (gray, Gy) is a measure of the amount of energy absorbed by the tissue. For melanoma either proton beam therapy or electromagnetic therapy (γ-rays) can be used at up to 110 Gy. For retinoblastoma X-rays and γ-rays are used with a level of 40–60 Gy.

There are many side effects of radiation to the eye. These include:

- endarteritis due to infiltration of the vessel wall by inflammatory cells and proliferation of spindle cells within the internal elastic lamina (this contributes to tumour necrosis but also can lead to telangiectasia and leakage of plasma into surrounding tissue)
- irradiation of the orbit, leading to dry eye due to lacrimal gland damage
- an increased risk of mutations, which can lead to second malignancy or congenital defects
- formation of cataract
- radiation retinopathy.

Chemical injury

Acid burns

Acids coagulate proteins. This means diffusion through the cornea and sclera is more limited. Necrosis of epithelium or stroma can lead to fibrosis and scarring of conjunctiva and cornea.

Alkali burns

Alkalis pass through tissue easily. Corneal damage can be severe and limbal ischaemia can result, preventing epithelial repair. The high pH destroys cells of the lens, uveal tract, and retina.

Corneal dystrophies and corneal endothelial dysfunction

Corneal dystrophies are a group of inherited, bilateral, and progressive diseases that lead to corneal opacification. Conditions are classed according to the layer affected.

Many of the corneal dystrophies (Table 4.3) are linked to mutations in the transforming growth factor-β-induced gene (*BIGH1*) on chromosome 5q31, particularly the dystrophies involving Bowman's layer and stromal layers. BIGH1 codes for a protein expressed on the cell membrane of corneal epithelium and stromal keratocytes, which aids with wound healing. Mutations cause abnormal folding for the proteins and amyloid or non-fibrillar deposits. All are autosomal dominant with complete penetrance.

Corneal endothelial dysfunction

In addition to the above dystrophies there are some conditions that are not classed as dystrophies but can lead to corneal endothelial dysfunction.

Iridocorneal endothelial syndrome

This is unilateral, sporadic, and occurs in adults. On specular microscopy the corneal endothelium has degenerate endothelial cells which may be surrounded by normal cells. The endothelial cells form blebs and can acquire numerous microvilli on their posterior surface.

Table 4.3 Summary of the corneal dystrophies

Affected layer	Specific pathophysiology
Epithelial	Cogan's microcystic dystrophy or map dot dystrophy • Sporadic or autosomal dominant—onset in second decade • Degeneration of cells with cyst formation leads to unstable epithelium • Histology shows thickened basement membrane, absent or abnormal hemidesmosomes, and fibrillar material between basement membrane and Bowman's layer Meesman's dystrophy—rare, autosomal dominant • Onset in early childhood • Separation of cells with deformation of neighbouring cells, leading to formation of loops of basement membrane and microcysts
Bowman's layer	Reis–Bücklers' dystrophy • AD • Mutation of the TGFB1 gene • Fine reticular opacity in the superficial cornea in early childhood or adult life • Histology shows nodules of fibrous tissue between Bowman's layer and epithelium • Electronmicroscopy gives electron-dense rods Thiel–Behnke dystrophy • AD • Mutation links to chromosome 10 • Usually presents in older patients • Histologically identical to Reis–Bücklers' but EM shows curly fibres within the superficial fibrous nodules

(*continued*)

Table 4.3 (*continued*)

Affected layer	Specific pathophysiology
Stromal	Granular and Avellino dystrophy • AD inheritance • Linked to TGFB1 gene • Onset is early life with white, well-demarcated stromal deposits • Histology shows discrete opaque granules in the anterior corneal stroma • The anterior stroma and Bowman's layer contain non-birefringent hyaline bodies with keratinoid Avellino or granular lattice dystrophy • Histologically and clinically the same as granular dystrophy with the addition of amyloid deposits seen in lattice dystrophy Lattice dystrophy • Normally AD or type 1 • Fine branching lattice lines are seen in the anterior stroma with a variable amount of stromal haze • Microscopy shows deposits of amyloid in the anterior stroma staining orange-red with congo red dye and exhibiting apple green birefringence with polarized light • Type 1 is AD-onset in first decade of life • Type 2 is AD and is rare; coexists with systemic amyloidosis; onset in third decade • Type 3 is AR; onset in fifth decade Macular dystrophy • Less common than the other stromal dystrophies • Autosomal recessive inheritance • Onset is in the first decade with focal grey-white anterior stromal opacities in the axial region • Mucopolysaccharide granules accumulate in the cytoplasm of keratocytes and in the adjacent interlamellar spaces when the cells rupture Schnyder crystalline dystrophy • Rare • AD • Local disorder of corneal lipid metabolism • Strongly associated with systemic hypercholesterolaemia • Fine polychromatic cholesterol crystals are deposited in the anterior stroma with a dense corneal arcus Gelatinous droplike dystrophy • Autosomal recessive corneal subepithelial amyloidosis • Presents in first decade with a band keratopathy
Endothelial	Congenital hereditary endothelial dystrophy • Primary dysfunction of corneal endothelial cells, leading to corneal opacification • The endothelium may be attenuated and vacuolated • Descemet's membrane exhibits fine lamination with an abnormal layer of collagen at the ultrastructural level • There are two types: CHED1 is autosomal dominant, presenting in the second year of life, and CHED2 is autosomal recessive, presenting from birth or within the first few weeks of life Fuchs' AD endothelial dystrophy • Very common—more common in females than males • Sporadic or inherited in an AD manner • Onset is usually after the age of 50 • Dystrophic endothelial cells result in the formation of guttata or excrescences as a result of abnormal collagen deposition on a thickened Descemet's membrane • A reduction in number and function of the cells and their pump action leads to progressive corneal oedema • Persistent epithelial oedema can lead to a bullous keratopathy Posterior polymorphous dystrophy • AD • Histology shows multilayered endothelial cells which have some of the features normally seen in epithelial cells, e.g. microvillae • In severe diffuse disease the posterior corneal surface is lined by stratified cells with prominent desmosomal attachments resembling corneal epithelial cells • The condition can lead to astigmatism and decompensation but is normally asymptomatic

These abnormal cells can act to form a membrane over the angle structures. The late outcome is corneal decompensation and oedema with or without glaucoma.

Iridocorneal endothelial syndrome is seen in association with progressive iris stromal atrophy, glaucoma due to endothelial sliding with a normal iris (Chandler syndrome), and the presence of an iris naevus (Cogan–Reese syndrome).

Bullous keratopathy

This is caused by Fuchs' endothelial dystrophy or other factors and occurs due to endothelial decompensation. This leads to corneal oedema, which progresses from stromal to epithelial microcystic and then epithelial macrocystic (bullous) oedema. Examination shows stromal oedema with Descemet's folds followed by epithelial oedema, subepithelial scarring, and corneal neovascularization.

Other causes of bullous keratopathy are:

- intraocular surgery, e.g. aphakic or pseudophakic bullous keratopathy more likely to occur following complications or placement of an anterior chamber lens
- endothelial cell inflammation due to herpes simplex or zoster
- corneal graft failure/rejection
- chronic anterior uveitis
- trauma.

CLINICAL TIP

Causes of cataract

Aetiology

Age related

Drug related
- Corticosteroid: topical or systemic
- Amiodarone
- Aspirin
- Chlorpromazine
- Topical glaucoma medication, e.g. pilocarpine

Traumatic
- During surgery, e.g. trabeculectomy
- Penetrating, blunt, or chemical injury
- Electrocution

Irradiation

Systemic disease
- Diabetes mellitus
- Myotonic dystrophy
- Wilson's disease
- Atropic dermatitis
- Neurofibromatosis type 2

Fabry's disease

Ocular disease
- Uveitis
- Myopia
- Acute angle closure glaucoma

Retinal dystrophy (retinitis pigmentosa)

Acquired cataract

Cataract can be divided into congenital and acquired. Acquired cataract has many possible causes.

Cataract can have a variety of morphologies, which are sometimes related to their aetiology.

Some of the more common morphologies include the following:

- Nuclear: this is the most common type of age-related cataract. It can cause increasing myopia as it matures.
- Cortical: these have a more radial appearance in the cortical zone of the lens.
- Subcapsular: posterior subcapsular can occur due to corticosteroid use.

In addition cataracts vary in their level of maturity:

- Immature: incomplete opacification of the lens.
- Mature: dense, white cataract which obstructs the red reflex and fundal view.
- Hypermature: mature cataract ages and leaks water, causing shrinkage of the lens capsule.
- Morgagnian: cortical lens matter liquefies, with inferior displacement of the nucleus in the capsular bag.

Age-related cataract

Age-related cataract is by far the most common form seen.

Ageing of the lens and cataract formation are two different entities. In ageing the level of oxidation is much lower than in age-related cataract, which exhibits extensive oxidation of lens proteins. In the normal ageing lens there is constant post-translational modification of lens proteins throughout life. γ-Crystallins are generally synthesized less and β-crystallins increase.

In age-related cataract there is:

- a breakdown of the antioxidant mechanisms in the lens, mainly reduced glutathione levels
- increased proteolytic activity
- UV light absorbed by lens tryptophan, which is then converted to compounds that act as photosensitizers and cause free radical formation—free radicals downregulate Na^+-/K^+-ATPases in the lens epithelium, leading to water influx.

There may also be decreased function of the MIP26 or aquaporin molecule, leading to water influx. In addition as α-cystallins denature they have reduced chaperone activity. They bind to unfolded proteins but lack the ability to refold βγ-crystallins. These changes lead to increased insoluble proteins, amino acid oxidation, increased chromophores, and loss of αA-crystalline and γS-crystalline.

Mechanisms of diabetic cataract formation

In diabetes-induced cataract it is thought that high glucose and galactose concentrations in the aqueous lead to an increase in intracellular glucose in the lens.

This increase overwhelms the anaerobic glycolysis pathway and hence prevents the normal aldose reductase inhibition of the polyol pathway. This leads to an upregulated polyol pathway and accumulation of polyols (sorbitol) in the cells. This acts to drag water into the cell. Aquaporin of MIP26 is then activated. This leads to reduced ATP and glutathione, and hence oxidative damage.

Whether age-related or diabetes-induced, loss of these mechanisms leads to the following cellular changes:

- loss of cell organization
- the formation of vacuoles within lens fibres
- water accumulation within the lens
- disruption of lens crystallin organization and formation of lens protein aggregates
- accumulation of protein aggregates and chromophores, which leads to changes in colour from yellow to red to black and reduced transmission of light.

Glaucoma

Glaucoma can be classified into primary or secondary types. In addition there are congenital forms of glaucoma that arise due to various malformations of the outflow system.

Primary open-angle glaucoma

Primary open-angle glaucoma (POAG) is the most common form of glaucoma. It is an age-related disease and is the leading cause of irreversible blindness in the developed world. Risk factors include:

- smoking
- diabetes
- hypertension
- hypercholesterolaemia
- myopia.

POAG has been linked to mutations in chromosomes 1 and 3p. The GLC1A gene (chromosome 1 open-angle glaucoma gene) codes for a protein known as TIGR (trabecular meshwork inducible glucocorticoid response) or MYOC (myocillin). Individuals with the mutation have an earlier age at onset, higher peak intraocular pressure (IOP), and are more likely to need surgery.

High IOP is a risk factor for POAG. The rise in IOP is thought to be due to an abnormal resistance to outflow of aqueous. The rise is usually slow and progressive. This results in a pressure-induced ischaemia of the capillary bed of the optic nerve and direct mechanical pressure reducing axoplasmic flow in the axons passing through the lamina cribrosa. The nerve fibre bundles passing into the optic nerve head above or below the horizontal line on the temporal side of the disc and the prelaminar optic nerve fibres are selectively damaged.

In addition this rise in IOP may be accompanied by occlusive disease, leading to poor perfusion of the posterior ciliary arteries, resulting in ischaemic optic atrophy.

The combination of poor perfusion due to occlusive disease and pressure-induced ischaemia causing prelaminar nerve fibre atrophy leads to enlargement of the optic nerve head in the vertical plane and field loss, for example a nasal step or arcuate scotoma.

Primary angle-closure glaucoma

This is, in part, an age-related disease. It is also more likely to occur in a hypermetrope (small eyes) or following blunt trauma, where the angle may already be damaged. It can occur acutely or on a more chronic and intermittent basis. As the globe becomes smaller with age and the lens enlarges, the anterior surface of the lens pushes the iris anteriorly, causing the drainage angle to narrow. As pressure builds up behind the iris it forces the iris more peripherally towards the trabecular meshwork, causing an acute rise in pressure.

Secondary open-angle glaucoma

The angle is obstructed by matter, for example:

- cells and protein in inflammation (uveitis)
- haemorrhage in trauma
- tumour cell infiltration
- lens matter or macrophages containing lens matter (phacolytic glaucoma).

Secondary angle-closure glaucoma

The chamber angle is displaced and closed mechanically, for example:

- tumours of the eye, e.g. melanoma, forcing the lens forward
- anterior or posterior synaechiae in uveitis—adhesion between iris and trabecular meshwork (anterior) and iris and lens (posterior) produces obstruction of flow at the pupil, otherwise known as iris bombe
- rubeotic glaucoma caused by fibrovascular proliferation in the angle due to retinal ischaemia.

Retinal microvascular occlusion

Diabetes mellitus is the most common cause of blindness in the working population. Diabetes predominantly affects the retinal circulation but can also affect other tissues, leading to vacuolation of iris pigment epithelium, thickening of ciliary process basement membranes, and cataract.

Diabetic retinopathy

Diabetic retinopathy is predominantly a microvascular disease. It is a microangiopathy affecting precapillary arterioles, capillaries, and postcapillary venules. The abnormality seen at the histological level in the smaller vessels is multilayering of the basement membrane and degeneration of endothelial cells and pericytes. This results in capillary non-perfusion and tissue ischaemia.

Characteristics of microvascular diease include the following:

- Micro-infarction—fluffy white swellings or cotton wool spots in the retina. These are swollen ends of interrupted axons. The infarct mainly involves the nerve fibre layer.
- Hard exudates—reduced perfusion of the vascular bed and damage to the endothelium of the deep capillaries causes plasma leakage into the outer plexiform layer. This results in a yellow, well-circumscribed area known as exudates. Histologically these are eosinophilic masses containing foamy macrophages with lipid.
- Microaneurysms—ischaemia of the capillary bed leads to weakening of the wall by necrosis of the pericyte. This leads to bulging of the vessel wall known as microaneurysms.
- Haemorrhage—breakdown of vessel walls leads to leakage of red cells into the retina, leading to various types of haemorrhage:
  - flame haemorrhage—rupture of a small arteriole leading to leakage into the nerve fibre layer
  - dot haemorrhage—rupture of capillaries in the outer plexiform layer
  - blot haemorrhage—larger than dot haemorrhages; bleeding from capillaries with tracking between photoreceptors and the RPE.
- Neovascularization:
  - newly formed vessels grow from the venous side of the capillary bed within an area of arteriorlar nonperfusion otherwise known as intraretinal microvascular abnormalities (IRMAs). These vessels leak on fluorescein angiography and can progress to vasoproliferative retinopathy.
  - vasoproliferative retinopathy—ischaemic areas of retina release vasoformative factors which diffuse into the retina and vitreous. These stimulate endothelial cells to proliferate at the edge of the ischaemic area. New vessels form in the prevenular capillaries and venules, and proliferate within and on the surface of the retina. If the vitreous is detached, the fibrovascular tissue grows on the inner surface of the retina. The membrane can then contract and lead to retinal detachment. Proliferation within vitreous leads to haemorrhage and formation of traction bands. Diffusion of vasoformative factors to the iris surface and trabecular meshwork can lead to rubeosis iridis.

Retinopathy of prematurity

In intrauterine life, blood vessels grow from the disc towards the periphery driven by a relative hypoxia.

In a premature infant on supplemental oxygen the hypoxic drive is reduced and hence the extension of the normal vascular bed is inhibited.

Excessive proliferation of blood vessels occurs when the infant returns to normal oxygen levels. In addition the peripheral non-vascularized retina is now ischaemic, hence further driving neovascularization from the peripheral vessels, which grow rapidly and in a disorganized manner within the retina and vitreous. This can lead to a bilateral retinal detachment if left untreated.

Retinal macrovascular occlusion

Macrovascular occlusion refers to obstruction of vessels equal to or greater than a medium-sized arteriole. Some examples of macrovascular occlusion are given below.

Hypertensive retinopathy

Hyalinization of blood vessels leads to the appearance of 'copper or silver wiring'. As disease becomes more advanced, narrowing of the vessels with spasm produces an ischaemic effect on the endothelial cells distal to the constriction. As the endothelium swells and degenerates, leakage of fibrin into the vessel wall occurs, leading to further narrowing of the lumen. This leads to a fibrinoid necrosis of the choroidal and retinal vessels. If the choriocapillaris is involved lobular infarcts form what is known as Eschnig's spots.

Accelerated or malignant hypertension is characterized by haemorrhage, exudates, cotton wool spots, and papilloedema.

Central retinal artery occlusion

The central retinal artery is an end artery. Central retinal artery occlusion (CRAO) may be the result of thrombosis in a degenerate central retinal artery but is usually due to embolus most typically from an atheromatous plaque of the carotid artery. It leads to loss of vision. Following infarction the retina becomes opaque, preventing transmission of the red reflex created by the choroidal vasculature. The macula is spared due to the thinness of the fovea and lack of neuronal tissue, allowing a view of choriocapillaris creating a cherry red spot.

Central retinal vein occlusion

The radius of the central vein in the lamina cribosa of the optic disc is around 50% of the prelaminar part of the disc and smaller than the retrolaminar region. The resistance provided by the venous narrowing increases flow in the narrowed segment, leading to turbulence and an increased risk of thrombosis. This in part contributes to the pathology behind central retinal vein occlusion (CRVO).

The difference between central retinal artery and vein occlusion on fundoscopy is the presence of extensive haemorrhages in vein occlusion. There can also be an improvement in vision. Patients are more likely to develop neovascularization and rubeotic glaucoma, usually within 3 months of a CRVO. Neovascularization and rubeotic glaucoma are extremely rare in cases of CRAO.

Posterior vitreous detachment and retinal detachment

Posterior vitreous detachment

Age-related degeneration of the vitreous commences in the teenage years. Liquefaction of the collagen gel leads to vitreous syneresis and is the cause of floaters. Eventually the vitreous gel comes away from the optic nerve head, creating considerable movement in the posterior gel. This

CLINICAL TIP

Classification of neural retinal detachment

Rhegmatogenous	Non-rhegmatogenous
Equatorial type	Transudative
Myopia	*Uveitis*
Lattice degeneration	
Horseshoe tears or round holes	
Oral type	Exudative
Aphakic	*Choroidal tumour*
Traumatic dialysis	*Posterior scleritis*
Giant neural retinal break	
Macular type	Haemorrhagic
High myopia	*Diabetic retinopathy*
Post-traumatic	

is a posterior vitreous detachment (PVD). The vitreous base provides the centre of energy whilst the posterior vitreous responds to the energy. If the pull on the peripheral retina is sufficient it can cause a retinal tear or hole. In a U-shaped tear the base of the tongue of the retina is anterior because the vitreous first separates posteriorly, tearing the retina at a point of adhesion. The action of the vitreous extends the tear anteriorly towards the vitreous base.

Retinal detachment

A neural retinal detachment is separation between the neural retina and the retinal pigment epithelium. It can be caused by (i) accumulation of fluid beneath the neural retina, e.g. choroidal malignant melanoma, (ii) traction bands in the vitreous, as mentioned above in diabetic retinopathy, and (iii) accumulation of fluid beneath a broken neural retina with vitreous traction, e.g. a rhegmatogenous neural retinal detachment.

Age-related macular degeneration and retinal dystrophies

Dry age-related macular degeneration

This accounts for 90% of age-related macular degeneration (ARMD). It is characterized by a gradual reduction in central vision. The risk increases with age and is most common over the age of 70 years. Retinal damage is limited to the foveomacular area. Risk factors include:

- smoking
- female
- hypertension
- high fat and cholesterol intake
- blue iris and abnormal skin sun sensitivity.

Clinical examination reveals the following factors:

- Pigment disturbances due to clumps of pigmented cells at the level of the RPE.
- Drusen—round yellow spots represent abnormal thickening of the inner aspect of Bruch's membrane. Hard drusen (small and well-defined) do not predispose to advanced ARMD. Soft drusen (larger than 63 μm with ill-defined borders) increase in size and number with age and are a risk factor for advanced ARMD.
- Geographic atrophy—late stages of dry ARMD representing atrophy of the RPE.
- Histological examination of the macula reveals atrophy of the photoreceptors over well-defined eosinophilic mounds beneath the RPE in hard drusen and more linear granular bands in diffuse drusen. These are situated between the cell basement membrane and Bruch's membrane.

A basal linear deposit refers to a deposit between the RPE cell membrane and its basement membrane. This deposit type has been linked to the start of neovascularization or wet ARMD.

In addition, Bruch's membrane is thickened or calcified, and occasionally choriocapillaris is replaced by degenerative fibrosis.

At the cellular level macrophages and endothelial cells proliferate in the deposits beneath the RPE.

Wet age-related macular degeneration

This leads to a sudden onset of central visual loss and the patient may have been aware of the fact they suffered from dry ARMD. Following a cellular infiltration by macrophages and endothelial cells a proportion of cases develop new capillaries and choroidal or subRPE neovascularization develops. Rupture of these vessels leads to oedema and haemorrhage, thus attracting more macrophages and further neovascularization. The RPE undergoes a fibrous metaplasia with deposition of collagen, leading to a disc-shaped mass beneath the macular known as disciform degeneration.

Retinitis pigmentosa

This affects individuals in early adult life. Retinitis pigmentosa can be inherited in an AD, AR, or X-linked recessive fashion.

AD forms are associated with mutations in the gene coding for rhodopsin on the long arm of chromosome 3q and in the peripherin gene on chromosome 6p.

The initial symptoms are of night blindness and a progressive reduction in visual field from the periphery towards the posterior pole. At the end stage retinal function is limited to the central macular, hence the term 'tunnel vision'. Fundus examination findings include retinal atrophy, narrowing and opacification of retinal vessels, and mixed, coarse strands of pigmentation.

Histologically the outer nuclear layer at the fovea appears as a single layer of cells with stunted photoreceptors. Towards the periphery the outer nuclear layer is replaced by Müller cells, which fuse with the RPE. RPE cells react by proliferation and migration into the retina to become distributed around the hyalinized vessels, giving a bone spicule appearance.

Best's disease

This is an AD heredomacular degeneration in which there is loss of central visual acuity associated with a disc of yellow tissue at the macula. Histologically there is accumulation of lipofuscin in the RPE cells and atrophy of the photoreceptor layer of the retina.

Stargardt's disease

This is an AR inherited macular dystrophy.

The photoreceptor gene ABCA4 or ATP binding cassette transporter-retina also known as STGD1 on chromosome 1p21 is mutated. This leads to abnormal transport of metabolites across the disc membrane of the photoreceptors and ultimately an accumulation of lipofuscin in rod and cone disc spaces, leading to destruction of RPE and hence photoreceptors. Interestingly around 18.7% of cases of dry ARMD have mutations in the ABCA4 gene; therefore the role of this set of genes in the pathogenesis of macular dystrophy and degeneration is thought to be extensive.

Clinically atrophy of the macula is associated with the appearance of small yellow flecks. At the early stages the RPE is enlarged by accumulation of lipofuscin and melanin. The outer layer of the retina is lost and the pigment epithelium is absent, causing fusion of gliotic retina with Bruch's membrane.

Ocular neoplasia

Numerous types of tumour affect the eye and orbit. This subsection will cover the following, with a particular emphasis on uveal melanoma and retinoblastoma:

- papillomas of the lids and conjunctiva
- adenomas of the lids
- basal cell carcinoma of the lids
- squamous cell carcinomas of the lids
- sebaceous gland carcinomas of the lids
- pleomorphic adenoma of the lacrimal gland
- adenoid cystic carcinoma of the lacrimal gland (see p. 141)
- teratomas of the orbit
- melanomas of the conjunctiva, uvea, iris, ciliary body, and choroid
- neural tumours, including retinoblastoma
- myomas and myosarcomas
- lymphomas of the ocular adnexa
- metastatic tumours.

Papillomas of the lids and conjunctiva

These are benign epithelial cell tumours. They can affect the eyelid and conjunctiva.

In the eyelid the most common tumours are basal cell and squamous cell papillomas.

In the conjunctiva they may have a pedunculated or sessile appearance. Papillomas for the most part are associated with the human papillomavirus. They can be associated with the poxvirus (molluscum contagiosum), which causes a benign squamous proliferation.

Adenomas of the lids

These are benign tumours from a gland. In the eyelid and conjunctiva these may be derived from sweat glands, pilosebaceous hair follicles, or sebaceous glands. They will be subclassified according to the characteristics of their differentiation, which may be towards acinar or ductular structures. For example, sebaceous adenomas are proliferations of lipid-laden sebaceous cells and most commonly occur as a yellow mass at the caruncle.

Basal cell carcinomas of the lids

This is the most common form of malignant tumour seen by ophthalmologists. It represents 90% of all eyelid tumours. Like basal cell carcinoma elsewhere it is more common in Caucasians over the age of 50 years and is directly associated with UV exposure. The classic appearance of a basal cell carcinoma is a nodular lesion, which can develop a central ulcer with a rolled edge. Histological subtypes are listed in Table 4.4. Basal cell carcinoma is locally aggressive and requires wide local excision to prevent recurrence or extension of the tumour into the orbit.

Squamous cell carcinomas of the lids

These are malignant epithelial cell tumours. They represent between 1 and 5% of all eyelid tumours. They are associated with UV exposure and immunosuppression. It is also important

Table 4.4 Histological subtypes of basal cell carcinoma

Histological subtype	Characteristics
Nodular	Well-circumscribed islands of proliferating basal cells Many mitotic figures At the edge of the tumour the cells are arranged as a palisade As it is well circumscribed surgical excision is usually successful
Superficial	This is less common than the nodular type and presents as a more scaly plaque Small nests of tumour cells bud from the undersurface of the epidermis as far as the superficial dermis Surgical excision is difficult as there are gaps between the nests of cells
Infiltrative	This is more aggressive Tumour cells grow in small strands and are embedded in a fibrous stoma There is a poor border, making surgical excision difficult
Micronodular	Also more aggressive Tumour forms small nodular aggregates of basaloid cells Excision may also be difficult as some nodules may not be visualized clinically

to remember that they can arise from conjunctiva and cornea. In the lids they present as a more rapidly growing nodular ulcer or can appear papillomatous with an overlying keratinous horn. Lymphatic spread can occur to preauricular (upper lids) and submandibular lymph nodes (lower lids).

Histologically squamous cell carcinomas are classified, like all tumours, according to their degree of differentiation. Well-differentiated tumours have a glassy, pink cytoplasm and intercellular bridges with keratin pearls. Poorly differentiated tumours lose these characteristics. The spindle cell morphology is sometimes seen and this is more aggressive.

Sebaceous gland carcinomas of the lids

These account for 1–5% of all eyelid tumours. They are most likely to be seen in older women. They most commonly originate from the meibomian glands but can also arise from the gland of Zeiss or other sebaceous glands of the lids.

Clinically they may look very much like a basal cell or squamous cell carcinoma, or can appear similar to chalazions or a chronic blepharoconjunctivitis. Remember the patient with chronic unilateral blepharitis! Prognosis tends to be poor due to the diffuse nature of these tumours. Early diagnosis obviously improves this.

Histological subtypes include:

- nodular: lobules of tumour cells with foamy or vacuolated cytoplasm
- diffuse: individual tumour cells spreading within the surface epithelium (pagetoid) and adnexal structures.

These are then graded according to their degree of differentiation.

Pleomorphic adenoma of the lacrimal gland

This is a benign mixed tumour and is the most common epithelial cell tumour of the lacrimal gland. It most commonly occurs in late to middle age. It is slow growing and pseudoencapsulated.

Histologically it consists of epithelial and mesenchymal elements, including myxoid tissue, cartilage fat, and sometimes even bone. These must be excised adequately as they can undergo malignant change to produce a pleomorphic carcinoma.

Adenoid cystic carcinoma of the lacrimal gland

This is also a common form of epithelial neoplasm of the lacrimal gland. It is normally seen in middle-aged or older patients but can be seen in younger patients. The tumour is more rapidly growing and proptosis, parasthesia, pain, and diplopia can occur due to local invasion of the orbit.

Histologically the most common form is a cribiform or Swiss cheese appearance. It is aggressive, requiring surgery with radiotherapy or chemotherapy.

Teratomas of the orbit

Orbital teratomas are rare and occur in neonates. The majority are benign. They are derived from totipotent germ cells and can occur at any site in the mid-line where germ cells have stopped on their migration to the gonads. They present as proptosis.

Histological examination will reveal tissue derived from the three embryonic germ cell layers such as respiratory or gastrointestinal epithelium, stroma containing fat, cartilage and bone, and neuroectodermal tissues.

Conjunctival melanoma

This malignant tumour may arise from primary acquired melanosis, a pre-existing naevus, or *de novo*.

Clinically primary acquired melanosis appears as unilateral or bilateral diffuse flat areas of conjunctival pigmentation in middle-aged to older patients. Primary acquired melanosis may be with (premalignant) or without (not premalignant) atypia.

Conjunctival melanoma presents as a raised, pigmented, or fleshy conjunctival lesion. It can metastasize to regional lymph nodes, the brain, and other organs. A poor prognosis is related to tumours thicker than 5 mm and is located in the fornix. Complete excision is required with or without topical chemotherapy such as mitomycin C.

Uveal melanomas

Malignant melanoma of the uveal tract can arise from melanocytes in the iris, ciliary body, and choroid. Choroidal tumours make up 80% of these. Tumours are usually unilateral and grow as pigmented or non-pigmented lesions. Metastatic spread is usually to the liver and occurs within 2–3 years.

Iris melanoma

These are usually slow-growing nodular tumours. Iris melanomas (Fig. 4.10) spread diffusely on the iris surface and around the chamber angle, which can result in secondary glaucoma as a result of infiltration of the trabecular meshwork. Histological examination reveals small, spindle-shaped cells with surface or stromal invasion.

Ciliary body and choroidal melanomas

Ciliary body and choroidal melanomas can grow in a variety of forms (Fig. 4.11):

- ovoid
- nodular

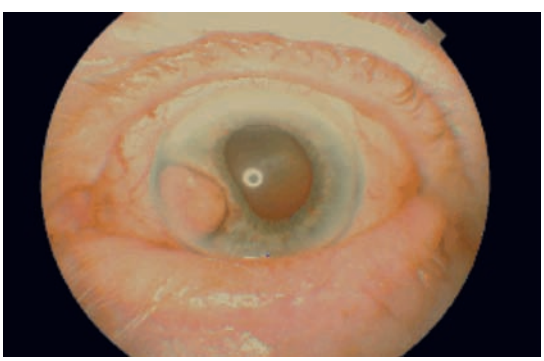


Fig. 4.10 Amelanotic iris melanoma.

Reproduced from Venki Sundaram, Allon Barsam, Amar Alwitry, and Peng Khaw, *Training in Ophthalmology*, 2009, Figure 1.81, p. 59, with permission from Oxford University Press.

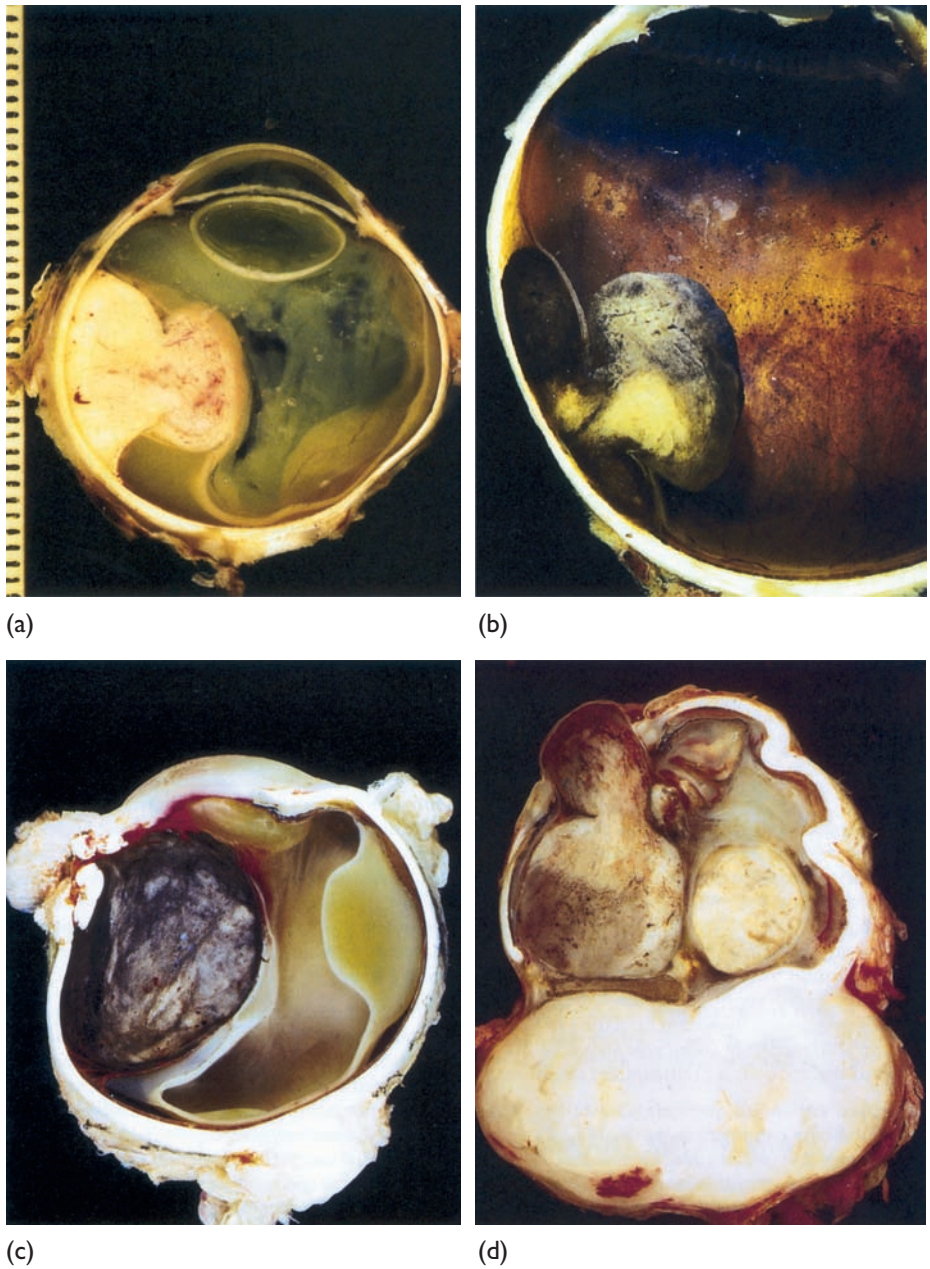


Fig. 4.11 Macroscopic appearances of uveal melanoma: (a) amelanotic and mushroom shape, (b) partially pigmented perforating retina, (c) ovoid black melanoma that has leaked fluid into the subretinal space, causing exudative retinal detachment, and (d) advanced melanoma perforating the anterior and posterior sclera.

- mushroom shape—due to tumour spread in the subretinal space after breaching Bruch's membrane.

Tumours can be amelanotic, light grey/brown, or heavily pigmented. Large tumours may undergo spontaneous necrosis. Haematogenous spread can be via collector channels, vortex veins, or short ciliary vessels.

Tumours are classified histologically according to their cell type (Fig. 4.12). They may be:

- spindle
- epitheloid
- mixed.

The majority of tumours are of mixed cell type. Vascular patterns are assessed using the periodic acid-Schiff (PAS) stain. There are nine patterns, including parallel, parallel with crosslinking, and closed vascular loops.

In addition to histology and vascular patterns immunohistochemistry will usually be positive for S100, HMB45, and Melan A.

Prognostic parameters are given in Table 4.5.

Treatment of uveal melanoma is enucleation with or without radiotherapy. Alternatively, proton beam therapy can be used for smaller tumours.

Neural tumours

Neurofibromas and schwannoma

These arise from within the orbit. They are normally benign. Neurofibromas are derived from endoneurium and schwannomas from the Schwann cells surrounding axons.

Histology of neurofibromas shows spindle cells with wavy nuclei and collagen. Occasional axons run through the tumour. Neurofibromas may be associated with neurofibromatosis type 1.

Histology of schwannomas shows a palisaded arrangement of spindle cells (Antoni A) and myxoid (Antoni B) areas and there are occasional axons running through the peripheral part of the tumour. Occasionally schwannomas contain melanin: making the distinction between them and spindle cell melanoma should be considered.

Optic nerve glioma

Juvenile and adult forms of optic nerve glioma can occur. Adult forms are rare and carry a poor prognosis due to extensive intracranial extension.

Fifty per cent of gliomas involve the orbital portion of the optic nerve. The intracranial or chiasmal portions can be involved. In the orbital portion the tumour can cause

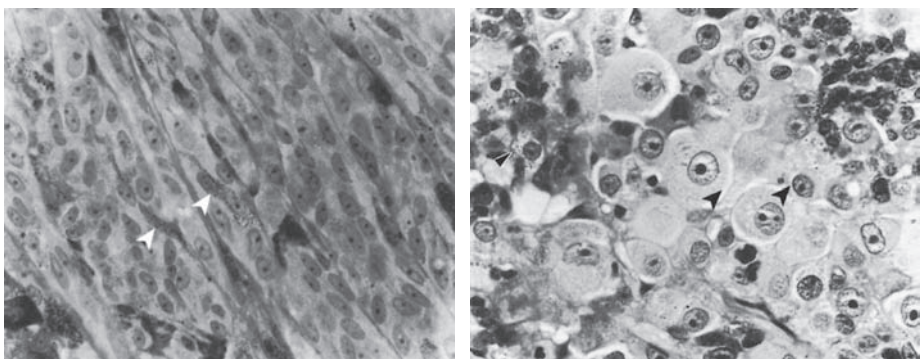


Fig. 4.12 (a) Malignant uveal melanoma of spindle B type (the cytoplasm of the cells contains melanin granules). (b) Epithelioid melanoma cells are much larger than spindle B cells and are separated from each other by prominent intercellular spaces.

Table 4.5 Prognostic parameters in uveal melanoma

Parameter	Prognosis
Age	Worse for older patients
Size	Larger tumours carry a worse prognosis
Location	Ciliary body location carries a worse prognosis compared with choroid
Cell type	Epithelioid cell type carries a poorer prognosis than those composed of only spindle cells
Vascular pattern	Closed loop vascular pattern on PAS stain carries a poorer prognosis

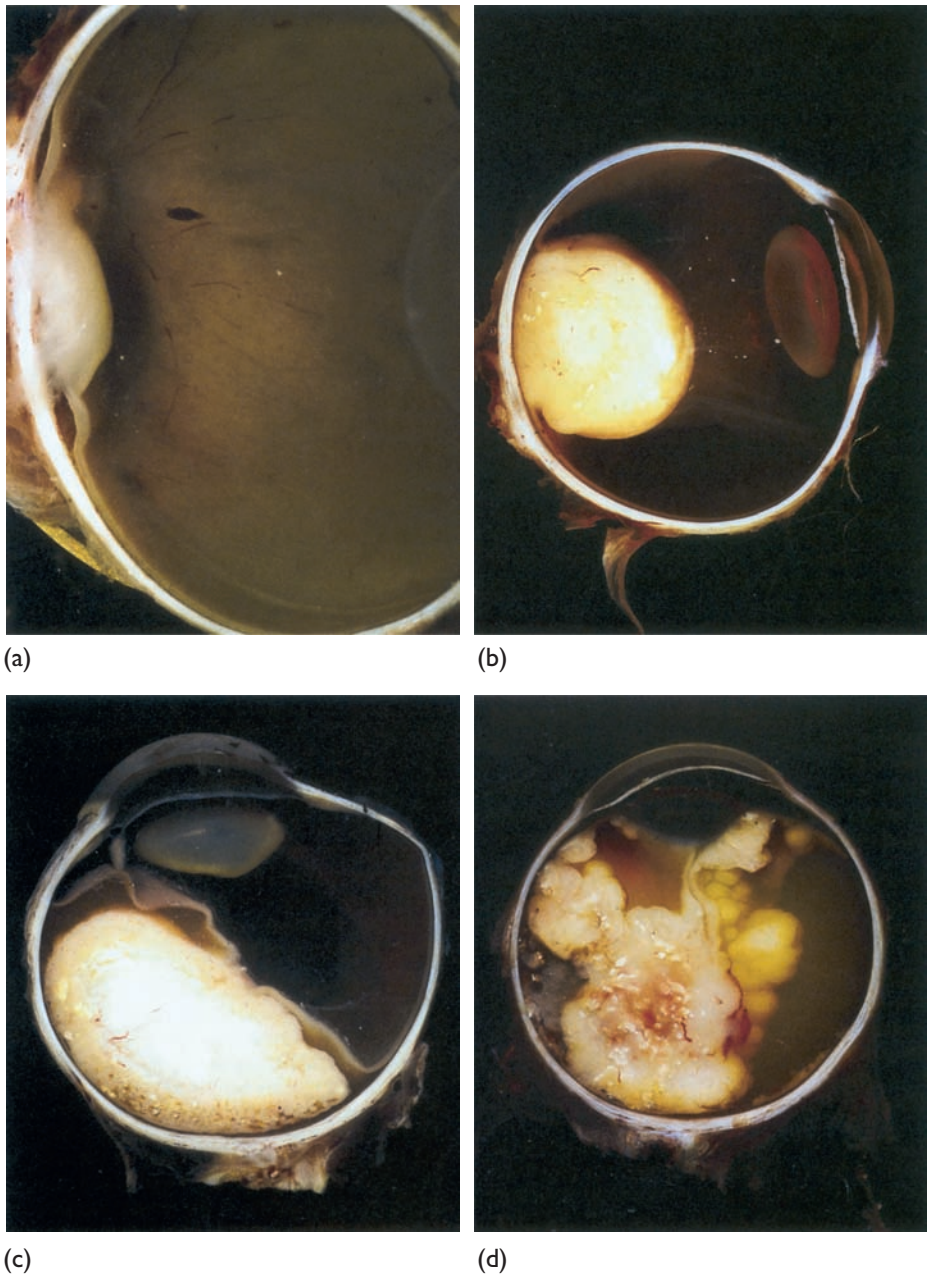


Fig. 4.13 Macroscopic appearances of retinoblastoma. (a) A small retinoblastoma overlying the optic nerve head containing a few flecks of calcium. (b) A large retinoblastoma with seedlings in the vitreous. (c) An exophytic retinoblastoma detaching the retina. (d) A large exophytic retinoblastoma with prominent calcification and funnel-shaped retinal detachment.

proptosis, optic disc swelling, and visual loss. In such patients imaging in the form of CT or MRI will show the extent of the tumour.

Histological examination shows areas of myxoid degeneration and eosinophilic masses representing a modified process of an astrocyte otherwise known as Rosenthal fibres.

If the tumour is causing visual loss it will require surgical excision, which if complete carries a good prognosis.

Meningioma of the optic nerve

Meningioma of the optic nerve may be primary or secondary due to extension of an intracranial meningioma. Unlike glioma, meningiomas tend to show slow progressive growth in adults and are more aggressive in children. Histological examination shows psammoma bodies with a transitional pattern.

Retinoblastoma

This is a malignant tumour of infancy affecting 1 in 20,000 live births. It is lethal if left untreated. The classical presentation is the finding of leukocoria, often picked up on photographs. The infant may otherwise appear well.

Differential diagnoses of leukocoria also include:

- Coats' disease
- astrocytic hamartoma
- retinopathy of prematurity
- persistent hyperplastic primary vitreous
- endophthalmitis
- toxocara or toxoplasma retinitis.

The tumour arises from embryonic retinal cells and can be unilateral or bilateral. Macroscopic appearance is of a smooth-surfaced white mass that can show endophytic growth into the vitreous or exophytic growth into the subretinal space. Yellow areas of necrosis or flecks of calcification may be seen (Fig. 4.13).

Histological examination reveals small cells with scanty cytoplasm. There is a high mitotic rate with prominent apoptosis and necrosis within the tumour, indicating high cell turnover. Differentiation may be seen in the form of:

- Homer–Wright rosettes: a multilayered circle of nuclei surrounding eosinophilic fibrillar material (Fig. 4.14)
- Flexner–Wintersteiner rosettes: a circle of cells limited internally by a continuous membrane (Fig. 4.15)
- fleurettes: primitive photoreceptor bodies in a 'fleur de lys' shape, which are usually found in irradiated tumours.

Prognostic indicators are tumour size, degree of differentiation, choroidal invasion, and optic nerve invasion.

Early diagnosis and modern treatment in the form of irradiation, chemotherapy, and enucleation give cure rates of 90%. However, it is important to remember that the abnormal gene carries a risk of pineal tumour in childhood, soft tissue and osteogenic sarcoma in early adult life, and carcinoma in later life.

Myomas and myosarcomas

Leiomyoma can arise from the smooth muscle of the iris and ciliary body. The malignant form leiomyosarcoma is rare.

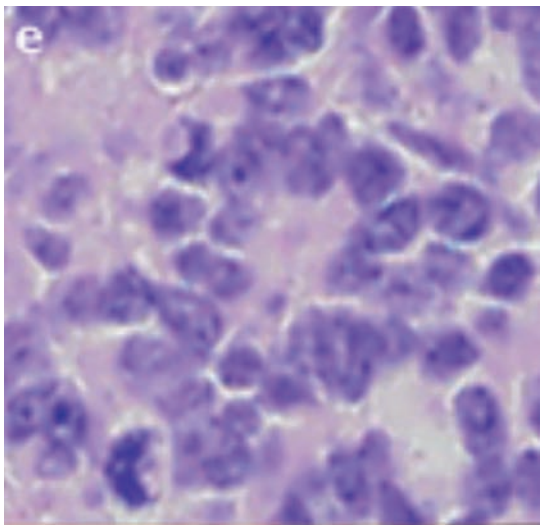


Fig. 4.14 Retinoblastoma cells forming a Homer–Wright rosette with lumen filled with eosinophilic cytoplasmic processes.

Reproduced from Venki Sundaram, Allon Barsam, Amar Alwitry, and Peng Khaw, *Training in Ophthalmology*, 2009, Figure 3.33, p. 129, with permission from Oxford University Press.

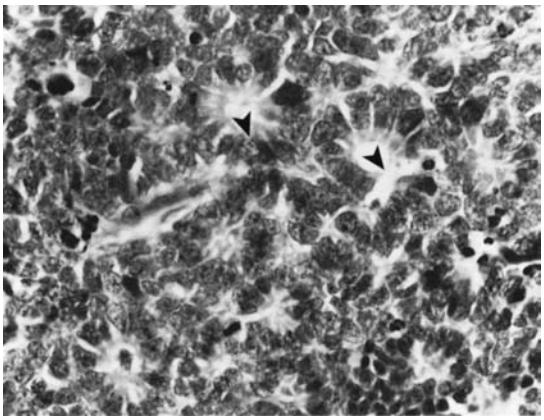


Fig. 4.15 Retinoblastoma characterized by Flexner–Wintersteiner rosettes, lined internally by a membrane similar to the outer limiting membrane of the normal retina.

This figure was published in *The Eye: Basic Sciences and Practice*, Third Edition, Forrester JV et al., Figure 9.34, Page 507, Copyright Elsevier (2008).

Rhabdomyoma and rhabdomyosarcoma can arise from striated muscle in the eyelid and orbit, although rhabdomyoma is extremely rare. Rhabdomyosarcoma is the most common orbital malignancy in childhood. It usually arises before the age of 20 and causes a proptosis and squint. Macroscopic examination reveals tan-coloured fleshy tissue. Histopathological examination reveals one of three subtypes:

- embryonal (most common)
- alveolar
- pleomorphic (rare, usually adults).

Immunohistochemisty for MyoD1, a muscle regulatory gene, can help with the diagnosis. Treatment is with a combination of chemotherapy and radiotherapy.

Lymphomas of the ocular adnexa

Lymphomas can originate from the conjunctiva, eyelids, lacrimal gland, and orbit. They may be the primary manifestation of disseminated disease and it is therefore essential to run full haematological investigations of these patients. Conjunctival lesions are associated with a lower incidence of systemic disease compared to the orbit (35%), lacrimal gland (40%), or eyelid (67%).

The most common lymphoproliferative lesions include:

- benign lymphoid hyperplasia
- extranodal marginal zone lymphoma (EMZL)—the most common type of ocular lymphoma this is a low-grade B-cell lymphoma derived from mucosal associated lymphoid tissue that may rarely transform to a high-grade lymphoma
- follicular lymphoma is usually part of systemic disease
- diffuse large B-cell lymphoma—these tend to be aggressive and can be associated with systemic disease
- primary intraocular lymphoma, which involves the retina, subretinal space, vitreous, and optic nerve—these are rare and are normally diffuse large B-cell lymphomas that are associated with primary CNS lymphoma in older people.

Metastatic tumours

In adults these usually affect the uveal tract. Most common metastatic tumours are carcinomas arising from the breast, prostate, lung, or gastrointestinal tract. In children orbital involvement is more common, for example neuroblastoma, Ewing sarcoma, Wilms' tumour, and rhabdomyosarcoma.

CHAPTER 5

Microbiology

Bacterial characteristics, virulence, and pathogenicity

Bacterial characteristics

Bacteria are a group of unicellular prokaryotes that contain DNA and RNA. The nuclear DNA is not enclosed in a membrane but lies free in the cytoplasm, which is devoid of other organelles. Bacteria reproduce by binary fission.

The basic structure of bacteria (Fig. 5.1) is as follows:

- The majority of bacteria have a cell wall. All walls contain a mucocomplex of muramic acid and are porous to all but very large molecules.
- The cell membrane is an important osmotic barrier and the site of a number of important enzymes.
- Plasmids are fragments of DNA that are found within bacteria and are thought to be involved in antibiotic resistance.
- Flagella allow for motility.
- Pili allow for conjugation to occur (transfer of DNA from one bacterium to another).

Bacteria have a diameter of roughly 1 μm and under the light microscope they are categorized morphologically as cocci (round) or bacilli (cylindrical). Further subdivisions into fusiform (tapered at both ends), filamentous (long threads), and vibrios (spiral) bacilli are also made.

Bacteria can be distinguished based on their metabolism and growth. Autotrophs are able to utilize simple inorganic compounds; heterotrophs cannot synthesize all their organic requirements. All bacteria need carbon dioxide to initiate growth. Oxidation in the form of respiration and fermentation provide energy. Aerobic bacteria obtain most of their energy by respiration. Most of these use fermentation as well and are therefore facultative anaerobes. Strict anaerobes rely entirely on fermentation for energy.

Bacterial virulence and pathogenicity

The human body has a large population of symbiotic commensal bacteria. However, all bacteria have the ability to become pathogenic or cause disease. The term 'virulence' describes the degree of pathogenicity of the individual organism.

Pathogenicity is dependent on several factors, including:

- the ability to withstand environmental stress and survive outside the host (for example spore formation: bacterial spores contain an outer coat, a cortex, and a core with a nucleus; when conditions are favourable the spore germinates; otherwise it can survive extreme changes in temperature and remain dormant for many years)
- the ability to mutate
- the ability to infect and incapacitate the host.

Tissue invasion can occur due to a direct cytotoxic action or via release of toxins.

Exotoxins

These are normally produced by Gram-positive bacteria. They are extremely potent proteins ranging from around 50 to 150,000 kDa which are released from bacteria, producing effects at sites distant to the primary site of infection, e.g. Botulinum toxin from *Clostridium botulinum*. Exotoxins are easily destroyed by heat. Exotoxins are immunogenic so can also be modified and used in vaccinations.

Endotoxins

Endotoxins are normally released by dead or lysed Gram-negative bacteria. They are lipopolysaccharides of 100–900 kDa derived from the cell wall and are required in large quantities to produce an effect (Table 5.1). They are heat stable.

Some bacteria produce enzymatic substances that aid their spread through the host (Table 5.2).

Pathogenicity is also dependent on the ability to overcome host defence mechanisms.

Bacteria have evolved to avoid complement-mediated killing mechanisms. Examples of these include:

- outer capsules that do not activate complement
- producing enzymes that can degrade complement components
- secreting proteins that bind complement and prevent complement being deposited on the bacteria directly.

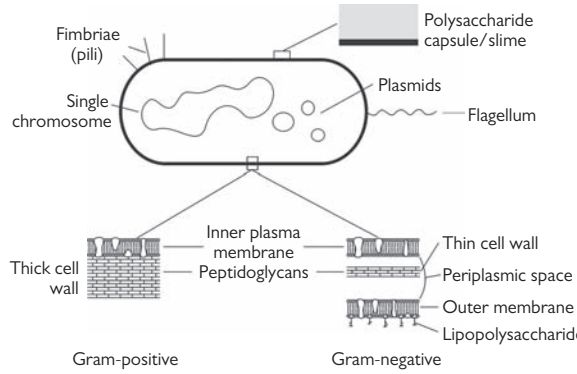


Fig. 5.1 Bacterial structure.

Reproduced from R. Wilkins et al., *Oxford Handbook of Medical Sciences, The Essential Guide to the Sciences that Underpin Medicine*, Figure 12.1, Page 799, 2011, with permission from Oxford University Press.

Table 5.1 The systemic and immunological effects of endotoxin

Fever
Tolerance—repeated exposure to the same endotoxin results in antibody production and active removal of antibody/endotoxin complexes with cessation of a febrile response
Lethal shock—can occur after large doses of toxin
Schwartzman phenomenon—intravascular coagulation due to endothelial cell damage and platelet aggregation; tends to affect the lungs and kidneys
Activation of complement—can activate the alternative complement pathway

The concept of synergy

In some circumstances combinations of microorganisms enhance the pathogenicity of each other. This is known as pathogenic synergy. For example, following a viral infection of an epithelial surface such as in herpes simplex keratitis, the commensal bacterial population can proliferate, leading to a bacterial keratitis.

Table 5.2 Enzymatic substances that enhance bacterial pathogenicity

Enzyme	Mechanism of action
Collagenase	Bacilli use this to disrupt collagen in connective tissue
Coagulase	Helps deposition of fibrin and coagulates plasma Fibrin coats the bacteria and therefore protects against complement
Hyaluronidase	Hydrolyses hyaluronate in the extracellular matrix of connective tissue and hence facilitates the spread of the organism
Streptokinase	Activates fibrinolysin and converts plasminogen to plasmin, which dissolves fibrin clots and aids spread through tissues Haemolytic streptococci produce this
Leukocidins	Group A haemolytic streptococci produce enzymes that lyse red blood cells and tissue cells, e.g. streptolysin O

Gram staining and classification of bacteria

Bacteria can be classified according to staining, morphology, and their culture requirements.

Gram staining

Gram staining is the most widely used stain and involves:

- staining with crystal violet (blue–black)
- staining with iodine
- decolourizing with acetone
- counterstaining with carbol-fuchsin (red).

A Gram-positive organism indicates the presence of a polysaccharide cell wall capsule that resists decolouration with acetone and hence stains crystal violet blue–black.

A Gram-negative organism has a three-layer cell wall that is susceptible to leaching of the stain with acetone (due to

HELPFUL HINT

Media used in bacterial and fungal culture

Solid media	Fluid media
Blood agar	Robertson's meat broth
Nutrient agar containing 5–10% horse blood	Nutrient broth with minced meat: meat extract enriches media because of the protein degradation products, carbohydrates, inorganic salts, and growth factors
Chocolate agar	Brain–heart infusion broth
Heated blood agar	Thioglycate media
Lowenstein–Jensen medium	
Glycerol, malachite green, and whole egg	
Sabouraud's agar	
Fungal growth plate containing glucose, peptone, and agar, enriched with yeast extract and chloramphenicol	
Theyer–Martin medium	
Selective, chemically enriched chocolate agar	
Non-nutrient *Escherichia coli*-enriched agar	
Acanthamoeba identification in bacterial keratitis	

their lipid content) and can be counterstained with carbol fuchsin and hence stain red.

Ziehl–Neelsen staining

Ziehl–Neelsen stain identifies acid-fast and alcohol-fast bacilli that resist decolouration with acid and then alcohol, e.g. *Mycobacterium* and *Nocardia*.

This method involves:

- staining with carbol-fuchsin
- gentle heating

Table 5.3 Microbiological investigation of bacterial keratitis

Investigation	Organism
Gram stain	Identify Gram-positive or negative bacteria
Modified Ziehl-Neelsen	*Nocardia*
Full Ziehl-Neelsen	*Mycobacterium*
Acridine orange PAS Immunofluorescence	Fungi
Blood or chocolate agar	Bacteria
Sabouraud's agar	Fungi
Meat broth	Anaerobes
Lowenstein–Jensen	*Mycobacterium*
Non-nutrient *E. coli* seeded agar	*Acanthamoeba*

With permission from Louise Bye.

- decolourization with acid and/or alcohol
- counterstaining with malachite green (modified Ziehl–Neelsen) or methylene blue (full Ziehl–Neelsen).

Other staining methods

Several other staining methods help in the differentiation of bacteria on microscopy:

- PAS stain and Gomori's methenamine silver stain are selective for fungal elements, such as hyphae.
- Acridine orange can detect fungi and bacteria such as *Nocardia*.
- Immunofluorescent methods detect *Chlamydia* and the protozoan *Acanthamoeba*.

Culture requirements

Bacteria vary in their growth requirements, including nutrient culture, atmospheric conditions, and temperatures.

Table 5.3 shows a common protocol for the investigation of microbial keratitis.

Commensal eye flora

The eye is sterile *in utero* and only acquires normal flora during birth. *Staphylococcus epidermidis* is found on 40–45% of eyes and *Staphylococcus aureus* on 25%. The diphtheroids are present in 25–40% of eyes. *Streptococcus viridans* and *Streptococcus pneumonia* are present in 2–3% of eyes. The most common anaerobe is *Propionibacterium acnes*, seen in chronic endophthalmitis and blepharitis. Protozoan commensals include *Demodex follicularum*, which is found on the lashes of most people over the age of 70 years. Up to 100 fungi are found on the lashes and lid margins. There are no viral commensals of the eye.

Important ocular bacterial pathogens

Gram-positive cocci

Staphylococci

Staphylococci are Gram-positive cocci (Table 5.4) that grow in clusters. They are aerobic, facultatively anaerobic, and produce catalase.

It is important to remember that the pathogenic *Staphylococcus aureus* is coagulase positive and the non-pathogenic *Staphylococcus epidermis* is coagulase negative.

Staphylococcus aureus enhances its pathogenicity by use of lysozymes, coagulase, and hyaluronidase. It adheres well to cell surfaces, has an opsonization-resistant capsule and produces a protein A that decreases phagocytosis and inhibits complement. It can also produce an epidermolytic toxin and an enterotoxin that causes food poisoning.

Meticillin-resistant *Staphylococcus aureus* (MRSA) is coagulase positive. MRSA can cause severe infection due to its resistance to conventional antibiotics. MRSA is often found in the anterior nares of asymptomatic carriers.

Streptococci

Streptococci are Gram-positive cocci that are catalase negative. They are uncommon ocular pathogens.

Streptococcus pyogenes accounts for 90% of streptococcal infections. It is a β-haemolytic streptococcus Lancefield group A. It produces an erythrogenic toxin responsible for the generalized erythematous rash seen in scarlet fever.

Streptococcus pneumonia is an α-haemolytic diplococcus that may cause meningitis or pneumonia.

Gram-positive rods

Bacillus

Bacilli are large, aerobic, and spore forming. Spores are visible as colourless refractile bodies. They consist of a central cortex surrounded by a layered outer coat made from laminated keratin. This is surrounded by a loose endospore. Spores are not a means of reproduction and are not metabolically active. Spores allow the organism to survive long periods and are resistant to heat, radiation, desiccation, and chemicals.

Examples of bacilli include:

- *Bacillus anthracis*—an encapsulated organism that produces exotoxin and causes anthrax
- *Bacillus cereus*—this produces a toxin that can lead to food poisoning (beware of cold/lukewarm rice)
- *Bacillus brevis*—produces the antibiotic bacitracin.

Table 5.4 Classification of Gram-positive bacteria

Cocci	Aerobes	Clusters, e.g. *Staphylococcus* Chains, e.g. *Streptococcus*
	Anaerobes	
Bacilli	Aerobes	Sporing, e.g. *Bacillus* Non-sporing, e.g. *Propionibacterium*
	Anaerobes	Sporing, e.g. *Clostridium* Non-sporing, e.g. *Actinomycetes*

With permission from Louise Bye.

Clostridium

Clostridia are anaerobic, spore-forming bacteria. They live commonly in soil, water, and decaying vegetation but can also be found in the human gut. Some produce powerful exotoxins.

Clostridium tetani causes tetanus. It produces large spherical terminal spores which have a drumstick appearance. Tetanus toxin is an exotoxin that targets the presynaptic terminals of inhibitory interneurons and hence produces a tonic spasm of the voluntary muscles. It is sensitive to penicillin.

C. perfringens is the most common cause of gas gangrene. It damages tissue with its powerful exotoxins and can occasionally lead to food poisoning via production of enterotoxin. It is also sensitive to penicillin.

The Nagler reaction is an identification technique that distinguishes different strains of *C. perfringens*.

Some strains of *C. perfringens* can lead to conjunctivitis, necrotizing keratitis, and a nasty suppurative panophthalmitis with retinal necrosis.

Corynebacteria

These are aerobic, non-spore-forming organisms that live as commensals on the skin and mucous membranes.

Propionibacterium

These are non-spore-forming.

Propionibacterium acnes is an anaerobic organism that forms part of the normal commensal of the eye. It resides on the eyelids and within the meibomian glands, and is a recognized cause of chronic blepharitis and low-grade endophthalmitis following cataract surgery.

Gram-positive filaments

Actinomycetes

These bacteria grow in the form of a mycelial network similar to filamentous fungi. Their hyphae are small (less than 1 μm in diameter).

Actinomyces israelii is a non-spore-forming anaerobic bacterium producing filaments, diphtheroid, and coccoid forms. This is a common cause of lacrimal canaliculitis and dacrocystitis.

Nocardia are branched filamentous aerobic actinomycetes. They are related to mycobacteria as they are weakly acid fast.

Nocardia asteroides is the most frequent cause of nocardial keratitis and endophthalmitis.

Gram-negative cocci

Neisseriae

Neisseria meningitides (meningococcus) and *Neisseria gonorrhoea* (gonococcus) are aerobic Gram-negative cocci (Table 5.5) that require special media such as heated blood agar with high carbon dioxide concentrations to grow. They possess endotoxin activity and IgA proteases, which allow cleavage of immunoglobulin. Gonococcus can be responsible for ophthalmia neonatorum (ocular infection within 28 days of birth). Growth of gonococcus is optimal on enriched chocolate media such as Thayer–Martin medium. Gonococcus normally infects the conjunctiva following contamination during vaginal delivery and usually commences 2–3 days post partum. Rapid diagnosis is essential to prevent blindness from corneal ulceration. Cephalosporin or penicillin is used to treat parent and child.

Gram-negative rods

Pseudomonas

Pseudomonas aeruginosa is an aerobic, non-motile, virulent ocular pathogen. It produces a water-soluble green pigment known as pyocin. *Pseudomonas* is dependent on iron for growth. It does not penetrate healthy corneal epithelium well, but the action of its numerous proteases allows it to pass through traumatized epithelium easily. This can lead to keratitis.

It produces toxin A, which acts to breakdown protein glycol matrices.

Risk factors for pseudomonal infection include:

- corneal trauma
- thermal burns
- vitamin A deficiency
- immune suppression.

Haemophilus

This is a small, non-motile, non-spore-forming aerobic bacterium. It requires haematin (X factor) and nicotinamide adenine dinucleotide (V factor), which is produced by other bacteria, to grow.

Haemophilus influenzae is cultured on chocolate or blood agar (low CO_2 concentration) and exhibits 'satellitism' around colonies of staphylococci grown in chocolate agar or brain–heart infusion agar. The presence of a capsule with a polysaccharide coat seen in serotypes A–F increases its virulence.

H. influenzae type b is a common cause of upper respiratory tract infections. It can lead to sinusitis with direct spread through the thin orbital walls, leading to orbital cellulitis. Haematogenous spread to the globe can lead to endophthalmitis, e.g. post intraocular surgery.

H. influenzae type b is a frequent cause of meningitis and epiglottitis. Most strains are sensitive to third-generation cephalosporins such as cefotaxime and to chloramphenicol.

Non-encapsulated *Haemophilus* include:

- *H. egyptius* (Koch–Weeks bacillus)—leads to conjunctivitis endemic in North Africa and the Gulf States of the USA
- *H. ducreyi*—cause of chancroid and Parinaud's ocular glandular syndrome.

Moraxella

Moraxella is a Gram-negative diplobacillus, similar to *Haemophilus*, which can cause purulent conjunctival infections and corneal ulceration.

Enterobacteria

These are small aerobic and facultative anaerobic bacteria (Table 5.6). They are part of the normal commensal flora of the gastrointestinal tract. For the most part they are opportunistic pathogens.

E. coli is the most common cause of urinary tract infection (UTI). *E. coli* are motile with polar flagella and grow well on MacConkey agar.

Serratia marcescens is an enteric organism that can contaminate contact lens solutions. It can cause infective keratitis and endophthalmitis.

Spirochaetes

Spirochaetes have a helical structure with flagellate structures allowing for spiral motility.

Table 5.5 Classification of Gram-negative bacteria

Cocci	Aerobes, e.g. *Neisseriae* Anaerobes
Bacilli	Aerobes, e.g. *Haemophilus* *Pseudomonas* Anaerobes, e.g. *Campylobacter*
Spirochaetes	*Treponema, Leptospira, Borrelia*

With permission from Louise Bye.

Table 5.6 Classification and pathogenicitiy of enterobacteria

Enterobacteria	Lactose fermenting	Usually motile, e.g. *Escheria* Non-motile, e.g. *Klebsiella*
	Non-lactose fermenting	Urease positive, e.g. *Proteus* Urease negative, e.g. *Salmonella*

With permission from Louise Bye.

Borrelia

Borrelia are large, slender organisms visualized by dark ground microscopy.

Borrelia burgdorferi is transmitted through the intermediate host of the tick *Ixodes ricinus* (a deer tick endemic in deer-populated areas such as the New Forest). The tick has three life cycle stages: larva, nymph, and adult. The primary host for *Borrelia burgdorferi* is the rodent that is required at the nymph stage of the tick's life cycle. Deer serve as the mating ground for the adult tick and provide a blood meal for egg production.

Humans acquire *Borrelia burgdorferi* from an infected nymph. Nymphs tend to feed in late spring and early summer. *Borrelia burgdorferi* causes Lyme disease (arthritis, conjunctivitis, and encephalomyelitis). Lyme borreliosis is detected by immunofluorescent assay or enzyme-linked immunosorbent assay (ELISA) to measure specific IgM and IgG antibodies.

Treponemes

Treponemes consist of an outer envelope, a peptidoglycan cell wall, and a cytoplasmic membrane.

Treponema pallidum is a motile, microaerophilic helical bacterium that is temperature sensitive and hence does not survive for long outside of the body and cannot grow on artificial media. It is the cause of syphilis.

Treponemes are visible on dark ground microscopy or by silver staining (Levaditi silver method). Serology is used as the gold standard for diagnosis.

The Venereal Diseases Research Laboratory (VDRL) or reagin test and the complement fixation tests return to normal after treatment. However, the VDRL test lacks specificity and hence false positives can occur. Specific antibody tests, including fluorescent treponemal antibody absorption (FTA-Abs), are now used instead. The FTA-Abs becomes positive first and remains positive after treatment.

Acid-fast bacilli

Mycobacteria are acid and alcohol fast. They are aerobic non-sporing rods with waxy cell walls that prevent them taking up the Gram stain. They stain with hot strong carbol fuchsin or the Ziehl–Neelson, stain which is retained despite attempts to remove it with acids and alcohols.

Mycobacteria will only grow on Lowenstein–Jensen medium which contains egg and have a very slow generation time of 12–24 hours, which means cultures can take up to 8 weeks to grow.

Only a few species of mycobacteria are pathogenic to humans:

- *Mycobacterium tuberculosis*—the characteristic lesion caused by this organism is a caseating granuloma made up of macrophages and macrophage-derived cells. Prophylaxis for tuberculosis is 75% successful with the live attenuated bovine strain known as bacilli Calmette–Guerin (BCG). Tuberculosis can cause a panuveitis by haematogenous spread to the choroid.
- *Mycobacterium leprae*—this is the cause of leprosy.
- *Mycobacterium avium* and *Mycobacterium fortuitium*—these are more commonly seen in immunosuppression, e.g. AIDS, and can lead to corneal ulceration and endogenous endophthalmitis.

HELPFUL HINT

Common ocular bacterial pathogens

Organism	Infection
Gram-positive cocci	
Staphylococcus aureus	Blepharoconjunctivitis
Staphylococcus epidermidis	Endogenous endophthalmitis
Streptococcus pneumonia	Corneal abscess and exogenous endophthalmitis
Streptococcus pyogenes	Endogenous endophthalmitis
	Corneal abscess/conjunctivitis
	Endogenous endophthalmitis
Gram-positive bacilli	
Propionibacterium	Postoperative endophthalmitis
	Chronic meibomianitis
Gram-positive filaments	
Actinomyces	Canaliculitis, dacrocystitis
Nocardia	Corneal abscess/endophthalmitis
Gram-negative bacteria	
Neisseria	Conjunctivitis/endophthalmitis
Moraxella	Conjunctivitis
Pseudomonas	Corneal abscess/endophthalmitis
Haemophilus	Conjunctivitis/orbital cellulitis

Mollicutes

Mollicutes are unique in that they lack a cell wall. *Mycoplasma pneumoniae*, *Mycoplasma hominis*, and *Ureaplasma urealytica* are examples of mollicutes. *Ureaplasma urealytica* has been associated with the aetiology of Reiter's syndrome.

Viral characteristics

Viruses are acellular and hence require the cellular material from other organisms in order to replicate. They are therefore described as obligate intracellular parasites.

A virus particle or virion consists of the viral genome (DNA or RNA) enclosed in a protein shell known as a capsid. Capsids vary in shape: they can be helical, cubid, or

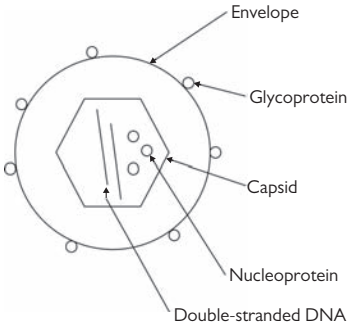


Fig. 5.2 Basic viral structure.

Reproduced from R. Wilkins et al., *Oxford Handbook of Medical Sciences, The Essential Guide to the Sciences that Underpin Medicine*, Figure 12.2, Page 808, 2011, with permission from Oxford University Press.

icosahedral. The term 'capsomere' refers to the morphological unit of viral genome and capsid together seen on electron microscopy (Fig. 5.2).

The capsid can be surrounded by a lipoprotein envelope (Fig. 5.2). Viruses vary in size from 10 to 300 nm in diameter.

Viruses are classified according to their constituent nucleic acid and morphology.

Viral replication and transmission

Viruses can replicate in the host nucleus or cytoplasm. Synthesized viral particles can sometimes be seen on light microscopy. These are known as inclusion bodies and represent sites of viral synthesis or replication.

The stages of viral replication are as follows:

- Adsorption—attachment of the virus particle to a cell by random collision, electrostatic attachment, or host cell receptors.
- Penetration—fusion of the viral envelope and cell membrane.
- Capsid removal—the capsid is removed by host cell enzymes.
- Nucleic acid replication—most DNA viruses replicate in the nucleus by host cell enzymes to make messenger RNA and to replicate DNA. DNA virus will therefore produce intranuclear inclusions.

Replication of RNA viruses occurs in the cytoplasm. They demonstrate cytoplasmic inclusions. RNA viruses are unable to borrow host cell enzymes because none exist for copying RNA from RNA. Use of reverse transcriptase to make DNA from an RNA template overcomes this.

Viral transmission can be vertical, horizontal, or a combination of the two.

Vertical transmission occurs as a result of a mother passing virus to her offspring. For example, cytomegalovirus and rubella can be transmitted vertically via a transplacental route.

Horizontal transmission occurs when a virus is passed from one individual to another by direct contact, faecoorally (enteroviruses), airborne-respiratory (paramyxoviruses), or parenteral routes such as in hepatitis B or rabies.

HELPFUL HINT

Cytomegalovirus produces both cytoplasmic and nuclear inclusions. The nuclear inclusions are referred to as 'owl's eye' inclusion bodies, in which enlargement of the cytoplasm and the nucleus occurs.

The cellular response to viral infection

The cellular effect of virus infection includes cell shrinking, rounding, inclusion bodies, and the formation of giant multinucleate cells. This can result in cell death due to cytolysis or inhibition of cell metabolism.

Chronic infection can result in:

- Latency—the viral genome is integrated into the host DNA. It does not replicate until a trigger occurs, e.g. stress. A classic example of this is the herpes simplex virus.
- Persistence—the virus replicates at a very low rate.
- Transformation—the virus initiates new cell properties, which lead to teratogenic change.

The host response to viral infection and viral pathogenicity

Inflammation and tissue damage can be due to the direct effects of the virus and also the host response.

Innate immunity in the form of structural barriers (cilia in the respiratory tract, acid in the stomach) can prevent viruses invading the epithelial surface. The presence of a fever inhibits viral replication.

The humoral response includes production of antibody, complement, and interferon, and the cellular production of T and B lymphocytes, macrophages, and polymorphonuclear leukocytes.

Viruses enhance their pathogenicity through evasion of immune mechanisms. They have a huge capacity to mutate and even minor changes in their antigenic surface prevent detection by immune cells. The term 'antigenic drift' refers to minor changes in antigenicity due to point mutations in the viral genome. Antigenic shift refers to a more major change such as when a virus swaps whole segments of the genome with reservoirs of different viruses.

Viruses have the ability to avoid complement. Herpes simplex type 1 produces a C3b-binding molecule that facilitates degradation of the alternative pathway molecule C3bBb convertase.

Diagnostic tests and serology

The diagnosis of viral infection can be via:

- direct viral isolation—use of electron microscopy to demonstrate virus in body fluids/tissue
- serological techniques for detection of antigen or antibody—complement fixation measures the reaction between viral antigen and a specific antibody, and gives a direct measure of the consumption of complement added. In a positive test the erythrocytes are not lysed by complement as it has already been bound by the antigen/antibody reaction.
- ELISA
- radioimmunoassay (RIA)
- reverse passive haemagglutination, such as in detection of hep B surface antigen
- organism-specific humoral responses can be detected in ocular fluids—Goldmann and Witmer reported a method to calculate whether antimicrobial antibodies are being produced within tissues and hence demonstrating local infection
- PCR—this can be used to detect genetic material via analysis of intraocular fluid or tissue, for example a vitreous biopsy in possible acute retinal necrosis or chronic endophthalmitis (see Chapter 3 for PCR methodology).

HELPFUL HINT

Goldmann and Witmer coefficient (C)

C = (antibody titre in ocular fluid/antibody titre in serum) × (total IgG in ocular fluid/total IgG in serum)

C ≥ 3, local antibody production present.

Important ocular viral pathogens

DNA viruses

Herpes viruses

These are double-stranded DNA viruses with an icosahedral capsid. An important feature of some of the herpes family, namely herpes simplex and varicella zoster, is that they possess the ability to lie dormant in neuronal sites and reactivate under certain conditions, e.g. stress or other illness. This is known as latency. Members of the herpes virus family are the causative agents of acute retinal necrosis.

The herpes virus family includes:

- herpes simplex
- varicella zoster
- cytomegalovirus
- Epstein–Barr virus (EBV).

Herpes simplex virus

Herpes simplex type 1(HSV-1) is most commonly associated with oral infection (cold sores). Normally the primary infection is in childhood and is subclinical, with the cold sore occurring as a consequence of viral reactivation. HSV-1 is relevant in ocular disease. It can give rise to conjunctivitis, keratitis, uveitis, and uveoretinitis. A culture or smear of the affected tissue shows intraepithelial vesicles and contains syncytial giant cells and eosinophilic intranuclear inclusions.

There are four stages in the life cycle of HSV-1 (Table 5.7):

1. Entry into the host and replication at this peripheral site (e.g. eyes, skin, or mucosae).
2. Spread to the axonal terminals of sensory neurons followed by retrograde intra-axonal transport to neuronal cell bodies in sensory and autonomic ganglia.

Table 5.7 The life cycle of herpes simplex virus, LATs, latency associated transcripts

Latent	Lytic
1. Entry of virus into nucleus	1. Entry of virus into nucleus
2. Reduced immediate early gene expression	2. Immediate early gene expression
3. Shut off by lytic genes	3. DNA synthesis
4. Latency	4. Late gene expression
– Viral DNA persistence	5. Assembly
– No virus production	6. Virus release
– Cell survival	7. Cell death
5. Chance of deactivation to enter lytic pathway.	

With permission from Louise Bye.

3. Latency in ganglia, e.g. trigerminal ganglion.
4. Reactivation with the production in the ganglia of infectious virus transported anterogradedly to the periphery, with further replication at the site of primary infection.

Expression of viral genes changes throughout the life cycle. A viral structural protein known as VP16 enhances transcription of these genes:

- immediate–early (regulatory genes)
- early (viral DNA replication)
- late (structural proteins of the virus)

The pathogenicity of herpes simplex is increased in immunosuppression, malignancy, and with the use of topical steroid. Reactivation itself can be triggered by these or hormonal changes, ultraviolet light, stress, and trauma.

The viral envelope is highly immunogenic and stromal disease, e.g. disciform keratitis, is due to a hypersensitivity reaction to viral antigen rather than the virus itself.

Varicella zoster virus

Varicella zoster virus (VZV) is the cause of chickenpox (collections of intraepidermal vesicles on the trunk, face, and mouth). Chickenpox is spread by the respiratory route. The virus can become latent in certain ganglia, most commonly the trigeminal, thoracic lumbar, and cervical nerve ganglia. Reactivation leads to shingles.

If reactivated in the trigeminal ganglion VZV can lead to ophthalmic shingles or herpes zoster ophthalmicus. Cell-mediated immunity maintains the virus in its latent state. Reactivation can be due to concurrent illness, immunosuppression, or radiotherapy.

Cytomegalovirus

Cytomegalovirus (CMV) infection is very common, but it is subclinical in 80% of cases. The virus is shed from the genital or urinary tracts and becomes latent in lymphocytes. Reactivation can occur during pregnancy, leading to asymptomatic infection of the foetus. If the primary infection occurs in pregnancy congenital anomalies can occur (this is more likely if primary infection is in the first trimester).

Congenital infection with CMV can cause cytomegalic inclusion disease, leading to:

- strabismus
- chorioretinitis
- microphthalmia
- childhood hepatitis
- post-transfusion mononucleosis.

Interestingly the virus is shed at birth by 1% of infants, with 10% of these having minor abnormalities, including hearing deficits.

CMV infection in the immunosuppressed can lead to:

- CMV retinitis
- transplant rejection
- CMV pneumonia.

Epstein–Barr virus

EBV is the causative agent of infectious mononucleosis or glandular fever. EBV also has the ability to transform B lymphoblasts and is associated with Burkitt's lymphoma and nasopharyngeal carcinoma.

Adenovirus

Adenovirus is a diverse group of double-stranded DNA viruses consisting of over 30 antigenic types or serotypes. The infective particle has an icosahedral capsid without an envelope. Transmission of the ocular infection is by direct contact with virus in ocular secretions. Contaminated instruments, eye drops, or the hands of health professionals can be the fomite. Adenovirus is highly prolific: a single infected cell can produce 10,000 virions per cycle of 30–36 hours.

Serotypes 1, 2, 3, 5, 7, and 14 are associated with pharyngoconjunctival fever. Serotypes 3, 7, 8, and 19 are associated with epidemic keratoconjunctivitis.

Adenovirus can produce proteins that interact with p53 and the retinoblastoma gene and hence some strains have oncogenic properties in animal models.

Adenovirus has the ability to avoid immune defences by suppressing transcription and presentation of major histocompatibility complex (MHC) Class I molecules on the infected cell surface.

Papovavirus

Human papillomavirus

Human papillomavirus (HPV) is a double-stranded DNA virus of which there are over 60 types. The DNA is arranged into a circular molecule with areas known as open reading frames. These are divided into early (E) and late (L) regions. E regions code for viral replication proteins.

HPV generally infects epithelial cells and some possess the ability to induce proliferation (benign papilloma). HPV 6 and HPV 11 are associated with benign conjunctival papillomata.

Malignant change is caused by insertion of a particular type of HPV DNA into the host genome. The E2 gene is an important regulator of HPV and disruption of this gene is found in all tumours that have HPV DNA integration.

Disruption of E2 leads to increased production of E6 and E7.

E6 forms complexes with p53 (tumour-suppressor gene) and hence promotes oncogenesis. This is seen in HPV 16 and HPV 18. They cause conjunctival dysplasia and carcinoma.

E7 inactivates the gene product of the retinoblastoma tumour-suppressor gene.

Pox virus

Pox viruses are DNA viruses that grow in the cytoplasm with very limited nuclear involvement. This is because they possess a DNA-dependent RNA polymerase, a transcript poly-A polymerase, a capping enzyme, and methylating enzymes, hence allowing them to replicate independently of the host cell.

Molluscum contagiosum

Molluscum contagiosum is the most common pox virus to cause ocular infection. It causes conjunctivitis and the appearance of pearly white or flesh-coloured lesions on the face, extremities, and trunk. It is commonly seen in children and is spread by direct contact or through contaminated fomites or water. Adult cases tend to be sexually transmitted.

Epidermal hyperplasia is due to the production of a protein related to the conserved domain of epidermal growth factor.

RNA viruses

Paramyxovirus

Measles

Measles infection is characterized by pyrexia, cough, coryza, and conjunctivitis. Complications include secondary bacterial respiratory infection, encephalitis, and rarely subacute sclerosing panencephalitis (SSPE), which can be associated with chorioretinitis and maculopathy.

Mumps

Mumps infection is characterized by fever and parotitis. In addition mumps can cause orchitis and reduced fertility, pancreatitis, and meningoencephalitis. Ocular complications include dacryoadentitis and extraocular muscle palsies.

Togavirus

Rubella

Rubella is subclinical in 80% of small children and 10% of adults. If a mother is infected in the first trimester of pregnancy congenital anomalies occur due to spread to the placenta and hence foetus. Miscarriage or stillbirth may occur. Congenital defects include cataract, microphthalmia, 'salt and pepper' retinitis, glaucoma, conductive deafness, and heart defects. For this reason a live attenuated vaccine is given to children under the age of 12 years.

Retrovirus

These RNA viruses cause leukaemia and lymphoma, and include human immunodeficiency virus (HIV).

Human T-cell lymphotrophic viruses

Human T-cell lymphotrophic viruses (HTLVs) are associated with T-cell lymphomas, including mycosis fungoides and Sézary syndrome. They are also implicated in the aetiology of progressive myelopathy and of uveitis in some ethnic groups. HTLV-1 is endemic in Japan, the West Indies, and central Africa.

HTLV-1 is transmitted vertically and horizontally through sexual contact or parenteral transmission and infects primarily $CD4^+$ lymphocytes.

HTLV-1 replication is regulated by two unique genes known as tax and rex. Tax transactivates genes for IL-2 and IL-2 receptors, hence making clonal expansion independent of IL-2 autocrine activity.

Human immunodeficiency virus

HIV-1 and HIV-2 are retroviruses possessing the ability to infect $CD4^+$ lymphocytes. They contain a single strand of RNA and reverse transcriptase. They differ in the structure of their glycoprotein envelopes.

The main genomic components of HIV include:

- structural genes—*gag* gene codes for p55, which is cleaved into p24, p18, and p15. The envelope gene codes for a glycoprotein enclosing the viral particle. This glycoprotein is cleaved to form two envelope proteins gp 4 and gp 120.
- regulatory genes—these act to stimulate viral transcription and cause proliferation of adjacent healthy cells, e.g. B cells and Kaposi's sarcoma.
- polymerase gene—codes for reverse transcriptase. This transcribes viral RNA into DNA and can be incorporated into the host genome.

Remember HIV can be killed by strong acid and alkali (pH < 1.0 and > 13) or by exposure to 10% bleach or 50% ethanol.

Immunological effects of HIV

HIV gains entry to $CD4^+$ lymphocytes via binding to CD4 antigen and CXC chemokine receptors on the cell surface. It therefore makes sense that this virus can infect other cells such as monocytes and microglial cells expressing these receptors (Table 5.8).

Viral infection of $CD4^+$ T cells leads to the formation of a multinucleate cell and cell death. The virus therefore diminishes the cell-mediated immune response predisposing to viral, protozoan, and some neoplastic conditions.

HIV is known to cause polyclonal antibody production (hypergammaglobulinaemia) and the production of autoantibodies.

The body's immune response against HIV is reduced because the latent virus is invisible to immune defences and mutates rapidly, making it very difficult for the immune system to keep up with this antigenic shift.

Diagnosis of HIV

HIV infection can manifest in several ways, from asymptomatic, acute febrile illness during seroconversion to persistent

Table 5.8 Immunological effects of HIV as a consequence of $CD4^+$ T-cell infection

Cell type	Effect
$CD4^+$ T cell	Reduced clonal expansion, reduced lymphokine production Reduced response to soluble antigen
CD8	Reduced clonal expansion, reduced cytotoxicitiy
Macrophage	Reduced chemotaxis, parasite killing
Natural killer cells	Reduced killing of tumour cells
B cells	Reduced Ig production to specific antigen Hypergammaglobulinaemia

generalized lymphadenopathy. Seroconversion to anti-HIV antibody occurs 4–12 weeks after acute infection.

HIV can be detected in bodily fluids. It is usually obtained from peripheral blood.

The tests for HIV infection include:

- HIV culture—the virus is isolated by co-culture with normal lymphocytes in the presence of IL-2. Multiplication is detected by reverse transcriptase assay or HIV antigen expression in culture.
- HIV antigen detection—the first detectable protein is actually p24 (core protein) at 2–3 weeks followed by the antibody response below.
- HIV antibody—diagnosis of HIV is based on detection of antibodies to HIV-1 by ELISA or Western blotting. Western blots demonstrate both IgG and IgM antibodies against envelope and structural proteins coded for by the gag gene (p55).
- HIV nucleic acid—PCR can be used to amplify HIV genome RNA. After PCR, ELISA estimation detects as little as 50 genome numbers per millilitre of blood. This can be used to estimate viral load. This can be used alongside $CD4^+$ T-cell count when monitoring response to treatment.

A fall in $CD4^+$ T-lymphocyte count and anti-p24 antibody titre with a rise in titre of core antigen can demonstrate a rise in viral load and can precede the onset of AIDs.

Progression to AIDS

Epidemiological data show that there are three types of AIDS transmission:

- Type 1—urban spread (USA and Europe) in homosexuals, heterosexuals, and intravenous drug users
- Type 2—African spread, mainly heterosexual
- Type 3—South-east Asia, yet to be fully defined.

AIDS is defined as an illness characterized by one or more of the listed Communicable Disease Centre (CDC) indicator diseases depending upon the status of laboratory evidence of HIV infection. In addition to this the definition of AIDS also includes all HIV-positive patients with $CD4^+$ counts of < 200 per µl.

Table 5.9 Ocular infections associated with AIDS

- CMV retinitis—treatment with ganciclovir or foscarnet is successful in up to 80% of cases; maintenance therapy is required; rate of relapse depends on the presence of positive leukocyte CMV cultures
- Cotton wool spots
- Toxoplasma chorioretinitis—treatment is with pyrimethamine and sulfadiazine or clindamycin; maintenance is usually required to prevent relapse
- Herpetic keratitis (simplex and zoster)
- Herpetic acute retinal necrosis and retinitis
- Candidal (in intravenous drug users) and cryptococcal endophthalmitis
- Tuberculous chorioretinitis
- Histoplasma retinitis
- *Pneumocystis carinii* choroidopathy

Some of the more common indicator diseases include:

- *Pneumocystis carinii* pneumonia
- cytomegalovirus retinitis
- cryptococcus
- primary lymphoma of the brain
- pulmonary tuberculosis (TB)
- invasive cervical carcinoma.

Before the advent of antiretroviral therapies, up to 25% of patients with AIDS presented with opportunistic ocular infection (Table 5.9). The most common of these is CMV retinitis.

Highly active antiretroviral therapy

Highly active antiretroviral therapy (HAART) is used in the treatment of both AIDS and HIV. HAART acts to increase $CD4^+$ T-cell counts and restores antigen-specific responses. It involves the use of a combination of agents, including:

- Reverse transcriptase inhibitors, e.g. zidovudine—zidovudine is phosphorylated in both infected and uninfected cells by thymidine kinase. This phosphorylation produces zidovudine-TP, which acts to inhibit viral reverse transcriptase and also promotes RNA chain termination prematurely.

HELPFUL HINT

Viruses in ocular disease

Virus	Infection
DNA viruses	
• Herpes virus Zoster	Keratoconjunctivitis
HSV-1 and 2	Uveitis/retinitis
Epstein–Barr	Conjunctivitis
CMV	Uveitis/retinitis
• Adenovirus Serotype 3	Pharyngeal conjunctival fever
Serotype 8	Endemic keratoconjunctivitis
Serotypes 1–11, 14–17	
• Pox virus Vaccinia	Conjunctivitis Blepharoconjunctivitis
Molluscum contagiosum	Conjunctivitis
RNA viruses	
• Orthovirus Influenza	Conjunctivitis Conjunctivitis/keratitis
• Paramyxovirus Mumps	Keratitis Conjunctivitis
Measles	Keratoconjunctivitis
• Picornavirus Coxsackie	Keratoconjunctivitis/ retinopathy
• Togavirus Rubella	
• Retrovirus HIV	

- Proteinase inhibitors, e.g. indinavir—indinavir inhibits recombinant HIV-1 and HIV-2 proteinase and therefore prevents cleavage of viral precursor proteins producing immature non-infectious particles. Indinavir is used in combination with a nucleoside analogue.
- Nucleoside analogues—chain terminators of HIV reverse transcriptase.

Ocular fungal pathogens

Fungi are eukaryotic organisms. They contain both DNA and RNA, and are able to reproduce sexually. They are dependent on exogenous sources of food and are therefore either parasitic or saprophytic.

Most fungi that cause orbital infections are normal commensals of the respiratory, gastrointestinal, and female genital tracts and can be found as part of the normal commensals of the conjunctival sac. The pathogenesis of fungi varies with the site and is affected by predisposing factors such as immunosuppression and trauma (Table 5.10).

Fungi are divided into three groups:

- Yeasts—unicellular.
- Filamentous fungi—contain branching filaments called mycelium. Mycelia absorb nutrients and produce reproductive spores.
- Dimorphic fungi (a mixture of both of the above)—microscopy and cytopathological techniques are used to distinguish the morphology of the fungi. Both yeasts and filamentous fungi are Gram-positive. Sabouraud's medium (glucose-peptone agar, pH 5.6) is used to culture fungi. Immunofluorescent techniques can be used to determine antibody titres for specific fungal infections. PCR can be used to detect fungal DNA from the vitreous in posterior intraocular inflammation.

Yeasts

Candida albicans

Candida albicans is an oval unicellular fungus that can reproduce by budding. It is a normal commensal of the gastrointestinal, genitourinary, and respiratory tracts, and is the most frequent cause of endogenous fungal endophthalmitis. Contaminated indwelling venous catheter tips are the most common source of infection in these patients.

Diabetes mellitus, malignancy, liver disease, prolonged antibiotic therapy, alcoholism, and intravenous drug use predispose to candidal infections.

Candida can be cultured on Sabouraud's glucose media or blood agar. It forms dome-shaped creamy white colonies at 24–48 hours. *Candida albicans* is a budding yeast form on Sabouraud's glucose medium. If aeration is reduced mycelium or pseudomycelium can form.

Cryptococcus neoformans

Cryptococcus neoformans is found in pigeon droppings. It is a true pathogen and causes lung infections following inhalation. If spread haematogenously it leads to meningoencephalitis and chronic endophthalmitis.

Filamentous fungi

Aspergillus fumigates

Aspergilli are found in decomposing plant debris. Their spores can lead to aspergillosis if inhaled.

Aspergillus fumigatus is an opportunistic pathogen. It can cause conjunctivitis, keratitis, endophthalmitis, and a granulomatous orbital inflammation.

Inhaled spores can germinate in the lumen of the bronchi and lead to an IgE-mediated allergic response or the production of IgG antibodies and complement activation. Aspergilloma refers to a mass of mycelia and is found in lung cavities after healed TB, bronchiectasis, or sarcoidosis.

Invasive aspergillosis is seen in immunocompromised patients.

Mucoraceae

Mucoraceae are found in soil, air, ventilation systems, and the nose and pharynx.

Iron is an important growth factor for the Mucorales. They are characterized by broad, irregular non-septate hyphae that branch at right angles. They stain best with Gomori methenamine-silver, haematoxylin and eosin, and PAS.

Inhaled spores are usually destroyed by macrophages but in an acidotic environment phagocytosis is compromised and infection can occur, leading to mucormycosis.

Mucormycosis is most commonly seen in non-ketotic diabetic ketoacidosis and in chronic illness such as metastatic neoplastic disease. It can be caused by three genera from the fungal class Zygomycetes, order Mucorales, and family Mucoraceae: *Rhizopus*, *Mucor,* and *Absidia*.

Table 5.10 Factors predisposing to fungal disease in the eye

Exogenous	Endogenous
• Local trauma—exogenous mycotic infection may follow a local corneal abrasion with vegetable matter, e.g. if a foreign body is present • Contact lens wear • Topical antibiotic and steroids	• Immunocompromised—haematogenous • Non-ketotic diabetic ketoacidosis from adjacent air sinuses • Contamination of indwelling catheters or intravenous lines

Mucormycosis may present as an orbital cellulitis and in fact leads to a cerebrorhinoorbital syndrome, which can be fatal.

Actinomyces israelii

Actinomyces israelii is a Gram-positive hyphal organism that can fragment into bacillary and coccoid forms. It is anaerobic or microaerophilic and is cultured in liquid media such as brain–heart infusion agar. Brown and Brenn Gram staining differentially stain the filaments.

Actinomyces forms white–yellow colonies known as sulphur granules that can be seen clinically.

Actinomyces can lead to canaliculitis and is sensitive to penicillin and cephalosporins.

Dimorphic fungi

The yeast form is found in infected tissues and can lead to intraocular inflammation, including:

- optic neuritis
- chorioretinitis
- panuveitis.

Histoplasma capsulatum

Histoplasma capsulatum is a soil fungus that is endemic in the Mississippi delta. Around 70% of people living in this area and the Ohio River valley have positive histoplasmin skin test reactivity.

Transmission is primarily by inhalation of mycelial fragments and/or spores with dust particles and often gives rise to a mild febrile illness, although it can be asymptomatic.

The organism reaches the choroid by haematogenous spread. Active choroiditis is not common. Presumed ocular histoplasmosis syndrome (POHS) is a multifocal choroiditis associated with the formation of subretinal neovascular membranes. It is thought to be immunologically mediated in individuals exposed to the fungus and is sensitive to amphotericin B and ketoconazole.

At 26°C POHS exists in a mycelia phase and at 37°C it is a yeast. Demonstration of this dimorphism is part of the necessary criteria for identification of *H. capsulatum* on brain–heart infusion agar.

Ocular intracellular parasites

Chlamydiae

Chlamydiae are small bacteria without a cell wall. They are unable to grow on normal media and are obligate intracellular bacteria that can only grow within a eukaryotic host cell. They grow well in cell culture or McCoy media.

Chlamydiae contain both DNA and RNA, and require host-derived ATP to survive and replicate in the cytoplasm of the host cell. They divide by binary fission.

The genus comprises three species:

- *Chlamydia trachomatis*
- *Chlamydia psittaci*
- *Chlamydia pneumonia*.

Chlamydia trachomatis

Chlamydia trachomatis is divided into serotypes by immune fluorescence tests: all are glycogen positive and have humans as their host. Different serotypes require different amino acids.

Serotypes include:

- A–C—these cause endemic trachoma and require tryptophan
- D–K—these cause adult inclusion conjunctivitis (sexually transmitted)
- L1, 2, 3—these cause lymphogranuloma venereum and require methionine.

Chlamydia trachomatis exists in two forms:

- The elemental body—this infectious form attaches to the host cell. It is enclosed in a cytoplasmic vesicle inside the host cell and is spared lysosomal degradation.
- The reticulate body—the replicative form. This forms after 6–8 hours inside the host cell. It divides for around 24 hours and then returns to the elemental form.

Diagnosis of chlamydia is based on Giemsa staining of smears, which identifies basophilic intracytoplasmic inclusion bodies. Chlamydia is isolated from swabs transported in sucrose phosphate medium and inoculated on to McCoy cells. ELISA can be used to detect inclusions within 24 hours. Antibody detection in the serum and tears can also be used and correlates well with clinical activity.

Treatment of chlamydia can be with tetracycline, erythromycin, rifampicin, azithromycin, and sulphonamides.

Trachoma control programmes use topical antibiotics to reduce the ocular reservoir. Systemic treatment is for more severe cases, usually oral doxycycline or erythromycin. A single dose of azithromycin can be used to treat chlamydial conjunctivitis.

Protozoa

Protozoa are unicellular eukaryotic parasites. Humans can act as an intermediate host in their life cycle.

Toxoplasma gondii

Toxoplasma gondii is an obligate intracellular parasite whose definitive host is the cat. Transmission can be via faecal spread of sporocysts from the cat host, ingestion of uncooked meats, or inhalation. Vertical transmission can occur if an antibody-negative mother is infected during pregnancy.

The first stage of infection is cell penetration dependent on the primitive mouthpiece or conoid of the parasite, which contains lytic enzymes. Once inside the cell

Toxoplasma gondii protects itself from lysosomal destruction using a vacuole.

Cell-mediated immunity with activated macrophages is the prime defence against toxoplasma.

The life cycle of toxoplasma is as follows:

- Conversion of the asexually produced merozoites to gametocytes occurs only in the cat intestinal epithelium.
- Microgametes produce oocysts, which are excreted in the faeces and spore in soil at 1–5 days.
- Eggs become infective after sporulation and ingestion by an intermediate host.
- Ingested sporozoites develop into tachyzoites in the intermediate host. These divide rapidly and travel to the lymph nodes inside white blood cells.
- Extraintestinal spread can be to skeletal muscle, heart, brain, and the eye. Cysts enclose slowly dividing bradyzoites at these sites.

To support the diagnosis PCR and detection of local antibody synthesis in the eye can be completed on intraocular fluid samples. However, in the majority of adult ocular toxoplasmosis the diagnosis is clinical.

Treatment of toxoplasma is with pyrimethamine, sulphadiazine, or clindamycin. Ocular toxoplasmosis may require steroid therapy in addition to these to assist with resolution of intraocular inflammation.

Acanthamoeba

Acanthamoeba are protozoa that can be found in public water supplies, swimming pools, hot tubs, fresh water ponds, lakes, bottled mineral water, and soil. They can cause indolent corneal ulceration, most commonly in soft contact lens wearers. Firstly the amoeba attaches to the epithelial surface; an epithelial defect may assist in the adherence. Stromal changes are thought to occur due to collagenase activity.

Acanthamoeba can only grow on non-nutrient agar covered with killed *E. coli*, which serves as its food source.

CLINICAL TIP

Clinical manifestations of toxoplasmosis

- In transplacental infection mothers acquire the parasite during pregnancy. If acquired in the first trimester the effects can be profound, including still birth, intracranial calcification, mental retardation, and chorioretinitis. Asymptomatic retinal inflammation and cyst formation may occur with chorioretinitis acquired at a later stage.
- Infection acquired in childhood can lead to meningoencephalitis.
- Infection acquired in adulthood can result in febrile illness with features similar to infectious mononucleosis. Adult-acquired chorioretinitis can occur. Ocular toxoplasmosis presents as a single discrete lesion in the choroid/retina.
- Recurrence presents as satellite lesions of the original scar.

Acanthamoeba species exist in two forms:

- active trophozoite
- dormant encysted form.

Treatment of *Acanthamoeba* is often long and may require more than one treatment. Agents used include aminoglycosides (neomycin), diamidines (brolenes), and imidazoles. These are active against trophozoites.

More commonly used treatments are antiseptic biocides, such as polyhexamethylene biguanide, as these are also active against the cysts.

Ocular helminth infections

Helminths are parasites; often the human is essential in order for the helminth to complete its life cycle. The immune response to helminth infection is harmful to the host. A helminth infection initiates antibody production and complement activation alongside a marked type I hypersensitivity reaction with an allergic IgE response, IgE-dependent degranulation of mast cells, and an eosinophilia.

Helminths are classified as follows:

- trematodes, e.g. *Schistosoma*
- cestodes, e.g. *Taenia*, *Echinococcus*
- nematodes, e.g. *Toxocara*, *Filaria*, *Onchocerca*, and *Trichinella*.

Some of the above cause ocular disease.

Trematodes (flukes)

Schistosomes

Schistosomes are the only trematodes that can cause ocular disease. Schistosomes are contracted from snails. The larvae known as cercariae penetrate the skin and lymphatics. Immature worms enter the lungs and mature here before spreading haematogenously. At this stage the schistosomes can cause a granulomatous reaction in the ocular adnexae and/or subconjunctivally.

Cestodes

Taenia solium

Taenia can be ingested from pork. The ingested larvae mature in the intestine and reproduce. They cross the mucosa and can spread haematogenously to the retina, causing chronic inflammation and fibrosis. This can lead to retinal detachment.

Echinococcus granulosus

This can cause hydatid disease and tends to be found in sheep-farming communities. Humans are infected through ingestion of ova from infected dog faeces. Larvae pass most

commonly to the liver to produce hydatid cysts. Ocular involvement can occur in the orbit, leading to proptosis.

Nematodes (round worms)

Toxocara canis *(dogs) and* Toxocara cati *(cats)*

The natural hosts of these nematodes are dogs and cats. Humans are an accidental host and are not necessary for completion of the life cycle.

Toxocara canis is endemic in dog populations. Adult dogs and humans are infected by ingestion of infective ova from soil or by ingesting uncooked meat with larvae.

First- and second-stage larvae remain within the eggshell. Third-stage larvae are released into the intestine when the eggshell ruptures. They migrate via the lymphatics to the blood. Haematogenous spread to lung, liver, brain, and eye can occur.

The larvae encyst in the end organ. The majority remain dormant here. In pregnant bitches the larvae are activated and can migrate transplacentally to puppies. Puppies and lactating bitches are the main source of infective ova. Ova remain infective for several months and are able to survive extremes of temperature (–25–35°C).

Toxocara canis causes a chorioretinitis and uveitis. Specific IgM antibodies can be detected in the acute phase of the infection. Serological tests can be negative and antibodies may be picked up in vitreous or aqueous samples. Treatment is with oral thiabendazole or albendazole with or without steroid.

Filaria

Filarial nematodes tend to be thread-like in appearance. They promote a specific IgG response and a marked IgE response and eosinophilia. Immunity is directed towards the microfilarial stage or first-stage larvae.

Onchocerciasis

Onchocerciasis (river blindness) is caused by *Onchocerca volvulus* and affects 30 million people worldwide. It is endemic in West and Central Africa and is the most frequent helminthic ophthalmic infection.

Onchocerciasis is spread by the black fly, with humans being the definitive host. Black fly are the intermediate host, contracting microfilaria from the human. These mature to infective larvae in the black fly and when the black fly bites a human the larvae are transferred into the skin. Here they mature for around 1 year. Haematogenous spread to the eye causes corneal and conjunctival infection. Dead or dying microfilarias cause an inflammatory response known as the Mazzotti reaction. This can be used to diagnose the condition.

Treatment is with ivermectin and prevention is focused at black fly eradication with DDT spraying of river basins.

Loa Loa

Loa Loa is transmitted by the deer fly. It can infect the eye in the microfilarial stage and mature to an adult worm subconjunctivally or subcutaneously. In addition it can cause a uveitis or subretinal lesions.

Trichinella spiralis

Human infection can occur due to ingestion of undercooked meat. Larvae are able to infect extraocular muscles, leading to a painful granulomatous reaction and inflammatory myositis. Intraocular spread can occur, leading to retinal haemorrhages, subretinal lesions, and papilloedema.

Principles of sterilization and disinfectants

Sterilization

The goal of sterilization is to destroy all microbes. It is performed for surgical instruments, intravenous and intraocular fluids, dressings, and eye drops.

Various methods can be used to perform sterilization (Table 5.11). The most commonly used method is wet heat in the form of autoclaves.

Disinfectants

The goal of disinfection is to destroy pathogenic microbes, for example prior to surgery the skin is disinfected but not sterile (Table 5.12).

Antimicrobials

General terminology

Antibiotics are divided into broad spectrum and narrow spectrum. A broad-spectrum antibiotic is active against many different bacteria and hence is often used as a first-line therapy prior to culture of a given organism. Following culture and sensitivities antibiotic may be changed to a narrower spectrum antibiotic as deemed appropriate. This principle is applied throughout clinical practice.

The minimal inhibitory concentration (MIC) of an antibiotic refers to the concentration required to inhibit bacterial growth. This is calculated by *in vivo* culture and may not always accurately represent concentrations in plasma. However, it can be useful in certain antibiotics, for example in gentamicin use where the MIC is close to the concentrations in the plasma that can cause toxicitiy.

Antimicrobials are generally termed bactericidal or fungicidal if the agents kill microorganisms in the absence of an

Table 5.11 Summary of the methods used for sterilization

Method used	Principles of use	Time taken
Wet heat autoclave	Exposure to steam at high pressures—a pressure of 100 kPa produces a temperature of 121°C Materials are then dried High-pressure vacuum pumps can reduce sterilization times Routine testing to ensure sterilization is effective is essential	120 minutes
Dry heat	Used for instruments that can lose their sharp edge through steam or powders and oils Dry heat is produced in a fan-assisted drying oven heated up to 160°C	Several hours
Ethylene oxide gas	Acts as a sterilizing agent below air temperature and therefore is good for instruments that are susceptible to heat, e.g. plastic and rubber Equipment has to be aerated afterwards to get rid of the gas	24–48 hours
γ-Radiation	Industrial sterilization of disposable medical items	

Table 5.12 Summary of the different types of disinfectant

Method used	Microbial target	Time taken
Phenol-based solutions, e.g. 20% formalin in ethanol	Bacteria, spores, fungi, and viruses	Hours
Glutaraldehyde	Vegetative bacteria, fungi, viruses Can be used for metal instruments and glass lenses, e.g. Goldmann tonometer prisms	Hours
Ethyl alcohol 70% in water	Non-sporing organisms but not in the presence of body fluid Active against viruses but not adenovirus	Seconds
Iodine in alcohol	Bacteria but not spores Can be irritant	Minutes
Chlorhexidine and hexachlorophene	Most skin pathogens Hexachlorophene is toxic if absorbed over large areas	Minutes

immune response, and bacteriostatic if they inhibit growth of the microorganisms. Some antibiotics are bacteriostatic at low concentrations and become bacteriocidal at higher concentrations.

Antibiotics and antifungals

Antibiotics exert their effect through various routes:

- inhibition of cell wall synthesis
- interruption of the cell membrane
- inhibition of protein synthesis
- inhibition of nucleic acid synthesis

The following section describes the principles behind these modes of action and provides specific examples relevant to ocular antibiotic therapy.

Inhibition of cell wall synthesis

A large proportion of antibiotics act via this route. Both Gram-positive and Gram-negative bacteria contain peptidoglycan chains held together by peptide chains. The antibiotic binds to proteins in the cell membrane, which inhibits crosslinking of peptidoglycan strands. The bacterial cell wall is now weak and is more susceptible to killing by complement/lysis.

Penicillin (antibiotic)

Pencillins are effective bactericidal antibiotics produced from fungi. Penicillins target the final stage of production of the bacterial cell wall. A key component of penicillins and cephalosporins is the β-lactam ring. Penicillins tend to be effective against streptococci and Gram-positive bacilli. Ampicillin is effective against Gram-negative bacilli such as *E. coli* and *H. influenza*.

Staphylococcus aureus is often resistant due to the presence of β-lactamases, which can destroy the β-lactam ring. For this reason flucloxacillin was produced. Its side chain stops it being susceptible to β-lactamases. Carbapenem and clavulanic acid are β-lactamase inhibitors. Clavulanic acid can be combined with other penicillins to provide a bactericidal effect.

β-lactams are generally of low toxicity and penetrate inflamed tissue well.

Cephalosporins (antibiotic)

Cephalosporins exert bactericidal effects and have a broader range of activity than penicillins, with an ability to act against β-lactamase-producing organisms, including staphylococci and Gram-negative bacilli. The Gram-negative targeting has been improved and extended in the third- and fourth-generation cephalosporins to include *Pseudomonas* and *Bacteroides*.

Vancomycin (antibiotic)

Vancomycin is a glycopeptide antibiotic. It binds to pentapeptide chains and prevents peptidoglycan assembly in the microbial cell wall, resulting in cell lysis. It has bactericidal activity against Gram-positive organisms. It is administered intravenously due to poor absorption. It can be used as an intravitreal preparation in the treatment of endophthalmitis.

Interruption of the cell membrane

The microbial cell membrane is a lipoprotein layer surrounding the cytoplasm. Antimicrobials that interrupt the cell membrane will cause cell lysis.

Polymyxins (antibiotic)

Polymyxins are broad-spectrum antibiotics that are active against Gram-negative bacteria, including *Pseudomonas*. They are cyclic peptides that adsorb to negatively charged lipids in the cell membrane, leading to disorganization of the membrane and loss of cell function. They lack selectivity, which can lead to nephrotoxicity and neurotoxicity.

Amphotericin and nystatin (antifungal)

Amphotericin and nystatin belong to the polyene group. They bind to ergosterol in fungal cell membranes, changing ionic transport and affecting permeability of the fungal cell wall, leading to cell lysis.

Amphotericin cannot be absorbed from the gut and is therefore administered intravenously. Ocular penetration is poor and when used topically it can be painful. It is active against *Candida*, *Aspergillus*, *Cryptococcus*, *Histoplasma*, *Coccidioides* and *Blastomyces*.

Side effects can include anaemia and nephrotoxicity.

Nystatin is isolated from *Streptomyces noursei*. It is active against *Candida*, *Cryptococcus,* and the *Trichophyton* group. It can be fungicidal or fungistatic.

Imidazoles (antifungals)

Imidazoles are a group of broad-spectrum synthetic antimycotic agents. They inhibit ergosterol production by acting on the cytochrome P_{450} pathway. Fluconazole is effective against *Candida*, *Cryptococcus neoformans,* and *Histoplasma capsulatum*. Fluconazole is ineffective in *Aspergillus* and *Mucor*. Itraconazole is effective in *Aspergillus* species. Newer triazole agents such as variconazole can now be used in immunocompromised patients in the treatment of invasive aspergillosis and *Candida* infections.

Azoles can cause abnormalities in liver function and hepatotoxicity (ketaconazole).

Inhibition of protein synthesis

Some antibiotics inhibit bacterial protein synthesis. Bacterial ribosomes are sufficiently different to avoid toxicity to humans.

Tetracyclines (antibiotic)

These are broad-spectrum bacteriostatic antibiotics active against Gram-positive and Gram-negative bacteria and *Chlamydia*. They interrupt the cycle of attachment of amino acids to the first binding site during protein synthesis by inhibition of aminoacyltransferase RNA and the mRNA–ribosome complex.

Side effects include gastrointestinal disturbance and teratogenicity.

Gentamicin and neomycin (antibiotics)

These are bactericidal aminoglycoside antibiotics. They prevent the binding of mRNA to the ribosome and can also cause mRNA to be misread and hence the incorporation of the wrong amino acid into the protein. They require aerobic transport to enter the cell and are therefore ineffective in strict anaerobes and streptococci.

They are most effective against staphylococci and aerobic Gram-negative bacilli. Intravenous administration requires strict monitoring to prevent ototoxicity and nephrotoxicity.

Topical gentamicin is used in microbial keratitis but does cause corneal epithelial toxicity. Intravitreal gentamicin leads to retinal toxicity.

Erythromycin (antibiotic)

This is a predominantly bacteriostatic macrolide antibiotic that binds to ribosomal subunits and interferes with translocation. At higher concentrations it can be bactericidal. It is used in penicillin-allergic patients in the treatment of orbital and preseptal cellulitis.

Chloramphenicol (antibiotic)

This is a potent broad-spectrum bacteriostatic antibiotic. It inhibits peptidyltransferase and hence prevents the transfer of the peptide chain to other amino acids. It is active against many Gram-positive and Gram-negative bacteria and is most commonly used in the treatment of conjunctivitis.

Systemic use is limited to life-threatening conditions such as *Haemophilus influenzae* meningitis or typhoid fever.

Chloramphenicol can lead to aplastic anaemia and 'grey baby' syndrome in neonates. It is contraindicated in pregnancy and breast feeding.

Inhibition of nucleic acid synthesis

Inhibition of nucleic acid synthesis by preventing purine and pyrimidine synthesis is used in the production of antibiotics, antifungals, and antivirals.

Ofloxacin/ciprofloxacin

This fluoroquinolone inhibits the action of bacterial DNA gyrase, which is ordinarily involved in the folding/unfolding of DNA during synthesis. It is a broad-spectrum antibiotic active against *Pseudomonas, Chlamydia*, Enterobacteriaceae, *Mycoplasma*, and *Rickettsia*. It is less effective against Gram-positive bacteria and anaerobes.

Ofloxacin is used topically to treat microbial keratitis and readily penetrates cornea with fewer side effects than topical gentamicin. It is well absorbed from the gut and intravitreal concentrations are as high following oral treatment as they are with intravenous therapy. It is used as part of endophthalmitis protocols.

Side effects include photosensitive rashes and gastrointestinal upset.

Sulphonamides (antibiotic)

These inhibit metabolism of *para*-aminobenzoic acid to folate, which is used in bacterial metabolism, DNA synthesis, and survival. They are bacteriostatic against Gram-positive and Gram-negative organisms. They tend to be used less now because of bacterial resistance.

Metronidazole (antibiotic)

Metronidazole disrupts DNA synthesis. It is effective against anaerobic bacteria and protozoan parasites such as *Giardia*, *Trichomonas*, and *Entamoeba*. It is used in ophthalmic practice for the treatment of orbital cellulitis in combination with cefuroxime.

Side effects include disrupted liver function and low platelets.

5′-fluocytosine (antifungal)

5′-fluocytosine is converted in fungal cells to 5-fluorouracil, which is incorporated instead of uracil into fungal RNA. This leads to inhibition of DNA synthesis. It is effective against *Candida* and *Cryptococcus*, and can be administered orally or intravenously. It penetrates cerebrospinal fluid and tissue well.

Antivirals

Effective antivirals are confined to the treatment of herpetic and DNA-related viruses.

Aciclovir

This is an acyclic analogue of guanosine. The mechanism of action is inhibition of viral DNA polymerase following phosphorylation by viral thymidine kinase. It is therefore only effective in infected cells.

Aciclovir is safe and effective against the herpes group of viruses, but it is not effective against CMV. It is used in the treatment of herpes simplex and varicella ointment.

Side effects can include renal insufficiency, gastrointestinal disturbance, and headache. Some resistance does exist, although it is rare in immunocompetent patients.

CLINICAL TIP

Summary of bacterial sensitivities of commonly used antibiotics

Antibiotic	Bacterial sensitivity
Penicillins	
Penicillin G	Gram-positive cocci
Penicillin V	*Gonococcus*
Ampicillin	*Haemophilus*
Carbenicillin	*Pseudomonas*, *Proteus*
Meticillin	*Staphylococcus*
Cephalosporins	
Cephalexin	Gram-positive cocci
Cefuroxime	*Staphylococcus*
Aminoglycosides	
Streptomycin	*Mycobacterium*
Neomycin	Gram-positive cocci
Gentamicin	*Escherichia coli*
Tobramycin	*Proteus*/*Pseudomonas*
Macrolides	
Erythromycin	*Staphylococcus aureus* Streptococci *Neisseria*, *Mycoplasma*, *Moraxella*
Chloramphenicol	*Haemophilus*, *Neisseria*, *Bordetella*, *Bacteroides*, *Proteus*, *Pseudomonas*, Streptococci
Ciprofloxacin	*Proteus*, *Pseudomonas*, *Moraxella*

Ganciclovir

This is an acyclic nucleotide analogue. It acts as an inhibitor of and a false substrate for CMV DNA polymerase. It is, therefore, far more effective than aciclovir in the treatment of CMV and is less active than aciclovir in the treatment of herpes simplex and VZV.

Treatment of CMV retinitis is with a loading course followed by maintenance therapy. Intravitreal doses can be given in refractory CMV retinitis.

Side effects include neutropenia, thrombocytopenia, and renal impairment.

Foscarnet

This is a paraphosphate analogue of phosphonoacetic acid. It selectively and reversibly inhibits viral-specific DNA polymerases and reverse transcriptases. Foscarnet is effective at the loading course stage of CMV retinitis and is used in acyclovir-resistant herpes simplex infections.

Foscarnet is also useful in AIDS patients with CMV retinitis as it can be used in conjunction with zidovudine without risk of profound neutropenia.

Side effects include renal impairment and proteinuria.

Antimicrobial resistance

Increasing use of antimicrobials can lead to resistance through mutation (as seen in resistance to antituberculous drugs such as rifampicin) or inheritance of 'resistance' genes which is either plasmid mediated by conjugation or via transduction through a bacteriophage.

Bacterial resistance is currently the most problematic form of resistance clinically.

Bacteria can develop resistance through various methods:

- altering the antibiotic target, for example changes to the bacterial cell wall or membrane by failure of ribosomes to bind erythromycin
- destruction or inactivation of the drug by enzymes, e.g. β-lactamase in penicillin resistance or by acetylation of aminoglycosides
- prevention of transport of the drug into the microbe
- use of alternative enzymic pathways that are resistant to the drug, e.g. enzymes resistant to sulphonamide and trimethoprim.

CHAPTER 6

Immunology

Principles of innate and acquired immunity

The immune system is designed to protect the host from infections and other stresses, and involves a variety of molecular and cellular responses, as well as physical barriers and defence mechanisms. Immune responses are broadly divided into innate and acquired.

The innate response is rapid and non-specific and the adaptive/acquired response is slower, is specific to the foreign body (antigen) such as an infectious agent (pathogen), and generates memory cells. The different types of immunity, immune responses, and the cells and molecules involved are described in this chapter, with reference to the eye when possible (Fig. 6.1).

Innate immunity

Innate immunity refers to the part of the immune system that provides rapid protection for the body and fights infections in a non-specific way, by providing general protection from a pathogen.

The innate system can be divided into three components:

1. Physical barriers, including the skin, eyelids, mucosal surfaces, mucous secretions, and tears.
2. Molecular—these can be divided into those that are:
   (a) normally present, e.g. complement factors, lysozyme, IgA found in tears, defensin in aqueous humour, and immunoglobulins in blood
   (b) produced by cells in response to injury, e.g. interleukins (cytokines) and TNF-α.

These will all be described in greater detail later in the chapter.

3. Cellular, e.g. phagocytic and cytotoxic cells such as polymorphonuclear leukocytes, macrophages, eosinophilic granulocytes, and natural killer (NK) cells.

Cells of the innate immune system

Neutrophilic granulocytes

Neutrophilic granulocytes, more commonly known as neutrophils, are the most commonly found white cell in the human circulation. They are polymorphonuclear leukocytes, developed from the common myeloid precursor, and are fully differentiated.

They act primarily as scavengers, ingesting foreign bodies (phagocytosis) and releasing free radicals and proteases from their numerous cytoplasmic granules and lysozymes, lactoferrin, and oxidative enzymes (e.g. NADPH-dependent oxidases, myeloperoxidases, and catalases). They have a half-life of 1–2 days and are extremely mobile, rapidly infiltrating areas of inflammation.

Myeloid mononuclear cells

These cells are derived from marrow stem cells. They differentiate into monocytes, macrophages, and myeloid dendritic cells.

Monocytes are mononuclear cells, with a distinctive kidney-shaped nucleus. They are involved in antibody- and complement-associated phagocytosis of antigens, but also can develop into either macrophages or dendritic cells.

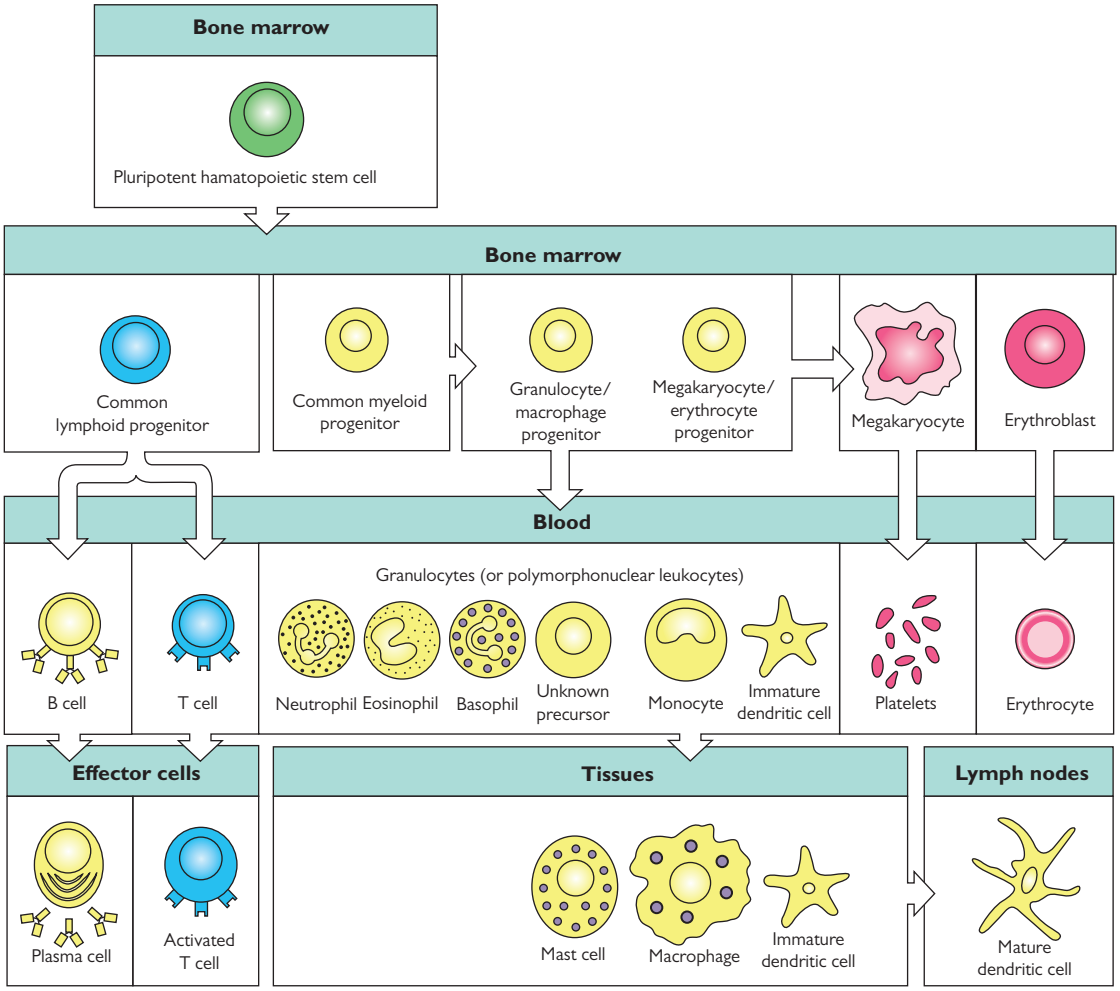


Fig. 6.1 Cells of the innate and adaptive immunity.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 1.3, 2008, with permission from Garland Science.

Macrophages are derived from monocytes and have many essential functions. They are present in tissues and phagocytose dead and damaged cells in an inflammatory environment. They also phagocytose opsonized (antibody- or complement-coated) antigens, and secrete endogenous enzymes to digest the antigens. They can also engulf antigens, process them via endocytosis, and subsequently present antigenic epitopes to T cells as part of the acquired immune response (see *Phagocytosis*, p. 171).

The most potent antigen-presenting cell (APC), the dendritic cell, provides the link between the innate and acquired cells by phagocytosing *de novo* foreign material/antigen via endocytosis and presenting the processed antigen on MHC class II molecules (see *Phagocytosis*, p. 171) to naïve T cells. The cell is therefore heavily involved in initiating an antigen-specific response and is often found in tissues that are areas of high antigen exposure, e.g. skin, conjunctiva, respiratory system, and gut.

Mast cells

Mast cells are a type of granulocyte found commonly in vascularized connective tissues. They are large cells characterized by distinctive cytoplasmic granules that stain positively with toluidine blue dye. They have cell surface receptors specific for Ig: IgE and IgG (see *Phagocytosis*, p. 171). These receptors attach to the Ig. If adjacent receptor-bound IgE or IgG is crosslinked by antigen, the mast cell will

undergo degranulation, meaning it will release the contents of its granules into the surrounding extracellular space. This process takes seconds, and in the case of mast cells results in the release of histamine and several other pro-inflammatory molecules. Histamine is a pro-inflammatory chemical that primarily induces vasodilation and increased vascular permeability as well as chemoattractants, causing an accumulation of tissue fluid and soluble blood components. This also attracts circulating leukocytes to migrate to the area of mast cell activation, and mast cells are a significant initiator of hypersensitivity responses (see Chapter 4, p. 123).

Basophils

Basophils are very similar to mast cells as they are large granular cells that release histamine in areas of inflammation. They are, however, found mostly in blood circulation and tend to migrate to areas of inflammation. The role of basophils has been difficult to assess due to their scarcity: they make up less than 1% of peripheral blood lymphocytes (PBLs) and have few cell surface markers. However, recent studies suggest not only that they have a role in initiating a T helper 2 cell (Th2) response, but also that they can act as APCs in areas of inflammation.

Eosinophils

Eosinophils are granulocytic leukocytes with cytoplasmic granules that stain bright orange by the acidic stain eosin. Eosinophils are found mainly in connective tissues below respiratory, gut, and urogenital epithelium. Once eosinophils are activated they release free radicals and highly toxic granules. These are very powerful agents that result in the death of microbes and parasites but also cause damage to surrounding tissue. Accordingly eosinophils have been implicated in chronic inflammatory diseases. To prevent eosinophilic tissue damage the body carefully regulates eosinophil production and activation through three mechanisms:

1. Eosinophil production in the bone marrow is very low, only upregulated by interleukin-5 (IL-5), produced by Th2 cells.
2. An eosinophil must become activated before it can degranulate. During a Th2-mediated response and/or inflammation, various chemokines are released which attract and activate the eosinophils—these are called eotaxins.
3. The antigen must crosslink specific IgE that is bound to the eosinophil, as with mast cells. This will result in eosinophil degranulation.

Natural killer cells

The primary function of NK cells is to directly kill infected or malignant target cells. If a cell is infected, it expresses increased levels of NK cell receptor ligands. If a cell has become malignant, it has a decreased expression of MHC class I, which acts as a trigger for NK cell lysis.

A summary of the innate cells is given in Fig. 6.2.

An innate response

We have discussed which cells are involved in the innate immune system, but not how they interact. The best way to explain their action is to imagine their function in the context of an infection by a microorganism.

If a microorganism manages to successfully penetrate the first line of defence, such as the skin, it then reaches the stromal layers of the tissues wherein it will begin to replicate and possibly release toxic substances that will damage the surrounding cells and tissue. This will attract leukocytes to the area of infection and innate immune cells have the ability to respond to foreign antigens via expression of specific molecules.

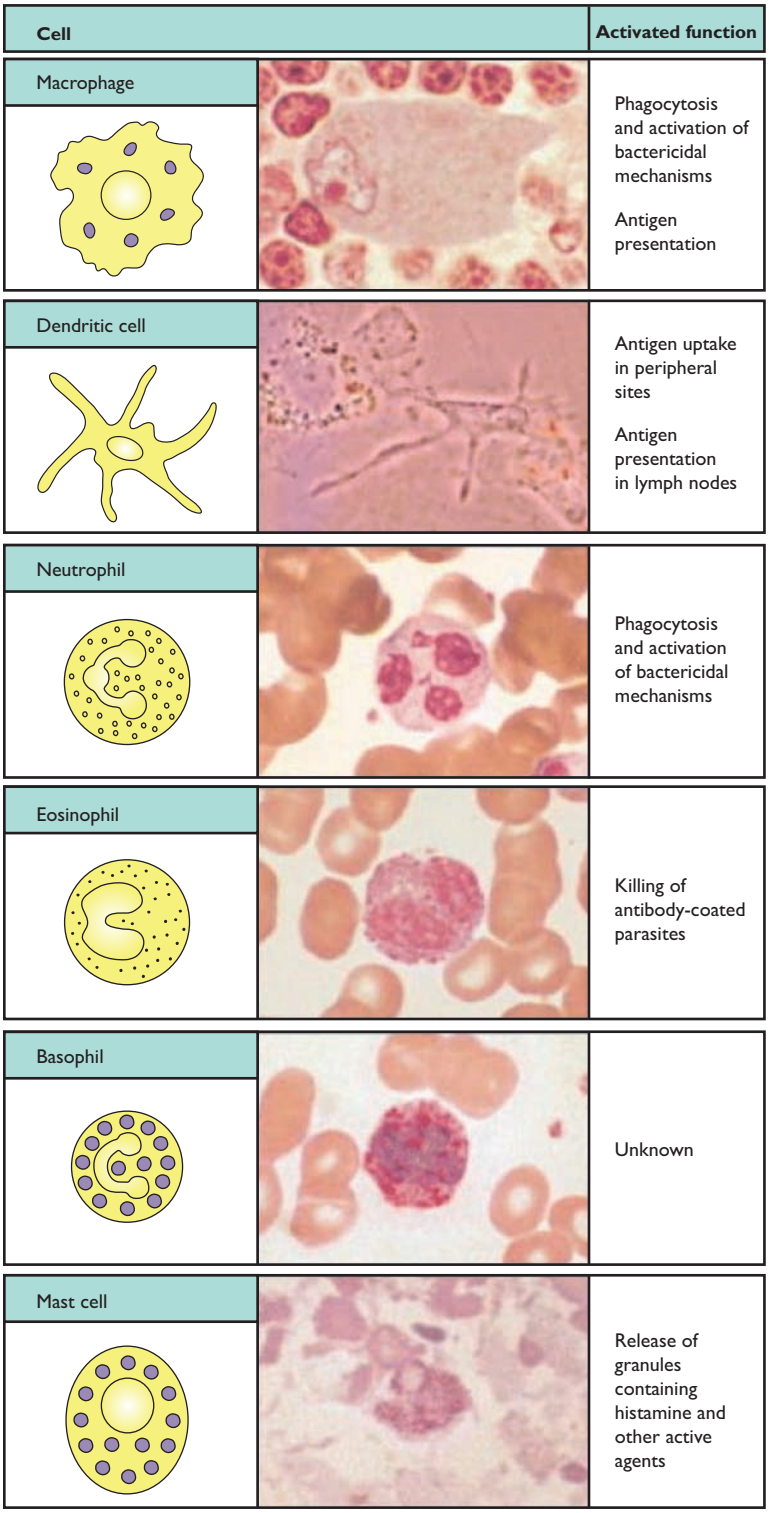
Innate effector mechanisms are regulated by germ-line encoded receptor systems that can discriminate between dangerous and harmless motifs. This was predicted by immunologist Charles Janeway, who suggested that prokaryotic pathogens expressed groups of molecules that were different to molecules in eukaryotes, thus making them easily distinguishable as pathogens. He termed these pathogen-associated molecular patterns (PAMPs). He also predicted that they would be recognized by pathogen recognition molecules (PRMs) or pattern recognition receptors (PRRs).

Janeway's theory was vindicated. It has been shown that, for example, lipopolysaccharide, a cell-wall component of Gram-negative bacteria such as *Salmonella*, acts as a PAMP and is recognized by a PRR on the surface of macrophages. This receptor is known as Toll-like receptor 4 (TLR-4) and is one of a group of PRRs.

The Toll-like receptors are a group of 10 cell surface proteins that act as receptors to trigger an innate cell response. They are named Toll-like as they are structurally homologous to Toll receptors, which are found in *Drosophila melanogaster*, a species of fruit-fly. The Toll receptors in the *Drosophila* (named after the German word for great or amazing—toll) are stimulated by various PAMPs in the fly and begin an immune response.

Once stimulated, Toll-like receptors upregulate downstream proteins to start production of cytokines and interleukins appropriate to fighting the pathogen type in question. For example, if TLR-3 is stimulated by virus-derived double-stranded RNA this leads to the production of interferon, which has antiviral effects. TLR-4 is stimulated on binding to lipopolysaccharide (LPS) and activates the NFκB pathway (pro-inflammatory pathway).

Fig. 6.2 Innate cells



Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 1.4, 2008, with permission from Garland Science.

Once the innate system has been stimulated, soluble factors or cytokines flood the immediate area and stimulate an inflammatory response. The resulting increased vascular flow and permeability encourages infiltration of leukocytes. This is discussed in greater detail in Chapter 4.

Leukocyte infiltration

Following the triple response (as described in Chapter 4, p. 121), leukocytes, including neutrophils and monocytes, migrate from the bloodstream into the site of injury. The key stimulus for cell migration to a site of infection is via chemotaxis, the process of attracting cells to a site by inducing them to migrate along a chemical gradient of chemokines. C5a, secreted by mast cells, is a key chemokine involved in trafficking of innate cells.

These leukocytes then enter the inflamed parenchyma through a four-stage process of cellular extravasation:

1. During an inflammatory event, soluble factors or cytokines such as TNF-α and interleukin-1 (IL-1) will be secreted. These cytokines will, in turn, upregulate expression of selectin ligand on endothelial cells lining the blood vessel walls and will lead to loose endothelial–neutrophil interactions during the rolling phase.
2. With increasing expression of adhesion molecules, the neutrophils become loosely adhered to the endothelium.
3. The neutrophil then becomes polarized and develops firm adhesion through integrin–cell adhesion molecule (CAM) interactions.
4. Extravasation of leukocytes through the endothelium is mediated by expression of PECAM-1 (CD31) on both the leukocyte and the endothelium, which has been hypothesized to disrupt the intercellular tight (occludin) junctions and adherens junctions.

Finally, leukocytes migrate through the endothelial cell layer to enter the extravascular tissues (Fig. 6.3).

Neutrophil cytotoxicity

Once the neutrophil has infiltrated the tissues via leukocyte activation, adhesion, and migration, as described above, it then rapidly phagocytoses the invading antigen. During this process, the cell undergoes neutrophil respiratory burst.

Cells such as neutrophils kill pathogens by producing a combination of reactive oxygen intermediates (ROIs) and reactive nitrogen intermediates (RNIs), and by exposure to lysosomal granule contents to digest the microorganisms.

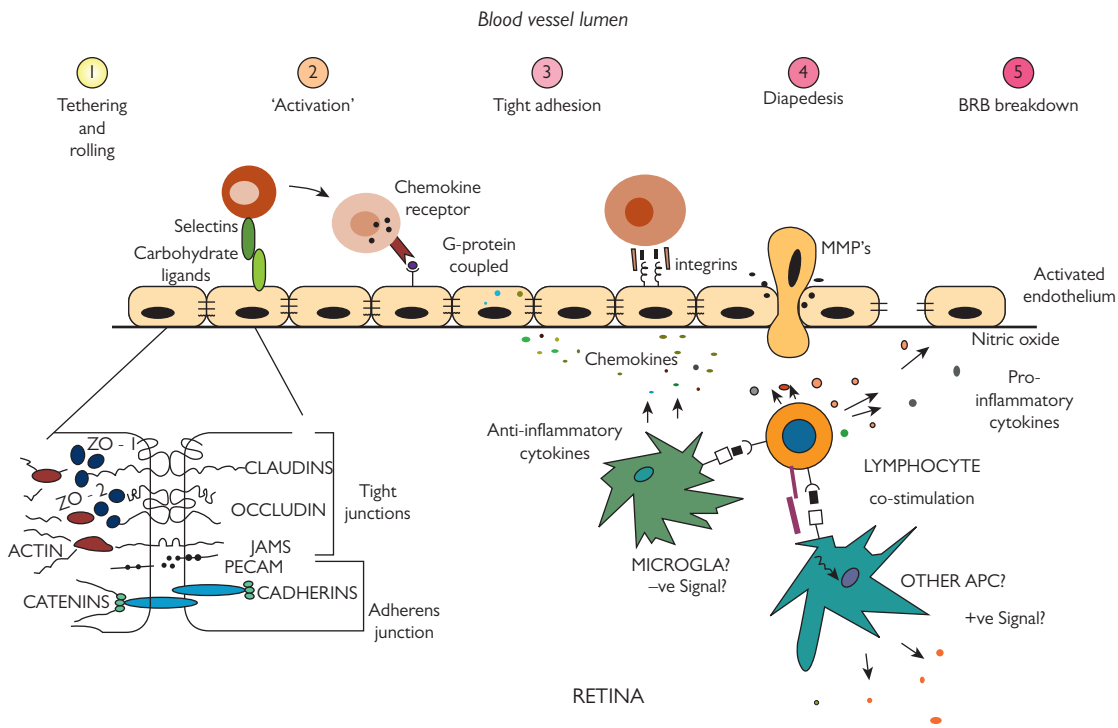


Fig. 6.3 Leukocyte trafficking, tethering, and rolling.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 2.36, 2008, with permission from Garland Science.

Neutrophils then die by apoptosis and are removed by tissue macrophages. Failure of the neutrophils to die by apoptosis (see Chapter 4) results in necrosis and tissue damage.

The processes involved during innate immune responses utilize a variety of rapid mechanisms to clear or hold in check a pathogen until it can be recognized by cells of the adaptive immune system (see *Complement*), which requires more time to develop as it involves induced effector mechanisms.

Complement

Complement is an essential part of the body's defence and works in conjunction with the innate and adaptive immune systems. It comprises a large family of plasma proteins, produced in the liver, and various epithelial cells in the body. There are a large number of proteins in the family and they make up to 15% of total serum proteins.

The proteins exist as pro-enzymes (zymogens) until they are activated. Once cleaved into their active form they act on the next complement protein and activate it. This creates a protein cascade. Complement has three actions:

1. Activated complement proteins bind to pathogens and opsonize them.
2. Small fragments of complement that have been cleaved off act as chemoattractants.
3. The final components of the cascade combine to create a protein structure, the membrane attack complex, which creates pores in bacterial cells.

There are three ways to initiate a complement cascade (Fig. 6.4):

1. The classical pathway—initiated by binding of C1q directly to:
   (a) the pathogen surface (Gram + ve)
   (b) bacteria and C-reactive protein complexes
   (c) bacteria and antibody complexes.
2. The alternative pathway—spontaneously activated complement binds directly to pathogen surfaces.
3. The lectin-binding pathway—complement binds directly to lectin molecules.

Membrane attack complex

The membrane attack complex is a protein structure formed from activated complement protein, creating a pore in the membrane of bacteria (Fig. 6.5). Component C5b binds to C6 and C7 to create C5b, 6, 7. The C7 component of this structure then binds to the bacterial cell wall. C8 molecules then bind to the complex, followed by multiple C9 molecules, which traverse the entire cell wall. Up to 16 C9 molecules then bind to create a pore in the membrane. This leads to a disruption in the proton gradient across the cell membrane and allows various enzymes, including lysozymes, to enter the cell.

However, inappropriate or excessive activation of complement can also lead to:

- inflammation
- anaphylaxis
- autoimmunity.

To avoid excessive complement activation, the system contains a number of regulatory complement components which act as a negative feedback loop at various stages of complement activation (see Fig. 6.6).

Phagocytosis

The process whereby cells engulf, digest, and remove antigens from a host is called phagocytosis. These cells, which have phagocytic functions, include macrophages and neutrophils.

The process of phagocytosis is greatly enhanced if the antigen has been previously coated with immunoglobulin or C3b, since phagocytes express cell surface receptors for these molecules, which act like a catalyst to promote adhesion and more effective digestion. This process is called opsonization:

- Pathogen-bound IgG is detected by phagocytes expressing Fc receptors.
- Pathogen-bound C3b (complement) is detected by phagocytes expressing CR1 receptors.

Certain bacteria have polysaccharide coatings, which prevent phagocytosis unless they are opsonized.

The phagocyte binds to the pathogen and begins to invaginate, enveloping it. The phagocyte membranes meet around the pathogen and fuse to create a vesicle. This process is called endocytosis and the resulting acidic vesicle is called a phagosome. The phagocyte then releases enzyme-containing granules into these vesicles, including lysosomes. The fusion of the vesicle with lysosomes is called a phagolysosome. This final stage of the process results in a complete digestion of the pathogen (Fig. 6.7).

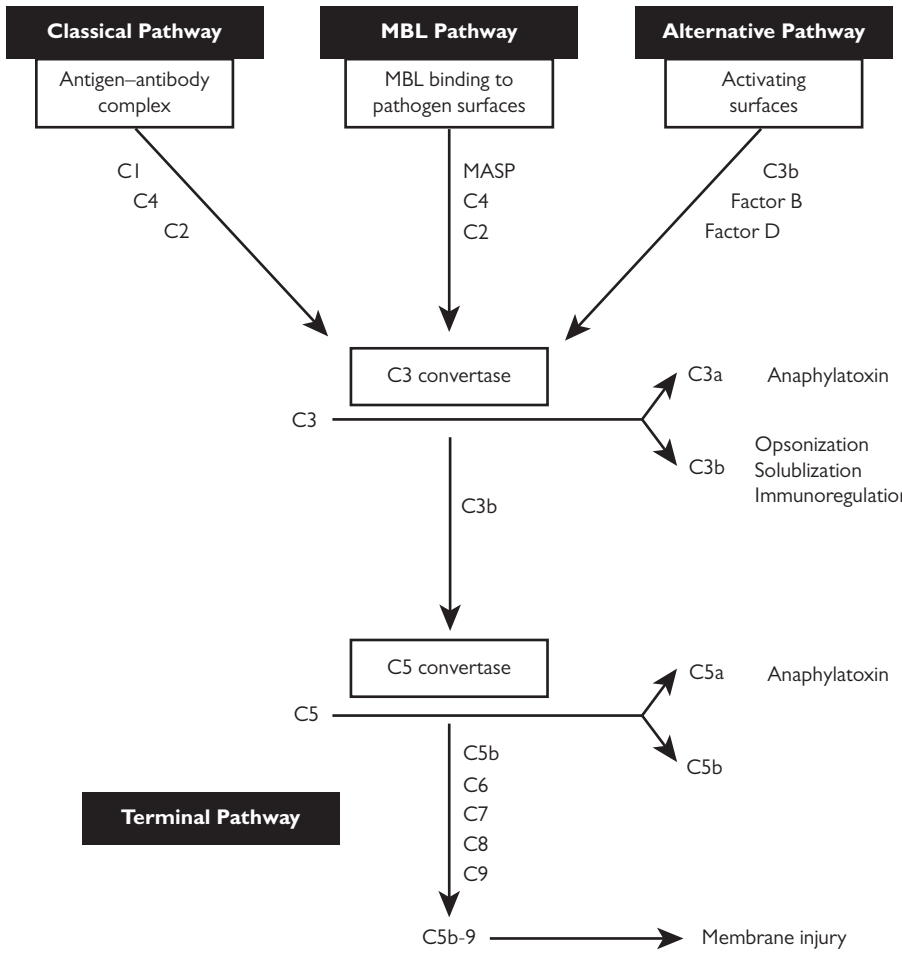


Fig. 6.4 Complement activation. MBL, mannose binding lectin.

Reprinted by permission from Macmillan Publishers Ltd: *Kidney International*, 'Intrarenal synthesis of complement', Wuding Zhou, James E Marsh and Steven H Sacks, 59, pp. 1227–1235; doi:10.1046/j.1523-1755.2001.0590041227.x, Copyright 2001.

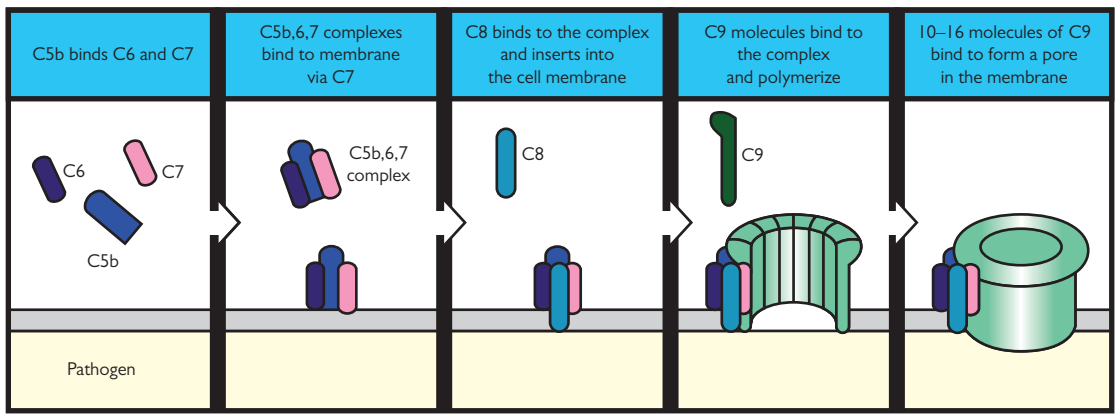


Fig. 6.5 Membrane attack complex.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 2.47, 2008, with permission from Garland Science.

Control proteins of the classical and alternative pathways	
Name (symbol)	**Role in the regulation of complement activation**
C1 inhibitor (C1INH)	Binds to activated C1r, C1s, removing it from C1q
C4-binding protein (C4BP)	Binds C4b, displacing C2b; cofactor for C4b cleavage by I
Complement receptor 1 (CR1)	Binds C4b, displacing C2b, or C3b displacing Bb; cofactor for I
Factor H (H)	Binds C3b, displacing Bb; cofactor for I
Factor I (I)	Serine protease that cleaves C3b and C4b; aided by H, MCP, C4BP, or CR1
Decay-accelerating factor (DAF)	Membrane protein that displaces Bb from C3b and C2b from C4b
Membrane cofactor protein (MCP)	Membrane protein that promotes C3b and C4b inactivation by I
CD59 (protectin)	Prevents formation of membrane-attack complex on autologous or allogenic cells. Widely expressed on membranes

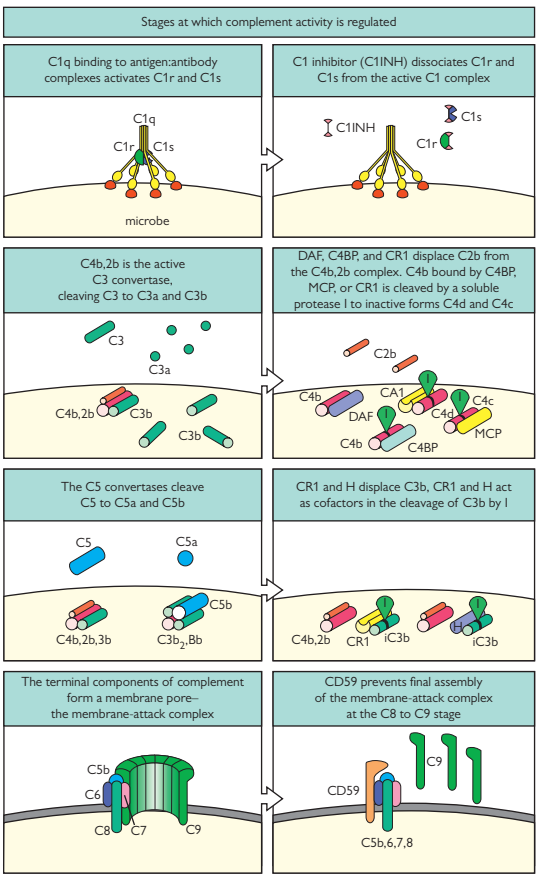


Fig. 6.6 Complement regulatory components.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 2.25, Figure 2.26, 2008, with permission from Garland Science.

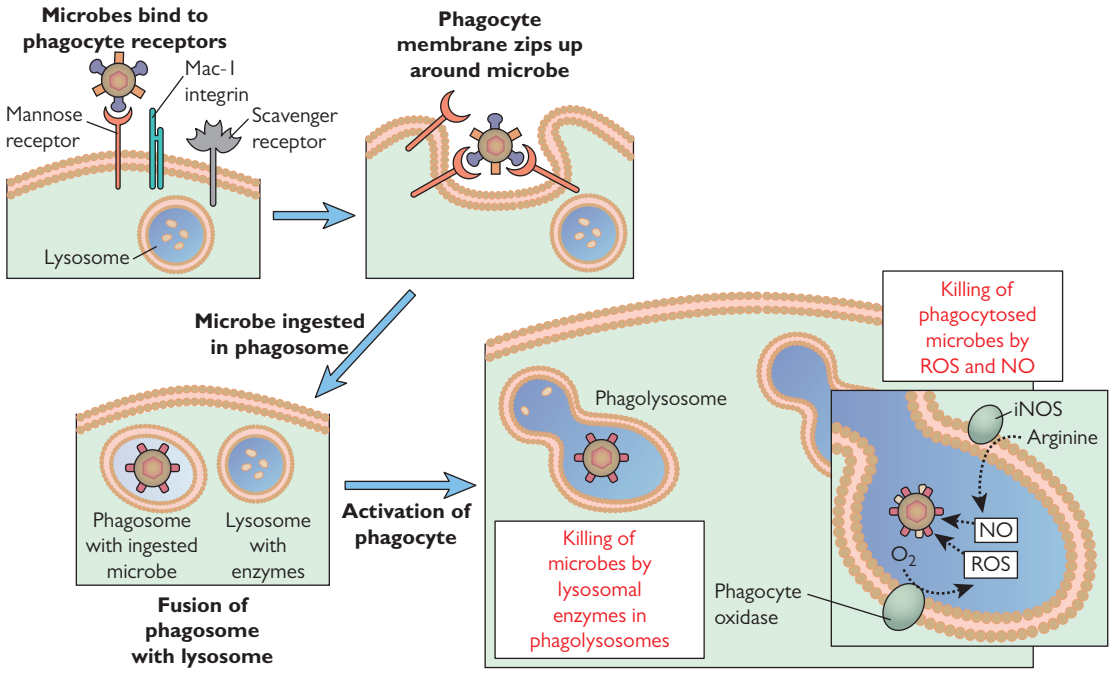


Fig. 6.7 Phagolysosome fusion.

This figure was published in *Cellular and Molecular Immunology*, Abbas AK, Lichtman AH, Pillai S, Figure 2.8 Phagolysosome fusion, p. 35, Copyright Elsevier, 2010.

Acquired immunity

Acquired immunity is the ability of the immune system to recognize an antigen it has previously encountered and respond more rapidly to eradicate the same antigen on subsequent exposure. It is the product of the adaptive immune system. This highly specialized form of immunity is mediated by T lymphocytes (T cells) and B lymphocytes (B cells), and is usually required for intracellular and/or complex antigens. The generation of immunological memory to that antigen involves the production of long-lived memory T and B cells. If there is an inappropriate response to an antigen, or the antigen is present in excessive amounts, this may lead to hypersensitivity responses (see Chapter 4, Type I to IV hypersensitivity reactions, pp 123–5).

Lymphoid system

The lymphoid system can be divided into T cells (and their subsets: $CD4^+$ and $CD8^+$), B cells, and NK cells. These three cell types arise from common lymphoid progenitors, which in turn arise from pluripotent haematopoietic stem cells in bone marrow.

T cells

T-cell precursors are produced in the bone marrow and migrate to the thymus, a central lymphoid organ in the upper chest, where they mature.

T-cell maturation is a complicated process. Naïve (immature, no antigen recognition) T cells migrate from the bone marrow to the thymus, where they continue to proliferate. The majority (98%) of these T cells will die and only 2% will survive as mature cells.

Achieving maturity is dependent on the development of the T-cell receptor (TCR). This is the cell surface-bound protein structure that binds to antigen. Stimulation via TCR binding results in T-cell activation and proliferation, and the beginning of an antigen-specific T-cell response. For this reason it is essential that T cells only recognize foreign/pathogenic antigen and not self-antigen (autoantigen). TCR recognition of autoantigen is the basis of some autoimmune diseases.

To prevent this, T-cell maturation in the thymus includes TCR testing. If a TCR is autoreactive it is stimulated to

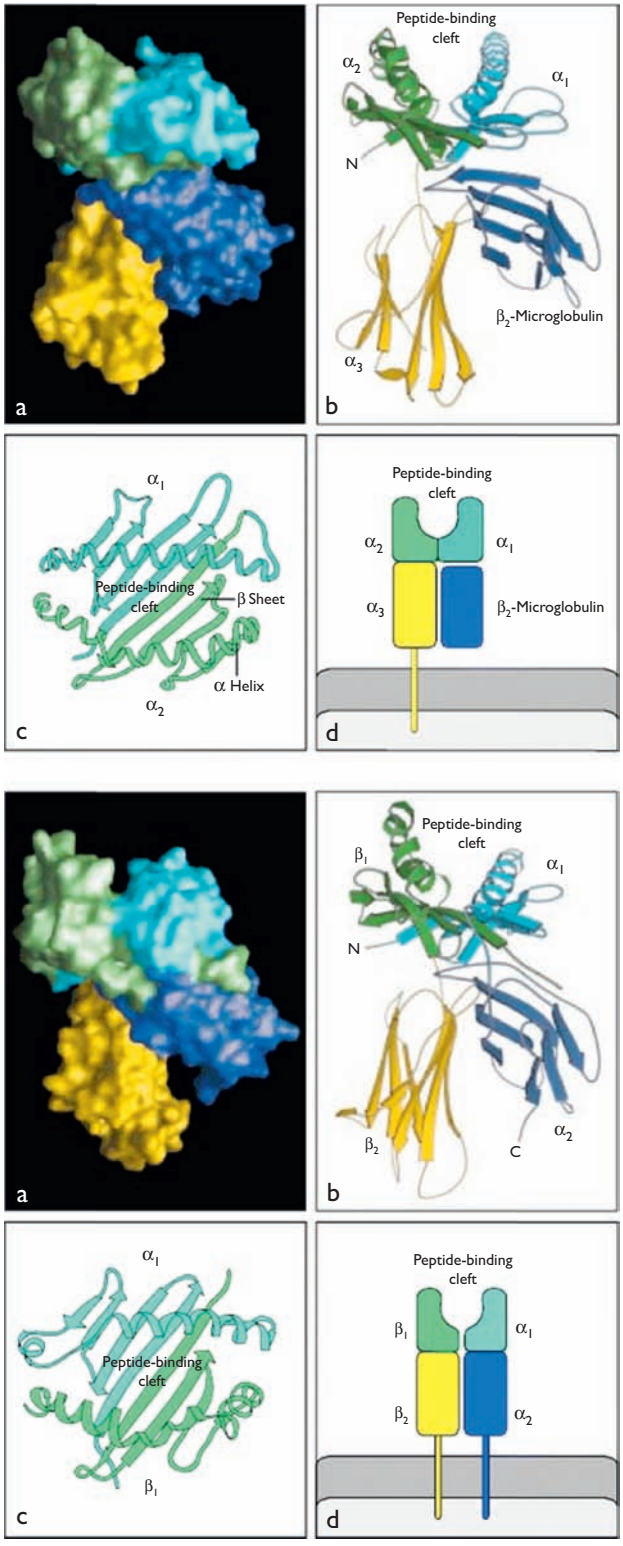


Fig. 6.8 MHC class I and II molecules.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 3.15, Figure 3.16, 2008, with permission from Garland Science.

either rearrange its TCR or undergo apoptosis. This process is called clonal deletion and it prevents the release of self-reactive T cells.

Clonal deletion

TCRs are divided into α/β and γ/δ TCRs, referring to the four possible protein structures that can make up a TCR. They are always paired as α/β and γ/δ.

α/β T cells make up the vast majority of T cells. They can be subdivided into CD4 and CD8 T cells, and are the main coordinators of an adaptive immune response. We will be focusing on the maturation of these T cells.

γ/δ T cells make up a very small percentage of all T lymphocytes and are often found in mucosal tissues. Their function is not fully understood, although there is some evidence that they are involved in innate immunity. The γ/δ T cells are clonally selected in a similar way to the α/β T cells.

When α/β T cells enter the thymus, they are lacking most cell surface proteins and have normal receptor genes. These genes then undergo rearrangement, as governed by expression of the recombination activating genes (RAGs). Once the genes are rearranged, a temporary TCR is expressed. The T cell then presents itself to self-antigen on self-major MHC. If stimulated, more gene recombination or death occurs. If the T cell is not stimulated it is able to mature and is safe to be released into the periphery.

T-cell recognition of antigen

The process of initiating an adaptive response requires the presentation of antigen to a T or B lymphocyte. This antigen is a part of the molecular structure of a pathogen, and when processed and presented to lymphocytes by an APC it activates lymphocytes against any cell with identical antigen. This is the basis of an antigen-specific response. A T cell can only be activated by MHC antigens, APCs, and antigen processing.

Specialized cell surface proteins carry out this presentation. These were first discovered on mouse APCs and were termed MHCs. There are two types, which differ in structure and function:

1. MHC class I—expressed on all cells, present to $CD8^+$ T cells
2. MHC class II—expressed on selected cells such as APCs (Fig. 6.8), present to $CD4^+$ T cells.

In humans, MHC antigens are called human leukocyte antigens (HLAs). HLAs are functionally and structurally similar but have a largely increased variability. We will first discuss antigen presentation and then HLA structure and variability.

MHC class I

MHC class I molecules are found on all cells and continuously present peptides broken down in the cytoplasm of the cell. This is a way of constantly proving to the immune system that it is a healthy cell (Fig. 6.9). The cytoplasmic proteosome degrades cytoplasmic proteins into peptides. These peptides are transported into the endoplasmic reticulum, where they are bound to MHC class I receptors, which then travel in vesicles to the cell surface.

However, if the cell becomes infected by a virus or intracellular pathogen, or becomes a tumour, it will start presenting foreign peptides on MHC I molecules, stimulating CD8 cells to destroy it.

Cross presentation

Cross presentation is the presentation of exogenous material on an MHC I receptor (which usually presents endogenous material). This occurs when an APC takes up a cell that contains foreign antigen, e.g. a virally infected cell. The material is broken down in an acidified phagolysosome and presented on an MHC II receptor, but some debris peptide material enters the cytoplasm. This is then presented on MHC I molecules and activates $CD8^+$ T cells, which then respond towards other cells with the same MHC–peptide complex. This increases the effectiveness of antiviral responses. It is also, unfortunately, a mechanism of graft rejection.

MHC class II

MHC class II molecules are presented by specialized APCs, which take up exogenous pathogens, phagocytose them, and present the peptide products (Fig. 6.10). The pathogen or macrophage–pathogen complex is digested in acidified endocytic vesicles. MHC class II molecules then bind with the peptide, travel to the cell surface, and the antigen/MHC is presented to T lymphocytes.

Human leukocyte antigen

The MHC molecule, or HLA as it is known in humans, binds peptide fragments from both pathogens and self, and presents them on cell surfaces to initiate immune responses. This results in the destruction of pathogens and the protection of normal healthy cells.

The number of different peptide combinations a single MHC molecule can present is limited. Structurally the MHC molecule has a 'pocket' in which the antigen sits. Certain peptides will have a high affinity for this 'pocket', whilst others will not be able to bind because of the incompatibility of their shapes at a molecular level. To maximize immunity the MHC molecule has developed to be:

1. polygenic—there are several different MHC class I and class II genes
2. polymorphic—there are multiple variants of each gene, increasing the potential number of peptide combinations that can be presented.

Genes for the HLA are located on various chromosomes (6, 15) and contain more than 200 genes. The HLA gene complexes have high variability (Fig. 6.11):

- three class I α genes—HLA A, B, and C
- three class II α and β genes—HLA DR, DP, and DQ.

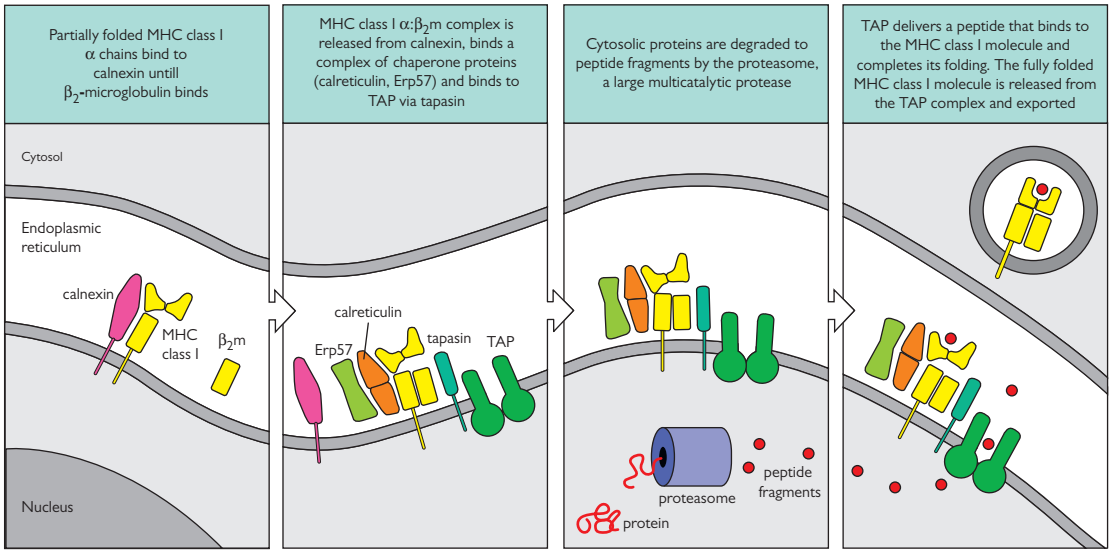


Fig. 6.9 MHC class I turnover.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 5.5, 2008, with permission from Garland Science.

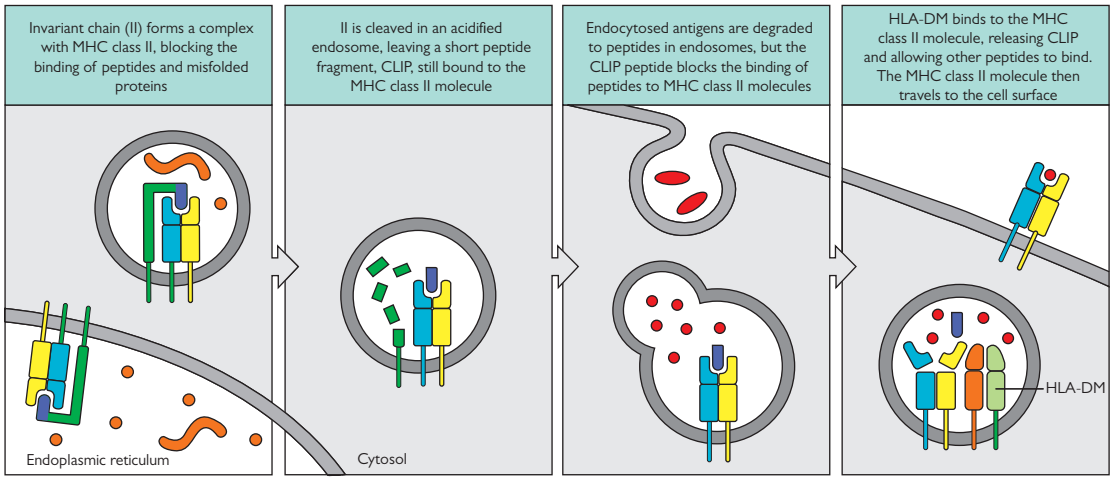


Fig. 6.10 MHC class II on APCs.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 5.11, 2008, with permission from Garland Science.

Each human expresses at least three different MHC class I and three MHC class II genes, so an individual often expresses six different MHC class I molecules and eight MHC class II molecules simultaneously, vastly increasing the number of different types of peptides they can present.

Alloreactivity

This is the recognition of non-self tissue from the same species (i.e. allogeneic) as foreign. There are two methods of alloreactivity:

1. recognition of foreign HLA molecule
2. recognition of the presented tissue-derived peptide as pathogenic.

This is the basis of the immune rejection of allografts and explains the importance of HLA matching.

Antigen presentation

Antigen presentation is carried out by specialized cells with the ability to process and present antigen on MHC II protein.

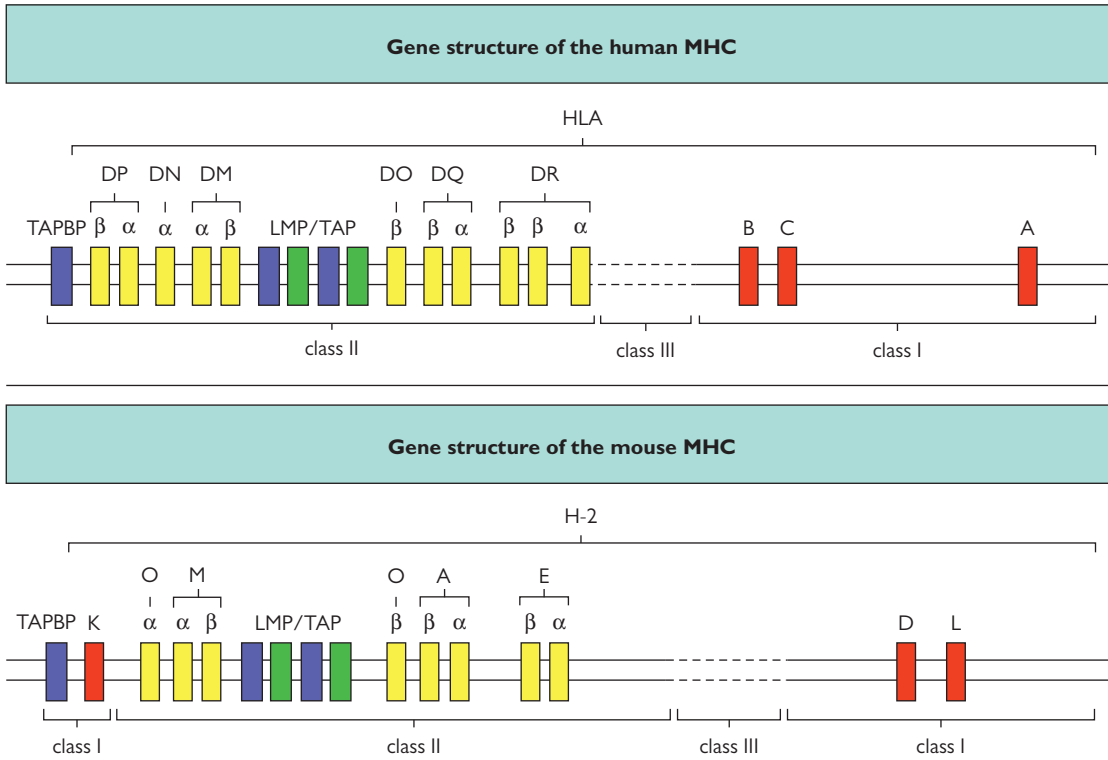


Fig. 6.11 HLA ABCD.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 5.12, 2008, with permission from Garland Science.

These cells are called APCs for short. They can be divided into classical APCs and non-classical APCs.

Classical APCs

The three classical APCs are:

1. dendritic cells (DCs)
2. macrophages
3. B cells.

DCs are the most important APC. DCs are named after their long finger-like processes, like the dendrites of nerve cells. They are constantly ingesting surrounding material through macropinocytosis. They then degrade the pathogen in acidic vesicles and mount the peptides on MHC II molecules, as discussed before. There are two main types of DC:

- myeloid DCs (DC1) produce (IL-12) and direct Th1 responses
- plasmacytoid DCs (DC2) produce type 1 interferon (IFN) and induce a Th2 response.

It is currently believed that human macrophages exist in distinct effector subpopulations, and these are the focus of intense investigation.

Effector T cells

Effector T cells can be divided into two classes, $CD8^+$ T cells and $CD4^+$ T cells, and the cells can be distinguished phenotypically, using antibodies to bind specifically to T-cell surface molecules. T cells are identified as being either $CD3^+$ $CD8^+$ (for $CD8^+$ T cells) or $CD3^+$ $CD4^+$ (for $CD4^+$ T cells).

The key difference between the two classes of T cell is in their recognition of antigen:

- $CD8^+$ T cells recognize antigen presented by MHC class I molecules
- $CD4^+$ T cells recognize antigen presented by MHC class II molecules.

Whilst the distinction between CD8 and CD4 T cells is based on phenotype, this does not provide information as to the function of these two types of T cell.

CD8+ T-cell funtction

CD8+ T cells are mainly cytotoxic T cells (Tc). Tcs can kill infected cells but leave adjacent non-infected cells intact, and this is a highly antigen-specific T-cell response (Fig. 6.12). Tc responses are particularly effective against virally infected cells and tumour cells, and require CD4+ T-cell help in most responses.

T cells are cytotoxic by inducing apoptosis in target cells. This is a very rapid response (within 5 minutes) caused by the release of preformed enzymes stored in lytic granules within the Tc.

The cytotoxic granules within Tc contain:

- membrane-disrupting proteins, e.g. perforin and granulysin
- serine proteases, e.g. granzymes A–H, K, and M: only forms A, B, and C are highly expressed in Tc, but other granzymes may be found in NK cells
- lysosomal enzymes, e.g. cathepsins
- stored effector molecules, e.g. Fas ligand (FasL), which binds to Fas-expressing cells to induce apoptosis in the target cells through a caspase pathway.

Granzymes exist in several forms:

- granzyme A disrupts mitochondrial membrane potential and induces caspase-independent cell death
- granzyme B cuts after aspartate residues and induces both caspase-dependent and caspase-independent cell death
- granzyme C is similar to A and induces caspase-independent death:
  - cathepsin C processes granzymes from inactive to active state

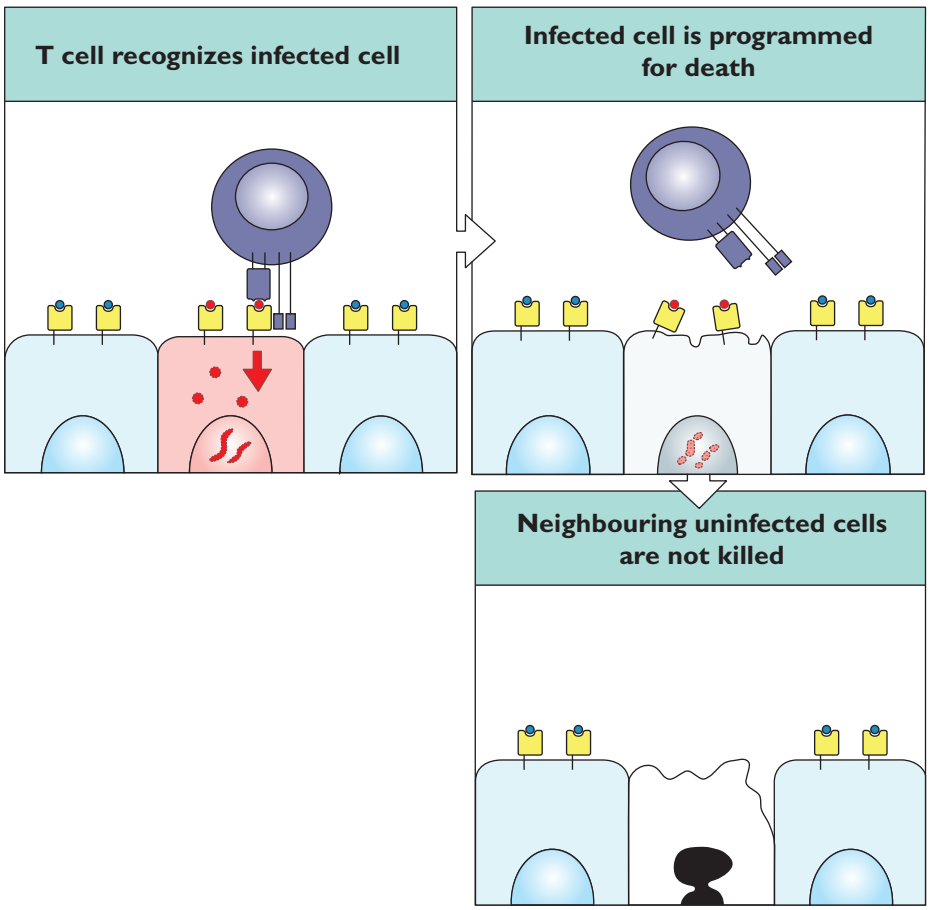


Fig. 6.12 CD8+ T-cell killing.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 8.39, 2001, with permission from Garland Science.

– cathepsin B protects Tc from perforin damage
– granule-induced cell death would appear to be the primary pathway.

CD4+ T-cell function

CD4+ T cells were originally referred to as 'helper' T cells (Th) since their main function is to help other immune cells (dendritic cells, B cells, macrophages, Tc) respond by secreting cytokines to upregulate responses.

Effector CD4+ T cells have been shown to exist in functional subsets, depending on their profiles of cytokine production:

- Th1 cells mainly secrete IFN-γ and IL-12
- Th2 cells secrete IL-4, IL-5, IL-6, and IL-13
- Th17 cells secrete IL-17
- regulatory T cells (Treg) secrete TGFβ and/or IL-10.

The different subsets of effector CD4+ T cells are affected by their cytokine environment (Fig. 6.13), the antigen dose, co-stimulatory molecules, and the site and type of infection.

Th1 cells produce IFN-γ, which activates macrophages, enhancing antigen presentation to T cells and increasing phagocytosis. Th1 cells also produce IL-12 to drive DC activation and presentation. Macrophage activation leads to the induction of several molecules, including nitrous oxide (NO) and cytokines. Since NO is an antimicrobial factor, activated macrophages have potent antibacterial and anti-protozoal activity.

When effector Th1 cells contact an infected macrophage, two signals are required for macrophage activation:

– T-cell secretion of IFN-γ
– CD40L/CD40 co-stimulatory molecule interaction, which delivers a sensitizing signal.

Increased expression of CD40 leads to an amplification of the response.

Th2 cells provide help to B cells through production of IL-4 and IL-6, and produce cytokines that stimulate antibody production and mucous secretion (IL-13). Th2 cells:

- do not activate macrophages
- produce IL-4 rather than IFN-γ
- do not express CD40L
- are particularly involved in allergy and parasitic infections.

Th17 cells have been identified by their production of IL-17, a potent neutrophil chemoattractant. This subset of effector CD4+ T cells can be pathogenic, as it has been demonstrated

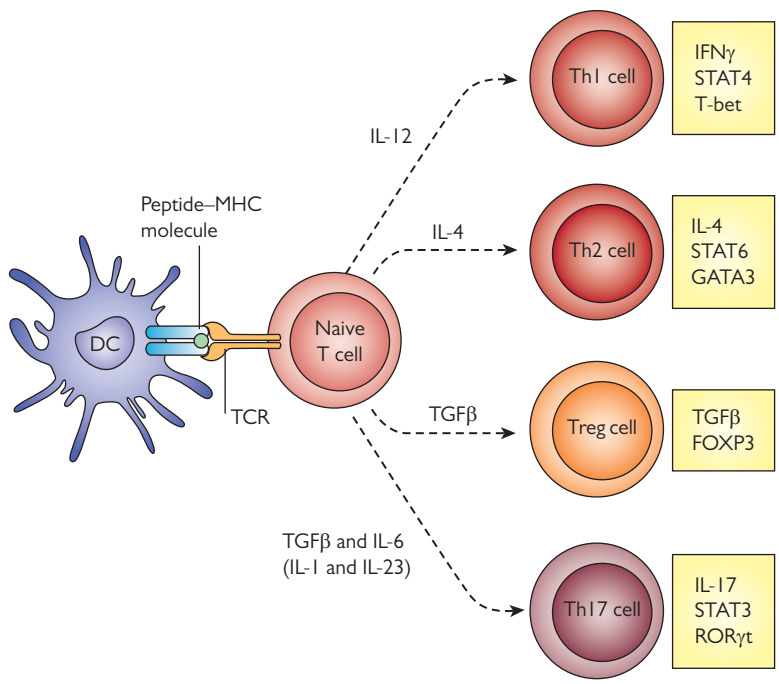


Fig. 6.13 Differentiation of helper CD4+ T-cell subsets.

to be able to induce experimental autoimmune diseases (experimental autoimmune encephalomyelitis and experimental autoimmune uveitis). Th17 cells:

- are effective in Th1-deficient hosts
- are driven by the cytokines TGF-β and IL-6
- become activated via TLR.

Treg cells are suppressive for T-cell responses, mainly via their secretion of anti-inflammatory cytokines. Treg cells exist in two forms:

- naturally occurring (nTreg)
- inducible Treg (iTreg).

The nTreg are thymus-derived, $CD4^+$ $CD25^+$ $FoxP3^+$ Treg and represent 5–10% of the peripheral $CD4^+$ T cells in humans. The iTreg include Th3 and Tr1 subsets, are $CD4^+$ CD25-Treg, and produce both Th1 and Th2 cytokines at low levels as well as high levels of IL-10 and TGF-β.

Other subsets

There are other subsets of effector $CD4^+$ T cells which have been identified and are under investigation, including Th9 and Th22. Their roles in human immune responses and in diseases affecting humans remain unclear.

Diversification of T-cell lineages

Effector $CD4^+$ T cells exist in several different subsets and have been demonstrated to be able to switch from one subset to another, depending on the cytokine environment in which they are found.

Cytokines

The term 'cytokine' is used to describe a cell-derived soluble mediator that can be secreted by a wide range of cells or by a few cell types. IL-1, IL-6, and TNF-α are all produced by a variety of different cell types during immune responses. In contrast, IFN-γ is produced predominantly by an activated effector $CD4^+$ T-cell subset (Th1, see *Acquired immunity*, p. 174).

Cytokines are produced by cells either as part of their normal turnover or by upregulation during both innate and adaptive immune responses. Pro-inflammatory cytokines describe a group of cytokines that promote inflammation by:

- activating cells directly
- upregulating the expression of proinflammatory molecules
- attracting other cells to the site of inflammation.

However, under certain situations, cytokines can also act as anti-inflammatory (regulatory) cytokines to downregulate immune responses by:

- suppressing immune cell responses (IL-10 and TGF-β both downregulate $CD4^+$ T-cell responses)
- antagonizing the function of another cytokine (IL-4 antagonizes the function of IFN-γ).

Although cytokines can exert many different effects, their function is highly dependent on:

- the type and local concentration of cytokine
- the local cytokine milieu
- the nature of the responding cell
- the levels of receptors expressed for that cytokine.

Cytokines can be:

- cytotoxic, e.g. TNF-α on epithelial cells
- growth factors, e.g. IL-4 for B cells
- synergistic with other cytokines to enhance responses, e.g. IL-1 augments the effects of IL-8
- clonally expanding cells, e.g. IL-2 promotes expansion of activated $CD4^+$ T cell and stem cell factor is a growth factor for mast cells
- cell attractants/chemokines, e.g. IL-8 attracts neutrophils and eotaxin attracts eosinophils.

Cytokines that are secreted during innate immune responses are usually preformed and stored within cells ready to be secreted immediately on activation of the cell. An example of this type is histamine, which is stored within mast cells, ready to be secreted rapidly to exert its biological functions. In contrast, B and T cells do not store cytokines and production of their cytokines requires *de novo* protein synthesis following activation.

Tumour necrosis factor α

There are two forms of TNF-α that exist: soluble and membrane-bound forms. This is a pro-inflammatory cytokine which plays a prominent role in inflammation. Of its many effects, it has been shown to:

- activate polymorphonuclear leukocytes
- be cytotoxic for target cells
- possess antiviral properties
- be an effective anticoagulant
- modulate haematopoiesis
- mediate several inflammatory effects.

There are two receptors for TNF-α (p55 and p75), which can be cleaved off the surface of cells.

IL-1

IL-1 was one of the first cytokines to be characterized and is produced in response to infection to raise body temperature

(pyrogen) in an attempt to kill the infection. It is produced by macrophages as well as many other cell types. It exists in two forms: IL-1α and IL-1β.

IL-2

IL-2 (originally named T-cell growth factor) is a key cytokine involved in T-cell responses. Upon activation, T cells transiently express IL-2 receptors, allowing them to bind to IL-2, which induces a clonal expansion of that activated T cell. Hence IL-2 is important in amplifying an antigen-specific T-cell response.

IL-4

IL-4 is a Th2 cytokine secreted by mast cells and Th2 T cells. It activates B cells to differentiate for the process of antibody production.

IL-10

Although it is secreted by a variety of different cell types, IL-10 possesses both pro-inflammatory and anti-inflammatory properties, depending on the target cells. IL-10 can promote B-cell responses but is downregulatory for human T-cell responses, APC function, and the production of pro-inflammatory cytokines, and can be secreted by Treg.

Transforming growth factor β

There are three isoforms of TGF β (TGFβ-1, -2, and -3), which are secreted by a variety of cell types. TGFβ can be pro-inflammatory by inducing fibrogenesis (fibroblast activation and collagen formation) and TGFβ-1 is suppressive for T-cell responses and is produced at high levels by Treg.

Chemokines

There are several cytokines that are involved in chemotaxis or signalling for directed migration. These molecules are still part of the cytokine superfamily, but have been classified as chemokines based on their function. They are listed in one of three groups depending on their molecular structure:

1. CXC chemokines, e.g. IL-8 (CXCL8)
2. CC chemokines, e.g. MCP-1, MCP-2, macrophage inflammatory protein-1α (MIP-1α), RANTES, eotaxin
3. C chemokines, e.g. XCL-1, XCL-2.

IL-8

One of the earliest chemokines to be described is CXCL8 (IL-8), a potent chemoattractant for neutrophils that also plays a role in angiogenesis. IL-8 is secreted by monocytes, macrophages, T cells, epithelial cells, and several other cell types. Cells known to respond to IL-8 include neutrophils, T cells, eosinophils, and keratinocytes.

Interferons

This is a group of cytokines that were originally described as antiviral agents, found within cells. These can act as pro-inflammatory or immunomodulatory cytokines.

Type 1 interferon (IFNα/β) can be used therapeutically in chronic hepatitis (IFNα) and multiple sclerosis (IFNβ).

Type 2 (immune) interferon (IFNy) upregulates expression of MHC class II (HLA-DR) on cells. It is secreted by Th1 cells. IFNy activates macrophages and induces NK cell activity.

Effector B cells

B lymphocytes (B cells) are derived from haematopoietic stem cells in the bone marrow and mature in the bone marrow. During maturation, B cells, like T cells, undergo gene rearrangement and receptor testing for autoantigen specificity. If the B cell is found to be autoreactive, it either undergoes receptor editing, again mediated by the RAG genes, or is deleted in the process of clonal deletion.

Upon activation, B cells mature into plasma cells and memory cells. The main function of plasma cells is to produce antibodies. Antibodies were the first components of the innate immune system to be discovered. They are antigen-specific protein structures that circulate in the bloodstream.

Their function is two-fold:

1. to bind specifically to a pathogen—antibodies disable viruses
2. to opsonize pathogens—antibodies enhance phagocytosis.

Antibody structure

There are five classes of antibody: IgM, IgG, IgD, IgE, and IgA. They have slightly different properties and consequently are produced in different areas of the body. All classes of antibody comprise single monomeric structures (Fig. 6.14) that are:

- approximately Y-shaped
- made from approximately three equally sized portions connected by a flexible hinge
- made up of two heavy chains and two light chains
- composed of variable (V) regions at the two distal arms of the Y which are involved in binding to antigen
- composed of a constant (C) region that is not associated with antigen binding—its function is to activate effector cells and complement.

Immunoglobulins can also be cleaved with the enzyme papain into two molecules:

- Fab fragments—antigen binding and contain the V regions
- Fc fragments—crystallizable and contain the C region.

Antibodies are structurally homologous to the B-cell receptor of the cell from which they are secreted.

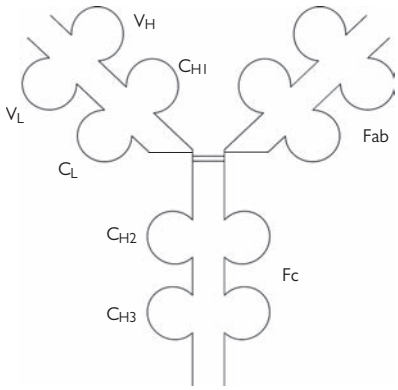
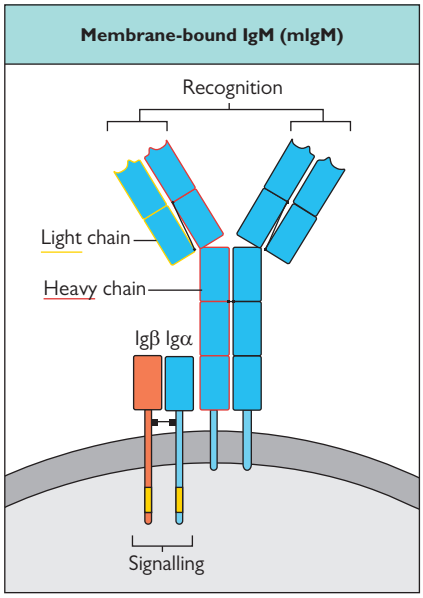


Fig. 6.14 Antibody structure.

Reproduced from R. Wilkins et al., *Oxford Handbook of Medical Sciences*, *The Essential Guide to the Sciences that Underpin Medicine*, Figure 12.6, Page 833, 2011, with permission from Oxford University Press.



B-cell activation

There are two ways to activate B cells:

1. the B cell takes up foreign material and presents it on MHC II antigens. An activated Th2 cell specific for that antigen recognizes the presented peptide. It then releases IL-4, which activates the B cell (Fig. 6.15)

or

2. by certain unprocessed antigens, if present in large amounts on, for example, the cell wall of bacteria. As these are not presented by a T lymphocyte, they are called thymus-independent (TI) antigens. For the B cell to become activated there must be significant crosslinking of B-cell receptors.

B cells recognize antigens either in three-dimensional conformations of native proteins or non-protein molecules, or in three-dimensional conformations of degraded or denatured antigens.

The dependence of B cells on T cells for activation is a protective mechanism. It prevents autoimmune B cells from becoming activated. There must be Th2 cells with the same antigen specificity to activate them. This interaction between T cells and B cells takes place in lymphoid tissues, e.g. the spleen and lymph nodes. Once activated, the B cell then divides into plasma blasts, which become plasma cells.

As the B cell divides it undergoes V region somatic hypermutation; often this is just one amino acid and has no improvement to affinity of the B-cell receptor to the antigen. The new B cell must show improved affinity to the antigen presented or it will lack co-stimulation and stimulation by IL-4 and will duly undergo apoptosis. Increased affinity will result in proliferation of the B cell and production of the new antibody. This process is called affinity maturation.

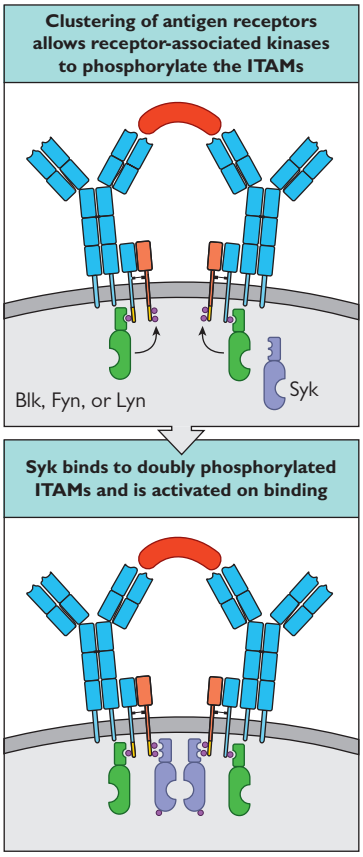


Fig. 6.15 B-cell activation.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 6.9, Figure 6.23, 2008, with permission from Garland Science.

Class switching

IgM is the first antibody produced by a plasma cell. It has a relatively low antigen affinity, but IgM exists in a pentameric form (five monomers), thus increasing its overall avidity. It is especially effective in defending against bacteria and in activating the complement system. However, because of its larger size (as a pentamer), it does not diffuse into the extravascular space particularly well, and therefore the body must produce other types of antibody.

Tolerance is the process that results in the immune system *not* responding to antigens. The tolerance achieved by clonal deletion is described as central tolerance. Sometimes autoreactive B cells may have escaped clonal deletion.

Transplantation immunology

The transplantation of organs or tissue from one organism to another is now a standard therapy in the treatment of various diseases and conditions. The main categories of transplants or grafts are:

- autografts—grafts from one organism to the same organism, e.g. a skin graft from the same individual
- allografts—grafts from one organism to another in the same species, e.g. kidney transplant
- xenografts—grafts from a different species, e.g. heart valve transplants from pig to human.

The main reason for graft failure is rejection by the immune system. This means the immune system recognizes the graft as foreign material and mounts an immune response, causing destruction of the tissue and failure of the graft. This does not occur with autografts as the tissue is recognized as self. However, in allografts, as the tissue presents foreign HLA molecules and foreign peptide in the HLA molecules, the graft can be recognized as foreign and destroyed. The HLA molecules themselves are referred to as major histocompatibility antigen. The antigen peptides they present are referred to as minor histocompatibility antigen.

Tissue compatibility matching is the single most effective way to increase graft survival rates. This means finding as close an HLA match as possible. This is very difficult because of the high polymorphism and its polygenic trait in the HLA code. Identical twins obviously provide a genetically identical tissue bank, but even non-identical siblings can prove to be non-compatible. For this reason finding a complete match across all the HLA genes can be difficult. Even if this were possible through a family donor the presentation of endogenous peptide as minor histocompatibility antigen can lead to rejection.

Tissue typing is carried out for HLA A, B, and DR loci. It is particularly important to match the DR loci as mismatch here increases the risk of graft rejection significantly. Well-matched transplants have 15% higher survival at 1 year. Some patients are particularly sensitized and therefore more likely to reject foreign tissue, e.g. women (due to pregnancy) and individuals who have previously received multiple blood transfusions.

Graft rejection

Corneal transplantation is one of the most common transplant operations worldwide and has an extremely high success rate. This is largely because the cornea anatomically is avascular and sits over the anterior chamber, an immunologically privileged site.

Immune privilege refers to areas that are excluded from normal immune system function and contain autoantigen that the immune system is not exposed to. Immune privileged areas differ from other tissues/sites in three main ways:

1. The immune communication with the rest of the body is limited as they are not supplied with lymphatic supply.
2. There is a high level of anti-inflammatory cytokines such as TGF-β.
3. There is an increased expression of FasL, which induces apoptosis of infiltrating Fas-expressing activated lymphocytes.

These privileged areas include the eyes, the brain, the gonads, the foetus, and the placenta.

Graft rejection is known as host-versus-graft disease. Graft-versus-host disease (GVHD) is the opposite of this and arises when the host tissue contains immune system cells that recognize the host as foreign. This is a severe condition that can cause a severe inflammatory response, including rashes, diarrhoea, and liver disease.

GVHD can be tested for before transplant. Lymphocytes from the donor are mixed with irradiated donors from the host. If the potential donor lymphocytes begin to replicate, it is clear that there are potentially alloreactive responses; the donor would then be discounted for the transplant operation. This is called a mixed lymphocyte reaction test and is especially important for bone marrow transplants.

Naïve alloreactive T cells are activated through two different methods:

1. Direct allorecognition—graft dendritic cells present the foreign MHC and co-stimulatory molecules to host T cells, thus initiating an alloreactive reaction.
2. Indirect allorecognition—uptake of foreign antigen by host dendritic cells and presentation to naïve alloreactive T cells.

Presentation of alloantigens to T cells in the context of self class I or II MHC molecules is called the indirect pathway of allorecognition. This is the only pathway available for presentation of minor transplantation antigens when donor and recipient share no MHC-encoded class I or II molecules. This pathway also accounts for the presentation of peptides derived from MHC alloantigens.

By contrast, intact MHC alloantigens can be recognized directly by T cells with a different set of receptors for antigen, and this is called the 'direct' pathway of allorecognition. In solid tissue allografts, numerous class-II-expressing,

bone-marrow-derived cells of the dendritic and macrophage type are present, and these cells, termed passenger leukocytes, are primarily responsible for the activation of alloreactive T cells through the direct pathway.

The absence of passenger cells within the normal cornea correlates very well with the relative inability of orthotopic corneal allografts to activate direct alloreactive T cells. Because corneal allografts activate and may eventually succumb to the effector function of indirect alloreactive T cells, APCs other than passenger leukocytes, that is, of recipient origin, must be responsible for initial T-cell activation.

Types of rejection

There are four types of rejection:

- Hyperacute: circulating antibodies are present to antigens on the graft tissue within minutes. If the recipient is sensitized to donor antigen, graft destruction via cell-mediated mechanisms will commence. Rejection occurs within 2–5 days.
- Acute: mainly a cell-mediated response. Occurs at 7–21 days.
- Chronic: occurs after 3 months. A disturbance in graft tolerance leads to rejection. Antibodies and complement are the main mediators.
- Acute on chronic: T cells are the instigators. This occurs if the recipient's immune system has been altered, e.g. immunosuppression.

Autoimmunity

Autoimmunity is the recognition of self-antigen by a person's immune system. When this results in tissue damage it is called an autoimmune disease.

Genetic rearrangement in lymphocyte development is an indiscriminate process and will lead to the production of some autoantigen-specific cells. These should be removed in the thymus and bone marrow through testing, genetic rearrangement, and clonal deletion (as discussed before). Sometimes these processes fail and potentially autoreactive cells are released into the periphery. These cells are most likely to become tolerized to the self-antigen as excessive presentation to autoreactive T cell will lead to anergy, especially in the absence of co-stimulatory molecules, which are rarely expressed in the absence of infection.

Nevertheless there is a potential for autoimmune disease. Excluding rheumatoid arthritis and thyroiditis, autoimmune diseases are not common. However, when combined, they affect approximately 5% of Western populations, often with severely debilitating diseases. The causes of autoimmunity are still being researched, but there are several theories on the loss of tolerance.

- Molecular mimicry—the antigen of a foreign pathogen that creates an immune response has a very similar molecular structure to a self-antigen. The body then mounts an immune response against itself.
- Epitope spread—unactivated autoimmune cells become activated in the presence of an inflammatory response because of the high levels of inflammatory cytokines and APCs. The autoantigen may be structurally different to the original antigen.
- Cytokine dysregulation—low levels of anti-inflammatory cytokine production.
- Failure of deletion—failure of the thymus or bone marrow to delete.

Autoimmune diseases can be broadly divided into organ-specific and systemic.

Organ specific

- Limited number of autoantigens.
- Autoantigens from one organ.
- Examples: diabetes mellitus type 1, Graves's disease.

Systemic specific

- Affect multiple organs.
- Chronic.
- Examples: systemic lupus erythematosis, rheumatoid arthritis.

Immunodeficiencies and immunosuppression

Immunodeficiencies are diseases where one or more of the mechanisms of the immune system are not functional. They are either primary immunodeficiencies or secondary immunodeficiencies:

- Primary immunodeficiencies—the product of mutations in any of a large number of genes that are involved in or control immune responses, resulting in one or more defective elements.
- Secondary immunodeficiencies—acquired immunodeficiencies, caused by other diseases, malnutrition, or iatrogenic causes.

Primary immunodeficiencies are caused by genetic defects. They can occur in any part of the immune system. For example, X-linked agammaglobulinaemia results in a lack of B cells, making the sufferer susceptible to extracellular bacteria and viruses. Most of the genetic mutations are recessive, and

many are sex chromosome linked. This leads to a higher prevalence of immunodeficiencies in males as they only have one X chromosome.

Severe combined immunodeficiency syndromes are very serious immunodeficiencies that can lead to patients suffering from a lack of specific humoral or cell-mediated immune responses; in addition they are unable to develop immunological memory. Without treatment patients are likely to suffer fatal infections. Management is with bone marrow transplant or gene therapy; both are risky, with mixed results.

Secondary immunodeficiencies are deficiencies caused by pathogens or malnutrition, the most significant of which is acquired immune deficiency syndrome (AIDS) as a result of infection with HIV. AIDS is characterized by susceptibility to opportunistic pathogens and rare cancers, mediated by a significant decrease in $CD4^+$ T cells.

HIV is a retrovirus that infects $CD4^+$ T cells, DC, and macrophages, leading to their destruction and a defective immune system (Fig. 6.16). HIV can be broadly divided into two subtypes, HIV-1 and HIV-2, although there is a possibility of more HIV subtypes. Initial infection is characterized by an influenza-like illness, which is accompanied by a massive increase in viral load and a decrease in circulating CD4 cells. This is associated with the production of antibodies and is called seroconversion.

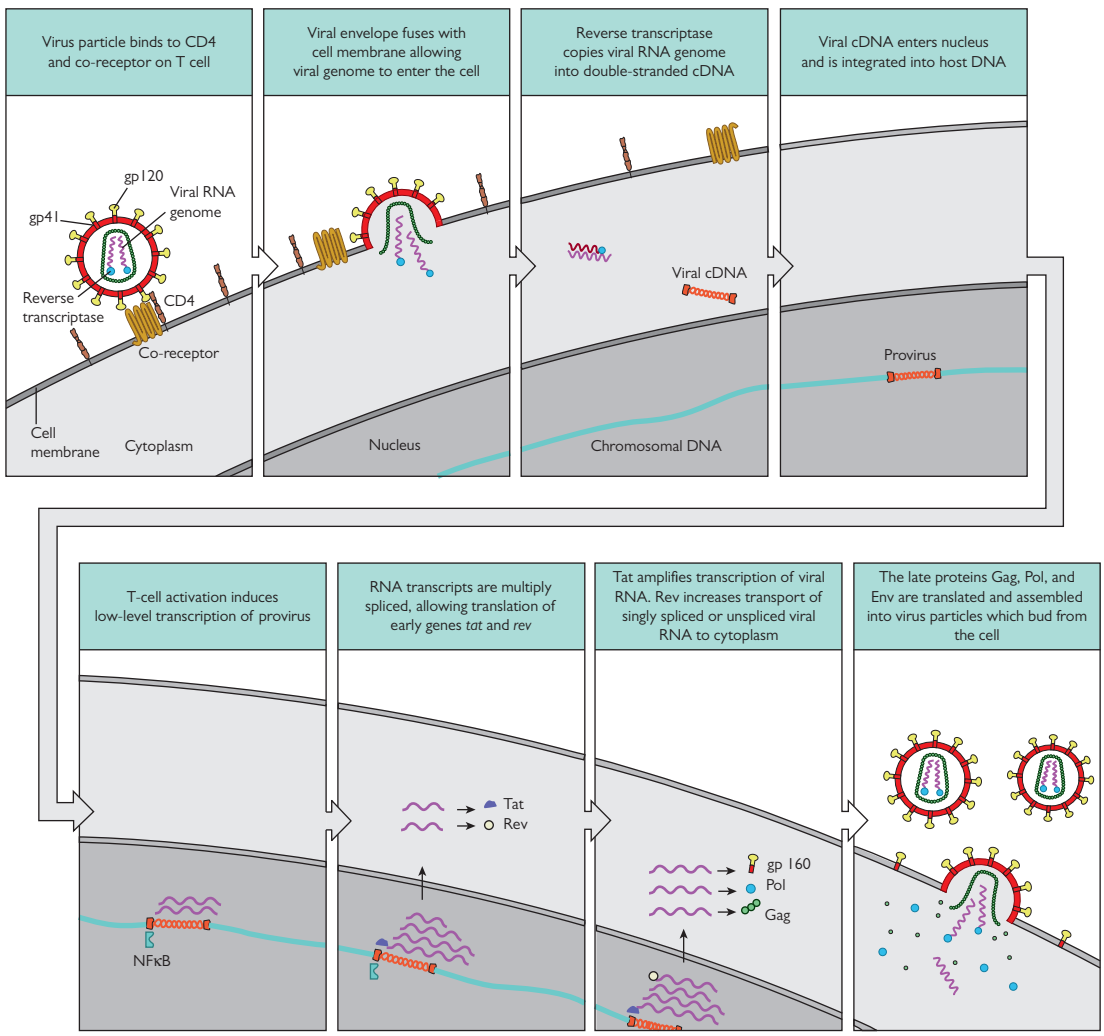


Fig. 6.16 HIV infecting a $CD4^+$ T cell.

Reproduced from Murphy KP, *Janeway's Immunobiology*, 7th edition, Figure 12.23, 2008, with permission from Garland Science.

CLINICAL TIP

HIV in the eye is a cause of some rare and opportunistic conditions, the most common being CMV retinitis. Therapy for HIV has become common in the West, most notably HAART. This is a combination of drugs, including reverse transcriptase inhibitors and viral protease inhibitors, that has been shown to significantly reduce mortality and morbidity in patients with advanced HIV infection.

There then follows a clinically asymptomatic period, where numbers of $CD4^+$ cells gradually reduce. Eventually patients have so few $CD4^+$ T cells left that they are unable to mount reasonable responses and now have AIDS. This can take from 6 months to 20 years, but eventually will occur in all HIV sufferers.

Immunosuppression is a reduction of the function of the immune system. It may be caused by infection (HIV) or may be iatrogenic. In the case of iatrogenic causes, it may be intentional or a side effect. For example, long-term steroid therapy (for treatment of inflammatory diseases) may leave the body immunosuppressed, increasing the risk of infections. The necessary use of immunosuppressants during transplantation, for example, increases the risk of opportunistic infections taking place. It is often therefore common to screen for viruses that may be dormant before embarking on transplant surgery.

Host defence and the eye

Innate defence

There are several innate immune mechanisms that provide protection at the ocular surface. Physically, the front of the eye is shielded by eyelids and eyelashes, and the ocular surface is bathed in tears to wash away potential microbes, preventing them from adhering to cells. More specifically the tear fluid contains mucins, which readily bind antigens.

Adaptive defence

Should an antigen successfully persist at the ocular surface, IgA is secreted within tear fluids and readily binds antigens and aids in their removal. The IgA exists in a dimeric form, which comprises two monomers joined together with a joining (J) sequence, and this molecule is protected from enzyme degradation by the presence of a secretory (S) piece. There are several cell types found at the ocular surface that are well placed to respond to antigens as part of a mucosal immunity:

- conjunctival epithelial cells
- gamma/delta T cells
- Langerhans' cells

Other posterior sections of the eye also possess an ability to respond to antigen, including retinal microglia and RPE cells, both of which are able to act as non-professional APCs. However, it is undesirable for immune responses to occur in the highly differentiated cells of the cornea, anterior segment, and retina since inappropriate or excessive activation of various immune responses, including complement, and antibody- and T-cell-mediated responses, could result in unwanted damage to the tissues. These parts of the eye therefore maintain a state of immunological tolerance, or immune privilege, which downregulates immune responses (see *Transplantation immunology*, p. 184). Although originally thought to be a form of immune ignorance, immune privilege is now understood to involve several cells and molecules that contribute to this active state of tolerance. The main characteristics of immune privilege are:

- a low level of leukocyte recruitment and traffic
- no professional APCs
- an environment that is immunologically downregulatory
- no conventional lymphatics.

The retina is protected from the systemic blood circulation by a selectively semi-permeable blood–retinal barrier, which is composed of a layer of endothelial cells formed by tight junctions, and this selectively allows cells and molecules to enter from the bloodstream into the retina, thus minimizing unwanted immune responses. The aqueous fluid contains anti-inflammatory molecules which are downregulatory for T-cell responses, including

- TGF-β
- vasointestinal protein (VIP)
- calcitonin gene-related protein
- α-melanocyte stimulating hormone
- somatostatin
- macrophage migration inhibitory factor
- soluble FasL.

Tissue-resident cells within these immune privileged sites express several molecules, such as FasL, which binds to Fas on activated immune cells to induce apoptosis. Hence any activated T cell that managed to cross the blood–retinal barrier could then be rapidly removed via FasL before it has time to secrete any molecules that could damage the local tissue cells.

Despite the many systems in place to protect the eye from both invading organisms and immune responses, inflammation does occur at the ocular surface and within the eye, often due to infections. Furthermore, as a result of an inflammatory response that has persisted due to either a lack of regulatory mechanisms or a failure to clear the

infection, immune-mediated tissue damage can occur. This can result in:

- uveitis, inflammation within the uvea which can result in loss of retinal photoreceptor cells
- corneal plaque formation, which can occur during chronic allergic conjunctivitis
- conjunctival scarring due to a fibrotic response following infection or inflammation.

CHAPTER 7

Growth and senescence

Ocular embryology

The human body develops in a complex sequence, each step facilitating the next and other parallel ongoing processes. Mastering ocular embryology requires an understanding of the terminology used and how congenital anomalies manifest themselves.

Basic embryology

A fertilized ovum is also called a zygote. Two identical cells termed blastomeres are produced by mitosis and undergo further embryonic cleavage.

By day 3, the 16-cell mass of identical totipotent blastomeres is called a morula.

Between days 4 and 5, the outer blastomeres of the morula thin to form the outer trophoblast layer with the inner blastomeres compacting to form the embryoblast. The trophoblast develops into the future placenta and the embryoblast, the embryo, surrounded by a fluid-filled space called the blastocoele. Implantation of the whole structure called a blastocyst in the uterine wall occurs on day 6.

The embryoblast then remodels into a bilaminar disc with the columnar epiblast layer outermost and the cuboidal hypoblast layer innermost. The space between the epiblast layer and trophoblast wall develops into the amniotic cavity.

By the third week, a narrow groove called the primitive streak is formed by adjacent epiblast cells proliferating on the caudal section of the midline. A pit at the rostral end of the primitive streak is called the primitive node. The primitive streak axially orientates the developing embryo.

HELPFUL HINT

Mesenchyme

Epithelial–mesenchymal transformation is a mechanism producing highly mobile multipotent cells called mesenchymes, which migrate and differentiate. Primary mesenchymal cells migrate from the primitive streak and differentiate into mesoderm and endoderm. Mesoderm further differentiates into definitive mesenchyme, which has muscle and connective tissue fates. Neural crest cells also produce mesenchyme.

The change from a bilayer to a three-layer embryonic disc is called gastrulation and occurs on day 14. Cells of the primitive streak migrate under the epiblast and differentiate into mesoderm, which fills the space between the ectoderm and endoderm—formerly known as the epiblast and hypoblast layers, respectively. Gastrulation is a key step because after this process is the earliest point that any tissue, including ocular, can be traced back meaningfully to a specific subset of cells.

HELPFUL HINT

Origins of ocular tissue

Neural ectoderm (optic cup)
- Neural retina
- Retinal pigment epithelium
- Pupillary sphincter and dilator muscles
- Posterior iris epithelium
- Ciliary body epithelium
- Optic nerve

Neural crest (connective tissue)
- Corneal endothelium
- Trabecular meshwork
- Stroma of cornea, iris, and ciliary body
- Ciliary muscle
- Choroid and sclera
- Perivascular connective tissue and smooth muscle cells
- Meninges of optic nerve
- Orbital cartilage and bone
- Connective tissue of the extrinsic ocular muscles
- Secondary vitreous
- Zonules

Surface ectoderm (epithelium)
- Corneal and conjunctival epithelium
- Lens
- Lacrimal gland
- Eyelid epidermis
- Eyelid cilia

- Epithelium of adnexa glands
- Epithelium of nasolacrimal duct

Mesoderm (muscle and vascular endothelium)

- Extraocular muscle cells
- Vascular endothelia
- Schlemm's canal endothelium
- Blood

Simultaneously, the notochordal process extends the primitive node, tunnelling rostrally. The lumen eventually collapses, forming the basis of the axial skeleton. The neural plate is a region of thickened ectoderm overlying the rostral embryonic disc.

As the neural plate invaginates, the ectoderm in regions known as the neural crests on either side of the neural folds begins to fuse (day 22) to form the upper plate of the neural tube. Fusion starts at the fourth somite and proceeds rostrally and caudally. The anterior neuropore closes at day 25 and the posterior neuropore on day 27.

Somites are the paired segments of mesoderm arranged adjacent to the neural tube which subsequently differentiate into myotome (muscle), sclerotome (bone and cartilage), and dermatome (dermis).

Formation of the optic cup

Dilatations of the cephalic neural tube form the brain vesicles. By day 25, outpouchings of the prosencephalon (forebrain) known as the optic vesicles are formed and go on to form the future eye and the vesicle stalk, the optic nerve (Fig. 7.1).

Thickened areas of ectoderm (lens placodes) induce the underlying optic vesicles (bilaterally) to invaginate whilst simultaneously invaginating themselves to form the lens vesicle, which then sinks into the mouth of the newly formed double-layered optic cup on day 27 (Fig. 7.2).

To form a double-layered optic stalk, the diencephalon wall wraps itself around a central axis occupied by the hyaloid artery, which gains access to supply the lens by entering the choroidal fissure proximally. The choroidal fissure is located inferonasally and closes on day 33 (Fig. 7.3).

Formation of ocular tissues

Retinal pigment epithelium

The outer neuroectodermal layer of the optic cup consists of columnar epithelial cells joined by zonulae occludens. Melanization starts posteriorly at week 6 and the RPE organizes into a single layer of hexagonal columnar cells by week 8.

The RPE forms the inner basement membrane of Bruch's membrane and is the first tissue in the body to be pigmented. It is thought to be fully functional at month 4 of gestation. At the anterior edge of the optic cup, the RPE (and retina) are continuous with the ciliary body and iris epithelium.

RPE induces the development of the sclera, choroid, and retina.

Retina

The inner neuroectodermal layer of the optic cup develops into the neurosensory retina. At 4 weeks' gestation, nuclei aggregate in the outer two-thirds of the developing retina in the primitive zone. No nuclei are found in the inner third, named the inner marginal zone, which goes on to form the nerve fibre layer.

Retinal cell differentiation starts adjacent to the optic nerve and spreads in a wave peripherally.

At week 6, developing retinal nerve fibres converge at the future optic disc and enter the optic stalk. At the same time, immature ganglion, amacrine, and bipolar cell bodies migrate from the primitive zone towards the surface of the retina and form the inner neuroblastic layer, which is seen at week 7.

The primitive zone, now renamed the outer neuroblastic layer, and inner neuroblastic layer are separated by a layer of tangled cell processes called the transient nerve fibre of Chievitz, which thins by 10 weeks to form the inner plexiform layer—except underlying the fovea where it persists until 4 years post partum.

By 4.5 months, the layers of the retina are largely established (Fig. 7.4).

The inner neuroblastic layer differentiates further to form the ganglion cells, amacrine cells, and Müller cells.

Bipolar and horizontal cells develop from the mid-zone of the outer neuroblastic layer. Photoreceptors differentiate from the outermost regions of the neuroblastic layer with cone outer segment differentiation commencing at 5 months and rods at 7 months.

Branch retinal arteries develop at 4 months' gestation.

The 'upside-down' arrangement of the retina means that photons must travel through most of the layers of the retina before striking a photoreceptor.

The ora serrata is a line posterior to the ciliary body demarcating the neurosensory retina's anterior border. Both inner and outer neuroectodermal layers of the optic cup anterior to the ora serrata extend forward and on to the ciliary body and posterior surface of the iris.

Macula

Macula development occurs relatively late in gestation and continues until 4 years post partum.

At 6 months, ganglion cells overlying the layer of Chievitz demarcating the future fovea thicken.

Subsequent displacement of the ganglion cells away from the centre of the fovea and thinning of the overlying inner nuclear layer forms the foveal depression by month 7.

Remodelling of the macula continues with only exposed cone nuclei present on the fovea at 4 months post partum.

Optic nerve

At week 6, migrating retinal ganglion cell nerve fibres leave the optic cup and enter vacuolations of the inner layer of the optic stalk where it joins the optic cup at the future optic disc (Fig. 7.5).

Axons migrate posteriorly and parallel to the hyaloid artery, and their number proliferates to 3.7 million at 16 weeks,

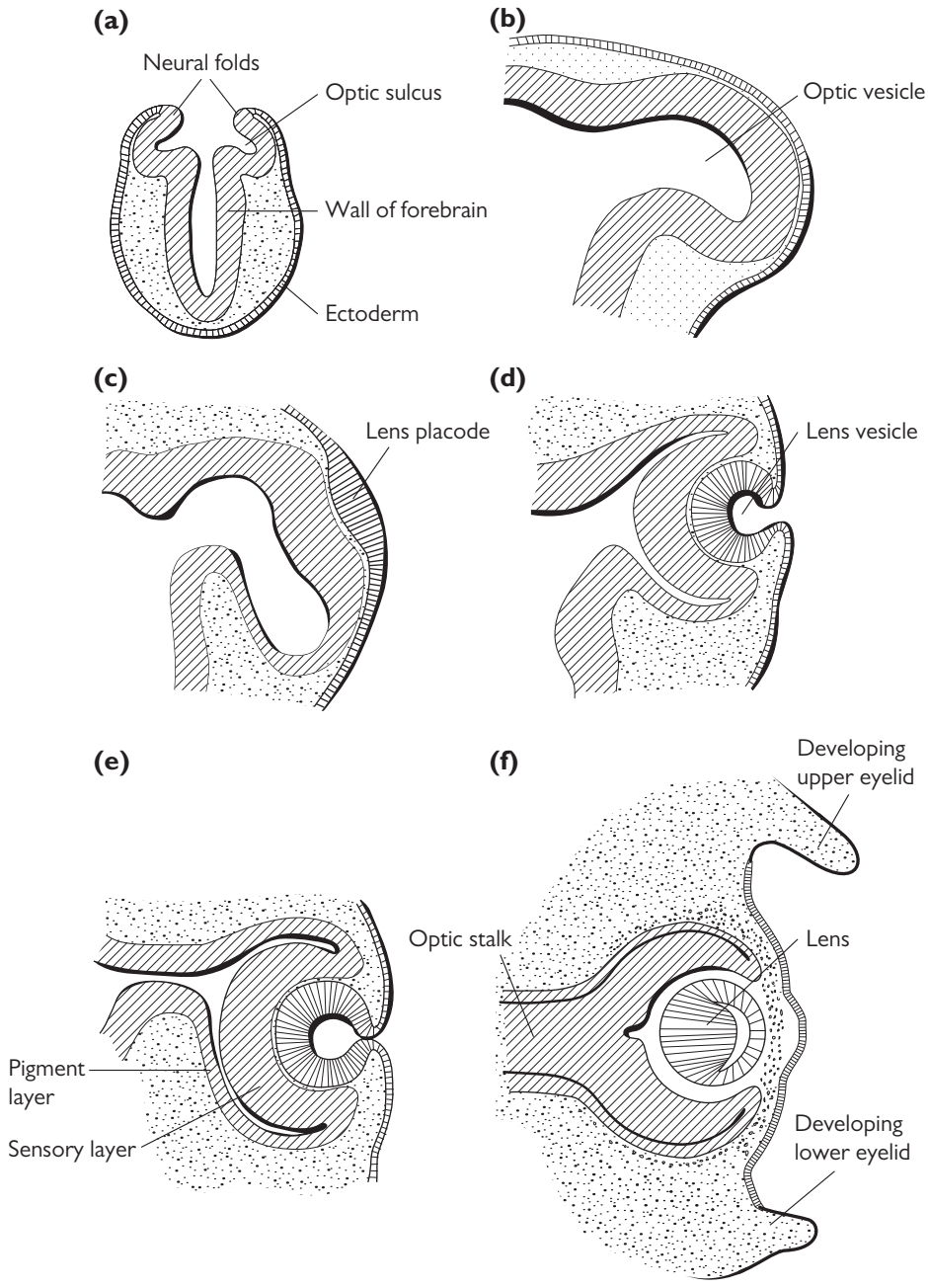


Fig. 7.1 Cross-section of developing eye with eyelids developing. (a)–(d) Developing upper eyelid. (e), (f) Developing lower eyelid.

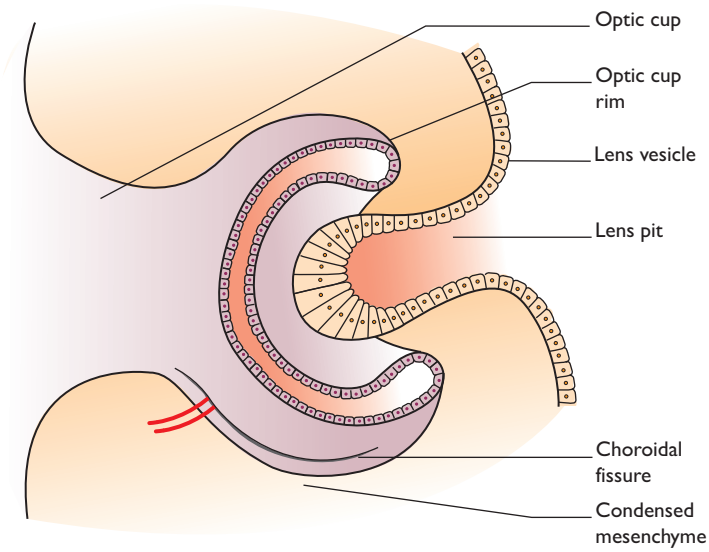


Fig. 7.2 Formation of the optic cup and lens.
With permission from Neil Modi.

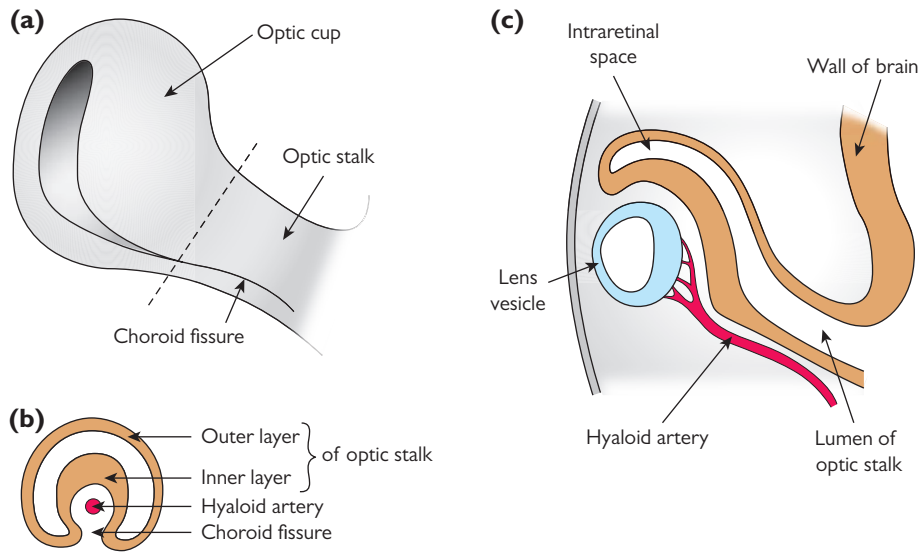


Fig. 7.3 Choroidal fissure formation (eccentric invagination) and hyaloid artery.
Reproduced from Thomas W. Sadler, *Langman's Medical Embryology*, 11th edition, figure 18.2, Wolters Kluwer with permission. Copyright 2009.

decreasing to 1.1 million at 33 weeks and term. Cells of the stalk's inner neuroectoderm layer differentiate into neuroglial cells, which support the axons and form the lamina cribrosa.

As axons extend posteriorly toward the lateral geniculate nuclei, partial decussation of nerve fibres forms the optic chiasm.

Outer layer neural crest cells form the overlying optic nerve pia, arachnoid, and dura mater at 4 months.

Myelination of the chiasm starts at month 7, then extends anteriorly and is not complete until after birth.

Regression of the distal hyaloid artery within the vitreous in month 4 leaves the proximal segment, which becomes the central retinal artery.

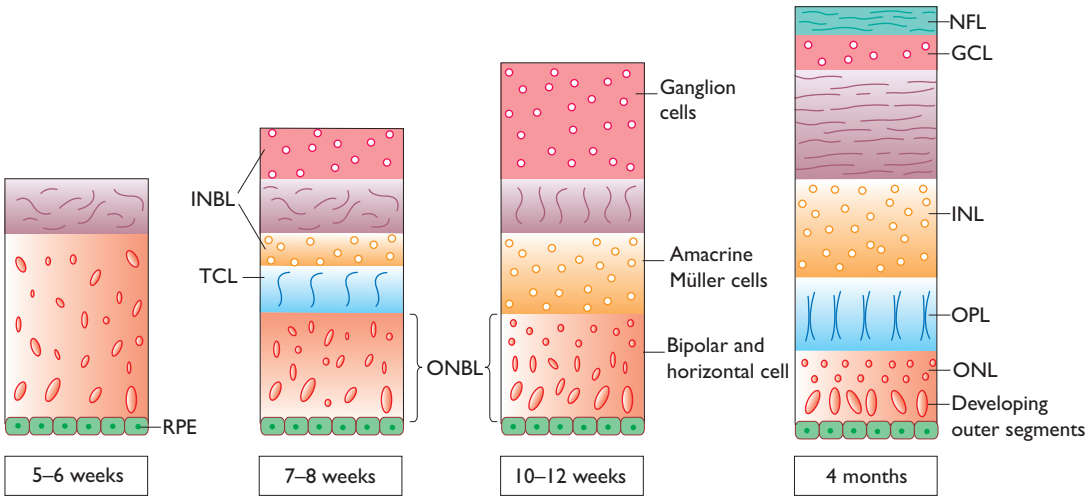


Fig. 7.4 Retinal layers.

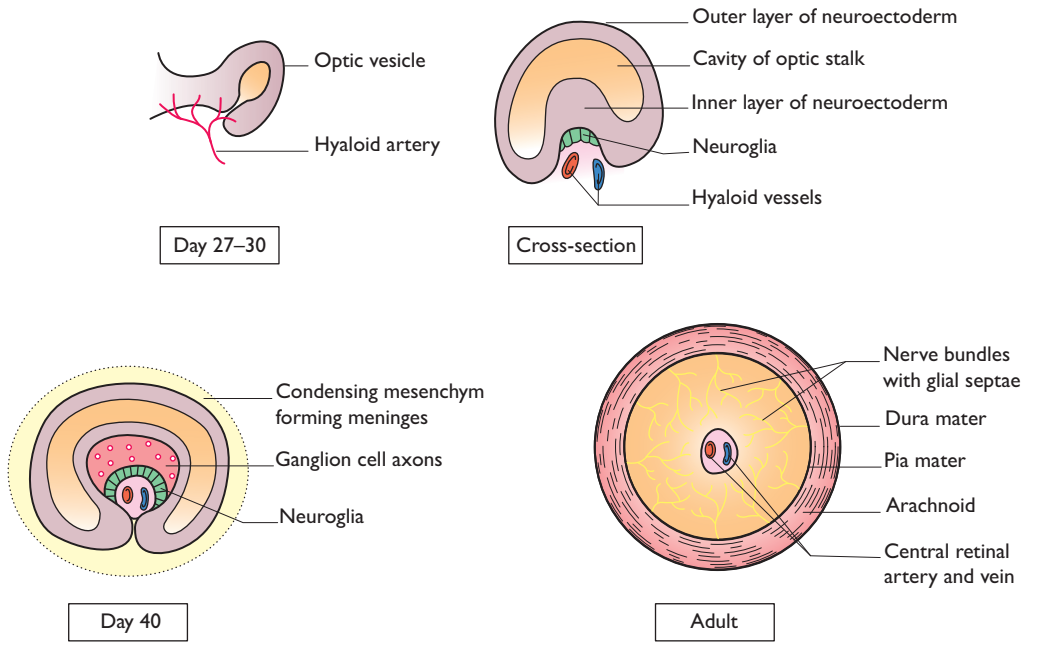


Fig. 7.5 Cross-section of the optic nerve.

Lens

The lens vesicle separates completely from the surface ectoderm layer at around day 33 of gestation. The lens vesicle is a single-cell layered sphere of epithelial cells enveloped by a basal lamina which eventually forms the lens capsule. The cells situated posteriorly elongate, laying down crystallins and obliterating the cavity. When the apices of the posterior cells attach to the anterior lens cells their nuclei degenerate, leaving the compact core of the adult lens known as the embryonic nucleus (Fig. 7.6).

The anterior epithelial cells continue to divide at the germinal zone at the equator of the lens. The bases of the secondary fibres remain attached to the basal lamina at the equator, with the apex extending either anteriorly or posteriorly. Anterior lens fibres meet at the most anterior point under the capsule at the anterior suture, which is an upright Y shape, and the lens fibres elongating posteriorly meet at the posterior suture shaped like an inverted Y.

New lens fibres are laid directly on top of older lens fibres concentrically, resulting in the lens's laminar 'onion-skin' structure. This ongoing pattern of development changes the sphere into its lentiform shape, as seen in adults (Fig. 7.6). Growth of the lens in the foetus is facilitated by the hyaloid artery, which forms a vascular plexus on the posterior lens surface. This hyaloid artery is a branch of the ophthalmic artery.

Cornea

The cornea starts developing after the lens vesicle detaches from surface ectoderm at day 33 (Fig. 7.7). Surface ectoderm forms the corneal epithelium. The first of three waves

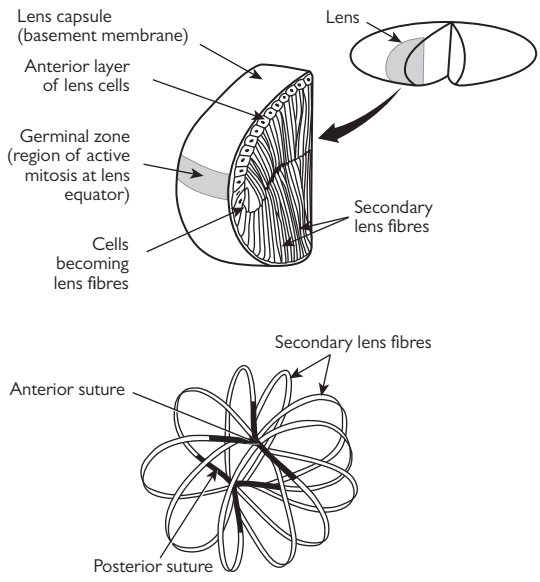


Fig. 7.6 Lens dividing zone at equator and laying down of lens fibres and sutures.

Reproduced from Snell RS and Lemp MA, *Clinical anatomy of the eye*, Second edition, Figure 1.7, p. 11, Wiley. Copyright 1998, with permission.

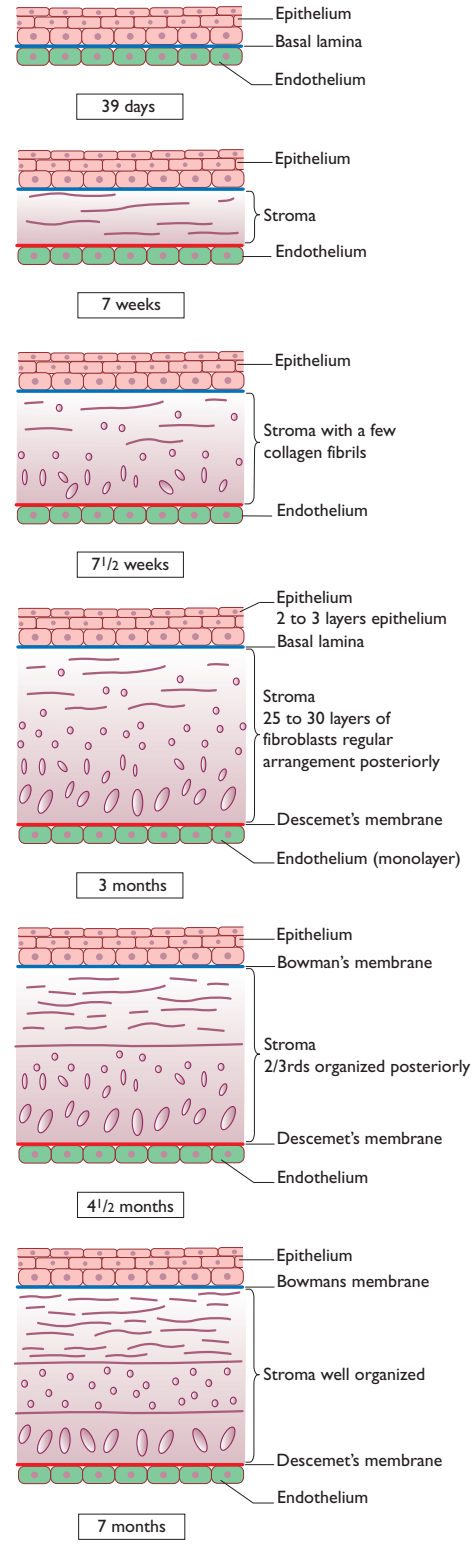


Fig. 7.7 The developing cornea.

of neural crest cell mesenchyme migration forms the corneal endothelium at day 39. A second wave of neural crest cell mesenchyme migrates from the edges of the optic cup to form the iris and the pupillary membrane. A third wave of neural crest mesenchyme at 7 weeks migrates into the space between the epithelium and endothelium, differentiating into keratocytes, which secrete type 1 collagen, forming the corneal stroma.

Descemet's membrane develops from the endothelium and becomes a monolayer by month 3.

Adjacent endothelial cells develop tight junctions at month 4, at the same time as aqueous humour production starts.

Bowman's layer is the last corneal layer to develop, at month 5.

Anterior chamber: angle, iris, and ciliary body

The anterior chamber begins to form at week 7 as a slit anterior to the lens vesicle. The iridocorneal angle begins to recess at month 3 and continues to deepen until 4 years post partum.

Neural crest mesenchyme around the rim of the optic cup differentiates to form the trabecular meshwork at 15 weeks and the ciliary body at 12 weeks.

Schlemm's canal starts off as a blood vessel (therefore derived from mesoderm) at the end of month 3 but functions as the angle drainage canal at month 5 when aqueous humour production starts.

At 6 weeks, neural crest mesenchyme forms the pupillary membrane in front of the lens vesicle.

The advancing bilayer edge of the optic cup (marginal sinus of von Szily) in month 3 coats the developing ciliary body and then the posterior surface of the developing iris stroma, which develops under the pupillary membrane from neural crest cell mesenchyme.

The marginal sinus of von Szily reaches the future pupil in month 4 and differentiates into the sphincter and dilator muscles, with myofibrils being laid down in month 5. (Both sphincter and dilator of the pupil are unlike any other muscle in the body, except arrectores pilorum, in being derived from ectoderm.)

The centre of the pupillary membrane degenerates at 8 months' gestation, forming the pupil and allowing communication between anterior and posterior chambers of the eye.

As it is derived from the (non-pigmented) inner ectoderm layer of the optic cup, pigmentation of the posterior layer of the iris starts at the pupil margin and reaches the iris root in the seventh month post partum.

Further growth of the iris stroma continues post partum.

Vitreous

The vitreous develops in the space between the lens vesicle and inner layer of the optic cup in three distinct stages.

Primary vitreous forms the developing vascular system and is derived from mesoderm mesenchyme, and to a small extent neural crest cells. The hyaloid artery is a branch of the primitive ophthalmic artery and enters the optic vesicle via the optic fissure and anastamoses with the posterior lens capsule, forming the tunica vasculosa lentis. Primary vitreous development peaks at 8 weeks.

Secondary vitreous is avascular and associated with hyalocytes, which secrete type 2 collagen fibrils, which expand the volume of the vitreous. Expansion causes the primary vitreous to form a funnel, axially enclosing Cloquet's canal as the intravitreal hyaloid artery begins to regress at month 4. Primary vitreous regresses at month 5.

Tertiary vitreous is a phase of development when the collagen fibrils specifically condense at the lens equator and contribute to formation of the lens zonules and the marginal bundle of Druault, which attaches to the internal limiting membrane of the retina to form the vitreous base.

Sclera and choroid

Sclera develops from aggregations of neural crest cell-derived mesenchyme at week 7, starting near the insertions of the recti muscles and spreading over the surface of the eye from anterior to posterior—except for the superotemporal section, which is derived from mesoderm. The lamina cribrosa is derived from sclera posteriorly, which is continuous with the dura mater around the optic nerve.

The choroid is a vascular layer between the sclera and retina that starts developing at the rim of the optic cup and progresses posteriorly. Neural crest cell mesenchyme forms the choroidal stroma, which is invaded by the mesoderm-derived fine capillary network called the choriocapillaris during weeks 4 and 5 of gestation.

The choriocapillaris is connected to the posterior ciliary arteries in month 8 of gestation.

Pigmentation of the choroid proceeds anteriorly from the disc at 6–7 months' gestation and is completed post partum.

The pia arachnoid layer in the optic nerve is the continuation of the choroid.

Eyelids

The eyelids start off as aggregations of surface ectoderm in front of the developing cornea on its superior and inferior edges (6–7 weeks). The two aggregations proliferate across the cornea until they fuse at 10 weeks.

Orbicularis develops in the lids at week 12. At 20 weeks, the lid divides to form the upper and lower eyelids at the same time as the meibomian glands start secreting lipid. Fusion is thought to prevent the underlying corneal epithelium from being keratinized during this stage of development.

Lacrimal gland

The lacrimal gland develops between weeks 6 and 7. Epithelial cells from the superior conjunctival fornix proliferate as cords superotemporally. The cords are tipped by neural crest mesenchymal cells, which differentiate into the acini. Cord lumen formation occurs in month 3.

During month 5, the developing levator palpebrae superioris bisects the lacrimal gland into orbital and palpebral sections. Different authorities state that lacrimal gland tear production does not occur until 20 days to 3 months post partum, with full maturation of the gland not completed until 3 years of age.

Orbit

Orbital bones are formed from neural crest mesenchyme and develop as ossified membranes, apart from the sphenoid, which has cartilaginous lineage.

Ossification starts in month 3 and bone fusion in month 6.

Extraocular muscles

At week 5, mesoderm mesenchyme differentiates into the extraocular muscles. The levator palpebrae superioris muscle develops from the dorsomedial aspect of the superior rectus at week 7.

Nasolacrimal duct

An ectoderm cord forms in the nasolacrimal groove between the lateral nasal and maxillary swellings on the developing face. The cord canalizes at 4 months' gestation to form the nasolacrimal duct, which is superiorly connected to the inferior meatus of the nasal cavity. Simultaneously, canalization of the lower and upper eyelid margins results in the formation of puncta and canaliculi, which connect to the nasolacrimal sac.

The membrane of Hasner, located at the nasolacrimal duct outlet, remains imperforate in 70% of newborns but normally regresses within the first month post partum.

Congenital anomalies

Organogenesis of the human eye occurs immediately after gastrulation, between 24 and 40 days' gestation, and is the most vulnerable period for malformation production.

Anophthalmia

Complete absence of ocular tissue is rare. Primary anophthalmia is bilateral and can manifest in otherwise healthy children. Secondary anophthalmia due to incomplete forebrain development is lethal and linked with SOX2 gene mutation.

Cyclopia

This is the formation of a single globe in the midline with absent optic nerve. It is lethal.

Synophthalmia

In synophthalmia two incomplete globes join as a single structure at the midline. This is a lethal condition.

Microphthalmia

Small disorganized globes can be associated with Patau's syndrome (trisomy 13), foetal-alcohol syndrome, and in-utero rubella, toxoplasma, and cytomegalovirus infections. This condition is linked with CHX10 gene mutation.

Nanophthalmia

Nanophthalmia is the development of small eyes with a normal size lens which are highly hypermetropic.

Cryptophthalmia

Abnormal eyelid development and fission results in abnormal development of the underlying eye, which remains buried at birth. The cornea and anterior chamber are absent. The retina and optic nerve are normal.

Sections of the underdeveloped eyelid adhere to the cornea in incomplete cryptophthalmos.

Anterior segment

Megalocornea

Usually X-linked recessive, 90% of corneas larger than 12 mm diameter at birth are found in males. Infants with megalocornea must be investigated for congenital glaucoma. Systemic associations include Marfan's syndrome, Alport's syndrome, Down's syndrome, and albinism.

Microcornea

Corneas smaller than 10 mm in diameter at birth are invariably associated with microphthalmia. This can be bilateral or unilateral.

Axenfeld–Rieger syndrome

This AD inherited range of congenital malformations of the anterior segment affects the development of the cornea, iris, and angle. Iris features include thinning of the cornea, correctopia, and polycoria. Posterior embryotoxon is the anterior displacement of Schwalbe's line, which can often be seen without gonioscopy as a white line near the limbus behind the cornea. The iridocorneal angle is bridged by iris strands and half of patients with this syndrome have secondary glaucoma. Transcription factor mutations of PITX2 and FOXC1 are thought to be causative. Systemic features may include dental abnormalities, craniofacial dysmorphisms, deafness, redundant periumbilical skin, and congenital heart defects.

Aniridia

Absence of neuroectoderm-derived iris pigment epithelium, sphincter, and dilator muscles results in absence of the pupil. Clinically, some iris can usually be identified because iris stroma is derived from neural crest cells and is variably present. Aniridia is associated with Wilm's tumour.

Ocular albinism

Defective melanin pigment production is termed albinisim. It can affect all the tissues in the body (oculocutaneous albinism) or the eyes alone (ocular albinism). Iris pigment deficiency is seen as transillumination defects. The result is poor, but stable, vision. Foveal hypoplasia and nystagmus are common.

Coloboma (absence of tissue)

Defects due to failure of the optic fissure to fuse on day 33 can occur at any given point along its tract. Clinical features of coloboma include a notch or hole defect of the inferior iris and ciliary body, choroid, and retina defects. Failure of retinal pigment epithelium development in colobomata leads to failure of induction of underlying choroid and retinal development. Macular coloboma is a misnomer and is thought to be due to congenital toxoplasmosis.

Coloboma can be an isolated finding or syndromal, e.g. CHARGE syndrome—Coloboma, Heart defects, Atresia of nasal choanae, Retardation of growth and development, Genitourinary defects, and Ear abnormalities, including hearing loss.

Peter's anomaly

Abnormal cleavage of the developing cornea from the lens vesicle at around day 33 results in abnormal development (and therefore opacity) of the central, paracentral, or entirety of the corneal endothelium and Descemet's membrane. In type 1 Peter's anomaly, cataract is not always present and the cornea does not adhere to the lens. Bilateral cataracts are present in the type 2 condition, with a keratolenticular stalk bridging the cornea and lens. PAX6 gene mutations and exposure to ethanol and isotretinoin in utero have been implicated as causative factors.

Cataract

Opacification of the lens in utero can be due to malnutrition, infection, e.g. rubella, or genetic predisposition, e.g. Down's syndrome. Rubella affects the developing lens before 6 weeks but infection thereafter does not cause cataract.

Posterior segment

Persistent foetal vasculature

Persistent hyperplastic primary vitreous

Failure of the primary vitreous to regress is associated with microphthalmia and causes leukocoria. Usually unilateral, a fibrovascular stalk links the posterior lens and optic nerve, elongating the ciliary processes, which can be seen as peripheral shadows on transillumination. Anterior and posterior vitreous disease subtypes have been described. Cataract and retinal dysplasia may be present.

Persistent hyaloid artery

There is some association of persistent hyaloid artery with premature low-weight babies. Complete Cloquet's canals seen post partum are rare.

Mittendorf dot

Failure of the anterior segment of the hyaloid artery's connection to the lens to degenerate is seen as a plaque on the posterior surface of the posterior lens capsule and is normally situated nasally.

Tunica vasculosa lentis

The developing lens is supplied by the hyaloid artery and its capillary meshwork on the posterior and lateral surface of the lens, which is seen at birth if incompletely regressed.

Bergmeister's papilla

Failure of the posterior section of the hyaloid artery to degenerate leaves a fibrous tuft and membrane attached to the optic disc. It does not affect vision and is usually an incidental finding.

Persistent pupillary membrane

Anteriorly, the pupil is formed by the regression of the iridopupillary membrane near term. Remnants are seen as connective tissue strands stretched across the pupil and are benign.

Optic pit

A localized defect in the optic fissure on the optic nerve head allows dysplastic retina to herniate into the subdural space through a defect in the lamina cribrosa, allowing cerebrospinal fluid to cause serous retinal detachments. Optic pits are usually located temporally or inferiorly and can cause arcuate field defects.

Atresia of the nasolacrimal duct

This is failure of the developing nasolacrimal duct to canalize, with or without punctual atresia.

Growth

Ocular development continues post partum. Presentation of a focused image onto the retina, controlled eye movements, and connection to the brain and higher processing centres are all prerequisites for adult levels of functioning vision.

Visual acuity in infants at birth is approximately 6/90, increasing to approximately 6/18 at 6 months and full adult acuity by age 10.

The eye itself is not fully developed at birth. The retina develops from 28 weeks' gestation and the macula from 32 weeks. At birth, peripheral retinal rods are present but the macular region undergoes further maturation. All three colour pigments are present but the cones themselves are spaced further apart than in adults and their photoreceptor outer segments are shorter. As the cones remodel, becoming more densely packed and lengthening individually, the increase in light-processing efficiency is thought to contribute to an increase in visual acuity. Moreover, cone nuclei in the foveal region have no ganglion cells overlying them by the fourth month post partum.

Unmyelinated nerve fibres are less efficient at transmitting action potentials; neurons in the brain and visual pathway do not fully myelinate until 4 months post partum.

The optical apparatus responsible for focusing images on the retina continues to develop after birth. Most infants have astigmatism at birth, which is eliminated by the end of the first year, as adult corneal dimensions are attained.

At birth, the eye is hypermetropic and its axial length (approximately 18 mm) continues to increase in the first few years of life. There is another increase in puberty, and children are therefore able to 'grow out' of hypermetropia when the adult ocular axial length settles at around 24 mm.

The lens is mostly developed at birth but continues to grow slowly throughout life. Later layers are laid down concentrically and overall the lens flattens with each addition.

The sclera develops from the insertions of the recti muscles and the posterior portions are the last to form. This may explain why the sclera is thinned in myopes with increased axial length.

The visual field at 3 months is 40° either side of the vertical meridian, extending to 80° by 10 months. Nasal field development lags behind temporal fields.

Iris stroma pigmentation occurs during the first year, with the ciliary muscle only reaching adult size in the late teens.

Infants have measurable contrast sensitivity at 6 weeks and spectral sensitivity at 2 months.

Spatial frequency channels, e.g. the magnocellular and parvocellular pathways, may exist in infants as young as 6 weeks and more convincingly at 12 weeks.

Spatial frequency is said to be attributable to the increased density of the photoreceptors in the fovea. Spectral sensitivity increases with photoreceptor length. Infants can distinguish red from white from the age of 2 months. Red, green, and blue cones and normal rod pigment are present but blue pigment function development lags behind the other photoreceptor pigments.

Stereopsis develops between 3.5 and 6 months.

At birth, most infants have exotropia, with the axes of the optic nerves forming an angle of 71°. (The adult angle of 68° is not achieved until 3 years.) Diplopia is unlikely to be troublesome as visual acuity at this stage is relatively poor. Orthotropia is the coupling of both eyes' movements so that they regard the same object and requires voluntary control.

Stereopsis is the result of functionally linking up the individual components of the visual system and develops between 3.5 and 6 months, when the visual apparatus has reached a mature stage of development. Only after ensuring the accurate acquisition of the visual input can the brain be in a position to process the differences between the images reaching both retinas in order to determine depth.

Literature suggests that there are pre- and post-stereoptic periods of development. Higher visual functions involving the brainstem, including saccades, pursuit, vergence, and accommodation, develop after stereopsis is established, and if not developed result in abnormal eye movements, e.g. nystagmus in its different forms.

Newborn infants have optokinetic nystagmus, supporting the hypothesis that this is a subcortical reflex, i.e. involuntary.

Senescence

Senescence is the process of ageing—the accumulated changes within an organism over time. The Hayflick limit theory observes that cells are not immortal and divide 40–60 times before proliferation ceases, thus limiting genetic instability and subsequent cancerous growth.

Degeneration is the deterioration in the structure and function of a cell or tissue.

Conjunctiva and cornea

Conjunctivochalasis

Mechanical stress on the inferior bulbar conjunctiva results in a fold of redundant conjunctiva protruding over the lower lid margin. This can obstruct the inferior punctum and disrupt the tear meniscus. Patients present with epiphora and foreign body sensation. Treatment is excision of the offending fold.

Involutional ptosis

Levator aponeurosis dehiscence, disinsertion, or stretching can lead to drooping of the upper lid.

Involutional ectropion

Horizontal lower eyelid laxity leads to eversion of the lower eyelid, which presents with epiphora and variable degrees of cicatricial changes.

Involutional entropion

Corneal scarring can occur when inverted eyelid margins and eyelashes abrade the ocular surface. Treatment is surgical repair.

Arcus senilis (gerontoxon)

Arcus senilis is bilateral deposition of cholesterol esters, triglycerides, and phospholipids in the peripheral cornea. Deposits are seen anterior to Descemet's membrane extending anteriorly and in Bowman's layer extending posteriorly, which meet, forming an hourglass cross-section. Separated from the limbus by a 0.2 mm zone of unaffected cornea, fine grey–yellow dots in the superior cornea spread circumferentially to involve the entire circumference. Arcus senilis is associated with hyperlipidaemia and atherosclerosis, but using it as a marker of vascular disease is controversial.

Limbal girdle of Vogt

White opacities in the medial and temporal limbal cornea may be confused with arcus senilis. Unlike arcus senilis, the limbal girdle does not progress and has been described as subepithelial elastosis—a process similar to pterygium and pingueculae formation.

Iron lines

Haemosiderin deposition in the basal corneal epithelium cells has been recognized in senescence but the pathophysiology remains unclear. The iron lines themselves do not affect visual acuity.

Common eponymous examples include:

- the Kayser–Fleisher ring—at the base of the keratoconus cone
- Ferry's line—adjacent to trabeculectomy blebs
- Stocker's line—precedes the migrating pterygium edge
- the Hudson–Stahli line—a 'tidemark' at the upper border of the tear film.

Pingueculae and pterygia

Pingueculae are yellow–white deposits adjacent to the cornea. Pterygia are wing-shaped vascularized overgrowths of conjunctival tissue which migrate on to the cornea.

Pingueculae and pterygia are benign neoplasms found in the nasal and temporal conjunctiva. Histologically both demonstrate actinic elastosis (degeneration of collagen fibres) but destruction of Bowman's layer is an additional feature in pterygia.

UV light is said to induce actinic elastosis, resulting in a higher prevalence of both conditions in sunnier climates.

Anterior chamber and lens

Presbyopia

The amplitude of accommodation is the amount that an eye can change its refractive power between focusing on the nearest and furthest objects possible.

With increasing age, the amplitude of accommodation decreases, preventing sufficient accommodation of the lens to focus on near objects.

There is no one single causal factor and several mechanisms have been suggested. Changes in the elasticity of the lens capsule may inhibit 'fattening' of the lens. Hardening of the lens may also occur, although this theory has been disproved as lenticular sclerosis predates the onset of presbyopia. Another hypothesis is that the ongoing growth of the eye and its components alters the geometric relationship between the ciliary muscle root and lens, which alters zonule tension with time.

Glaucoma

Primary open-angle glaucoma is a chronic progressive optic neuropathy with characteristic morphological optic nerve head and retinal nerve fibre layer changes in the absence of other ocular disease or congenital anomalies.

Known risk factors are age, raised intraocular pressure, race, positive family history, diabetes, and myopia.

Cataract

An opacified lens is known as a cataract—derived from the Latin for waterfall 'cataracta'. Cataract formation can be primary due to senescence or secondary, i.e. acquired.

Cataracts are partially opaque when immature and opaque when mature. Hypermature cataracts are opaque and begin to liquefy, causing wrinkling of the anterior lens capsule.

In Morgagnian cataracts the cortex is so liquefied that the hard nucleus can be seen sinking in the lens capsule.

Age-related cataract

Cataracts are classified by describing the location of the opacifying pathological process in the lens but often have mixed features. They can be predominantly anterior or posterior, involve the capsule (capsular), or affect the lens adjacent to the capsule (subcapsular), or involve the cortex (cortical) or nucleus (nuclear).

Cortical

Osmotic shifts affect the cortical lens fibres in the anterior lens. Depletion of intracellular potassium and retention of sodium, chloride, and calcium ions leads to the influx of water into the cortex between lens fibres. This is seen clinically as fluid-filled spokes or bubbles. Because fluid has a lower refractive index than cortical lens matter, patients complain of glare as light is scattered before reaching the retina.

Nuclear

Nuclear cataracts result from post-translational modification of the proteins in the lens fibres. The oldest lens fibres are the embryonic nucleus, and oxidation leads to aggregation by crosslinking and precipitation. Early nuclear sclerotic cataracts are yellow–green due to urochrome deposition and with progression become brown–black when mature—a brunescent cataract.

If the lens changes increase the refractive index, myopia is induced and patients for a while can read without glasses again. This index myopia is also known as the 'second sight of the aged'. Other types of cataract are discussed in Chapter 4.

Acute angle-closure glaucoma

Grey–white anterior subcapsular or anterior capsular opacities are seen within the pupil in patients with acute or recurrent angle-closure glaucoma. These focal infarcts of the anterior lens capsule epithelium are termed 'Glaukomflecken'.

Vitreous

Posterior vitreous detachment

With ageing, free radicals cause the vitreous to liquefy (synchysis) and collapse (syneresis). Separation of the posterior vitreous from the internal limiting membrane of the retina starts perifoveally and extends anteriorly towards the vitreous base.

When the vitreous cortex overlying the optic nerve head detaches, a hole (Weiss ring) is created through which liquefied vitreous drains out, filling the retrohyaloid space, further facilitating the vitreous detachment. PVD is a common cause of 'floaters' and is benign but can be associated with horseshoe retinal tear, retinal detachment, and vitreous haemorrhage.

Asteroid hyalosis

Age-related deposition of calcium hydroxyapatite, calcium phosphate, and phospholipids in the vitreous is seen clinically as yellow–white opacities scattered throughout the vitreous gel which move with ocular movements. Patients are usually asymptomatic.

Retina

Retinoschisis

Acid mucopolysaccharide-filled cavities split the neurosensory retina at the outer plexiform layer (typical retinoschisis) or the nerve fibre layer (reticular retinoschisis). Most retinoschises occur in the inferotemporal retina and are bilateral, with the majority of retinoschises being incidental findings and asymptomatic. A visual field defect can occur if the split extends posteriorly towards the macula but most extensions stop spontaneously. Retinoschisis is associated with retinal detachment. Only symptomatic patients should be treated.

Lattice degeneration

Lattice degeneration is characterized by demarcated, circumferentially orientated areas of retinal thinning usually adjacent to the vertical meridian peripherally. Atrophic small and round retinal holes can be present. There is an associated risk of retinal detachment.

Age-related macula degeneration

ARMD is the leading cause of blindness in the developed world and the main risk factor is age.

Dry (non-neovascular) ARMD is characterized by loss of RPE and photoreceptor layers, thinning of the outer plexiform layer and Bruch's membrane, and atrophy of the choriocapillaris layer. Wet (neovascular) ARMD constitutes 10–15% of ARMD and is characterized by ingrowth of choroidal vessels into the RPE and subretinal space. If the choroidal neovascular membrane vessels rupture and bleeding occurs, the architecture of the macula can be severely disrupted.

CHAPTER 8

Optics

Properties of light

Physical optics studies interference, diffraction, polarization, and other phenomena in which light is treated as a wave. In contrast, geometric optics approximates light as a ray and is primarily concerned with the imaging properties of lenses and mirrors.

Light may be described as discrete packets (quanta) of energy or a wave.

The human eye is sensitive to a small proportion of the electromagnetic spectrum, lying between X-rays and microwaves, termed visible light. This area is subdivided into seven wavebands:

1. ultraviolet C (UVC, 200–280 nm)
2. ultraviolet B (UVB, 280–315 nm)
3. ultraviolet A (UVA, 315–400 nm)
4. visible light (400–780 nm)
5. infrared A (IRA, 780–1400 nm)
6. infrared B (IRB, 1400–3000 nm)
7. infrared C (IRC, 3000–10,000 nm)

The cornea and sclera absorb all the incident radiation at short wavelengths (UVB and UVC) and long wavelengths (IRB and IRC). UVA is strongly absorbed by the lens. Visible light freely passes through the ocular media. The longer wavelength IR that reaches the retina can have thermal effects, e.g. in a solar eclipse, causing retina burns.

There is also evidence that the shorter wavelength UVA light can cause damage to the retina in aphakic or pseudophakic patients.

Colour vision is perceived by three populations of cone receptors, each with a different peak spectral sensitivity. It is the relative stimulation of these that results in the perception of colour vision. This is described in Chapter 2.

Light can be described as having two defining characteristics: wave behaviour and particle behaviour. More likely, however, there is no discrimination between the two, but rather a flux between two states, also known as the wave-particle duality.

Wave theory of light

Light travels as waves, although its path is often represented as a 'ray'.

The wavelength, λ, of a wave is the distance between two successive peaks of propagation. One oscillation is defined as the segment of the wave where it returns to the start, which is also known as one cycle. The amplitude is the maximum displacement of the wave from its baseline.

If two waves are travelling together but their waves do not line up, they are 'out of phase'. Light waves that are out of phase are termed 'incoherent' whereas those that are in phase are termed 'coherent'.

Interference

Interference is the phenomenon where two or more waves interact with one another.

Constructive interference occurs when two or more waves that are in phase summate to form a new wave of amplitude equal to the sum of the initial waves.

Destructive interference occurs when the waves are out of phase by half a cycle. Here two waves of equal amplitude will summate to produce a resultant wave of zero amplitude. This occurs in the corneal stroma where the collagen fibrils are evenly spaced such that light deviated by them is eliminated by destructive interference.

Diffraction

Diffraction is described as the apparent bending of a wave around an obstacle. This causes a secondary wave front that is of a different phase and lower intensity to that of the initial wave prior to diffraction.

Diffraction through a circular aperture causes a diffraction pattern where there is a central bright disc (the Airy disc) that is surrounded by alternating light and dark rings, together known as the Airy pattern.

The size of the aperture affects the size of the Airy disc: the smaller the aperture the larger the disc. This is because diffraction is most prominent in smaller apertures.

Diffraction occurs in all optical instruments, for example cameras and the human eye. Thus, when the pupil of the eye is at its smallest, diffraction is the main cause of image imperfection.

Limit of resolution and resolving power

The limit of resolution is defined as the smallest angle of separation between two points such that the two points are clearly visible as separate. The limit of resolution is achieved when two Airy discs are separated so that the middle of the second disc falls on top of the first dark ring of the first disc. The resolving power of the eye is measured with visual acuity tests. This is discussed in Chapter 2.

Polarization of light

Light emitted from a source occurs in all directions at random. Polarization refers to the unification of light waves perpendicular to the direction of propagation. Polarized light can be formed from non-polarized substances, e.g. crystalline minerals. Light is also polarized on reflection off a plane surface, e.g. water, if the angle of incidence equals the polarizing angle of the plane surface.

Birefringence

This is the property whereby light waves may pass undeviated if in one plane, but not in a plane perpendicular to this. Quartz crystals show this property. Because they split incident unpolarized light into two polarized beams travelling in different directions, they have two refractive indices.

Dichroism

This is the property whereby light waves not aligned with the structure of a molecule are completely blocked. Only one beam of polarized light emerges. Tourmaline and Polaroid are dichroic substances, Polaroid being used in sunglasses. Polarizing glasses may be used to dissociate the eyes when assessing binocular vision (the Titmus test).

The measurement of light

Consider that light travels from a point source and is reflected off a surface. There are several ways to quantify light at various stages along this path. The radiometric system quantifies light in the entire electromagnetic spectrum as an absolute value. The photometric system is only concerned with quantifying light in the visual spectrum as a function of the visual response it produces. The different ways of measuring light are:

- Radiant flux (measured in watts) is the radiometric equivalent of luminous flux (measured in lumens), which is photometric, and quantifies how much light is emitted from a point source. A luminous flux of one lumen per square metre corresponds to a luminance of one apostilb.
- Radiant intensity (the equivalent of luminous intensity) considers the intensity of light in a given solid angle.
- Irradiance (the equivalent of illuminance) is the amount of light falling on a given surface area.
- Radiance (measured in watts per steradian) and luminance (measured in candelas) is the amount of light reflected per solid angle per area.

The unit known as the troland is a measure of retinal illumination when a surface luminance of one candela per square metre is viewed through an entrance pupil that measures one square millimetre after correction for the Stiles–Crawford effect.

Geometric optics describes light propagation as a 'ray'. The concept of light as a ray allows for the construction of ray models, which may be used to describe how light propagates through an optical system and ignores the effect of diffraction. Geometric optics uses geometric principles to describe images and the properties of lenses and mirrors.

Reflection

There are two important laws that define reflection (Fig. 8.1):

1. The incident ray, the normal (the imaginary line that is perpendicular to the reflecting surface at the point of reflection), and the reflected ray all occur in the same plane or medium.
2. The angle of incidence (i) = the angle of reflection (r).

Reflection at a plane mirror

The brain assumes that an object is positioned in the direction from which the light enters the eye. In Fig. 8.2 it can be seen that light from the object appears to come from the image when reflected in a plane mirror. However, this is an image that is erect, virtual, and laterally inverted.

Rotation of a plane mirror

If a mirror is rotated, the angle of reflection moves by twice the angle of rotation of the mirror.

Reflection on an uneven surface

A lot of the materials we come into contact with are not perfect mirrors, but they still reflect light so that we are able to see them; their surface is irregular and causes diffuse reflection.

Reflection involving spherical reflecting surfaces

Spherical mirrors present themselves in a variety of everyday situations as well as within ophthalmological instruments. The *Helpful hint* box runs through some of the terminology used when describing spherical reflecting surfaces.

A spherical mirror may be convex or concave

Concave mirrors

Parallel rays represent the object at a distance of infinity.

Parallel rays to the principal axis are reflected towards the principal focus and the images produced are always real (Fig. 8.3).

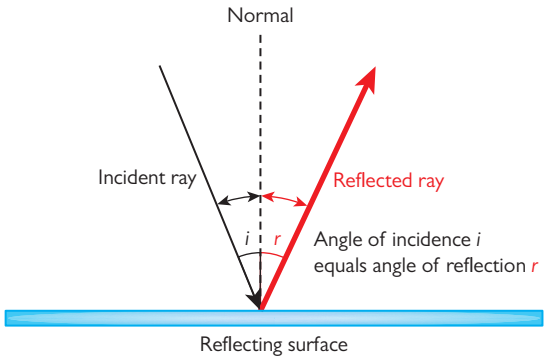


Fig. 8.1 Reflection at a plane surface.
With permission from Louise Bye.

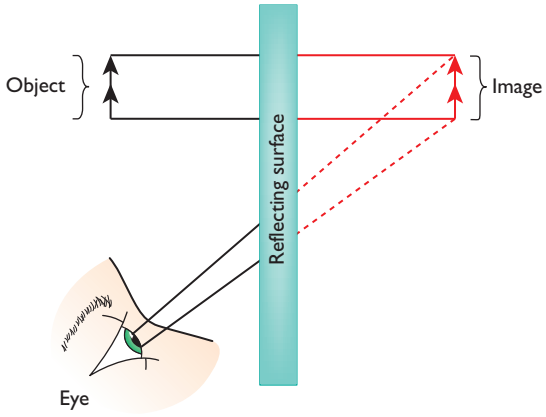


Fig. 8.2 Reflection at a plane surface: extended object.
With permission from Louise Bye.

HELPFUL HINT

Terminology for spherical reflecting surfaces

The centre of curvature (C) represents the centre of the sphere of which the mirror is made.

The pole of the mirror (P) is the centre of the reflecting surface.

The principal axis is the radius of curvature (CP), which is a line passing through the centre of curvature to the pole of the mirror.

The principal focus (F) is the real or virtual intersect of parallel rays (to the principal axis) reflecting off the mirror.

The object may be outside the centre of curvature, at the centre of curvature, or within the centre of curvature (Fig. 8.4).

Convex mirrors

Parallel rays to the principal axis are reflected away from the principal focus and a virtual image is produced behind the mirror by extrapolating the reflected rays.

Focal length

In both convex and concave mirrors the distance between the principal focus and the pole of the mirror is the focal length, FP. In perfectly spherical mirrors the focal length, *f*, is half the radius.

Calculating the focal length, *f*:

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f} = \frac{2}{r}$$

where *v* is the image distance, *u* is the object distance, *f* is the focal length, and *r* is the radius.

As a general rule of thumb, any focal point that is inferred by re-tracing a ray diagram usually produces a virtual image.

Magnification

Magnification, mathematically, is the ratio between the object size (*o*) and the image size (*i*); therefore $\frac{i}{o} = \frac{-v}{u}$.

Constructing ray diagrams

In all cases the same principles are true:

1. Draw a line from the top of the object through the mirror to the centre of curvature in one continuous straight line (O–C–M).
2. Draw a parallel line (to the principal axis) to the mirror, and depending on the nature of the mirror (convex or concave) draw the reflected ray (away or towards the principal axis respectively).
3. At the point of reflection draw a straight line through the principal focus, F. Where the two rays intersect is the top of the image (M–F–1).

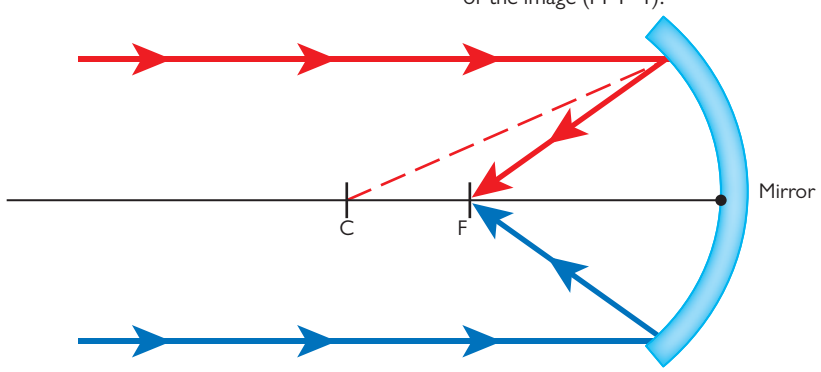


Fig. 8.3 Concave mirror: reflection of parallel light.
With permission from Louise Bye.

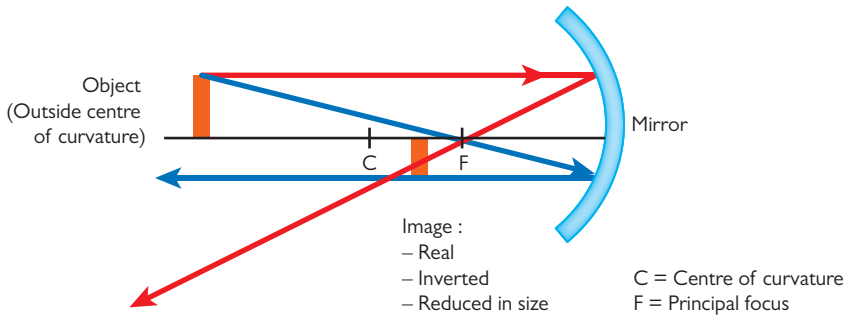


Fig. 8.4 Image formation by the concave mirror.
With permission from Louise Bye.

Refraction

Refraction is a change in the direction of light when it passes from one transparent medium to another. The denser the medium, the slower the light passes through it. On striking an interface, the first edge of the beam to reach the dense medium is slowed first and this causes the beam to deviate. The last edge of the beam to reach the denser medium continues at the same velocity. As with reflections, the incident ray and the refracted ray all lie in the same plane.

A classic example of refraction is the straw in a glass of water—the image looks jilted due to refraction at the water–air interface.

Absolute refractive index

The absolute refractive index *n* is defined as:

$$\text{absolute refractive index} = \frac{\text{velocity of light in vacuum}}{\text{velocity of light in medium}}$$

When light enters a medium of higher refractive index it is 'bent' towards the normal; the opposite will occur if the light were to leave the denser medium for a less dense medium.

Snell's law

The incident light makes an angle (*i*) with the normal and the refracted ray forms an angle (*r*) with the normal. These angles can be worked out if the refractive indices (RIs) are known using the following formula according to Snell's law:

$$\text{RI} = \frac{\sin i}{\sin r} = \frac{n2}{n1}$$

where *i* is the angle of incidence, *r* is the angle of refraction, and *n* is the refractive index of the media involved when light travels through *n*1 into *n*2.

In fact, at each interface, around 4% of incident light is reflected and 96% is transmitted. Hence, only around 92% of light passes through a pane of glass. Glass can therefore be used as an image splitter (Fig. 8.5). Reflected light is used in the teaching mirror of an indirect ophthalmoscope.

Refraction at a curved interface

Refraction at a curved interface obeys Snell's law, but the action can either converge or diverge light depending on the ratio of *n*1 to *n*2.

For example, in a convex interface where $n1 < n2$ (such as air:cornea) the light is focused onto a point (Fig. 8.6). However, if $n1 > n2$ (glass:air) then light diverges away from the principal axis.

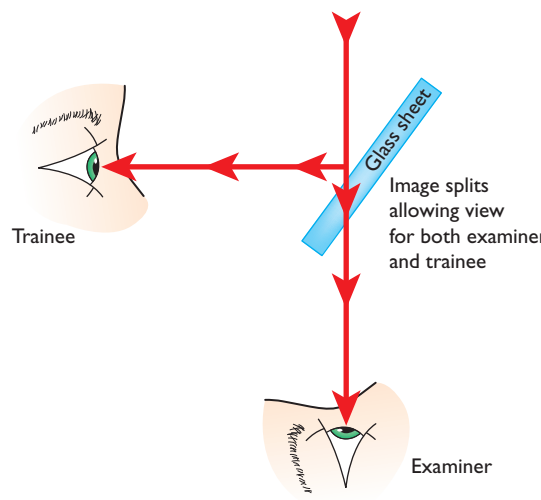


Fig. 8.5 Parallel-sided glass sheet used as an image splitter.
With permission from Louise Bye.

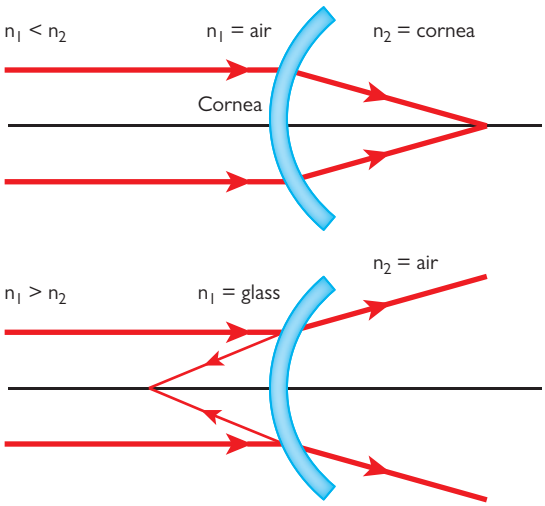


Fig. 8.6 Refraction of light at a convex refracting interface.
With permission from Louise Bye.

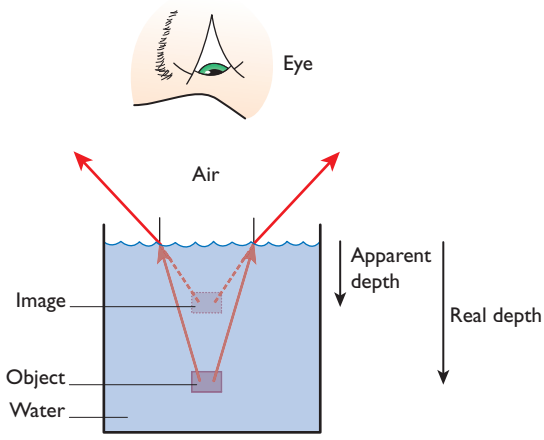


Fig. 8.7 Real and apparent depth.
With permission from Louise Bye.

The refracting power of a spherical surface is given by the equation:

$$\text{surface power (dioptres)} = \frac{n2 - n1}{r}$$

where r is the radius of curvature.

Surface power is seen in the air:cornea example and accounts for the majority of the refraction occurring in the eye.

Objects placed in an optically dense medium are viewed as distorted when seen from a less dense medium, like that of a straw in a glass of water; the image seems closer (Fig. 8.7).

The critical angle and total internal reflection

A ray that strikes the surface of the water from underneath passes through undeviated. As the rays pass through more obliquely, they become more deviated. Eventually, the emerging ray runs parallel to the surface of the water. The incident angle for light rays when this happens is called the critical angle. It depends partly on the refractive indices of the two media. At larger incident angles, light will fail to emerge from the water and this is termed 'total internal reflection'. For the water:air interface, the critical angle is 48.5° and for the glass:air interface the critical angle is 41°.

Optical instruments rely on this principle in prisms, fibre-optic cables, and gonioscopy. Gonioscopy can allow visualization of the drainage angles. Usually, rays from the angle do not leave the eye because of total internal reflection. This is overcome by using a contact lens with a coupling agent such as saline or viscotears. It can also be used to view the peripheral retina. Chapter 10 contains images of ophthalmic instruments.

Dispersion

Light is made up of many component wavelengths. The RI of any medium differs slightly for light of different wavelengths. Shorter wavelengths (blue) are deviated more than longer ones (red). The angle between the red and blue light is the dispersive power of a material and is not related to the RI.

This is the basis for the chromatic aberration that affects optical systems and the eye.

Prisms

A prism is a transparent optical element with flat, polished surfaces that are inclined at an angle to each other such that the prism refracts light. The angle (α) between the two surfaces is called the refracting angle or apical angle of the prism.

The orientation of the prism is described by the position of its base. Light passing through the prism obeys Snell's law and thus light is deviated towards the base (Fig. 8.8).

The total change in angle is the angle of deviation (D). D depends on:

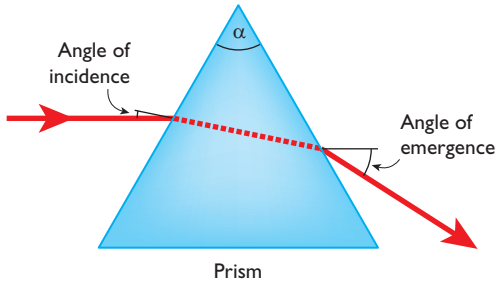


Fig. 8.8 Passage of light through a prism.
With permission from Louise Bye.

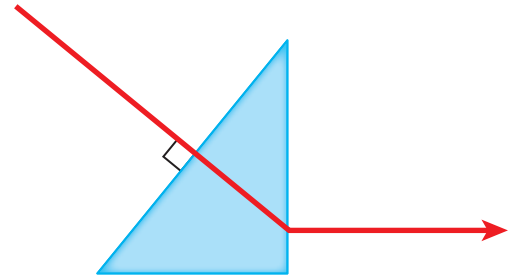


Fig. 8.10 The Prentice position of a prism.
With permission from Louise Bye.

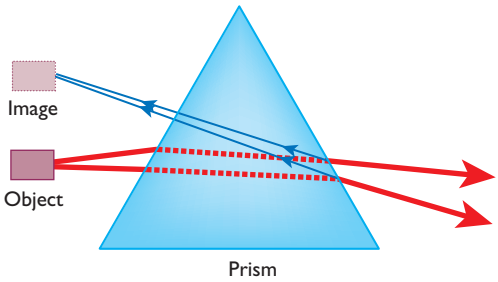


Fig. 8.9 Image formation by a prism.
With permission from Louise Bye.

- the refractive index of the prism material
- the refracting angle, α, of the prism
- the angle of incidence (*i*) of the ray considered.

For any prism, *D* is least when the angle of incidence equals the angle of emergence.

Refraction is now symmetrical and the angle is known as the angle of minimum deviation. Here the angle of deviation = $(n-1)\alpha$, where *n* is the refractive index of the medium. This means that the angle of deviation will be half that of the refracting angle for a glass prism with refractive index 1.5. A 1 prism dioptre power prism with a refracting angle of 1° produces an angle of apparent deviation of 1/2°.

The image formed by a prism is erect, virtual, and displaced toward the apex of the prism (Fig. 8.9).

The prism can also be put in the Prentice position. Here, one surface is perpendicular to the incident ray of light, so all deviation takes place at the second surface of the prism. In this position, the prism is of greater power than in the position of minimum deviation because in the Prentice position the angle of incidence does not equal the angle of emergence (Fig. 8.10).

Glass ophthalmic and trial frame lenses are normally prescribed in the Prentice position. However, for plastic prisms such as prism bars, it is the power in the position of minimum deviation that is specified.

Prisms cannot be stacked as the incident light will not enter at the correct angle and the sum of the prisms will not equal the effective power of the prisms together. A horizontal and vertical prism may be stacked, however, and the diagonal power is calculated using Pythagoras' theorem (the square of the diagonal equals the sum of the squares of the vertical and horizontal sides). A Risley prism used in conjunction with a Maddox rod to measure phorias is an example of this.

Notation of prisms

The power of any prism can be expressed in various ways.

Use of prisms

Prisms are used extensively in the diagnosis and measurement of ocular motility problems.

Therapeutic prisms can be used for convergence insufficiency to improve fusional reserve. Prisms are base-out during exercise periods.

Prisms are also used to relieve diplopia in decompensated heterophorias, small vertical squints, and some paralytic squints. Prisms can be permanently incorporated into glasses or Fresnel prisms can be stuck to an existing pair of glasses for immediate or temporary relief of diplopia.

HELPFUL HINT

Notation of prisms

The Prism dioptre (Δ): a prism of 1 prism dioptre power (1^{Δ}) produces a linear apparent displacement of 1 cm of an object situated at 1 m.

The angle of apparent deviation is the angle of apparent displacement of the object. A prism of 1 prism dioptre power produces an angle of apparent deviation of 1/2°.

CLINICAL TIP

Diagnostic prisms

- Measure angle objectively by prism cover test.
- Measure angle subjectively by Maddox rod.
- Measure fusional reserve to assess binocular vision.
- Measure microtropias using the 4-dioptre prism test.
- Assessment of simulated blindness.

CLINICAL TIP

Prismatic correction

When prescribing prisms the correction is normally split between both eyes.

- To correct convergence (esotropia) = base-out prism.
- To correct divergence (exotropia) = base-in prism.
- To correct vertical deviation = prisms are opposite for the two eyes, e.g. RE base-down, LE base-up for an R hypertropia.

Spherical lenses

A lens can be defined as a transparent optical medium that allows for the convergence or divergence of light rays to form an image. Lenses can take several forms, but if a refracting surface forms part of a sphere they are known as spherical lenses.

Essentially, spherical lenses can fall into two major categories:

1. Convex lenses: these cause the convergence of incident light.
2. Concave lenses: these cause the divergence of light.

A plane surface can be thought of as a sphere with a radius of curvature of infinity.

The total vergence power of a *thin* lens is the magnitude by which the lens can either converge or diverge light. It is defined as the sum of the two surface vergence powers. In thick lenses, the thickness of the lens must also be taken into account. By convention the vergence powers (Fig. 8.11) are designated as positive for convex lenses and negative for concave lenses.

Fig. 8.11 is a ray diagram illustrating the vergence power of thin convex and concave lenses.

In Fig. 8.12 the nature of the lenses AB is simplified to arrows and brackets. The principal axis passes through the nodal/principal point, N, and the principal focus, F. Because the lenses are transparent and have the same medium either side of them, any light travelling in the opposite direction to convention will get refracted and have a principal focus. Therefore, in Fig. 8.13 there are two principal foci, F_1 and F_2.

The first principal focus (F_1) is the point of origin of rays of light which, after refraction, are rendered parallel. The distance from F_1 to N is the first focal length, f_1.

Incident light parallel to the principal axis is converged to or diverged from the second principal focus, F_2. The distance from F_1 to N is the second focal length, f_2.

The position of F_2 is positive if it lies to the right of N and negative if it lies to the left of N. Hence lenses are numbered by their second focal length, so convex or converging lenses are plus lenses (+) and concave or diverging lenses are minus (–) lenses.

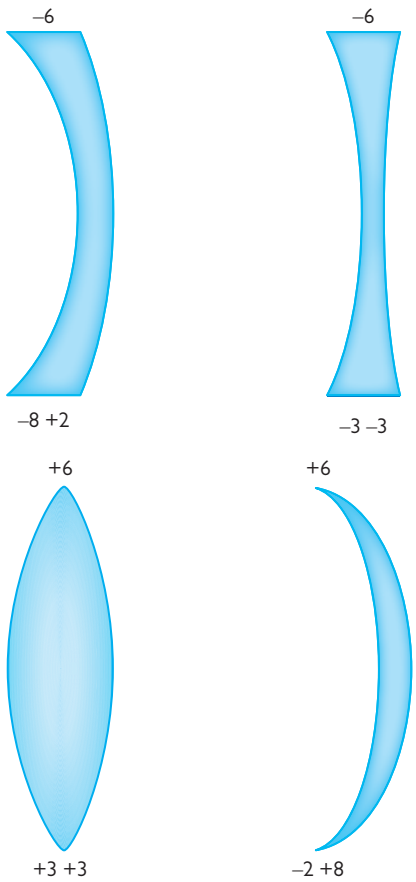


Fig. 8.11 Vergence power of thin spherical lenses.
With permission from Louise Bye.

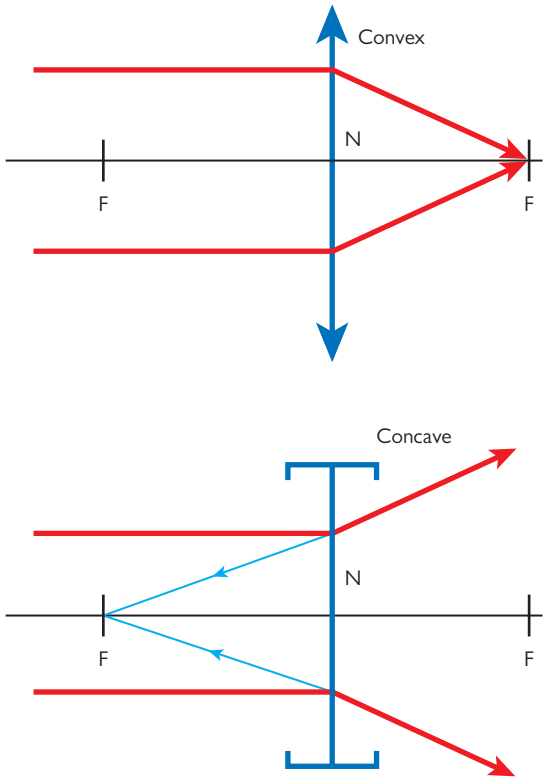


Fig. 8.12 Cardinal points of thin spherical lenses.
With permission from Louise Bye.

The thin lens formula

The position of the image can be calculated as follows:

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f_2}$$

Here, v is the distance of the image from the principal point, u is the distance of the object from the principal point, and f_2 is the second focal length.

The image produced by a lens for a given object (Figs 8.14 and 8.15) can be constructed as follows:

1. A ray of light passes from the object through the nodal point, N, undeviated.
2. A ray of light also passes along the principal axis undeviated.
3. A ray of light parallel to the principal axis, which reaches the principal plane and after this passes through (+) or away from (–) the second principal focus, F_2.

Lens power

A more powerful lens has a greater ability to converge or diverge light. The vergence power in dioptres (D) is the reciprocal of the second focal length (metres):

$$F = \frac{1}{f_2}$$

For example, for a convex lens of second focal length 20 cm:

$$F = \frac{1}{f_2} = \frac{1}{0.20} = +5D$$

whereas for a concave lens of focal length 30 cm:

$$F = \frac{1}{0.30} = -3.33D$$

Linear magnification

It can also be seen above that a convex lens has magnification. The image, I, subtends a larger angle at the eye than the object, so when the object is placed between the nodal point and the first principal focus the magnification is:

$$\text{magnification} = \frac{I}{O} = \frac{v}{Ou}$$

where I is the image size, O is the object size, v is the distance of the image from the principal plane, and u is the distance of the object from the principal plane.

If the object is placed exactly at the first principal focus, the image is infinitely distant to the object. Both subtend the same angle at the lens. The angular magnification is therefore 1.

This can be used practically and is known as a 'loupe' (Fig. 8.16). It allows viewing at a much closer distance than would be possible unaided because the near point of the eye in most adults is 25 cm because of a limitation of accommodation. As a result of this apparent closer viewing, the object subtends a greater angle at the eye and a larger image is thrown onto the retina.

Here, an object can be moved from 25 cm to a position f:

magnifying power (M) = apparent size of image/apparent size of object (at 25 cm):

$$M = \tan\theta_2 / \tan\theta_1$$

- $\tan\theta_1 = O/0.25$
- $\tan\theta_2 = O/f$

therefore:

$$M = O/f \times 0.25/O = 0.25/f$$

- $1/f$ = F dioptres

therefore:

$$M = F/4$$

Hence an 8× loupe must have a power of 32 D.

Lens decentration

Rays of light that approach a lens at its periphery are refracted towards or away from the principal axis. This is not constant

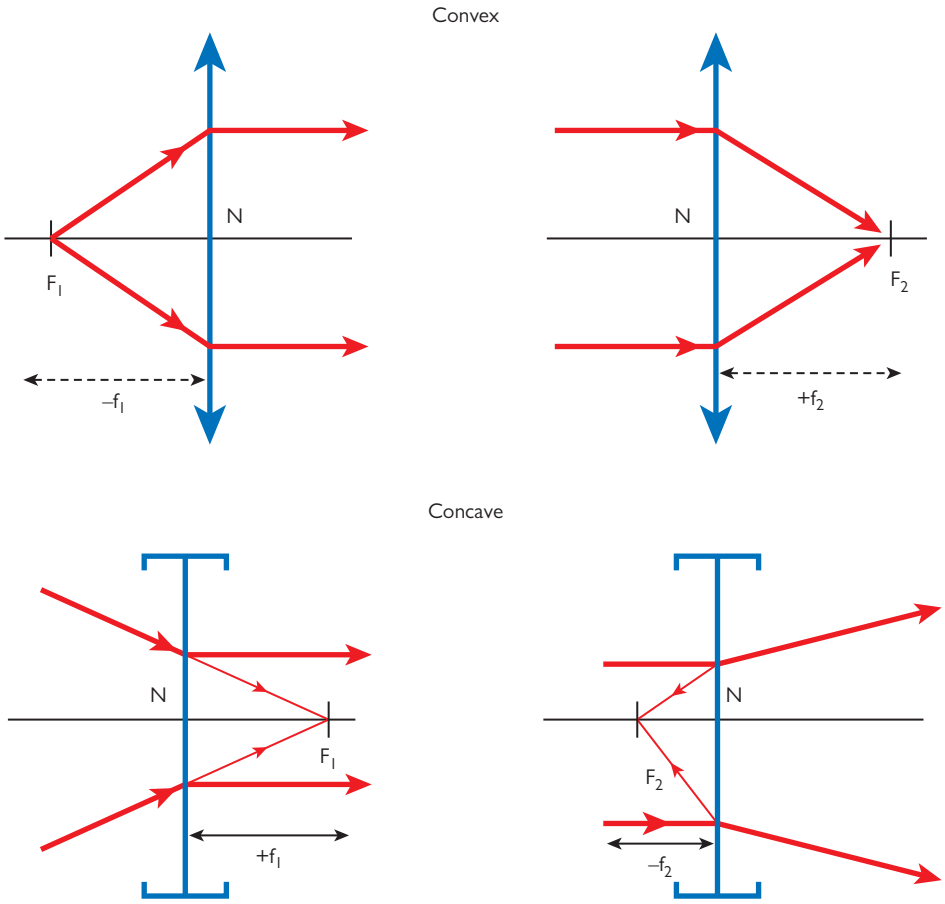


Fig. 8.13 The principal foci of thin spherical lenses.
With permission from Louise Bye.

across the lens and the prismatic effect increases further out in the periphery of the lens. Use of a non-axial portion of the lens for prismatic effect is called decentration. This is the underlying mechanism for spherical aberration.

The prismatic power can be calculated as follows:

$$P = F \times D$$

where P is the prismatic power in dioptres, F is the lens power in dioptres, and D is the decentration in centimetres away from the visual axis.

Astigmatic lenses

Meridians are used in optics to define a reference to the direction of astigmatism.

Astigmatism is a difference in the degree of refraction in two or more meridians, producing two or more foci.

Cylindrical and toric lenses are the two types of astigmatic lens.

Cylindrical lenses

This type of lens contains part of a cylinder and one plane surface. A plane of light traversing the lens in the same plane as the axis would pass through the lens unaltered. This is because a cylindrical lens has no power along its axis. Instead, the power of a cylindrical lens exists 90° to its axis.

What is produced by these cylindrical lenses is a line image of a point object, O, which is known as the focal line (Fig. 8.17).

The Maddox rod

The Maddox rod may be used in the diagnosis of extraocular imbalance. It constitutes a series of convex cylindrical

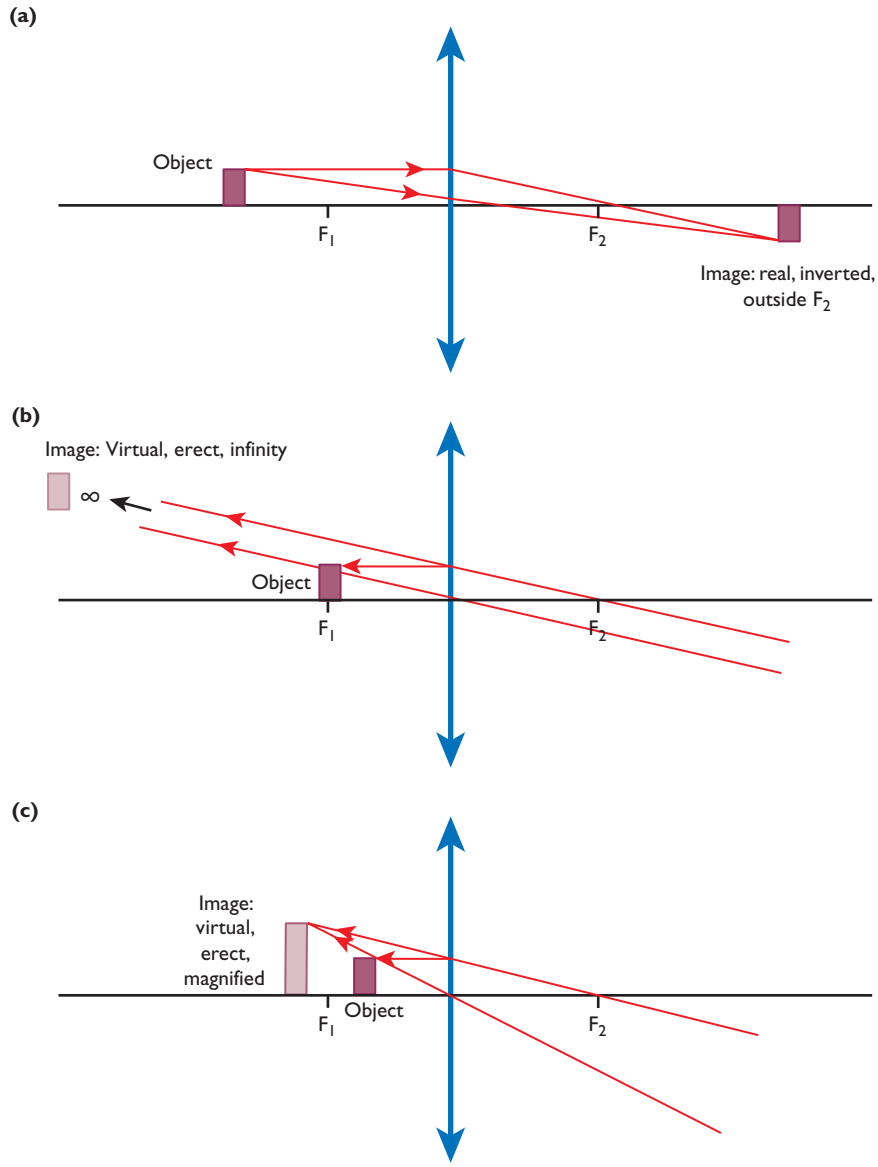


Fig. 8.14 Image formation by a thin convex lens.
With permission from Louise Bye.

lenses placed next to each other in a row. These cylindrical lenses may be put into a trial lens frame.

To use the Maddox rod, a point source of light or spot light is placed in the distance so that incident light is parallel. The Maddox rod is then placed in front of the eye. The image that the patient sees is described as follows.

Light in the plane of the axis of the cylinder is brought to a point focus by the eye.

Several planes of light in the same cylindrical axis are brought to multiple foci by each cylinder creating a row of dots. This row of foci is interpreted by the eye as a line of light lying 90° to the axis of the Maddox rod (Fig. 8.18).

Light in the plane 90° to the cylinder power is brought into focus between the rod and the eye, and multiple planes of light are brought to multiple foci between the eye and the rod. This real image is well within the near point of the eye and the patient is therefore unable to distinguish a clear image. The light instead is scattered over a wide area in the retina and is interpreted as diffuse illumination.

To summarize, placing the Maddox rod over an eye will cause the eye to see the distant spot light as a streak 90° to the cylindrical axes of the Maddox rod.

The Maddox rod is used to test muscle imbalance by dissociating the eyes by presenting dissimilar images of the spot light in the background and the streak in the foreground. This establishes a muscular resting state to test.

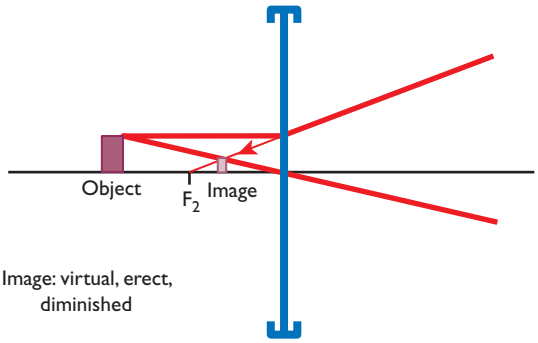


Fig. 8.15 Image formation by a thin concave lens.
With permission from Louise Bye.

Toric lenses

A toric surface is curved in two meridians. The meridians of minimal and maximal curvature are called the principal meridians. In ophthalmic lenses these lie exactly perpendicular to each other. The principal meridian of minimum curvature is the base curve.

Toric lenses form two foci of images in planes that are perpendicular to each other. The image lies somewhere between the two in an arrangement called Sturm's conoid. The plane where the two rods of light intersect is called the circle of least confusion.

The lenses can be thought of as a sphere with an overlying cylinder. They can be written as a fraction with the spherical power as the numerator and the cylindrical power the denominator, e.g. a toric lens with power +2.0D

CLINICAL TIP

Use of Maddox rod to test muscle balance

The Maddox rod is placed in front of the right eye (in the trial frame).

A white spot of light is used in the light box and the ambient lighting is dimmed.

The right eye will see a line of light that appears as a straight line perpendicular to the Maddox rod.

The left uncovered eye will see a white spot of light in the light box.

This in effect has dissociated the eyes as they see dissimilar images.

To test for horizontal imbalance, the rod must be horizontal to give a vertical line of light.

If the patient sees the line going through the spot, then s/he is orthophoric. If the line is to the left of the spot, the patient has exophoria (not tropia as the eyes are dissociated). If the line is to the right of the spot, the patient is esophoric. If the line is below the spot the patient has a right hyperphoria and if it is above the line the patient has a right hypophoria or left hyperphoria.

It must be remembered that the eye behind the Maddox rod is moving in the opposite direction to that indicated by the red line. So, if it turns out in exophoria, the line will be seen in the nasal field and will appear to the left of the spot.

Similarly, vertical imbalance may be measured. If the line is below the spot, the right eye is hyperphoric and vice versa.

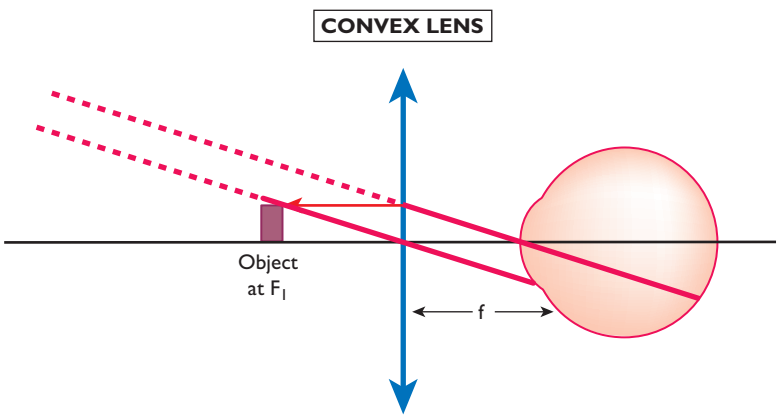


Fig. 8.16 The simple magnifying glass (the loupe).
With permission from Louise Bye.

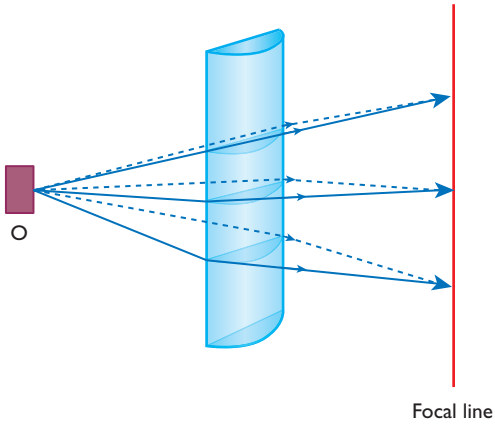


Fig. 8.17 Image formation by convex cylindrical lens of point object, O.
With permission from Louise Bye.

in one meridian and +6.0D in the other can be written as +2.0DS/+4.0DC.

The spherical equivalent is helpful to know the closest overall effect and it coincides with the circle of least confusion. It is simply the sum of the spherical power plus half the cylindrical power. In the above example, the spherical equivalent is +4.0DS. A +2.0DS/–2.0DC toric lens has a spherical equivalent of +1.0DS.

The Jackson's cross-cylinder (Fig. 8.19) is used during refraction to refine the power and axis of a cylinder. It is a sphero-cylindrical lens in which the power of the cylinder is twice the power of the sphere. This is the same as superimposing two cylinders of opposite but equal power perpendicularly onto each other. The handle is placed at 45° to the axes.

Note that the axes marked on the cylinder are perpendicular to the refractive power, as is the convention with all cylinders.

CLINICAL TIP

Use of Jackson's cross-cylinder

The cross-cylinder is used during subjective refraction. It is held first with its axis along the line of the trial cylinder in the frame as determined by retinoscopy. The cross-cylinder is rotated and the patient is asked which gives the clearer image. The axis of the trial cylinder is rotated in the trial frame slightly towards the axis of the same sign in the cross-cylinder. This is repeated until the rotation gives equally different vision in either direction.

This cross-cylinder can also be used to work out the power of the cylinder. The cross-cylinder is first held with one axis and then the other in line with the trial cylinder. This increases or decreases the power of the trial cylinder.

If retinoscopy does not reveal astigmatism, the cross-cylinder can be used to confirm this. It is held in four orientations: 90°, 180°, 45°, and 135°. If the patient does not volunteer an improvement in vision in any orientation, it can be assumed that no cylindrical correction is necessary.

Notation of lenses

A spherical lens can be written simply as +3.00DS.

The lens power is the sum of the surface powers and a +4.0D spherical lens can be in any of the forms shown in Fig. 8.20.

A cylinder is recorded by its axis, which is marked on the trial frame.

A cylinder placed vertically therefore of –1D is written as –1DC axis 90°.

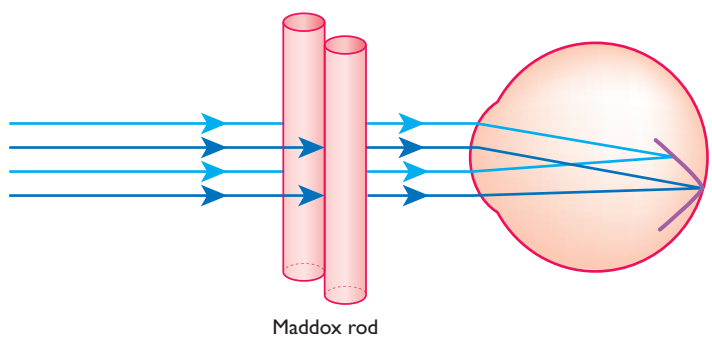


Fig. 8.18 Optics of Maddox rod. Formation of line of foci by adjacent elements of Maddox rod.
With permission from Louise Bye.

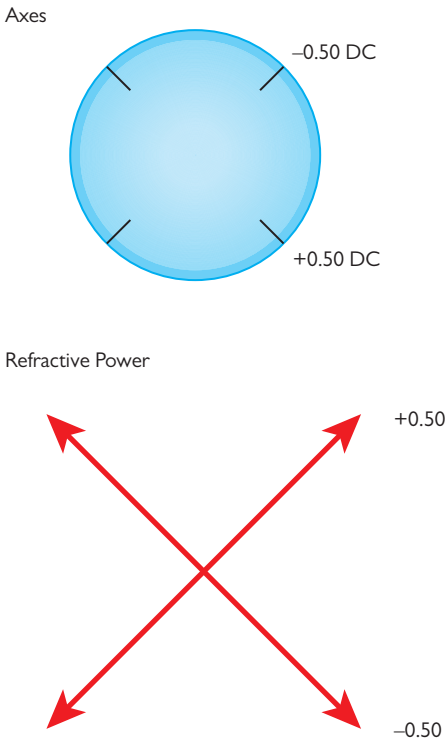


Fig. 8.19 The cross-cylinder showing axes as marked on the lens and refractive power in the principal meridians.
With permission from Louise Bye.

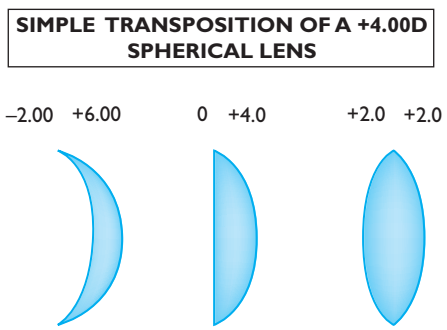


Fig. 8.20 Simple transposition of a +4.0D spherical lens.
With permission from Louise Bye.

Both the spherical and cylindrical properties of a lens must be written. A combination of the above two lenses would be written as +3.00DS/–1.00DC × 90°.

This lens form can be written in two ways (transposition of lenses) as the cylinder can be considered to act as a –1.00DC at 90° or a +1.00DC at 180° with different values of DS. This lens has a power of +2.0D with an axis of 90° (i.e. acting horizontally) and +3.00D with an axis of 180° (i.e. acting vertically). It can also therefore be written as +2.00DS/ +1.00DS × 180°.

A toric lens has one spherical surface and one toric surface. The weaker power of the toric surface is known as the base curve of the lens.

Toric transposition

If the denominator is given in a sign opposite to that of the base curve, then the formula must be rewritten because a plus cylinder cannot be transposed onto a minus cylinder. In this example, the cylinder is minus, as it is the base curve, so there is no problem.

The formula +3.0DS/–2.0DC axis 180° with base curve +6D must be transposed because the base curve is the opposite sign to the cylinder that is to be projected onto it.

To achieve this, the prescription must first be transposed to give:

+1.0DS/+3.0DC axis 90° with base curve +6D

The base curve must remain the same, so the new prescription must be fitted to it.

The power of the anterior spherical surface can be worked out by subtracting the base curve from the spherical power required of the lens:

+1.0 – (+6.0) = –5.0DS

The axis of the base curve can be specified. It must remain fixed at +6D and its axis will be perpendicular to the axis of the steeper meridian (because it is by definition the flatter meridian):

+6.0D axis 180°

The axis of the cylinder is added to the power of the base curve with its axis as above:

+9.0D axis 90°

This complete formula is therefore:

–5.0DS
+6.0DC axis 180°/+9.0D axis 90°

Identification and neutralization of lenses

Identifying lenses without equipment is frequently useful in clinical practice. When a lens is moved from side to side, the image moves with or against according to the optics of the lens, as shown above. A convex lens will give an 'against movement' and a concave lens will give a 'with movement'. If looking at a cross, astigmatic lenses will cause the lines of the cross to 'scissor'.

Once the unknown lens is determined to be a plus (convex) or minus (concave) lens, another lens of opposite sign is put in front of it. Looking at an image through the two lenses, the known lens of opposite power is changed until there is no movement of the image. The sign of the known lens is inverted and the number is recorded. With astigmatic lenses, this can be performed in the same way but whilst moving the lens in two different meridian to test the different axes of the lens.

Spectacle lenses are designated by their back vertex power, so the known lens must be put to the back surface of the unknown spectacle lens. This may be difficult with curved lenses and may be more accurate if the known lens is placed at the front of the spectacle lens in some cases.

Lens measure

As thin lenses are a function of their surface powers, a lens measure may be used to determine the surface curvature and thus the power of the lens.

The Geneva lens measure (Fig. 8.21) uses three pins as points of contact, where the middle pin is attached to a spring-loaded measure. By pressing the Geneva lens measure on the lens the power is indicated by an arrow. Note that the Geneva lens measure was specifically designed to measure thin crown glass lenses, and thus a correction must be applied in the case of other materials.

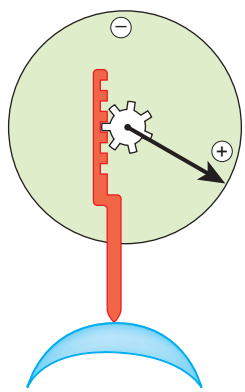


Fig. 8.21 Principle of Geneva lens measure.
With permission from Louise Bye.

The focimeter

The focimeter (Fig. 8.22) has five components:

1. viewing telescope
2. collimating lens (a lens that renders light parallel)
3. green light source
4. moveable target, along a measure rule (in dioptres)
5. the lens to be tested.

The green source retro-illuminates the target, which is usually a disc containing a ring of perforations. The green light is used to minimize chromatic aberration.

The rays of light transmitted by the ring of dots are made parallel by a collimating lens if the target is situated at the first principal focus of the collimating lens. This indicates that the focimeter is ready to use, as the position of the target should be zero on the rule if the focimeter is correctly calibrated.

In this position, the observer should be able to view the target through the collimating lens as sharp, if the telescope is set to focus at infinity (i.e. parallel rays).

The eyepiece within the telescope contains a rotatable graticule and a protractor to help identify the axis of a cylinder. At the beginning all components should be zero whilst maintaining a sharp view of the target.

The test lens is placed in a frame, which coincides with the second principal focus of the collimating lens.

Subsequent displacement of the target will change the vergence of light arriving at the test lens. If the test lens transmits parallel light, the target is observed as sharp and the end point has been reached. The position of the target may then be noted. This position is calibrated to indicate the dioptric power of the lens tested.

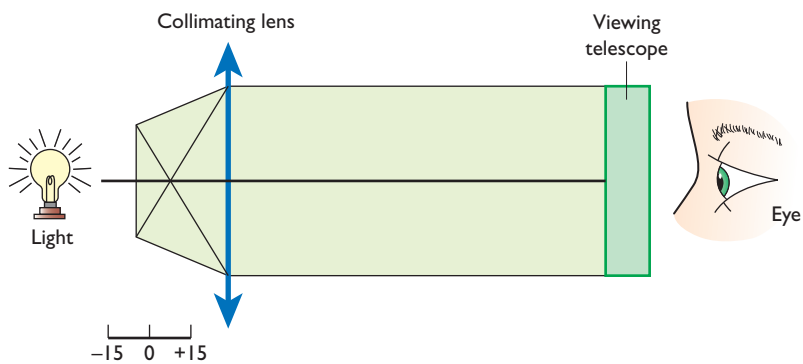


Fig. 8.22 The focimeter.
With permission from Louise Bye.

Remember that the spectacles or test lens must be placed with the back surface towards the target as focimeters measure the back vertex power.

Focimeters and astigmatism

In spherical lenses, the end point is seen as a focused ring of dots. The end point of one meridian in an astigmatic lens is seen as a focused collection of parallel lines (a stack of sticks). This represents the elongation of the ring of dots by the cylinder. The orientation of the focused parallel lines signifies the direction of the power as opposed to the axis of the cylinder.

The instrument should then be adjusted to find the power and direction of the remaining meridian. This allows the user to create a power cross. For example, the above power cross is obtained by focimetry. The resultant prescription must then be:

$$+2.0DS/-1DC \times 180$$

Focimeters and bifocals

The front vertex distance should in both near and distance zones be measured in bifocals, as the power of the near add is a function of the anterior surface. The near add is calculated as the positive difference between the power recorded at the near and distance zones.

Focimeters and prisms

If a test lens contains a prism, the appearance of the target will be displaced because the image is inverted by the telescope.

Thus, provided the instrument was correctly set at zero from the start, even if the target is sharp, the ring of dots will be displaced. The graticule within the telescope contains a rule, which may be aligned with the direction of displacement. Each step of displacement on the rule corresponds to one prism dioptre.

Prismatic effect may also be produced by decentration of the optical centres. Decentration is very easy to miss if the optical centre is not marked. Here, the optical centre of the test lens must be clearly marked and compared with the patient's inter-pupillary distance (IPD) in order to establish the magnitude of decentration prescribed. Otherwise, the spectacles will appear as any normal prescription until placed on the patient, where the line of sight doesn't coincide with the optical centres of the lens.

Prism power in optical decentration is given by the formula:

$$P = F \times D$$

where P is the prismatic power (in prism dioptres), F is the power of the test lens, and D is the distance of decentration (in centimetres).

Automated focimetry

Here a computer measures the level of deflection of rays as they traverse the lens in order to establish the prescription. This deflection is dependent on the focal and prismatic power of the lens and the level of decentration from the optical centre.

Four lines are projected through the lens and captured on a sensor. Knowing the original path and the level of deflection of the rays allows for the calculation of the spectacle prescription.

Tinted lenses and filters

Tinted lenses can either reflect or absorb unwanted light in specific wavelengths. A good example is of UVA-, UVB-, and UVC-protected sunglasses or protective goggles in laser therapy. This technique can be applied to any undesired wavelengths.

A neutral density tint reduces the transmission of all wavelengths by absorption and so does not alter the spectrum of light.

Filters apply the same principle and wavelength selection and can be added to different optical devices as an adjunct. For example, most slit lamps contain a heat filter to protect the patient against the thermal effects of the infrared spectrum.

Photochromic lenses change their transmission characteristics depending upon the intensity of incident radiation: the lens becomes darker in brighter light. Glass photochromic lenses comprise colourless silver halide crystals suspended in borosilicate. Electromagnetic energy dissociates the silver and halogen to cause darkening.

Antireflective coatings

Coatings of different refractive index to the lens may be used to cancel out reflection.

These coatings adjust reflected light such that they emerge at one half of the wavelength out of the phase than that of the incident ray, resulting in destructive interference. This is achieved by a coating of thickness that is one quarter of the wavelength of the undesired light. This produces a quarter phase shift on incidence and a further quarter shift on emergence.

A coating with the thickness of half the undesired wavelength will produce constructive interference and a mirror effect.

Aberrations of lenses

In every optical system there is imperfection. It is important we understand these aberrations so that we may construct solutions.

Chromatic aberrations

Remember that white light consists of many wavelengths and that the longer the wavelength, the less it is deviated by

a refracting medium. This dispersion phenomenon is known as chromatic aberration.

When white light is incident on a simple spherical lens, the same chromatic aberration occurs and is more apparent the further away from the principal axis.

Achromatic lens systems

The properties of dispersion and refraction are mutually exclusive and thus we may select materials that are of an appropriate refractive index but low in dispersion.

By selecting a lens with a high refractive power and low dispersive power to counter a lens with a low refractive power and high dispersion, it is possible to cancel out the dispersive effects of the total optical system whilst maintaining the desired refractive properties.

Chromatic aberration in the eye

In the eye, the dioptric difference between the dispersion of red versus blue light is approximately 2D (Fig. 8.23). Remembering that the human eye's peak spectral sensitivity lies at 550 nm (green), the human retina is deliberately placed such that it is in between the dispersion of white light, i.e. between red and blue. This optimizes the best level of focus specifically for its peak spectral sensitivity. The pupil and the nucleus of the lens also help to minimize chromatic aberration.

Duochrome test

The duochrome is a plate seen on many visual acuity charts. It appears as two coloured backlit panels one on top of the other, each with black letters. The top panel is red and the bottom is green. If you are emmetropic or rendered emmetropic, placing a +1D lens over the eye will shift the dispersed spectrum forward and the letters on the red panel become clearer, and vice versa with a –1D lens.

We may test the accuracy of a spectacle correction by ensuring that the appropriate colour of light is focused onto the retina. This is especially the case when too strong a negative lens has been used for a myopic eye, thus rendering the patient more hypermetropic. Patients often dislike this as they are forced to accommodate to overcome this and exhibit symptoms of strain. In this case, the duochrome test would have shown that the green panel was far clearer than the red and the necessary adjustments could then be made to allow the myope to see the red panel more clearly.

Given that the patient needs only to indicate the position of the panel that is clearest, the duochrome can even be used with those who are colour blind.

Spherical aberration

This refers to the prismatic effect of a spherical lens as discussed in optical decentration on p. 215. Here, if the incident light is further away from the principal axis, a stronger prismatic effect is exerted. This causes the peripherally dispersed light to disturb the clarity of the centrally focused rays, thus blurring the image.

Correction of spherical aberration

Aplanatic surfaces are lenses by which the anterior, peripheral refracting curved surface is less than the central optical zone. This attempts to lessen the prismatic effect where it is most needed.

Aspherical lenses employ the use of a weaker opposing lens of a different refractive index fused with a main lens. This way the refractive properties are relatively maintained but the peripheral prismatic effect, and therefore spherical aberration, is neutralized. Because doublets make use of different refractive media a combination of aspheric and achromatic lenses is common.

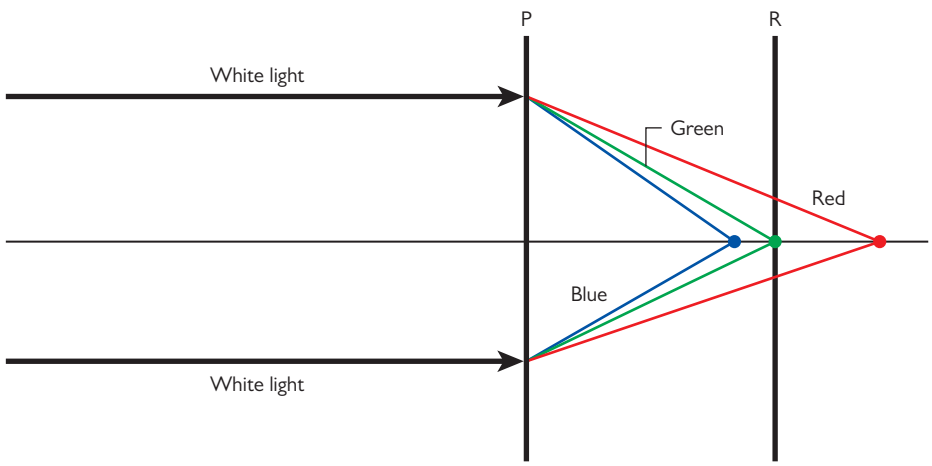


Fig. 8.23 Chromatic aberration, emmetropic eye. P = principal plane, R = retina.
With permission from Louise Bye.

Spherical aberration in the eye

There are four main ways that the eye can reduce spherical aberration.

- *Corneal solutions:* the cornea itself acts as an aplanatic surface whereby the central optical zone is steeper than that of the periphery. This decreases the vergence power at the periphery and lessens the prismatic effect.
- *Aperture solutions:* the pupil helps restrict the incident light into the para-axial zone of the lens, thus preventing the more aberrant prone peripheral rays. The pupil is most effective at 2–2.5 mm in diameter.
- *Lenticular solutions:* the natural lens has a central nucleus, which is of a higher refractive index than the cortical matter. This means that at the periphery, rays that are typically affected by a greater prismatic effect are balanced with a lessened refractive effect, thus reducing spherical aberration at the level of the lens.
- *Minimizing aberrant rays with the Stiles–Crawford effect:* the photoreceptors have a decreased response to light which is obliquely incident upon them. This means that the greatest response is likely to come from correctly refracted para-axial rays. This phenomenon is known as the Stiles–Crawford effect.

Oblique astigmatism

All optical systems are typically designed to refract incident light, which occurs along the principal axis. If the incident light is oblique to the principal axis, a toric effect results. This is, naturally, directly proportional to the magnitude of the lens power and, as we know, this leads to the production of Sturm's conoid (Fig. 8.24).

We can replicate oblique astigmatism in real terms by making the line of sight oblique to the principal axis on a pair of glasses.

This is to be distinguished from oblique astigmatism of the eye, in which the principal meridians do not lie at 90° and 180°, but remain mutually perpendicular.

This effect is greater in more powerful lenses as they have a greater cylindrical effect. To overcome this, spectacles are made with a downward tilt (pantoscopic tilt), especially in the case of high prescriptions and in the near-reading portion, where the oblique astigmatism is more profound.

Oblique astigmatism is also noted to be worse in biconvex or biconcave spectacle prescriptions. By using the equivalent meniscus form of the lens system, the oblique astigmatism is reduced.

Best form spectacles employ vast data and tables to give the optimum configuration of anterior and posterior curvatures to reduce both spherical and oblique astigmatism.

The eye reduces oblique astigmatism by:

- the aplanatic nature of the cornea
- the natural curve of the retina—this allows obliquely focused light to form a circle of least confusion on the retina in most angles

The fact that the peripheral retina is not intended for fine detail means that its oblique astigmatism has a minimal effect on vision.

Coma

Coma is, loosely, a combination of oblique rays and spherical aberration. Light originating from points away from the visual axis pass obliquely through the centre and the periphery of a lens, causing two focal points to occur. This gives rise to the superimposition of a main image and a blurred, elongated image. A circular target will appear as an elongated shape, or coma.

Coma is reduced in the eye by the restriction of rays to the principal axis and, like oblique astigmatism in the eye, the effect is often negligible.

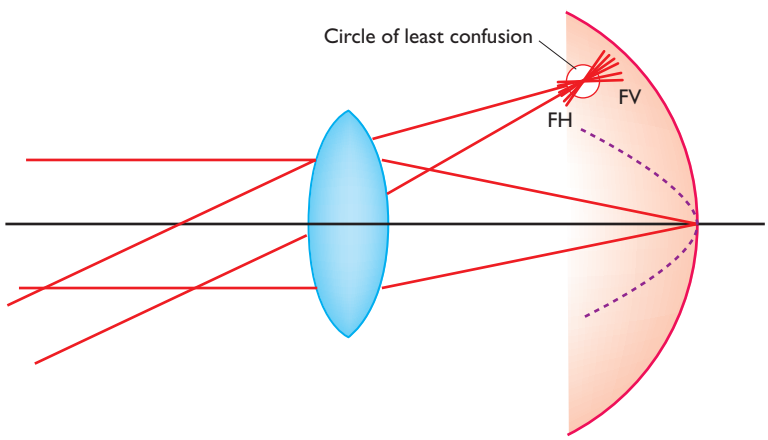


Fig. 8.24 Oblique astigmatism. FH and FV represent the horizontal and vertical line foci of a Sturm's conoid, respectively.
With permission from Louise Bye.

Distortion

The same increased prismatic effect in the periphery of a lens can account for distortion of the edges of an object. When viewing a square, high-power convex lenses classically give rise to a pin-cushion distortion, forcing the edges of the square to bow inwards.

A concave lens will produce the opposite effect, forcing the edges outward, which is known as barrel distortion.

Curvature of field

This is where a flat or plane target produces a three-dimensionally bowled image. The retina ingeniously accommodates for this by also being bowled. Note that this effect may take place in spite of having accounted for spherical aberration, oblique astigmatism, and coma.

Optics of the eye

The eye consists of several refracting surfaces and therefore the thin lens formulae as described above are insufficient to deal with the complexity of the eye as they consider refraction only at two lens surfaces. Thick lens calculations are required because the combined power of a thick lens is not simply the sum of the individual thin lenses. Thick lens theory is simplified here and a few new concepts must be introduced (Fig. 8.25).

- The principal axis is the horizontal line all the way across.
- P_1 and P_2 are the principal planes and the points at which they cross the principal axis are known as the principal points. The principal planes are perpendicular to the principal axis. A ray of light incident at a principal plane (P_1) leaves the second principal plane (P_2) at the same vertical distance from the principal axis. The exact position of the principal planes is derived from calculations using lens position, curvature, thickness, and refractive index.
- There are two nodal points, N_1 and N_2, in any optical system (which may lie at the same point). These lie on the optical axis and have the property that a ray passing through the anterior nodal point (N_1) leaves the lens as if it has passed directly through the second nodal point (N_2). It emerges from N_2 at the same angle from the principal axis to that at which it originally approached N_1. The nodal points lie on the principal points only when media on opposite sides of the lens are the same.

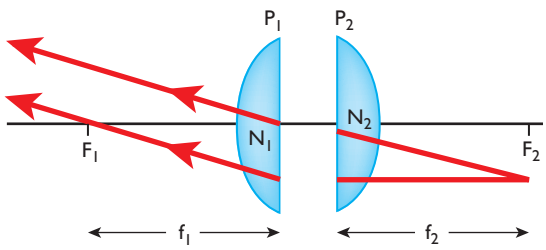


Fig. 8.25 Thick lens. Cardinal points.
With permission from Louise Bye.

- The principal foci (F_1 and F_2) mean the same as for thin lenses and the focal lengths (f_1 and f_2) are the distances between the principal planes to the focal points.
- The nodal points, focal points, and principal points are described as the cardinal points because together they fully describe an optical system.

Equivalent power and back vertex power

It can be seen that the reciprocal of the posterior vertex power of the lens (PVFL) gives the posterior vertex power (or back vertex power). This differs from the equivalent power of the lens, which is calculated from the two surface powers and includes a correction for lens thickness. This difference between back vertex power and equivalent power is a cause of error when prescribing high-power spectacles or contact lenses.

The back vertex power is important in the prescription of spectacles because this is the point at which parallel rays are brought to a focus—which should also correspond to the plane of the retina. Dispensing opticians correct for back vertex power when prescribing glasses, using mathematical tables that exist for this purpose.

Refraction by the eye

Many attempts have been made to understand and model the optics of the human eye. Gullstrand (Professor of Ophthalmology, Sweden) developed a schematic eye that closely modelled the human eye, and won a Nobel Prize in 1911 for his work. The refracting surfaces that need to be considered in the eye are the precorneal tear film/anterior corneal surface, and the anterior and posterior surfaces of the lens. The posterior corneal surface is much less important because of the very small difference in refractive index between the cornea and the aqueous. The *Helpful hint* box summarizes the refractive indices of the media of the eye and the schematic eye cardinal points (Fig. 8.26).

Looking at the lowest part of the table in the *Helpful hint* box, the nodal points and the principal points do not lie at the same point. This is because the media on either side of the lens are not the same. The nodal points are placed around the posterior pole of the lens and, because the pupil only allows a small circular cylinder of light to enter, even a small posterior polar or subcapsular cataract will cause significant symptoms.

HELPFUL HINT

Refractive indices of the transparent media of the eye (Gullstrand)

Air	1.000
Cornea	1.376
Aqueous humour	1.336
Lens (cortex–core)	1.386–1.406
Vitreous humour	1.336
Schematic eye, cardinal points (distance in millimetres behind anterior corneal surface) (Gullstrand)	
First principal point P1	1.35
Second principal point P2	1.60
First nodal point N1	7.08
Second nodal point N2	7.33
First focal point	–15.7
Second focal point	24.4
Refractive power	+58.64D

The reduced schematic eye

As all the points lie close together, many can be combined to form a single intermediate point. This calculation is known as the reduced schematic eye (Fig. 8.27). This has only one nodal point and a single principal plane. Anterior and posterior focal lengths are adjusted around this new principal plane. Note that the posterior focal point (F_2) lies on the retina. Here, the eye is treated as having a single refracting power of +58.6D.

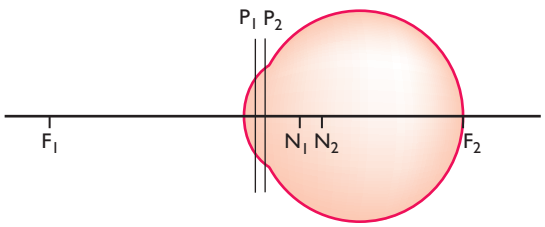


Fig. 8.26 The schematic eye.
With permission from Louise Bye.

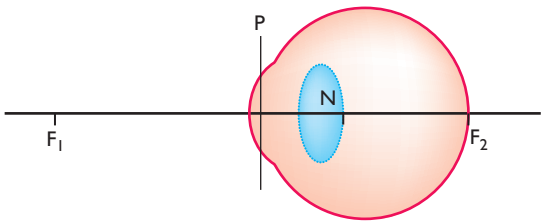


Fig. 8.27 The reduced eye.
With permission from Louise Bye.

HELPFUL HINT

The reduced eye cardinal points (distance in millimetres behind anterior corneal surface) (after Gullstrand)

Principal point P	1.35
Nodal point N	7.08
First focal point	–15.7
Second focal point	24.13

In the aphakic eye, where the cornea is the only refractive element, the calculated power is +43D. The crystalline lens therefore has a power of +15D (58 – 43 = 15D). The cornea is much more powerful than the lens because of the air–fluid interface and the greater difference between the refractive index at either side of the cornea.

Using the reduced eye, a retinal image can be drawn (Fig. 8.27). P is the principal plane and R is the retina. The principal axis corresponds to the optical axis and is perpendicular to P and R. N is the nodal point and lies at the back surface of the lens. F_1 is the anterior focal point of this optical system. F_2 is the second principal focus and falls on the retina in the emmetropic eye.

Two rays of light form this image: first, a ray passing through the one nodal point, undeviated; second, a ray parallel to the first, through the anterior focal point, F_1, which continues parallel to the principal axis after refraction at the principal plane.

The image size can be seen to be proportional to the angle subtended by the object at the nodal point (visual angle). As the object approaches, it will subtend a greater visual angle.

Accommodation and accommodative convergence/accommodation ratio

The crystalline lens can increase its dioptric power by reducing ciliary muscle contraction on the zonular fibres. This allows the lens to assume a more spherical shape. Most change occurs at the anterior lens surface, which therefore moves forward slightly. This ability is called accommodation.

- The far point (FP) is the position of an object where its image falls on the retina in a relaxed eye. In emmetropia, this is infinity.
- The near point (NP) is the nearest point at which an image can be seen with maximal accommodation.
- The range of accommodation is the difference between the two.
- The amplitude is the difference in diopteric power. The diopteric value is the reciprocal of the near and far point distance. The amplitude is given by:

amplitude of accommodation (A) = 1/NP – 1/FP

So, if NP = 5 (reciprocal of 0.25 m) and

FP = 0 (reciprocal of infinity), then A = 4D.

Therefore to focus from infinity to a point 25 cm from the eye, the eye must exert 4D of accommodation.

The accommodative reflex involves accommodation as well as convergence and miosis. There is a link between accommodation and convergence—the accommodative convergence/accommodation ratio (AC/A ratio). The normal AC/A ratio is around 4:1. There are various ways of measuring this:

- the heterophoria method
- the gradient method
- the haploscopic method
- the fixation disparity method.

For simplicity we shall describe the first two, which are the most popular.

The heterophoria method

This method measures the horizontal deviation in distance fixation (optical infinity) with any refractive error fully corrected, on the assumption that no accommodation is present under these conditions.

The horizontal deviation is then measured at near distance (1/3 m or 3D) on the assumption that the convergence exerted is entirely by accommodation alone. Here, the AC/A ratio is given by the following equation:

$$AC/A = IPD + (\Delta d - \Delta n)/D$$

where IPD is the inter-pupillary distance in centimetres, Δd is the prismatic horizontal deviation at distance (prism dioptres), Δn is the prismatic horizontal deviation at distance (prism dioptres), and D is the dioptric equivalent of the near distance.

Remember that the convention of sign for an exodeviation is (–) and for an esodeviation is (+). This is true for heterophoric and gradient methods.

The gradient method

Arguably the most preferred method of AC/A measurement is via the so-called gradient method. The essential points are as follows:

- The drive to change accommodation is given by ophthalmic lenses as opposed to a change in viewing distance.
- It is assumed that the accommodative response to the lenses (and therefore the accommodative convergence produced) is linear within a certain range.
- For a fixed distance, a minus lens diverges light, forcing the lens to adapt by means of accommodation in order to maintain a sharp image, whilst plus lenses relax accommodation.
- It is also assumed that using a –1D lens is equivalent to 1D of accommodation, whereas a +1D lens would reduce accommodation by 1D.

For a given fixation distance the AC/A ratio inferred from the effect of ophthalmic lenses may be readily ascertained from the simple formula:

$$AC/A = (\Delta_o - \Delta_L)/P$$

where Δ_o is the original horizontal deviation without a lens in place (in prism dioptres), Δ_L is the horizontal deviation with a lens in place (in prism dioptres), and P is the power of the accommodation-inducing lens.

For example, supposing the original deviation for a given fixation distance was an exodeviation of 2Δ. If –2D lenses induced an esodeviation of 8Δ, the AC/A ratio would be:

$$AC/A = 8 - (-2)/2 = 5/2$$

The gradient method gives an AC/A ratio that is less than the heterophoria method. This is largely due to the effects of proximal convergence.

AC/A and age

AC/A ratios tend to increase with age given the decline of accommodation. However, patients tend to adapt by stopping short of overconverging as the stimulus of diplopia probably overrides the blur stimulus. Instead, the target is placed further and further away to account for the distancing near point.

Clinically, a high AC/A ratio will present as a convergence excess esotropia where the eyes are straight for distance but converge excessively for near. This can be controlled with bifocal spectacles, which reduce the accommodation required for near and therefore the convergence. Surgical correction may also be performed in the most severe cases.

Catoptric images (Purkinje–Sanson images)

These are an indirect method of measuring changes in lens configuration. They depend on the reflection from each refracting element. Four images are seen: anterior (I) and posterior corneal (II) surfaces and anterior (III) and posterior lens (IV) surfaces.

In Fig. 8.28 the images are shown as seen by the examiner.

Images I, II, and III are real, erect, virtual images as they reflect from a convex surface. Image IV is a real, inverted image as it is formed from a concave surface.

The actual images appear further back/posterior to where they appear because of refraction of the reflected light.

The first image is often used to study corneal curvature, using Placido's disc or keratometry. It is also used when assessing strabismus.

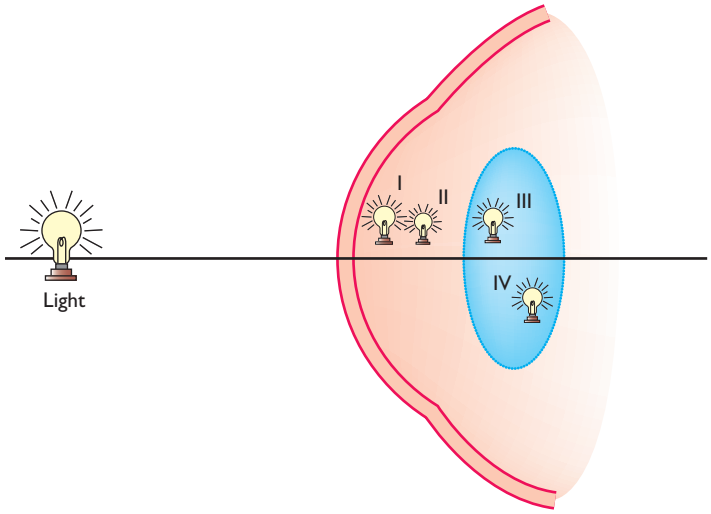


Fig. 8.28 Purkinje–Sanson images. Apparent positions as seen by the observer.
With permission from Louise Bye.

Ametropia

In the ametropic eye, light fails to focus correctly on the retina, i.e. the second principal focus of the eye does not fall on the retina. The focus may be in front of the retina (myopia) or behind the retina (hypermetropia).

Myopia

In myopia, the second principal focus lies in front of the retina (Fig. 8.29) because:

- either the eye is too long (axial myopia)—most common form of myopia
- or the dioptric value is high although the length is normal (refractive/index myopia), e.g. keratoconus, nucleosclerosis.

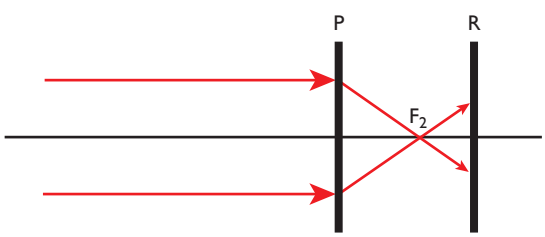


Fig. 8.29 Myopia.
With permission from Louise Bye.

Hypermetropia

In hypermetropia, the second principal focus lies behind the retina (Fig. 8.30) because:

- either the eye is too short (axial hypermetropia)—most common
- or the refractive power of the eye is inadequate (refractive hypermetropia), e.g. aphakia.

Hypermetropia can be overcome by increasing the refractive power of the lens (accommodation). However, the ability to accommodate declines with age and may be lost. As a consequence, these patients may need refractive correction for presbyopia at an earlier age than emmetropic patients.

As with squints, hypermetropia can be classified as latent or manifest. Manifest hypermetropia is spectacle correction that is required to achieve clear distance vision, whereas latent hypermetropia is additional refraction that is required to eliminate the remaining ciliary tone/accommodation. In children cycloplegic refraction is required to overcome all of their hypermetropia and give the full spectacle correction. Adults can do this voluntarily by gazing into the distance.

Astigmatism

In astigmatism, the refractive power of the eye varies in different meridians (Fig. 8.31). The image from an astigmatic eye is not focused on one plane and forms an area of

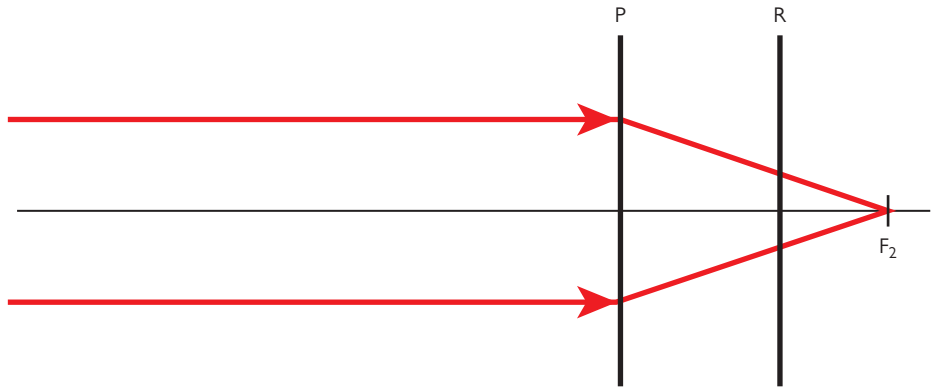


Fig. 8.30 Hypermetropia.
With permission from Louise Bye.

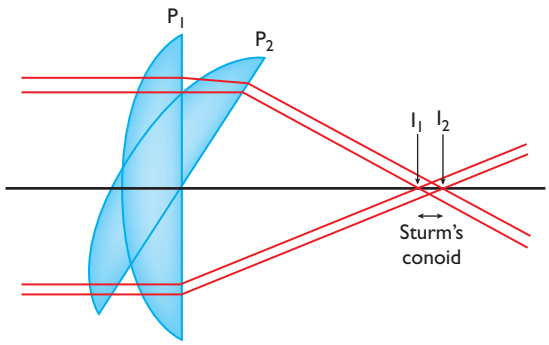


Fig. 8.31 Astigmatism—image formation.
With permission from Louise Bye.

blur over which the image is focused, known as Sturm's conoid.

If the principal meridians are at 90° to each other, this is known as regular astigmatism. If the principal meridians are at 90° to each other but do not lie at 90° and 180°, this is known as oblique astigmatism. If the principal meridians do not lie at 90° to each other, this is known as irregular astigmatism and cannot be corrected by spectacles.

Astigmatism can be classified according to the position of light rays relative to the retina:

- compound hypermetropic astigmatism—all light rays focus behind the retina
- simple hypermetropic astigmatism—rays in one meridian focus on the retina and in the other focus lies behind the retina
- mixed astigmatism—rays in one meridian lie in front of the retina and in the other behind the retina
- simple myopic astigmatism—rays in one meridian lie on the retina and in the other in front of the retina
- compound myopic astigmatism—rays in all meridians focus in front of the retina.

Anisometropia

When the refraction of two eyes is different, the condition is known as anisometropia. Large degrees of anisometropia (usually > 1D) are a common cause of amblyopia because accommodation is a binocular process and the eyes cannot accommodate by different amounts. The more hypermetropic eye usually becomes amblyopic. Myopic patients tolerate greater degrees of anisometropia.

Far point

The far point of the eye is that position in which the image of an object falls on the retina of a relaxed eye.

In emmetropia, parallel rays are focused to a point on the retina.

In myopia, the eye is longer and light needs to be divergent on reaching the principal plane to be refracted to a point that is further behind the normal emmetropic principal plane, given the same refractive power of the eye.

In hypermetropia, the eye is shorter and light needs to be convergent on reaching the principal plane to be refracted to a point that is in front of the normal emmetropic principal plane, given the same refractive power of the eye.

Optical correction

Optical correction of ametropia is necessary to deviate the incoming parallel rays of light so that these rays may be focused correctly upon the retina.

In Fig. 8.32 it can be seen that with correction, parallel rays of light are brought to correct focus on the retina when the convex (+) lens is placed in front of the eye. It can be seen that the far point of the eye must coincide with the focal point of the lens, i.e. the lens converges the incoming light so parallel rays converge to give incident light that

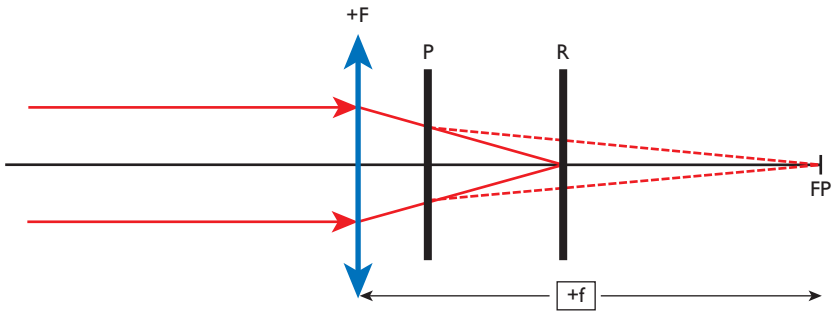


Fig. 8.32 Correction of hypermetropia.
With permission from Louise Bye.

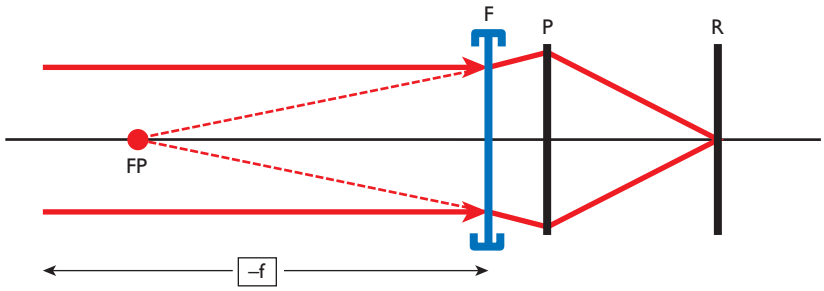


Fig. 8.33 Correction of myopia.
With permission from Louise Bye.

approaches the eye at the same incident angle. The lens refracts light so parallel rays converge sufficiently on passing through the lens to be directed to the far point rather than rays that were already converging.

In myopia, a concave lens is used (–) so parallel rays are brought to correct focus on the retina when the lens is placed in front of the eye. It can be seen that the far point of the eye must also coincide with the focal point of the eye. The lens refracts light so parallel rays diverge sufficiently on passing through the lens, so that they appear as if they are coming from the far point of the eye as they were without correction (Fig. 8.33).

In practice, lenses are held in spectacles. These usually sit around 14 mm in front of the cornea. Moving the lens to this position has a consequent effect on the position of the focal point of the lens. The lens power for spectacles has to be calculated based on the position at which they are placed.

Back vertex distance

If a lens is moved away from the eye, the image moves forward in both convex and concave lenses. In a hypermetrope, the image falls behind the eye and the purpose of the lens is to move the image forward. Hence, when the lens is moved away from the eye, the image is moved further forward and the lens effectively increases. In myopes, however, the image falls before the eye and the lens throws the image backwards onto the retina. If the lens slips forward, the patient becomes myopic again and, for each forward movement, a stronger concave lens is needed. Myopes do not like their glasses slipping forward for this reason.

A general formula to correct for the position of the lens relative to the eye applies to both concave and convex lenses as follows:

$$F_2 = 1/f_1 - d$$

where F_2 is the power of the lens (in dioptres) required at the new position, f_1 is the focal length of the original lens in metres, and d is the distance moved in metres.

This is equivalent to the following formula:

$$F_2 = F_1/1 - dF_1$$

where F_1 is the dioptric power of the original lens.

Practically speaking, a lens of power greater than 5D must be placed in the back of the trial frame because the

distance from the cornea makes a significant difference. If more than one lens is used, the higher power lens must be placed at the back of the trial frame. The examiner must state the distance of the lens in front of the eye (there is a scale on the side of the trial frame to help measure this), which is called the back vertex distance.

For example, refraction shows that an aphakic patient requires a +15D lens at back vertex distance 14 mm. He needs a contact lens. What power should this be?

$$F_2 = F_1/1 - dF_1$$

$$F_2 = +15/1 - (+0.014) \times (+15)$$

$$F_2 = 18.98\text{D (rounded to 19D)}$$

Spectacle magnification

Spectacle magnification (SM) is the corrected image size divided by the uncorrected image size.

Relative spectacle magnification (RSM) is the corrected ametropic image size divided by the uncorrected emmetropic image size.

In *axial* myopia, if the lens is worn at the anterior focal point of the eye, the image size is increased. If the spectacles are brought nearer to the eye, the RSM is greater than 1.

In *refractive* ametropia, the image size with spectacles is never at unity, even at the anterior focal point of the eye. In refractive hypermetropia, the image size is increased (RSM > 1) and in myopia it is decreased (RSM < 1).

The image can be made larger by bringing the lens nearer to the eye and made smaller by bringing it further from the eye.

Aphakia and intraocular lens selection

In aphakia, there is no convex lens in the eye. Light is focused behind the eye and this state is equivalent to having a high refractive hypermetropia. The aphakic anterior focal length (FaphD = 23.23 mm) is greater than the emmetropic anterior focal length (Faph = 17.05 mm).

Spectacle correction leads to an RSM of 1.36 when placed at the anterior focal point of the eye (23.2 mm in front of the principal plane). Spectacles are usually worn around 14 mm in front of the cornea and the RSM at this distance is 1.33. Contact lenses produce a relative spectacle magnification of 1.1.

The RSM can be calculated as follows:

$$AB = DE = \text{emmetropic image size}$$

$$AC = DG = \text{corrected aphakic image size}$$

$$RSM = AC/AB = DG/DE$$

As rays FemE and FaphG are parallel

$$\text{angle DFaphG} = \text{angle DFemE}$$

and

$$DG/FaphD = DE/FemD$$

$$DG/DE = FaphD/FemD = 23.23/17.05$$

$$\text{therefore } 23.23/17.05 = 1.36.$$

An RSM of 1.33 means that the image will appear one-third larger than normal. Objects appear closer because of the increased angle of the image subtended by the eye. As a result, Snellen acuity will improve. Contact lenses will improve this to a magnification of 1.1 and an intraocular lens to a magnification of 1.0. In a patient with unilateral aphakia, spectacles can be problematic because of the imbalance in magnification. Here contact lenses or a secondary intraocular lens is required. In both, the images can be fused to restore binocular vision.

Aphakic spectacles (usually +10.0D or more) cause aberrations, image distortion (pin-cushion effect), and a prismatic effect. Around the edge of their glasses, patients will have a ring scotoma and objects may appear and disappear into and out of this scotoma ('jack in the box' phenomenon) (Fig. 8.34).

As the eye moves, the scotoma will also move.

Patients cope with this problem by moving their heads instead of their eyes and restricting their gaze to a small axial zone.

If an intraocular lens is to be inserted, several theoretical or empirical formulae have been used to predict intraocular

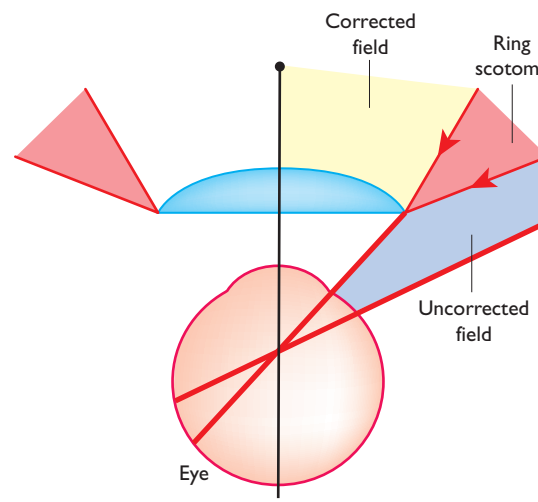


Fig. 8.34 Ring scotoma in corrected aphakia.
With permission from Louise Bye.

lens (IOL) power. The commonly used SRK empirically derived formula states that:

$$P = A - B(AL) - C(K)$$

where *P* is IOL power in dioptres, *A* is a constant which denotes the position of the lens in the eye (each make of lens is different), *B* is the multiplication constant for axial length, *AL* is the axial length in millimetres, *C* is the multiplication constant for the average keratometry reading, and *K* is the average keratometry reading in dioptres. The values of the multiplication constants are $B = 2.5$ and $C = 2.9$.

If a refractive condition (R) different to emmetropia is required, this formula is changed to:

$$P = A - B(AL) - C(K) - D(R)$$

D is 1.25 if required IOL power is greater than 14D, or 1.0 if IOL power required is less than 14D.

If a posterior chamber lens was planned but an anterior chamber lens had to be inserted for clinical reasons, the IOL power must be changed. A less powerful lens is needed because the lens is further forward. The IOL power must be reduced by the difference between the *A* constants of the two eyes.

The original SRK formula was not accurate for eyes of varying lengths. This was corrected for in the SRK II formula.

For these formulae, the axial length of the eye and keratometry must be known. Corneal curvature may be measured with a keratometer (Chapter 10, p. 255) while the axial length may be measured with an A-scan (ultrasound). Axial length measurements must be performed along the visual axis of the eye. Errors will be magnified as the axial length is multiplied by 2.5 in the SRK formula. If axial length is different between the two eyes by more than 0.5 mm then this should be rechecked.

Intraocular lens selection is a combination of science and art. Not all patients will be happy as emmetropes or –0.5 to –0.3 myopes. Important factors are the patient's previous refractive experience (myopes prefer to remain slightly myopic) and the refractive condition of the fellow eye (anisometropia must be avoided).

Accommodative problems

Accommodation is the mechanism by which we are able to adapt our focus from distant to near targets. We all require around 30% of accommodation to be kept in reserve in order to view near targets comfortably.

As a person grows older, the degree to which s/he can accommodate lessens, and this is termed presbyopia. As an infant the magnitude of accommodation is up to 14D. This declines to around 1D by around 60 years of age with a more rapid decline noted at the age of 45 (here the adult is capable of around 4D of accommodation).

The reason for this decline is likely to be twofold:

1. Lenticular and capsular rigidity reduces the degree by which the natural lens becomes convex.
2. Ciliary body weakens with age (noted at around 40 years of age).

The patient's near point is the minimum distance a target may be viewed in metres and the reciprocal of this distance infers the amount of accommodation in dioptres the patient is using.

Working out the presbyopic correction

There are many ways to quantify how much near prescription to give. One of the simplest methods is as follows.

By using the near point of a patient, one can work out the necessary presbyopic correction.

Remembering that in order for a patient to see a near target comfortably s/he must have a quarter of her/his accommodation in reserve, the presbyopic prescription becomes:

$$1/3 \times 1/\text{patient near point (m)} = 1/3 \times \text{near point (m)}$$

Thus if a person's near point is 33 cm, $1/3 \times 0.33$ m = 1D

It is important to remember that a simple near add correction must be task specific and does not translate to a multitude of activities.

Ascertainment of range is a useful method to prevent over addition. We first establish that the patient is able to read N5 at a comfortable distance, then establish whether a slightly larger font size is clear at arm's length. If the patient is able to see clearly for both distances, the range is likely to provide satisfactory vision for most important distances.

How presbyopia affects people with different refractive states

Hypermetropes are essentially always in a state of accommodation in order to maintain a focused target as without accommodation their focal point would lie behind the retina. It is easy to see why presbyopia affects the hypermetrope more profoundly.

The amplitude of accommodation for hypermetropes is 3D. A hypermetrope of +3D prescription has a magnitude of accommodation of 6D to (i) neutralize the hypermetropia and (ii) bring a target at 33 cm into focus.

Conversely, a –3D myope requires no accommodation to achieve clarity at 33 cm.

Bifocals

There are many types of bifocal lens, which differ in construction and form.

The near addition relies on either an increased convexity or a higher refractive index to achieve an increased vergence for near objects. The near addition is often placed in a position that is inferonasal to the optical centre of the main lens.

Remember that before prescribing a near addition, a careful history detailing the needs and activities of the patient is required. For example, a computer worker may require a large distance portion with an intermediate power near addition. A myope may wish to have a distance prescription and leave the near add as plano as s/he may be perfectly capable of reading without glasses.

HELPFUL HINT

Types of bifocal lens

Bifocal type	Construction
Split bifocals	• Two-piece lens of same refractive index • Essentially two half lenses of a near and distant prescription which are fixed together
Cemented	• Two-piece lens of same refractive index whereby a sliver of glass is cemented to the front of a main distance prescription
Fused	• Two-piece lens whereby a supplementary button is heat-fused into a void in the main distance prescription • The button is of a higher refractive index (usually flint glass), which achieves the necessary vergence for near addition • The button is ground flush with the surface of the main lens • Chromatic aberration may occur at the edges of the segment
Solid bifocal	• One-piece lens • The near zone employs a thicker section of a steeper curvature to achieve the required vergence • Executive bifocals belong in this category. Here, the near add occupies the entire lower half of the main distance correction • Executive bifocals are used in the treatment of convergence excess esotropia. Here, the top of the near segment must be aligned to the base of the pupil to ensure the maximal amount of time using the near add • For reasons of ease of production plastic bifocals are often solid bifocals

The typical configuration of a bifocal lens notes two key positions on the lens:

- the assumed intersection for the line of sight point for distance (LOSD)
- the assumed intersection for the line of sight point for near (LOSN).

Problems with bifocals

There are many reasons why intolerance occurs in bifocals.

Poorly fitted frames

Occasionally a pair of spectacles will not be tolerated as the near segment has been placed in an inappropriate position.

The bifocal is usually created such that the LOSN is around 8 mm inferior and 2 mm nasal to the LOSD. The near segment itself should be placed so that it is comfortable to access but far enough away from the visual axis in primary gaze that it does not interfere with distance vision. The edge of the near section is usually lined up with the inferonasal corneal limbus.

Prismatic effect

As with lens decentration, the prismatic effect is directly proportional to the power of the lens and how peripheral the incident light lies.

The mutually distant zone from the two optical centres on a bifocal prescription is therefore where the most prismatic effect is exerted. We shall term this the near–distant junction.

The near segment itself has its own independent prismatic effect as it carries the sum of the near correction and the main lens.

Solutions to prismatic effect

If a spectacle lens provides any solution to a prismatic effect, the lens is known as a prism-controlled prescription.

By bringing the two optical centres closer together, the prismatic effect can be reduced. In fused bifocals the button is placed near the LOSD, thus reducing the distance between the two optical centres. Furthermore the button may be shaped (as in B, C, and D shapes) to further reduce the distance between the optical centres.

In monocentric glasses, the optical centres of the distant and near portions coincide at the near–distant junction, and thus prismatic jump is eliminated.

In a hypermetropic correction the lower half may be considered as a base-up prism. A down curve segment exerts a base-down prism effect. When the two are brought together, they cancel out each other's prismatic effect.

Prismatic jump

Owing to the previously described prismatic effect, a patient will notice a sudden change or 'jump' when moving from the distance segment to the near segment. This is because there is a sudden change in prismatic effect, which leads to an

abrupt change in image position. As prismatic jump is essentially a function of the prismatic effect, it follows that the higher the refractive change the worse the jump.

Anisometropia

If there is a significant discrepancy between the prescription of one eye and that of another, it follows that there will be a discrepancy between the resulting prismatic effects of each lens.

For example, remember that the average patient is only able to tolerate up to 1.5D of vertical prismatic imbalance. If the patient is pushed beyond this due to high anisometropia and thus high prismatic imbalance, discomfort and diplopia will occur.

Similarly, caution must be used in patients with asymmetrical vertical muscle restriction/imbalance.

Prismatic correction in near adds may take two construction processes, either pre-cast or biprism. The biprism or 'slab off' process is particularly useful in high anisometriopia that induces vertical imbalance. To reduce an asymmetric prismatic effect, first add a corrective prism to both near and distance anterior surfaces in the least prismatic lens so that right and left are equal. The prism is then removed from the distance portion to leave a prism-corrected near segment.

Oblique astigmatism

Remember that oblique lens astigmatism occurs when light travels through the lens obliquely (non-paraxial), creating an optical aberration.

Oblique lens astigmatism is corrected for by accounting for the pantoscopic tilt.

This is to be distinguished from oblique astigmatism of the eye, in which the principal meridians do not lie at 90° and 180° but remain mutually perpendicular.

Hazards associated with bifocal use

The hazards of bifocals are primarily concerned with dangerous activities; these include:

- working at heights, which is a direct contraindication
- working with dangerous chemicals or hot objects
- vertigo.

Trifocal and progressive addition lenses (varifocal) lenses

Nowadays the daily activities of the average patient are varied and so require a multitude of different viewing distances. Some patients may require a 'one lens fits all' solution. It is here that multifocal spectacles become relevant.

Trifocal lenses

Trifocals employ the use of an intermediate prescription catering for middle distances that exists as an extension of the near segment on the near–distant junction.

However, the viewing distances tend to be a little rigid and there are now two prismatic jumps. Sometimes trifocals may be created to provide some degree of overlap but a perfect solution is never entirely possible. Similarly to ordinary bifocals, trifocals are not often tolerated in the cases of anisometropia or muscle imbalance, where prisms are required.

Progressive addition lenses (varifocal)

These lenses provide a more continuous solution between the near and distance segments, allowing for all distances to be catered for. They employ a 'strip' of lens that is optically true and gradually alters in power when the eye moves from the distance segment to the near segment—this strip is also known as a power progression corridor (PPC).

Progressive addition lenses vary in design between 'hard' and 'soft' designs. These terms refer to the width of the PPC in relation to the size of the distance and near zones, where the 'softer' the design, the smaller the distance and near segment, and the wider the PPC.

Advantages of varifocal lenses

- There is no visible change between distance and near segments, which provides a better cosmetic appearance.
- The prismatic jump that is problematic in bifocals and trifocals is eliminated.
- All distances are catered for in seamless fashion.

Disadvantages of varifocal lenses

- Optical aberration does occur in varifocal lenses. Astigmatism and aberration arise on either side of the PPC and become more profound when the line of sight is more horizontally deviated. This is especially difficult to neutralize in patients with already high astigmatism or in those requiring a more powerful near addition and provides a major reason for intolerance in these cases. In 'soft' and 'ultrasoft' designs, a wider PPC is afforded, which reduces size of the peripheral segments on either side. This, in turn, reduces the amount of aberration and astigmatism associated with the peripheral segments.
- The configuration or rate of graduation of the PPC is also important to take into account. Some users may find that the power of the near addition they often use for close work is uncomfortably low.

Low visual aids

Most low visual aids consist of some sort of magnifying aid. They act by increasing the angle subtended by the object at the eye and therefore produce an enlarged retinal image.

The convex lens

The magnification is achieved by allowing the eye to view the object at closer range than would be possible unaided

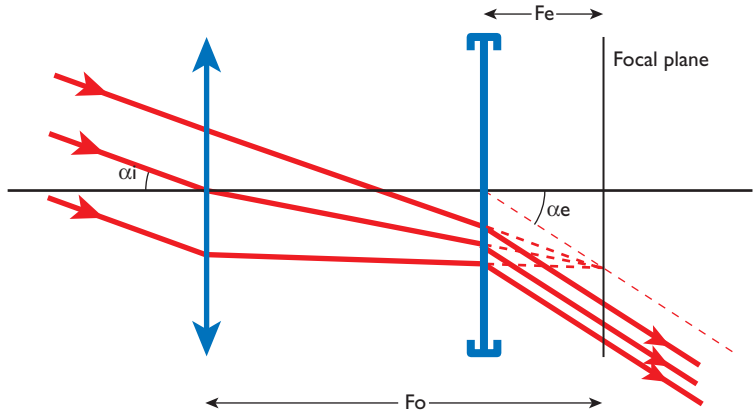


Fig. 8.35 Galilean telescope.
With permission from Louise Bye.

or with standard presbyopic reading correction. The power of the convex lens used (~ +5D) is less than a standard loupe (×8 = +32DS) and it acts simply as an augmented presbyopic correction.

The object is placed between the lens and the first principal focus (F1). A magnified, virtual image is formed.

The Galilean system

This consists of a convex (+) objective and a concave eyepiece (–), which are separated by the difference between their focal lengths (see Fig. 8.35). It produces an erect magnified image that is not greatly affected in the periphery or by astigmatism.

Limitations of optical magnifying devices

- High magnification results in a reduced field of view.
- The object has to be held close to the eye.
- Magnification means that depth of focus is reduced.

CHAPTER 9

Therapeutics

All complex multicellular organisms rely on chemical communication. This is mediated by hormones and neurotransmitters which act on protein receptors. The presence of protein receptors gives the system both sensitivity and specificity.

A receptor is a cellular macromolecule that is concerned with chemical signalling between and within cells. A hormone, neurotransmitter, or intracellular messenger, when combined with a receptor, results in a change in cellular activity, e.g. by a change in membrane permeability or alteration in gene transcription.

The action of a drug at a receptor depends on:

1. binding of a drug to a receptor
2. generation of a response by the drug–receptor complex (activation).

Pharmacokinetics and pharmacodynamics

Pharmacokinetics

Pharmacokinetics describes the time course of drug absorption, distribution, metabolism, and excretion. One of the simplest measures of this is the half-life of a drug ($t_{1/2}$), which is the time taken for the drug plasma concentration to fall by half after administration. The $t_{1/2}$ can be determined from a plasma concentration–time curve. This is a simplistic model based on the assumption of a single compartment or that the drug assumes a uniform concentration throughout all compartments (intra- and extracellular). Elimination by metabolism and excretion is assumed to be directly proportional to the drug concentration.

Bioavailability describes the oral dose that reaches systemic circulation and becomes available to the site of drug action. However, a rapidly absorbed drug will reach a higher concentration, but rapid elimination may lead to low bioavailability. Hence, bioavailability is defined as the area under the curve of log plasma concentration against time. Bioavailability may be affected by incomplete absorption of the drug or destruction by first-pass metabolism. Drugs instilled into the eye may be absorbed by the nasal and nasopharyngeal mucosae directly into the systemic circulation; therefore they escape first-pass metabolism and have a high bioavailability. Thus, topically administered agents can give rise to quite marked systemic effects.

Topically administered drugs, as well as being eliminated through the systemic circulation, may pass into the aqueous into the canal of Schlemm. They penetrate poorly into the posterior segment (vitreous, retina/choroid, optic nerve).

Drug kinetics can be described as first-order (linear) or zero-order (non-linear, saturation).

- First-order kinetics describes where the rate of drug transfer is proportional to the amount of drug present (i.e. half-life is constant) and can be defined by linear differential equations. Most drugs exhibit first-order kinetics.
- Zero-order kinetics refers to transfer related to a functional capacity of the body rather than drug concentration. In zero-order kinetics, drug dynamics show saturation at high drug concentrations. Saturation may occur when the capacity of drug-metabolizing liver enzymes is surpassed, leading to unmetabolized drugs in the circulation for longer periods. The administering dose is therefore important in saturation kinetics. Active transport systems change drug kinetics from zero order to first order.

'Tissue binding' prolongs the retention of a drug in a compartment by rendering it unavailable for tissue kinetics. The efficacy of the drug is reduced, but duration of action is increased, for example melanin binds atropine and pilocarpine.

Metabolic exchange within the eye is so rapid that independent compartments do not exist. There are active transport systems in the ciliary epithelium, lens epithelium, RPE, and cornea.

The four major processes by which the body handles a drug are absorption, distribution, metabolism, and excretion.

These processes are important when developing and using drugs in the eye. A drug that exists only in an ionized form will not penetrate the cornea unless it is damaged. Once absorbed, it may bind and become sequestered by melanin within the pigment epithelium of the iris and ciliary body. This may reduce bioavailability and increase clearance time. Later, the drug may be modified by metabolism (e.g. monoamine oxidase). The anterior segment is highly metabolically active and any drug that penetrates the eye and may be a substrate for these drugs will be substantially degraded during absorption.

Drug absorption

Generally, ophthalmic drug penetration increases with lipid solubility. The cornea has particular structure in this respect.

Drugs can pass through cell membranes by:

- diffusion directly through lipid
- diffusion through aqueous pores that traverse the lipid
- using carrier molecules
- pinocytosis by the cell.

Fick's law is used to determine the rate of diffusion:

$$\text{rate of diffusion} = KA(x_1 - x_2)/D$$

where K is the diffusion coefficient, A is the diffusion area, $x_1 - x_2$ is the concentration difference between plasma and intracellular compartments, and D is the thickness of the membrane. The diffusion coefficient is determined by lipid solubility, degree of ionization, and molecular size of the drug.

The absorption depends on a drug's lipid solubility and inversely on its polarity or degree of ionization. The more unionized a drug is, the more likely it is to be lipid soluble and the more likely it is to be transferred by passive diffusion through the membrane. The pK_a is the point at which the compound is 50% ionized. For a weak acid or base, the pK_a value determines the degree of ionization. The Henderson–Hasselbach equation can be used to predict the degree of ionization.

For a weak acid the ionizing reaction is:

$$pH = pK_a + \log\{[A^-]/[HA]\}$$

and for a weak base it is:

$$pH = pK_a + \log\{[B]/[BH^+]\}$$

For example, local anaesthetics are weak bases, i.e. at acid pH they are mostly ionized (BH^+) and at alkaline pH they are mostly unionized (B).

Active transport requires energy-dependent carrier-mediated mechanisms. Facilitated diffusion is carrier-mediated transport that does not require energy because it does not proceed against a concentration gradient.

Protein binding after absorption is important because it keeps the free concentration low, maintaining the concentration gradient (sink effect).

Drug distribution

Most drugs do not distribute uniformly in the bloodstream. Initially, well-perfused tissues show the highest concentration of the drug. The concentration rises more slowly in poorly perfused tissues.

Eventually, in the steady state, the concentration varies markedly between tissues. This is because of variation between tissues of lipid solubility and variation within tissue of affinity to particular tissue components. For example:

- low lipid solubility (e.g. poor penetration of the brain)
- high lipid solubility (e.g. accumulation in fat, e.g. halothane)
- high affinity for particular tissue components.

Diffusion of drugs from the blood to the eye is similar to the passage into the brain. The blood–retinal barrier is similar in many ways to the blood–brain barrier, but, comparatively, lipid-soluble drugs will penetrate the eye more readily.

Most drugs bind to plasma or tissue protein to some extent. Bound drug is restricted to the distribution of the binding protein. The bound drug is inactive (e.g. warfarin). Albumin is the major plasma protein and is responsible for the binding of many drugs. In most cases, the binding does not involve covalent bonds but is readily reversible. Each protein molecule may possess many different binding sites for acids and bases, which may be of different binding capacity. For highly bound drugs, the bound fraction represents a store. As free drug is eliminated, dissociation will tend to minimize the fall in concentration.

Drug metabolism

Metabolism mainly occurs in the liver but other tissues contribute. The molecules are enzymatically altered to increase their water solubility in preparation for excretion, usually in the urine and sometimes in the bile. The compound is simultaneously made metabolically and pharmacologically active or inactive. The parent drug (prodrug) may be inactive and can be metabolized to form active metabolites (e.g. diamorphine) or toxic metabolites, which may persist longer in the circulation than the parent molecule (e.g. lignocaine).

Biotransformations may be:

- phase 1—aromatic hydroxylation, aliphatic hydroxylation, N-deamination, N-dealkylation, S-oxidation, desulphation, reduction, hydrolysis
- phase 2—conjugations (e.g. with glucuronic acid, glycine, glutamine, sulphate, acetate).

Phase 1 reactions are carried out by a heterogenous group of enzymes called cytochrome P_{450}. These exist in the liver and in some peripheral organs, including the eye.

Table 9.1 Enzyme inducers and inhibitors

Enzyme inducers	Enzyme inhibitors
Barbiturates	Isoniazid
Phenytoin	Chloramphenicol
Phenothiazines	Metronidazole
Rifampicin	Warfarin
Nicotine	Carbon monoxide

There are also other enzyme systems that are involved in drug metabolism, e.g. xanthine oxidase (purine metabolism), alcohol dehydrogenase, and monoamine oxidase (catecholamine metabolism).

Many drugs have inducing or inhibiting effects on these enzymes (Table 9.1).

Metabolism is also affected by age, smoking (induces enzymes), alcohol (enzyme inhibitor), genetic control, and nutritional status.

Drug excretion

Drugs are eliminated by biotransformation or excretion in an unchanged form. Major routes for drug excretion are via the kidneys and liver. Volatile gasses may be lost via the lungs and minor amounts are lost into sweat and milk.

Renal excretion has four components:

- Filtration in the glomerulus—most drugs have a low molecular weight and will leave the circulation in the ultrafiltrate irrespective of charge, unless they are bound to plasma proteins. Clearance is therefore related to the unbound fraction of the drug and depends on the glomerular filtration rate.
- Secretion—drugs may be secreted into the tubules by the glomerular filtrate. There is competition for active sites in the carrier systems, hence secretion of drugs can be blocked by other drugs.
- Reabsorption—large amounts of water are reabsorbed by the nephron. The countercurrent multiplier system leads to an increasing concentration gradient that drives the drug back into the plasma deeper into the loop of Henle. Lipid-soluble drugs pass back into the plasma more rapidly and so are more slowly excreted from the body. Drugs that are weak acids or bases must be in their unionized form; therefore altering the pH of the urine will increase the elimination of the drug, so acidic drugs are excreted more rapidly in alkaline urine.
- Biliary excretion—liver cells can also transfer drug metabolites from blood to bile. Conjugated drugs are concentrated in the bile and delivered to the intestine, where the conjugate may be hydrolysed, releasing the active drug. This drug may then be reabsorbed and recirculated (enterohepatic circulation).

Ocular drug absorption

CLINICAL TIP

Factors influencing topical drug absorption

- Environmental conditions: temperature and humidity
- Volume of drug administered
- Drug formulation: pH, preservative, vehicle
- Blink rate, tear film stability, corneal surface health
- Nasolacrimal tear drainage
- Absorption through conjunctival vessels

The rate of drug absorption varies with the route of administration. In the eye, the permeability of ocular structures varies.

Cornea and tear film

See Chapter 1.

Drugs that must pass through the cornea can only do so by diffusion and must therefore go down an electrochemical gradient. The cornea has three layers for molecules to pass through. They form a sandwich of fat (epithelium), water (stroma), and fat again (endothelium).

The epithelium is the most important barrier to intraocular transport of drugs via this route for three reasons:

1. The epithelium and endothelium both contain 100 times as much phospholipid as the stroma and are therefore hydrophobic.
2. The epithelial cells are bound together by desmosomes.
3. Bowman's layer presents a further barrier to diffusion. The epithelium is therefore impermeable to electrolytes, but permeable to fat-soluble substances.

The opposite is true of the stroma, which accounts for 90% of corneal substance. Its ground substance (glycosaminoglycans and water) permits ionized water-soluble drugs to pass more freely than lipid-soluble drugs.

The endothelium provides little resistance to drug transport and the aqueous and stroma can be considered as one single compartment.

Drugs must therefore exist in both ionized and unionized forms. The pH of the tear (6.5–7.6) film may affect the ionization of the drug and thus its diffusion capacity. Many ocular drugs are weak bases (e.g. tropicamide, cyclopentolate, and atropine) and occur in both ionized and unionized forms within the pH range of normal tears.

Because lacrimation increases the clearance time of a drug, pH must be adjusted to near that of the tear film. If a drug is too acidic or alkaline it can be an irritant, which increases blink rate and lacrimation.

Some drugs have a direct effect on the lacrimal gland (e.g. muscarinic agonists increase lacrimation), which increases the clearance rate as well.

Benzalkonium chloride (a commonly used preservative in ophthalmic drugs) and other cationic surfactants increase

the ocular absorption of drugs by increasing corneal permeability by compromising corneal integrity.

Conjunctiva

See Chapter 1, p. 42.

The conjunctival epithelium is similar to the corneal epithelium. Topical administration into the inferior fornix of the conjunctiva is the most common route of ocular drug delivery. Blinking and lacrimation affect the residence time of fluid in the inferior fornix. The anatomy and physiology of the lids, precorneal tear film, and the health of the local tissues affect the efficacy of such delivery systems. Environmental factors are also important and include temperature and humidity.

Drugs administered topically to the inferior conjunctival fornix must first pass through the epithelium. Once the conjunctiva has been traversed, there is an extensive subconjunctival venous plexus within the stroma. Drugs can easily penetrate the circulation by this route. Also, drugs can drain directly into the nasolacrimal duct and into the nose, where further systemic absorption occurs through the mucosae.

The alteration and efficacy of drugs can be altered by changing local conditions. The residence in the conjunctival sac can be increased and this in turn produces greater ocular drug absorption, especially for lipid-soluble drugs. Polymers that increase viscosity include polyvinyl alcohol, hydroxypropylcellulose, and other cellulose derivatives.

Sclera

See Chapter 1, p. 47.

This is a tough but porous barrier, even to larger molecules. Penetration of drugs by sub-Tenon's or subconjunctival injection is limited by trans-scleral outflow, which is dependent on IOP.

Iris vessels

See Chapter 1, p. 49.

A blood–ocular barrier prevents leakage of protein (and fluorescein) under normal circumstances.

Ciliary epithelium

See Chapter 1, p. 51.

The ciliary epithelium actively secretes some ions and permits free passage of others. Under normal circumstance it blocks the entry of most molecules, e.g. proteins and antibiotics.

Drug-delivery vehicles

Most pharmacological ocular disease management depends on the use of drops despite significant advances in drug-delivery systems. Topical drug application is limited by the significant barrier to solute flux presented by the corneal epithelium and the rapid loss as a result of drainage and tear film turnover.

The principal aims of new drug-delivery systems are to increase drug–ocular contact time, increase drug delivery, and reduce systemic absorption and side effects.

CLINICAL TIP

Reducing systemic toxicity from eyedrops

Human tear film volume = 7 µl. This can increase momentarily to 30 µl.

The conjunctival sac has a capacity of 15–30 µl.

The tear film turns over at a rate of 15% per minute with a normal blink rate (15–20 per minute).

One eyedrop = 10–75 µl, so most is lost in spillage.

Administration of more than one eyedrop will not increase bioavailability but can lead to increased systemic toxicity via the nasolacrimal duct mucosal absorption.

Methods to reduce systemic absorption include:

- use of more viscous therapy, e.g. ointment
- digital pressure over lacrimal sac
- pouch method—evert the lower lid and close the eye for 2 minutes
- deliver smaller amounts of drug by reducing drop size or using sprays.

When applying two different drops, a 'washout effect' should be avoided by leaving a 5-minute interval between the two types of drop.

Solutions, suspensions, and colloids

A solution is a homogenous mixture composed of two or more substances. A solute (which can be solid, liquid, or gas) is dissolved in another substance, known as a solvent (which can be solid, liquid, or gas).

Solutions and suspensions are common modes of delivery. These possess a short contact time, which can be increased by adding polyvinyl alcohol or methylcellulose. These substances increase viscosity and lower the surface tension.

A suspension is a heterogenous fluid containing solid particles that are large enough for sedimentation. Usually these particles are larger than 1 µm in size and the internal phase (solid) is dispersed through the external phase (liquid) through mechanical agitation. These particles will eventually settle with time.

A colloid is a type of mixture in which one substance is dispersed (rather than dissolved) evenly throughout another. In a colloid, the particles are smaller and do not settle (Table 9.2).

Suspensions and colloids are different to solutions in that a solute does not exist as a solid when in solution.

Semi-solids (ointments)

An ointment is a homogenous, viscous, semi-solid preparation. Drugs are made up with a thick, greasy oil which is usually a combination of hydrocarbons, mineral oils, lanolin, and polymers, e.g. polyvinyl alcohol and methylcellulose. These are retained on the ocular surface longer than other preparations. They are more likely to cause visual blurring and contact dermatitis.

Table 9.2 Classification of colloids

		Dispersed phase		
		Gas	Liquid	Solid
Continuous phase	Gas	None (gasses are mutually miscible)	Aerosol (e.g. hair spray)	Solid aerosol (e.g. smoke)
	Liquid	Foam (e.g. shaving cream)	Emulsion (e.g. milk, mayonnaise)	Sol (e.g. blood)
	Solid	Solid foam (e.g. pumice)	Gel (e.g. agar, jelly)	Solid sol (e.g. cranberry glass)

Slow-release preparations

These allow the constant release of drug whilst minimizing the drainage rate of the drug. Ocular inserts are flexible, elliptical devices that are made of three layers. The two outer layers are ethylene vinyl acetate enclosed in an inner coat of drug/alginate mix. Their main use is in the treatment of dry eye. Collagen shields are made up of chains of alpha helices. Their crosslinking can be controlled, which in turn affects the time taken for these devices to dissolve when placed on the cornea. Soft contact lenses are made of polymer, which is hydrophilic. Water-soluble drugs are therefore absorbed into the lens. When the lens is placed onto the ocular surface, drug is released until an equilibrium is reached between drug concentration in the contact lens and that in the conjunctival sac.

Advances in ocular delivery systems

New ophthalmic delivery systems are constantly under development. Few are currently in routine ophthalmic use. Liposomes are vesicles that consist of a lipid membrane with an aqueous interior compartment. They carry lipophilic drugs more easily than hydrophilic drugs and share the properties of the outer cell membrane. They have the advantage that they are easily prepared, non-irritant, and cause no blurring of vision.

Intracameral and intravitreal administration

As a result of the integrity of the blood–retinal and blood–aqueous barriers, intraocular drug penetration is often not sufficient even with systemic therapy.

Intracameral cefuroxime is commonly used after cataract surgery as prophylaxis.

Intravitreal antibiotics are used for the treatment of endophthalmitis to administer a sufficient dose of antibiotic.

Water-soluble antibiotics such as penicillins have a prolonged half-life in the eye as the retinal pump mechanism is damaged. Aminoglycosides (e.g. gentamicin) are more rapidly cleared from inflamed eyes, as they are eliminated through aqueous circulation.

Recently there has been an explosion in the use of intravitreal vascular endothelial growth factor (VEGF)-inhibitors to treat wet age-related macular degeneration. A series of injections is often needed. There are a number of complications, including damage to the lens, haemorrhage, infection, and retinal damage.

Systemic drugs

Generally the intraocular concentration of drugs given systemically is not known. Acetazolamide is given orally or intravenously to reduce IOP. Ciprofloxacin penetrates the aqueous humour after oral administration. Ciprofloxacin and moxifloxacin have been shown to have a good effect on intraocular infection when given systemically. Steroids and non-steroidal anti-inflammatory drugs also penetrate the eye when given orally. In severe uveitis, a number of immunosuppressant drugs are often given systemically, for example methotrexate, azathioprine, ciclosporin, mycophenolate mofetil, and tacrolimus.

Conversely, topically applied drugs can reach the systemic circulation and affect the contralateral eye. This is well characterized with the use of timolol, in which the pressure reduction when used in one eye is associated with a pressure reduction in the contralateral untreated eye. It can also cause an exacerbation of asthma and cardiac conditions in susceptible individuals.

Pharmacodynamics

Pharmacodynamics deals with the effects of a drug and considers the relationship between drug concentration and response.

A drug that acts on a receptor may act as an agonist or antagonist. Drugs often give a graded dose–response curve and increasing the drug concentration will increase the drug effect. This only occurs when drugs are not permanently bound to the receptor. A drug that can elicit the maximal tissue response is termed a full agonist, whereas a drug that cannot elicit the maximal response is termed a partial agonist (Table 9.3).

The efficacy of a drug is defined as the maximal response, whereas the potency describes the amount of drug necessary to give the desired response. A drug can therefore have efficacy and low potency, and hence need high doses to produce an adequate response. If the drug is irreversibly bound to the receptor, its effect will continue after it has been eliminated from the bloodstream. On the other hand, it may diminish (tolerance) as a result of downregulation of specific drug receptors.

Many of the most useful pharmacological drugs are antagonists at receptors, i.e. they are substances that reduce the action of another agent, which is often an endogenous agonist (e.g. hormone or neurotransmitter).

Table 9.3 Agonist activity

Partial agonist	Full agonist	Acting at
Prenalterol	Adrenaline, isoprenaline	β-adrenoceptors
Pilocarpine	Acetylcholine	Muscarinic receptors
Impromidine	Histamine	Histamine H_2 receptors

Antagonists may be divided into competitive and non-competitive antagonists:

- Competitive: binding of the agonist and antagonist is mutually exclusive.
  - The agonist and antagonist compete for the same binding site, or adjacent sites which overlap, and form short-lasting combinations with the receptor, so that equilibrium between agonist, antagonist, and receptors can be reached in the presence of the agonist. The blocking action can be surmounted by increasing the concentration of the agonist, which will occupy a higher proportion of the binding sites (reversible competitive antagonism), e.g. atropine competitively blocks the action of acetylcholine on muscarinic receptors.
  - The antagonist may compete irreversibly with the binding site for the agonist. When enough receptors are blocked, antagonism is insurmountable and a full response cannot be achieved (irreversible competitive antagonism).
- Non-competitive: the agonist and antagonist can be bound at the same time to different regions of the receptor molecule.

For the non-competitive antagonists, antagonism may occur in a number of ways:

1. Chemical antagonism: the antagonist combines directly with the substance being antagonized, e.g. ethylenediaminetetra-acetic acid (EDTA) is used to treat lead poisoning, forming a less toxic chelete.
2. Functional or physiological antagonism: the antagonist is an agonist to its own separate receptor but produces an opposing effect to the substance being antagonized, e.g. adrenaline relaxes bronchial smooth muscle and reduces bronchoconstriction caused by histamine and leukotrienes.
3. Pharmacokinetic antagonism: the antagonist reduces the concentration of the active drug at its site of action, e.g. phenobarbital induces hepatic enzymes and thus inactivates warfarin by reducing its plasma concentration.
4. Indirect antagonism:
   a. The antagonist acts as a second downstream receptor which links the action of the agonist to the final response observed, e.g. β-blockers reduce the rise in heart rate caused by sympathomimetic amines (e.g. tyramine) because the amine causes release of noradrenaline, which then goes on to act at β-receptors.
   b. Another mechanism is that the agonist may interfere with other post-receptor events that result in the tissue response, e.g. calcium-channel receptors (e.g. verapamil) block the influx of calcium needed for maintained smooth muscle contraction and therefore reduce the contractile response to acetylcholine.

Drug–receptor interactions

Ionotropic receptors

Ion channels are pore-forming transmembrane proteins that help to establish and control the small voltage gradient across the plasma membrane of all living cells by allowing the flow of ions down their electrochemical gradient. They are present in the membranes that surround all biological cells. Biochemically, they are integral membrane proteins or an assembly of several proteins ('multi-subunit assemblies') involving a multi-subunit assembly of several identical or homologous proteins packed around a water-filled pore through the lipid bilayer.

Ion channels are responsible for fast synaptic transmission at the neuromuscular junction, peripheral autonomic neuroeffector junctions, autonomic ganglia, and central synapses. Because voltage-activated channels underlie the nerve impulse and transmitter-activated channels mediate conduction across the synapses, channels are especially prominent components of the nervous system. They are also involved in a wide variety of biological processes that involve rapid changes in cells, such as cardiac, skeletal, and smooth muscle contraction, epithelial transport of nutrients and ions, T-cell activation, and pancreatic β-cell insulin release.

Ion channels can be classified by gating. They may be voltage gated, ligand gated, or gated by other mechanisms, e.g. second messengers, light, or mechanical stimuli.

Voltage-gated ion channels are activated by changes in electrical potential difference near the channel. It is thought that the electromagnetic field induces a conformational change in the channel, which opens the pore to allow ion influx or efflux to occur across the membrane, down its electrochemical gradient. This subsequently generates an electrical current sufficient to depolarize the cell membrane. Channels may be ion specific, although similarly sized and charged ions may sometimes travel through them.

Examples include:

- sodium and potassium voltage-gated channels of nerve and muscle
- voltage-gated calcium channels that play a role in neurotransmitter release in presynaptic nerve endings.

Ligand-gated ion channels (also known as ionotropic receptors) open in response to a specific chemical messenger or ligand molecule binding to the extracellular domain of the receptor protein. The binding site of endogenous ligands on ligand-gated ion channel (LGIC) protein complexes are normally located on a different portion of the protein (an allosteric binding site) compared to where the ion conduction pore is located. Ligand binding causes a conformational change in the structure of the channel protein that ultimately leads to the opening of the channel gate and subsequent ion flux across the plasma membrane. These channels are usually very selective to one or more ions, such as Na^+, K^+, Ca^{2+}, or Cl^-.

LGICs are classified into three superfamilies:

- cys-loop receptors, e.g. anion-permeable γ-aminobutyric acid-gated ($GABA_A$) receptor, glycine receptor, and the cation-permeable 'nicotinic' acetylcholine receptor
- ionotropic glutamate receptors, e.g. α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA), kainate, and *N*-methyl-D-aspartate (NMDA) receptors
- ATP-gated channels, e.g. P2X receptor.

Metabotropic receptors

Neurotransmitter receptors are divided into two groups: ionotropic and metabotropic. Both receptors are activated by a specific neurotransmitter ligand. Metabotropic receptors do not form an ion channel themselves. When a metabotropic receptor is activated by a ligand (primary messenger), a primary effector is activated. This activates a secondary messenger, often a G-protein, that leads to a cascade of intracellular events that results in ion channel opening or closing. Hence, metabotropic receptors are a type of G-protein-coupled receptor. Others are tyrosine kinases and guanylyl cyclase receptors.

Since there are a number of steps involved in the opening of ion channels via metabotropic receptors, these channels take longer to open and are thus not involved in mechanisms that require quick responses. Once they are open, they may remain in this configuration for seconds to minutes. Their effects are therefore longer-lasting than those of ionotropic receptors. Their effects are also more widespread within the cell than those of ionotropic receptors.

G-proteins (guanine nucleotide-binding proteins)

G-proteins are important signal transducing molecules in cells. G-protein coupled receptors are the drug target for 40% of medicines in the USA. G-protein linked receptors are receptors that pass through the cell membrane and have seven hydrophobic transmembrane domains. They all possess:

- a ligand binding domain
- a G-protein binding domain on the third intracellular loop.

Signal molecules bind to the extracellular receptor. The receptor inside the cell activates a G-protein. The G-protein acts as a transducer between the receptor and effector systems, and activates a cascade of further compounds, finally causing a change downstream in the cell.

The human genome encodes 950 such receptors, which are important in detecting photons (light), hormones, growth factors, drugs, and other endogenous ligands (see Table 9.4).

Table 9.4 G-protein coupled receptors

Receptor class	Neurotransmitter or hormone receptor
Amino acid receptors	Metabotropic glutamate and $GABA_B$ receptors
Monoamine receptors	Adrenoceptors, dopamine, and 5-hydroxy tryptophane receptors
Lipid receptors	Prostaglandin, thromboxane, and platelet activating factor receptors
Purine receptors	Adenosine and ATP receptors
Neuropeptide receptors	Neuropeptide Y, opiate, cholecystokinin, VIP, etc.
Peptide hormone receptors	Angiotensin, bradykinin, glucagon, calcitonin, parathyroid, etc.
Chemokine receptors	Interleukin-8
Glycoprotein receptors	Thyroid stimulating hormone, luteinizing hormone, follicle stimulating hormone, chorionic gonadotrophin, etc.
Protease receptors	Thrombin

HELPFUL HINT

Local anaesthetic agents, e.g. lidocaine and bupivacaine, block the increase in Na^+ ion channel permeability which underlies the upstroke of the action potential in nerves.

Small fibres are blocked before large fibres and myelinated fibres are more resistant than non-myelinated fibres.

Pain sensation is carried mainly by small myelinated and non-myelinated fibres and hence disappears first. This is followed by temperature, touch, and deep pressure. There is great individual variation.

In the anaesthetized portion, the nerve impulse is slowed down and reduced in amplitude. It may stop altogether. The compound action potential is reduced in size because fewer nerve fibres contribute to this potential and the impulses are more temporally spread out.

Receptor tyrosine kinase

Receptor tyrosine kinases (RTKs) are the high-affinity cell surface receptors for many polypeptide growth factors, cytokines, and hormones. RTKs have been shown not only to be key regulators of normal cellular processes but also to have a critical role in the development and progression of many types of cancer.

A kinase is a type of enzyme that transfers phosphate groups from high-energy donor molecules, such as ATP, to specific target molecules (substrates), a process termed phosphorylation. The opposite of a kinase, an enzyme that removes phosphate groups from targets, is known as a phosphatase.

Kinase enzymes that specifically phosphorylate tyrosine amino acids are termed tyrosine kinases. When a growth factor binds to the extracellular domain of an RTK, its dimerization is triggered with other adjacent RTKs. Dimerization leads to rapid activation of the protein's cytoplasmic kinase domains, the first substrate for these domains being the receptor itself. The activated receptor as a result then becomes autophosphorylated on multiple specific intracellular tyrosine residues.

Src and phospholipase Cγ contain specific domains that bind to the receptor on phosphorylation of specific tyrosine residues within the activated receptor. Phosphorylation and activation of these two lead to signal transduction.

Examples of RTK signalling pathways include the epidermal growth factor receptor family, the fibroblast growth factor receptor family, and the vascular endothelial growth factor receptor family.

Drug actions

Ocular drugs and the autonomic nervous system

Many drugs used in ophthalmology act on the autonomic nervous system (ANS). The ANS is a peripheral, efferent, involuntary system. It is divided into sympathetic and parasympathetic systems.

In the sympathetic system, the cell body originates in the central nervous system (CNS). The preganglionic neuron synapses in the superior cervical ganglion and produces acetylcholine, which acts on nicotinic receptors. The postganglionic neuron synapses at the effector cell, producing noradrenaline, which acts on α_1, α_2, β_1, and β_2 receptors (Fig. 9.1).

In the parasympathetic system, the cell body again originates in the CNS. The preganglionic neuron synapses in the ciliary ganglion and also produces acetylcholine, which acts on the nicotinic receptors. The postganglionic neuron synapses at the effector cell, producing acetylcholine, which acts on muscarinic receptors (Fig. 9.2).

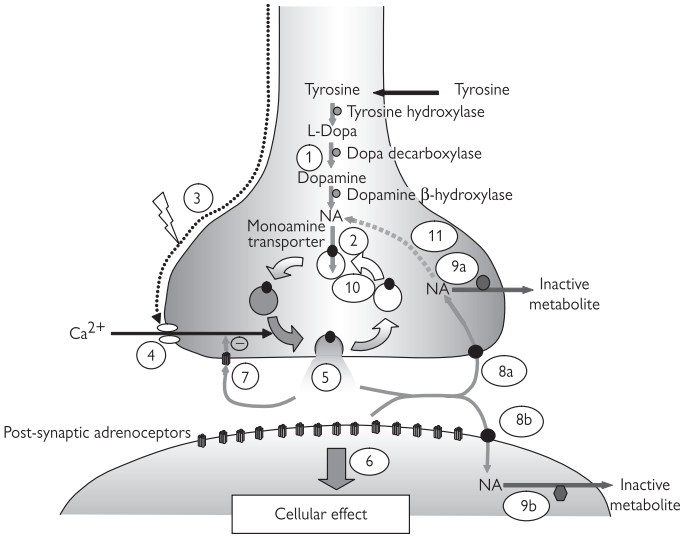
Drugs may target receptor sites by competing with the transmitter or affecting the concentration of transmitter in the neuro-effector synapse. If drugs or molecules act in a similar way to a neurotransmitter they are known as either adrenoceptor agonists (sympathomimetics) or muscarinic agonists (parasympathomimetics). If drugs or molecules bind with a receptor and block the access of a neurotransmitter, thus inhibiting a response, they are termed adrenoceptor antagonists (α-/β-blockers) or muscarinic antagonists. If drugs or molecules are similar to a neurotransmitter and bind with a receptor to cause a partial response, they may be termed partial agonists.

In the following section, drugs are discussed according to their role in clinical practice rather than in order of their effect on receptors (Table 9.5). *See The pupil*, p. 74.

Certain antimuscarinic drugs are commonly used for cycloplegia and mydriasis. Their side effects are the result of blockade of muscarinic receptors.

- Atropine may rarely cause the following:
  - dry mouth, dry skin, thirst, difficulty swallowing, disturbed speech
  - rapid heart rate, palpitations, arrhythmias, tachycardia
  - flushing
  - decreased tone and motility of the gastrointestinal tract
  - retrosternal pain due to gastric reflux
  - occasional vomiting, giddiness, staggering, and restlessness
  - rapid respiration, rash on face and upper trunk, hyperpyrexia, confusion, excitement, hallucinations, delirium, depression of CNS, hypertension with circulatory failure, coma, and death.
- Hyoscine is similar to atropine but has more CNS reactions. It has no advantages over atropine and is only occasionally used when patients exhibit atropine allergy. It is no longer commercially available.
- Homatropine is again similar to atropine but shorter acting. It is reported to have fewer side effects.
- Cyclopentolate has not been reported as having peripheral effects. CNS effects are rare but have been reported. They include hallucinations, drowsiness, ataxia, and emotional disturbances.
- Tropicamide has very few side effects and is reported to be safe.
- Sympathomimetic drugs are used for pupillary dilatation. They have no cycloplegic activity.
- Phenylephrine has α1 activity and is used as a mydriatic.
- Adrenaline/epinephrine has both α and β activity (only the α activity is required in dilating the pupil as this leads to unwanted side effects elsewhere—hence adrenaline is not used for mydriasis in ophthalmology).

The dilator pupillae is a weak muscle, so giving only a sympathomimetic may give incomplete dilatation.



1. Transmitter synthesis
2. Transmitter uptake into vesicles
3. Action potential in presynaptic neurone
4. Voltage-gated Ca^{2+} channel activation
5. Exocytosis
6. Receptor activation in post-synaptic neurone
7. Feedback inhibition of Ca^{2+} channels (α_2 receptors)
8. (a) Uptake 1 and (b) uptake 2
9. Transmitter deactivation by
   (a) Monoamine oxidase (MAO) and
   (b) Catechol-O-methyltransferase (COMT)
10. Vesicular recycling
11. Transmitter recycling

Fig. 9.1 Noradrenergic receptors and metabolism of adrenergic neurotransmitters.

Reproduced from R. Wilkins et al, *Oxford Handbook of Medical Sciences, The Essential Guide to the Sciences that Underpin Medicine*, Figure 4.8, Page 241, 2011, with permission from Oxford University Press.

Cycloplegia

Cycloplegic drugs block the muscarinic receptors on the ciliary muscle, causing relaxation and hence paralysis of accommodation (Tables 9.6 and 9.7). They also may cause mydriasis by their action in the pupil.

The residual accommodation varies with:

- number of drops applied (i.e. concentration)
- age of patient (less cycloplegia in younger patients, thus tropicamide not used in children under 18 years)
- colour of irides (cyclopentolate is less effective in dark irides and takes longer to act).

These drugs can be used in the management of anterior segment inflammation, e.g. iritis, by three mechanisms:

- paralysing inflamed muscle (antimuscarinic)
- dilating the pupil → synaechiae are broken (antimuscarinic and sympathomimetic)
- reducing the permeability of inflamed vessels → decreasing outpour of inflammatory cells and protein (antimuscarinics).

Mydriasis

Mydriasis may be brought about by drugs that block muscarinic receptors on the iris sphincter (antimuscarinics) or by drugs that stimulate α_1 receptors on iris dilator muscle (sympathomimetics), e.g. phenylephrine (Table 9.8).

The antimuscarinic drugs are described under 'Ocular drugs and the autonomic nervous system'. Phenylephrine is the only sympathomimetic used clinically and it has only α_1 activity. Adrenaline/epinephrine has both α and β activity. The β activity is not required for pupil dilatation and it results in more systemic side effects. Also, the dilator pupillae is not a powerful muscle, so use of a sympathomimetic alone may result in incomplete dilatation.

Vasoconstrictor sympathomimetics

These drugs are used to relieve conjunctival congestion, bleeding from small capillaries, and arterioles and their duration of action is shown in Table 9.9. They combine with α_1 receptors on blood vessels and therefore cause vasoconstriction. Repeated instillations may cause a reactive hyperaemia or mask underlying disease.

The main effects seen with phenylephrine are hypertension and cardiac arrhythmias. Great care should be taken in patients with cardiac disease, aneurysms, thyrotoxicosis, hypertension, sympathetic denervation, and neuropathies.

Parasympathomimetics (cholinergic agonists/miotics)

Acetylcholinesterase is found in the cornea and breaks down acetylcholine. By blocking acetylcholinesterase, both physostigmine and ecothiopate potentiate the effects of acetylcholine and cause miosis (see Table 9.10).

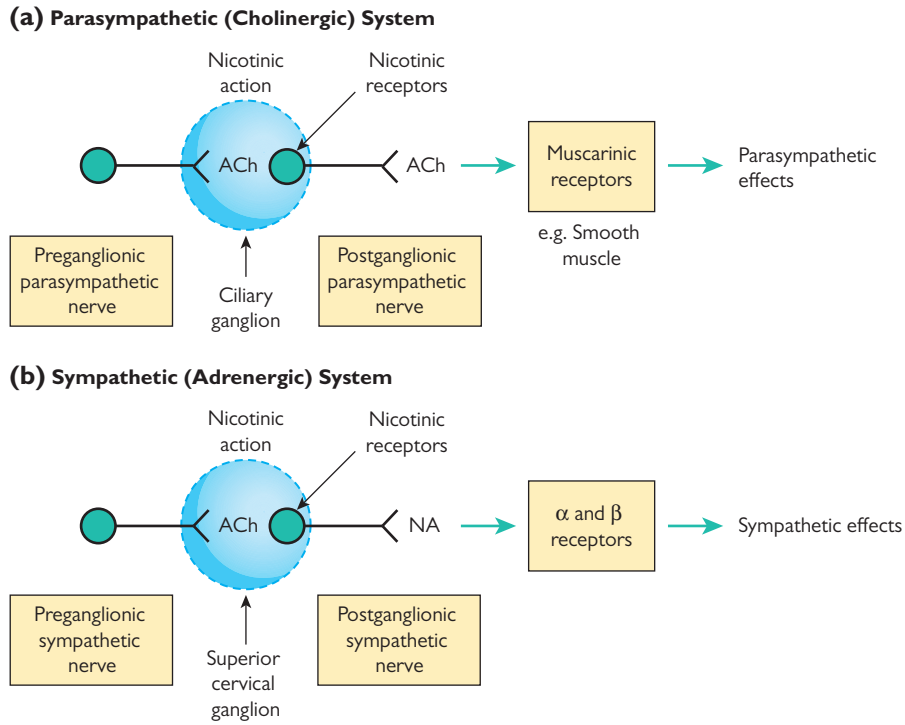


Fig. 9.2 Synaptic connections of the nervous system. (a) Parasympathetic (cholinergic) system. (b) Sympathetic (adrenergic) system. Ach: acetylcholine; NA: noradrenaline.

Table 9.5 Ocular autonomic receptors

Eye effector cell	Sympathetic/adrenergic receptors (stimulation)	Parasympathetic/cholinergic receptors (stimulation)
Iris dilator	$\alpha_1 \rightarrow$ mydriasis	
Iris sphincter	Mostly parasympathetic	Muscarinic → miosis
Ciliary muscle	Mostly parasympathetic	Muscarinic → accommodation
Ciliary processes	$\beta_2 \rightarrow \uparrow$ and $\downarrow$ aqueous formation	
Outflow channels	$\beta_2 \rightarrow \uparrow$ outflow	
Outflow channels	$\alpha_1 \rightarrow \uparrow$ outflow	
Blood vessels	$\alpha_1 \rightarrow$ vasoconstriction	Muscarinic → vasodilatation
Blood vessels	$\beta_2 \rightarrow$ vasodilatation	
Lacrimal gland	*See Fig. 1.31*	Muscarinic → secretion
Müller's muscle	$\alpha_1 \rightarrow$ constriction (lid retraction)	

Parasympathetic stimulation of the muscarinic receptors located on longitudinal muscle fibres of the ciliary muscle produces a pull on the scleral spur, both posteriorly and internally. This causes the trabecular meshwork to be pulled open and leads to an increase in aqueous outflow. Pilocarpine in particular is therefore used in the treatment of glaucoma and ocular hypertension. This effect may be limited because theoretically these drugs may also reduce drainage by the uveoscleral pathway by contraction of the ciliary muscle and obliterating spaces between the muscle bundles.

The actions of these agents on the iris have no effect on IOP (as the iris is not directly attached to the sclera spur or trabecular meshwork) and miosis can be thought of as an unwanted effect in this process. The miosis itself causes darkening or dimming of vision, especially at night.

Pilocarpine has no serious side effects, but it causes a number of annoying ocular effects due to unwanted stimulation of muscarinic receptors at other sites, e.g. iris sphincter and blood vessels. Ciliary spasm is the most troublesome effect and is more pronounced in younger patients. It commences soon after instillation of the drops and lasts 2–3 hours. There can be an increase in accommodation of up to –5D. Headache and ciliary spasm are worst for the first 1–2 weeks of treatment.

Twitching and spasmodic contractions of the lids occur in patients using the cholinesterasae-inhibiting miotics as a result of the action on nicotinic receptors (somatic system).

Vasodilation and hyperaemia occur on instillation and last around 30 minutes. Cholinergic miotics cause a breakdown of the blood–aqueous barrier and therefore should be avoided in patients with neovascular glaucoma, uveitic glaucoma, or any condition with ocular inflammation.

Systemic adverse events are rare. They are occasionally seen with irreversible cholinesterase inhibitors and include head and brow ache, salivation and perspiration, increased gastric acid secretion, nausea and vomiting, increased tone in the gastrointestinal tract that results in abdominal pain and diarrhoea, bronchospasm, pulmonary oedema and respiratory paralysis, hypotension and bradycardia, and muscle weakness.

Table 9.6 Cycloplegic drugs and residual accommodation remaining after their usage

Drug	Residual accommodation (dioptres)
Atropine	Nil
Hyoscine	Less than 2D
Homatropine	0.5–2D
Cyclopentolate	1–2D
Tropicamide	1.6–3.2D

Sympatholytics (adrenergic antagonists)

Guanethidine (which is no longer available) causes depletion of stores of noradrenaline from the nerve terminals. This gives rise to a transient increase in outflow facility and mydriasis. When stores of noradrenaline are depleted,

Table 9.7 Cycloplegic drugs and their action

Drug/eyedrop	Onset of action	Onset of maximal cycloplegia	Recovery begins	Ability to read fine print	Full recovery
Atropine 1%	30 minutes → slow	1–6 hours	2–3 days	3–5 days	7–14 days
Hyoscine 0.25%	15 minutes	30–60 minutes	4D at 3–4 hours		3–7 days
Homatropine 2%	15 minutes → rapid	30–90 minutes	3.5–4.5 hours	5–6 hours	10–48 hours
Cyclopentolate 1%	Rapid 1–2D at 20 minutes	30 minutes (can be up to 75 minutes)	1–1.5 hours	3–8 hours	24 hours
Tropicamide 1%	Rapid	2D at 20–25 minutes	Not reliable after 35 minutes	2–4 hours	6 hours

Table 9.8 Drugs that cause mydriasis

Drug	Onset of action (minutes)	Maximum effect within (minutes)	Recovery to normal
Atropine 1%	10–15	30–40	7–14 days
Hyoscine 0.5%		20–30	3–7 days
Homatropine 1–2%	10–20	30–60	5 hours to 4 days
Cyclopentolate 0.5 and 1%	10–30	20–60	Less than 2 days
Tropicamide 1% Tropicamide 0.5%	10–30	20–40	6–9 hours 4–5 hours
Phenylephrine 2.5 and 10%	10–20	30–90	5–12 hours

Table 9.9 Duration of action of vasoconstrictor sympathomimetics

Adrenaline 0.1% (1:1000)	Duration 1 hour
Phenylephrine 0.125–0.25%	Duration 4 hours
Xylometazoline 0.05%	Duration 8 hours

aqueous humour formation is reduced and the pupil becomes miosed.

Thymoxamine is an α_1 antagonist and is used to reverse sympathetic mydriasis.

Management of IOP

Glaucoma management depends on the ability to lower IOP either pharmacologically or surgically. Over the last two decades there have been significant advances in the development of IOP-lowering agents.

β-*Adrenoceptor antagonists (*β-*blockers)*

These drugs block β_2-receptors in the ciliary epithelium and prevent formation of aqueous humour.

The commonly used β-blockers vary by:

- sympathomimetic activity: carteolol (teoptic) is the only topical β-blocker with intrinsic sympathomimetic activity (ISA). This may be a theoretical advantage in that there is less chance of developing bradycardia or other systemic side effects.
- selectivity: the ability to block β_1 or both β_1 and β_2 receptors. Timolol, carteolol, levobunolol, and metipranolol block both β_1 and β_2 receptors. Betaxolol predominantly blocks β_1 receptors and is termed 'cardioselective'. This, however, does not explain why it is effective at lowering IOP.

There are no ocular adverse effects, but systemically these drugs may have some effects. Bradycardia, reduced cardiac contractility, cardiac failure, arrhythmias, syncope, and hypotension may occur due to β_1 effects on the heart. Bronchospasm may occur due to β_2 effects on bronchial smooth muscle in the lung. These drugs are therefore contraindicated in patients with sinus bradycardia, second or third degree atrioventricular block, cardiac failure, cardiogenic shock, asthma, or chronic obstructive pulmonary disease.

The ophthalmic β-blockers are highly lipophilic and may cross the blood–brain barrier. This may lead to depression, fatigue, confusion, lethargy, and hallucinations. These effects are rare.

Adrenoceptor agonists (sympathomimetic agonists)

These drugs act by increasing the aqueous outflow facility. Adrenaline acts on α_1, α_2, β_1, and β_2 receptors. This has three actions:

- increased conventional aqueous humour outflow
- increased uveoscleral outflow
- increased aqueous humour formation.

It is thought that the most significant effect is the increase in conventional aqueous humour outflow, stimulated by β_2 and α receptors in the outflow channels.

Dipivefrine is a prodrug of adrenaline. It is lipophilic and therefore crosses the cornea more readily than adrenaline. This means a low concentration can be used, which should theoretically give rise to low incidence of adverse events. It is converted to adrenaline by ocular tissue enzymes. Its side effects include reactive hyperaemia from repeated vasoconstriction of the blood vessels. Allergic skin reactions and conjunctival, corneal, and eyelid deposits are common after several years. Soft contact lenses may become discoloured. In aphakes it may give rise to macular oedema. Systemically, these drugs can give rise to cardiovascular side effects. These include tachycardia, palpitations, and arrhythmias. There may be interactions with monoamine-oxidase inhibitors, with which they can cause a dangerous rise in blood pressure. Their effects may also be potentiated by the use of systemic sympathomimetic agents, e.g. tricyclic antidepressants.

α_2 Agonists act principally by reducing aqueous humour formation. Stimulation of α_2 receptors prevents further release of noradrenaline into the synapse, which in turn

Table 9.10 Action of parasympathomimetics

Drug	Commercial product	Action
Pilocarpine	Generic 0.5, 1, 2, 3, 4% (preserved) Minims 2, 4% (unpreserved) Pilogel 4% (preserved)	Directly acting muscarinic agonist (action similar to acetylcholine)
Carbachol	No longer available	Directly acting muscarinic agonist
Eserine/physostigmine	No longer available	Indirectly acting agonist (reversibly blocks acetylcholinesterase)
Ecothiopate/phospholine iodide	Named patient only 0.125% (preserved)	Indirectly acting agonist (irreversibly blocks acetylcholinesterase)

decreases aqueous humour formation. Apraclonidine has a slight α_1 activity, but brimonidine is predominantly α_2 selective. Both cause a decrease in aqueous humour formation but brimonidine has an additional effect increasing uveoscleral outflow facility. Apraclonidine is licenced for short-term use to lower IOP (its IOP-lowering effect diminishes over around 4 weeks). Brimonidine is licenced for long-term IOP control.

The main side effects are hyperaemia, pruritus, dry mouth, and taste disturbance. These drugs only have minimal cardiovascular side effects.

Prostaglandin $P_{2\alpha}$ analogues

Prostaglandin analogues are thought to lower IOP by increasing the non-conventional (uveoscleral) outflow facility of aqueous humour. It is thought that bimatoprost (lumigan 0.03%) also increases trabecular outflow.

Prostaglandin analogues occur naturally in the body in response to tissue injury, where they have a role in the inflammatory response.

Travoprost and latanoprost are prodrugs with good corneal penetration and are converted to active acid forms. There are different prostaglandin receptors and these two drugs act on the prostaglandin F receptor. The mechanism of how this affects the uveoscleral outflow is not fully understood.

Adverse effects are likely to be due to the role of prostaglandins in the inflammatory response (even though these molecules were modified to limit these effects). Foreign body sensation and hyperaemia are common initially, but these subside with time. There may be an increase in melanin in the stromal melanocytes of the iris. Eyelashes may also grow darker, longer, and thicker. There have been rare causes of macular oedema, iritis, and uveitis reported.

There are few reported systemic adverse effects from prostaglandin analogues, but exacerbation of asthma has been reported.

Carbonic anhydrase inhibitors

These act by blocking the enzyme carbonic anhydrase (CA) in the ciliary epithelial cells. CA is an enzyme found in:

- corneal endothelium
- pigmented and non-pigmented ciliary epithelium
- Müller cells
- retinal pigment epithelium.

It facilitates the following equilibrium:

carbon dioxide + water ↔ carbonic acid ↔ bicarbonate ion + hydrogen ion carbonic anhydrase

Ciliary epithelial cells lose bicarbonate and sodium ions into the posterior chamber. This causes an osmotic gradient into the posterior chamber. CA catalyses the conversion of carbon dioxide to bicarbonate ions. The remaining hydrogen ions result in an acidic environment. CA re-establishes the equilibrium by producing more bicarbonate. If CA is inhibited, however, the cell (ciliary epithelium, corneal endothelium, etc.) becomes more acidic. For this reason, CA inhibitors are not recommended when the corneal endothelium is compromised.

The structures of acetazolamide, brinzolamide, and dorzolamide are chemically similar to sulphonamides and therefore should not be given to patients with sulphonamide allergy. CA inhibitors are also contraindicated where there is low serum sodium or potassium, especially in liver or kidney failure.

CA inhibitors can be given topically, orally, or intravenously. There are fewest side effects with topical therapy although patients still complain of ocular hyperaemia and a bitter taste in the mouth.

Systemically, acetazolamide has a diuretic effect for the first few weeks of treatment, after which the effects wear off gradually. Patients complain of paraesthesia (hands and feet), taste disturbance, headache, and fatigue. Less commonly, patients may complain of dizziness, tinnitus, depression, gastrointestinal disturbance, blood dyscrasisa, renal colic, and calculi. There is also a risk of Stevens–Johnson syndrome.

Systemic osmotic agents

These increase the osmolarity of the extracellular fluid, allowing water to leave the eye down an osmotic gradient. They are effective for short-term reduction of an elevated IOP, e.g. in an attack of acute closed-angle glaucoma.

Oral glycerol lowers IOP within 1 hour, but this effect is short lived and may need to be repeated. It may cause dehydration, headache, nausea, and vomiting, limiting the possible dose used. It cannot be used in patients with diabetes because it has a high calorific value.

Mannitol can be used as an intravenous infusion. This also takes effect within 1 hour and lasts 5–6 hours.

Non-steroidal anti-inflammatory drugs

Eicosanoids are generated *de novo* from cellular phospholipids after a number of stimuli. They are mediators of the inflammatory response.

The two principal eicosanoids are thromboxanes and leukotrienes. The main source of these is the unsaturated fatty acid arachidonic acid (Fig. 9.3). This is found esterified in the cell membrane.

Stimuli that result in the generation of these include cell damage, thrombin from platelets, C5a from complement, and bradykinin. Non-steroidal anti-inflammatory drugs (NSAIDs) act by inhibiting cyclo-oxygenase and thus inhibiting the synthesis of eicosanoids.

NSAIDs can be administered:

- topically—treatment of cystoid macular oedema
- systemically—anticoagulation.

There are only a few side effects:

- aspirin may cause mental confusion (exacerbated by carbonic anhydrase inhibitors causing metabolic acidosis)
- exacerbation of asthma (increased leukotriene production).

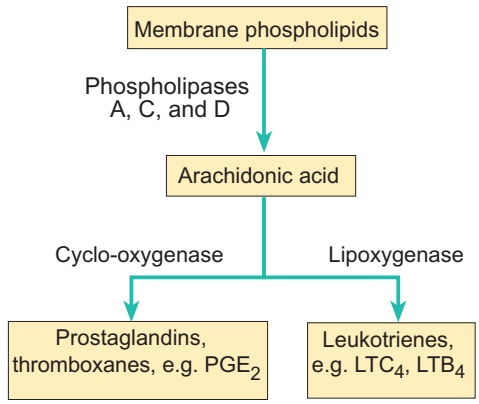


Fig. 9.3 Synthesis of eicosanoids.

Reproduced from Gillian Pocock and Christopher D. Richards, *Human Physiology, The Basis of Medicine*, Third edition, Figure 5.8, Page 56, 2006, with permission from Oxford University Press.

Glucocorticoids

Inflammation is characterized by:

- local hyperthermia
- capillary dilatation
- increased vessel permeability
- oedema
- protein leakage into the anterior chamber.

The inflammatory reaction can help to eradicate infection, but may also result in damage to tissue.

Glucocorticoids have different effects on lymphocytes, polymorphonuclear lymphocytes, vascular endothelial cells, fibroblasts, and other cells. They decrease mobilization of macrophages, cause sequestration of lymphocytes, decrease antibody production, decrease fibroblast and keratocyte activity, hence reducing scar formation, and finally inhibit neovascularizaton.

Inflammatory stimuli cause breakdown of phospholipids within the leukocyte cell membrane, liberating arachidonic acid, which is converted to endoperoxides. These quickly break down to prostaglandins (mediators of acute inflammation) and thromboxane (which causes platelet aggregation). Steroids prevent breakdown of these cell membrane phospholipids.

Glucocorticoids may be administered:

- topically, e.g. to prevent/suppress corneal graft rejection, anterior uveitis
- by periocular injection, e.g. intermediate uveitis
- systemically, e.g. for giant cell arteritis and posterior segment inflammation.

Local side effects:

- activation of collagenase precursors in keratocytes → precipitating corneal melting
- decrease of fibroblast and keratocyte activity, thus slowing epithelial healing
- impaired mobilization of leukocytes—eye (particularly left) more susceptible to HSV and fungal infections
- increased IOP
- cataract formation
- rebound leukocyte invasion after steroid withdrawal
- ptosis
- scleral melt.

Systemic side effects:

- suppression of pituitary–adrenal axis
- osteoporosis
- hyperglycaemia
- psychiatric effects, e.g. euphoria, psychosis
- aseptic necrosis of the hip
- peptic ulcer
- electrolyte imbalance.

Immunosuppressive agents

Immune-mediated disease in the eye may require systemic immunosuppression. The local and systemic conditions that require such therapy include allograft rejection, allergy, scleritis, thyroid eye disease, uveitis, and many more.

Steroids are still the mainstay of treatment, especially in the first instance. More recently, a number of other drugs have also been developed to treat such conditions. These have come about as our understanding of the immune system has moved forward.

Commonly used agents are:

- corticosteroids—act on cytosolic receptors and block transcription of cytokine genes
- methotrexate—folic acid antagonist, inhibiting dihydrofolate reductase and suppressing DNA synthesis
- tacrolimus—inhibits IL-2
- azathioprine—inhibits purine synthesis, which blocks RNA and DNA synthesis
- ciclosporin—inhibits IL-2 production and stimulates transforming growth factor (TGF)-β production
- mycophenolate mofetil—blocks *de novo* purine synthesis (selective for lymphocytes).

The main limitations of these drugs are their side effects and their lack of specificity. In response to this, modern molecular biology has generated molecules that can specifically target receptors, membrane proteins, or soluble proteins. The anti-angiogenic drugs in the next section are an example of these. Another example is infliximab, which is a chimeric antibody (human with mouse variable region specific for TNF-α) that inhibits TNF-α. Etanercept also binds TNF and thereby prevents further binding.

Anti-angiogenic drugs

An increased expression of VEGF has been found in neovascular membranes. Inhibitors of VEGF have been developed that are now established in the treatment of neovascular age-related macular degeneration (AMD).

Pegaptanib sodium slows visual loss in patients with wet AMD, and was the first drug to be approved for clinical use in patients with sight-threatening choroidal neovascularization (CNV).

Subsequently, ranibizumab, has been the subject of large phase III, multi-centre, randomized, double-masked clinical trials that have not only resulted in stabilization of AMD CNV, but also been associated with an improvement of vision for a significant number of patients. Based on subsequent cost-effectiveness analysis, ranibizumab was approved in the UK by the National Institute for Health and Clinical Excellence (NICE) for the treatment of this group of patients.

Bevacizumab is an anti-VEGF agent that has been approved for intravenous use as an adjuvant agent in the treatment of metastatic colorectal carcinoma. Because ranibizumab is relatively expensive, off-label bevacizumab was used for CNV treatment. Although at this stage published studies have all been uncontrolled it has been found to be well tolerated, with a similar side-effect profile to ranibizumab.

Intravitreal injection of anti-VEGF therapies carries a risk of developing cataract, retinal detachment, and endophthalmitis. Intravitreal injections of anti-VEGF therapies can result in some systemic absorption of the drug. This has been linked to an increased risk of stroke. By targeting the VEGF 165 isoform only, pegaptanib sodium is thought to reduce the possibility of major systemic vascular accidents compared to ranibizumab and bevacizumab, which target all VEGF isoforms. However, the visual results with bevacizumab and ranibizumab are better than with pegaptanib sodium: as yet, there are no controlled trials comparing ranibizumab and bevacizumab.

Anti-angiogenic therapies are currently the subject of much research and in future more targeted therapies will emerge.

Tear substitutes

These aim to replicate the action of the mucin component of the tear film, thus increasing wettability and tear retention time. They consist of a solution of inorganic ions (0.9% NaCl) and polymers. Polymers commonly used include polyvinyl alcohol and semi-synthetic celluloses (e.g. methylcellulose, hydroxypropylmethylcellulose, and hydroxycellulose). Hyaluronic acid has even greater retention times and improves tear film stability. Polyacrylic acid (carbomer) is a hydrophilic gel that forms a stable tear film.

Recently, tear substitutes have become available in preservative-free formulations. Preservatives affect corneal epithelial stability and ability to maintain precorneal tear film.

Topical anaesthetic agents

Local anaesthetic agents (Table 9.11) produce a reversible blockade of nervous impulses in a restricted region of the body. They act by blocking sodium transportation via ion channels, into the nerve fibre, thus preventing generation and conduction of the action potential. These agents are active in their cationic state (positively charged). As the concentration increases, the amplitude and rate of rise decreases. Thus, conduction velocity slows and refractory period lengthens. The firing threshold increases and eventually the neuron becomes inexcitable. After this, increasing local anaesthetic concentration has no further anaesthetic effect. To increase depth of anaesthesia and duration of action, repeated instillations at intervals are required.

Tetracaine/amethocaine is a potent anaesthetic that has some action on lids and is therefore useful for removing foreign bodies. It may cause irregularities of the corneal surface, however, so is not preferred for contact tonometry.

Oxybuprocaine (benoxinate) can be used for contact tonometry.

Lidocaine (lignocaine) contains an amide link (instead of an ester link as in tetracaine (amethocaine), oxybuprocaine (benoxinate), proxymetacaine, and cocaine), making it more resistant to hydrolysis and hence longer acting. It is also useful in cases of allergy to the other ester-linked anaesthetics.

Proxymetacaine is the most commonly used topical anesthetic for contact tonometry. It stings the least of all the anaesthetics and is therefore useful in children. However, it is the least stable and must be stored in the fridge.

Cocaine is an alkaloid from *Erythroxylon coca*. It was the first successfully used topical anaesthetic in 1818. It is rarely used in ophthalmic clinics due to toxicity and legal restrictions. It has indirect sympathomimetic activity by preventing the reuptake of noradrenaline into neuronal endings. This results in pupillary dilatation.

By virtue of their toxicity, topical anaesthetics may damage the epithelium by decreasing the motility of epithelial cells. They also increase the permeability of corneal epithelium to other drugs and chemicals, and therefore increase

Table 9.11 Action and recovery of topical local anaesthetics

Drug	Duration of onset (seconds)	Duration of action (minutes)	Total recovery (minutes)
Tetracaine (amethocaine)	10–30	10–20	~30
Oxybuprocaine (benoxinate)	30	10–20	30–60
Cocaine	30	10–15	30–60
Lidocaine (lignocaine)		30–45	~45
Proxymetacaine	10–14	10–20	~30

CLINICAL TIP

All topical anaesthetic agents have side effects of stinging. Cocaine is the worst and proxymetacaine is the best. All eliminate the blink reflex, so care must be taken to protect the ocular surface from foreign body contamination and patients must be warned not to rub their lids. Topical anaesthetics also reduce wetting and therefore affect the result of tear flow test strips due to loss of blink reflex.

their toxicity. As well as being toxic to humans, they reduce the replication rate of bacteria and viruses, which may affect swab results. For this reason, local anaesthetic agents may not be prescribed for prolonged use. Cocaine has the worst toxicity profile and proxymetacaine is the best with regard to toxicity. Clinically this may be seen as a punctuate keratitis, then epithelial loss in the palpebral aperture. The stroma may have a grey ground-glass appearance with infiltration in the deeper layers. Most cases make a full recovery within weeks.

Diagnostic staining agents

Fluorescein sodium is commonly used in ophthalmic clinics. It is a water-soluble dye that can be found neat or mixed with lidocaine or proxymetacaine. In alkaline solution (pH > 7) fluorescein sodium exhibits green fluorescence, whereby it emits light of a longer wavelength when stimulated by light of shorter wavelength. Healthy corneal epithelium (with high lipid content) resists penetratioin of water-soluble fluorescein and is therefore not coloured by it. Epithelial defects stain bright green when using a cobalt blue light. This is because epithelial loss permits rapid fluorescein entry to Bowman's layer, stroma, and even the anterior chamber.

Intravenous fluorescein can be used to detect leakage from blood vessels in the retina. Fluorescein is 80% bound to plasma proteins and red blood cells. Under normal circumstances, the choriocapillaris is leaky to fluorescein (bound and unbound). From the retinal vessels, only the unbound portion enters the eye, unless the blood–ocular barrier is damaged such that it permits proteins to leak out of blood vessels.

Oral fluorescein is excreted in the urine within 48 hours. The main side effect is that urine is discoloured. There is also a risk of anaphylaxis and teratogenicity.

Rose Bengal (1% aqueous solution) stains devitalized (dead) corneal and conjunctival cells red in dry eye conditions and other ocular disorders. It also stains mucus. However, it may be very irritant or even painful and therefore topical anaesthetic should be used beforehand.

Indocyanine green is used in a new angiography technique. It is retained to a large extent not only in retinal vessels but also in the choriocapillaris. On absorption of light, it also has the ability to fluoresce through pigment and haemorrhage (unlike fluorescein).

Antibiotics and antivirals

These agents are covered in Chapter 5, pp. 162–4.

Anti-allergy drugs

Allergic eye disease is a hypersensitivity reaction in which antigen combines with IgE on the surface of mast cells, causing liberation of histamine. Clinically, this causes itching, hyperaemia, and pain.

Histamine is derived from the amino acid histidine, is synthesized and stored in most tissue, and excreted in the urine after being degraded in the liver by histaminidase. It is a mediator of the inflammatory response, especially in the type 1 hypersensitivity response. It also plays an integral response in neurotransmission, e.g. in regulating gastric acid secretion.

Two common histamine receptors are H_1 and H_2, based on the antagonist that can bind to them. H_1 activation results in an increase in intracellular calcium, and this type of receptor is found in human bronchial muscle. H_2 activation results in activation of adenyl cyclase and second messenger production, and this type of receptor is found in the stomach, heart, and uterus.

H_1 antihistamines are used topically in ophthalmology (Table 9.12). They inhibit increased vascular permeability caused by histamine. Their side effects may be as a depressant of the CNS and they may have an anti-emetic effect.

Mast cell stabilizers (Table 9.13) inhibit the release of histamine from the mast cells within the mucosa by stabilizing mast cell membranes. These must therefore be administered before mast cell priming with IgE and allergen as there will be no effect after the mast cells have already degranulated.

Agents used to treat allergic eye disease include:

- antihistamines
  - block H_1 histamine receptors (Table 9.12)
  - mast cell stabilizers (Table 9.13)—prevent histamine release from mast cells
- NSAIDs—arachidonic acid metabolites are found in higher quantities in the tears of allergy patients
- steroids—last resort as a temporary measure.

Table 9.12 Antihistamine drugs

Drug
Antazoline 0.5%
Azelastine 0.05%
Emedastine difumarate 0.05%
Epinastine 0.05%
Levocabastine hydrochloride 0.05%
Olopatidane hydrochloride 0.1%

Drug toxicity

Adverse drug reactions are unwanted effects seen at therapeutic doses.

Side effects are non-elective actions of the drug that are unavoidable and quite predictable, e.g. respiratory depression with morphine or sedation with antihistamines taken for hay fever.

Secondary effects are unwanted effects resulting from the action of the drug, e.g. opportunistic fungal infections (thrush) subsequent to the administration of antibiotics.

Idiosyncratic effects occur only in susceptible individuals on an allergic or genetic basis (affected individuals are termed as hypersensitive or intolerant).

Types of toxic effect:

- cellular damage (cytotoxicity), e.g. paracetamol
- carcinogenicity, e.g. diethylstilbestrol
- mutagenicity, e.g. anticancer drugs
- teratogenicity, e.g. thalidomide
- drug allergy, e.g. penicillin.

Targets/mechanisms of toxic agents:

- corrosive actions, e.g. acids/alkalis
- receptors/ion channels, e.g. morphine
- biochemical pathways/energy supply, e.g. salicylate
- calcium homeostasis—activates Ca^{2+}-dependent degradative enzymes (glutamate excitotoxicity)
- DNA—initiation and promotion phases for carcinogenesis by activation of oncogenes
- immune system—drugs (more commonly a reactive metabolite) usually combine covalently with endogenous protein to form an effective antigen
- foetus—teratogenesis depends on the stage or organogenesis when exposure occurred.

Quantitative aspects

The LD_{50} is a simple and convenient measure of the acute toxic potential of a drug. It is defined as the dose of a drug that kills 50% of treated animals within a specified short period of time. It gives no information on non-lethal effects or long-term effects. However, the toxic effects in humans are not the same as in animals, e.g. the limb abnormailities caused by thalidomide could not have been predicted by tests on rats. Conversely, drugs that are safe in humans may have produced abnormalities in rats.

Table 9.13 Mast cell stabilizer drugs

Drug
Sodium cromglicate 2%
Lodoxamide trometamol 0.1%
Nedocromil 2%
Ketotifen 0.025%

Ocular toxicity

Preservatives in topical medication

To keep ophthalmic solutions sterile, preservatives are added. Most are toxic to the corneal epithelium. This may assist in ocular absorption of drugs.

Benzalkonium chloride is a commonly used cationic surfactant preservative. It attains its bactericidal properties by attaching to the bacterial cell wall, increasing permeability and eventually rupturing the cell wall. It is most effective at an alkaline pH (pH 8.0), but is inactivated in the presence of soaps and salts of, for example, magnesium and calcium. EDTA (a chelating agent) is therefore often added. Other preservatives include thiomersal, chlorhexidine, and chlorbutol.

Severe toxicity may result from direct cellular damage or hypersensitivity reaction to components of the drug. This may give rise to a papillary conjunctivitis, punctuate keratitis, and corneal oedema.

Drugs themselves may be the cause of ocular toxicity:

- Antimalarials
  1. Chloroquine—binds melanin in RPE and uveal tract
     - Causes maculopathy, verticillata, and lens deposits
  2. Hydroxychloroquine
     - Fewer side effects than chloroquine
     - May cause verticillata
  3. Quinine
     - Damage to RPE/photoreceptors, optic nerve, retinal vasculature
- Antimycobacterials
  1. Ethambutol
     - Retrobulbar optic neuritis
     - Red–green colour affected
     - Safe dose 15 mg/kg/per day
  2. Rifampicin
     - Discolours tears and soft contact lenses (orange)
     - Uveitis
     - Thickened conjunctival secretions
- Antibiotics
  1. Chloramphenicol—optic neuritis
  2. Sulphonamides—Stevens–Johnson syndrome
  3. Metronidazole
     - Transient visual loss
     - Oculogyric crisis
     - Optic neuritis
  4. Tetracyclines—conjunctival deposits

- Tamoxifen (> 40 mg/day over 1 year)
  - White/yellow macular deposits
  - Cystoid macular oedema
  - +/– Optic neuritis/optic oedema
- Digoxin
  - decreased visual acuity and colour vision
- Amiodarone
  - Corneal verticillata
  - Photosensitivity
- Steroids
  - Cataract
  - IOP increase
- NSAIDs
  1. Indometacin
     - verticillata
  2. Ibuprofen
     - toxic amblyopia
- Gold
  - Corneal verticillata
- Phenothiazines

  Chlorpromazine
  - Pigmentary granules in conjunctiva, cornea, and sclera
- Oral contraceptive pill
  - Retinal artery/vein occlusion
- Penicillamine
  - Myasthenia → ptosis, diplopia
- Anticholinergics
- Tricyclic antidepressants, antihistamines, ipatropium, scopolamine, atropine
  - Dry eyes
  - May precipitate acute angle-closure glaucoma

CHAPTER 10

Lasers and instrument technology

Lasers

A laser (Light Amplification by the Stimulated Emission of Radiation) is an instrument that produces coherent light (all photons have the same wavelength and are in phase) that is collimated (waves are in parallel).

Atoms are most stable in their lowest energy state (ground state). Energy is delivered to these atoms by a process called pumping. On absorbing energy, an atom elevates its electrons to a higher energy level. The electrons may remain at an upper energy level without returning the atom to its ground state. If there are more atoms with electrons in the excited state, this is termed population inversion. On returning to their ground state, electrons from these atoms release light energy. This energy is incoherent and travels in all directions.

If an electron at a higher energy level is further stimulated by a photon of light that is at a wavelength the atom would normally produce, the atom will drop to a lower energy level and the emitted light will be coherent with the stimulating photon. Most of the light emitted by the active medium is incoherent spontaneous emission, but a small amount of that released by stimulated emission can be amplified.

The active laser medium is contained in a tube with a mirror placed at each end. The distance between the two mirrors is important because if it is a multiple of the wavelength of the emitted light then resonance can occur. The stimulated emission is reflected and re-reflected at both mirrors. The resulting light is exactly in phase and reinforces itself (resonance). As further stimulated emissions occur, the light becomes stronger and stronger while remaining in phase (coherent). If one mirror is made partially transparent, some light may escape the tube. This light will be coherent (in phase), monochromatic (one wavelength), and collimated (rays parallel). If the light is produced continuously, the laser is said to be operating in continuous-wave mode.

The actual luminous flux emitted is only a small proportion of the total energy input as a great deal of energy has to be pumped to get a small number of coherent waves. Because the luminous flux is concentrated in a fine parallel beam this beam is extremely bright.

Lasers are named after their active medium, which contains all the atoms or molecules that will undergo stimulated emission. This may be gas (argon, krypton, carbon dioxide), liquid (dye), or solid (neodymium doped yttrium aluminium garnet crystal, Nd:YAG). The source of energy pumped into the laser medium may be electrical, incoherent light, or light from a second laser.

Laser light is not precisely parallel in practice, however. In fact, it is slightly divergent and more intense at particular points (transverse electromagnetic modes). For continuous diffuse delivery, such as argon laser photocoagulation, this is not significant. However, for photodisruption, e.g. YAG capsulotomy, it is important to focus energy at a point to achieve a disruptive effect. Energy is most concentrated at the centre of the laser beam at any point and diminished peripherally. Newer lasers are able to increase the distribution of energy towards the centre of the beam.

A continuous beam of light is produced by an argon laser, for which output is measured in watts (energy (joules) per unit time). A YAG laser has a peak of power that is delivered over a fixed period of time and is therefore measured in joules. Power delivered may be increased by 'Q-switching' (aka giant pulse formation), which produces a brief pulsed output of light. Here, a shutter is placed before one of the two mirrors in the laser tube. When this shutter is closed, the laser medium is pumped, leading to population inversion. Energy stored in this way eventually reaches a maximum, at which point the Q-switch is activated. Lasering cannot occur until the shutter opens and light can start resonating. Because a large amount of light is stored in the laser medium, the laser light resonating builds up quickly and a strong short pulse of light is produced. In the Nd:YAG laser, the Q-switch is made of a material whose transmission increases when light intensity exceeds a certain threshold. After the pulse, the material of which the switch is made returns to its closed conformation, so the next pulse is delayed until the energy in the gain medium is fully replenished. Control of the laser can be achieved by using a pulsed pump source.

In practice, there are several frequencies and wavelengths of light amplified in a laser. These depend on the:

- gain medium: each medium produces a range of frequencies or bandwidths

- resonant cavity: light reflecting between two mirrors will produce constructive and destructive interference, resulting in the formation of standing waves.

The standing waves form a discrete set of frequencies, known as the longitudinal modes of the cavity. These modes are the only frequencies of light that are self-regenerating and allowed to oscillate by the resonant cavity; all other frequencies of light are suppressed by destructive interference.

Each of these modes can oscillate between the two mirrors independently, with no relationship with each other. In effect, they act like a series of independent lasers within the same tube and medium that emit light at slightly different frequencies.

The individual phase of the light waves in each mode is not fixed and therefore may vary depending on external conditions such as temperature. In lasers with only a few oscillating modes, interference between the modes may result in random fluctuations in light intensity. In lasers with large numbers of modes, the interference averages out and the laser operation is termed continuous wave.

Instead of oscillating independently, if each longitudinal mode operates with a fixed phase relative to the modes, they will periodically all constructively interfere with one another, producing an intense burst of light. Such a laser is said to be mode locked or phase locked.

Mode locking can be done actively by modulating the signal inputted into the laser cavity, e.g. by placing a standing wave acousto-optic modulator into the laser cavity, which produces a sinusoidal amplitude modulation of the light in the cavity. It can be achieved passively, e.g. by introducing a saturable absorber into the laser cavity, which behaves differently based on the intensity of light passing through it.

Effect of lasers on the eye

Radiation of wavelength from 4 to 1400 nm can reach the retina through the ocular media. What effect the laser has on the tissue depends on the wavelength, pulse duration, and nature of the tissue to which it is applied.

Laser damage may be due to:

- Thermal effects: photocoagulation produces a rise in temperature of 10–20°C, causing photocoagulation and a localized burn. CO_2 lasers raise the tissue temperature to over 100°C and thus cause vaporization of the tissue to which they are applied. Thermal damage or burn occurs when tissues are heated to the point of denaturation. The thermal effect depends on the pigment present. Melanin absorbs most of the visible spectrum of light, xanthophyll at the macula absorbs blue light and shorter wavelengths more than the longer red or green wavelengths, and haemoglobin absorbs blue, green, and yellow light. Photocoagulation of choice therefore would avoid the shorter wavelengths (which are also more absorbed by the cornea and lens). Ideally a red or green laser should be used in the eye.
- Ionization: photon energy may strip electrons from molecules, forming ions. A plasma is formed, where these ions are dispersed in the gaseous phase. This plasma is of high temperature and expands rapidly, producing mechanical shock and displacing tissue. An example is the Nd:YAG and argon–fluoride excimer lasers.
- Photochemical effects: with pulse duration greater than 10 seconds, free radicals are formed which are toxic to cells. Shorter wavelengths (blue, UV) cause damage at lower levels of irradiance and therefore are more harmful.

Therapeutic lasers in ophthalmology

Lasers are now used universally in ophthalmic departments for a number of conditions and the scope for lasers in clinical practice is expanding. Laser light at present can be delivered via a fibre-optic cable, by slit-lamp, by indirect ophthalmoscopy, or by a hand-held probe during vitreoretinal procedures.

- The argon blue–green gas laser: these are most commonly used for argon laser photocoagulation. They operate at 488 nm (blue) or 514 nm (green) or a mixture of the two. Photocoagulation should act on the outer retina and avoid the nerve fibre layer of the inner retina. Green light is well absorbed by melanin and haemoglobin. Blue light is absorbed by xanthophylls in the inner layer of the macula and thus is contraindicated in this area.
- The He:Ne gas laser: this is a low power laser with a visible red light (633 nm) that is used as an aiming beam for lasers with an invisible output.
- The diode laser: this emits infrared light (810 nm) in continuous wave mode. Diode laser light is absorbed only by melanin. This laser is efficient, generates little excess heat, is portable, and there is little laser scatter. It can be used for retinal photocoagulation (even if the retina is obscured) and photocycloablation for glaucoma as this wavelength penetrates sclera.
- The Nd:YAG laser: this emits 1064 nm infrared radiation. This is a power continuous wave laser that is Q-switched and is often used to perform a capsulotomy in cases of posterior capsule opacification or iridotomy in shallow angle glaucoma. An He–Ne laser produces a red aiming beam. The YAG laser beam must be focused before use.
- The excimer laser: this uses an argon–fluorine (Ar–F) dimer laser medium to emit 193 nm ultraviolet radiation. This is highly absorbed by the cornea and causes ablation (tissue removal), which can be precisely determined. This laser is well suited to performing photorefractive keratectomy (PRK), laser intrastromal keratomileusis (LASIK) to reshape the cornea surface, and phototherapeutic keratectomy (PTK) to remove abnormal corneal surface tissue.

Diagnostic lasers in ophthalmology

Lasers can be used to image and scan parts of the eye. There are currently several diagnostic applications of lasers:

- Confocal microscopy: confocal optics are used to increase the contrast and resolution of the image by minimizing the amount of scattered light. Instead of flooding

the object to be viewed from every angle, as in the conventional microscope, the confocal microscope uses point illumination and a pinhole in an optically conjugate plane in front of the detector to eliminate out-of-focus signal. In practice this is achieved by using a laser source of illumination. The field being observed is increased by scanning across the area being examined. In ophthalmology this technique is used practically to view layers of the cornea.

- Optical coherence tomography (OCT): OCT is an interferometric technique using near-infrared light. Here, the optical beam of light directed at a tissue results in most of the light scattering, but a small amount of light being reflected. The scattered light does not contribute to the image but instead contributes to glare. The interferometer uses the principle of superimposition to combine separate coherent light waves together such that their combination has constructive or destructive interference to distinguish scattered light from reflected light. OCT captures micrometer-resolution, three-dimensional images from optical scattering media. The use of relatively long wavelength light allows it to penetrate more deeply into the sample than confocal microscopy (1–2 mm beneath the surface). It has a better resolution (around 10 μM) than other forms of imaging, including magnetic resonance imaging (MRI) and ultrasound. In time-domain OCT the path length of the reference arm is translated longitudinally in time. In frequency-domain OCT the broadband interference is acquired with spectrally separated detectors. This improves the signal-to-noise ratio. Spatially encoded frequency domain OCT (spectral or fourier domain) extracts spectral information by distributing different optical frequencies onto a detector stripe via a dispersive element and therefore obtaining all information in one exposure.
- Scanning laser polarimetry: this technique uses polarized light to measure the thickness of the retinal nerve fibre layer (RNFL) based on its birefringent properties. A confocal scanning laser ophthalmoscope (CSLO), which uses confocal optics, projects a spot of polarized laser light into the eye. As light passes through birefringent tissue (cornea and nerve fibre layer), it slows and changes (retardation). Change caused by the cornea is 'compensated' for. The light detectors sense this change and convert it into thickness units.

An example of an instrument that uses scanning laser polarimetry is the GDx Nerve fiber analyzer (Laser Diagnostic Technologies Inc.), which is used clinically to measure nerve fibre layer thickness, which is the first structure to be damaged by glaucoma.

Laser safety

Lasers are labelled by safety class number:

Class 1: safe, light contained in an enclosure, e.g. CD player

Class 2: safe during normal use, e.g. laser pointer—up to 1 mW power

Class 3a: small risk of damage within the time of the blink reflex—up to 5 mW power

Class 3b: may cause severe eye damage on exposure, e.g. CD and DVD writers—up to 500 mW power

Class 4: can burn skin and cause eye/skin damage, e.g. industrial and scientific damage.

People working with class 3b (including ophthalmic lasers) and class 4 lasers can protect their eyes with safety goggles which are designed to absorb light of a particular wavelength.

Optics of ophthalmic instruments

Direct ophthalmoscope

Hermann von Helmholtz (1821–94), a German physician, invented the direct ophthalmoscope in 1851. This gives a virtual erect image without the need for a condensing lens. At any one time, an area of about two disc diameters is in view (field of view) and the retina can be examined to the equator. There is no stereopsis because it is a monocular viewing system.

Fig. 10.1 shows that light is condensed to a point, which is then reflected by a mirror onto the retina. An image is returned to the retina. The mirror has a hole in it which allows the image to be viewed.

The area of retina that corresponds with the hole in the mirror (or the size of the pupil—whichever is smaller) is the field of view.

Light passing through the viewing hole must be parallel and this light then will be focused onto the retina of the observer using the ophthalmoscope by her or his own cornea and lens, assuming s/he is emmetropic.

CLINICAL TIP

The field of view is enlarged when the pupil of either the subject or the examiner is dilated and also when the distance between the subject and the examiner is decreased.

The image size is smaller in hypermetropia and larger in myopia than in emmetropia.

Indirect ophthalmoscope

The indirect ophthalmoscope (Fig. 10.2) consists of a light attached to a headband and a powerful hand-held convex (condensing) lens (usually around 20D). This provides a wider and indirect view of the inside of the eye, which allows a view of the peripheral retina. This image is vertically and horizontally inverted.

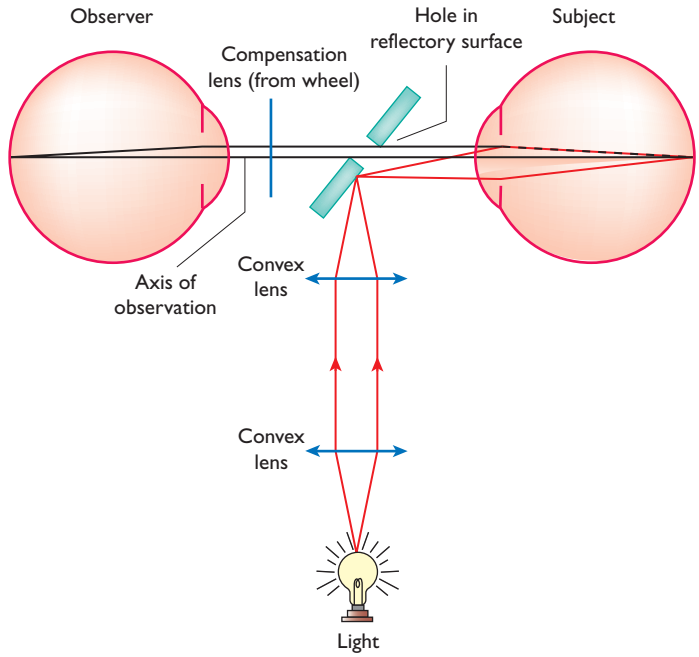


Fig. 10.1 The direct ophthalmoscope.
With permission from Neil Modi.

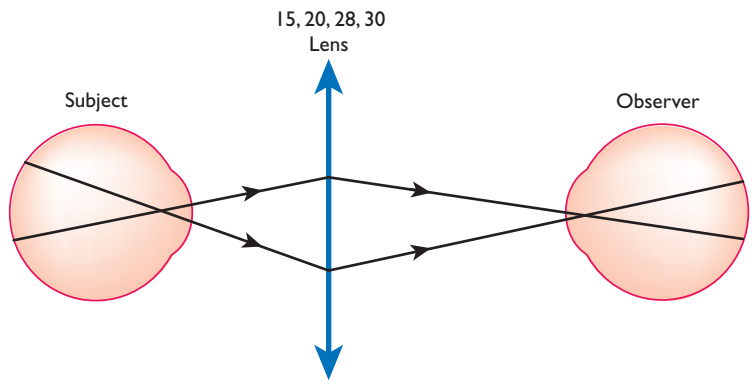


Fig. 10.2 The indirect ophthalmoscope.
With permission from Neil Modi.

The light shines through the headlight, through the condensing lens, and into the eye of the subject. Light shines back from the retina and is refracted by the condensing lens to form a real image of the retina between the condensing lens and the observer. Fig. 10.2 illustrates the optics of the indirect ophthalmoscope. The real image is viewed directly by the observer.

The condensing lens is held by the examiner at such a distance that both the patient's and examiner's pupil are conjugate foci. This means that light arising from one pupil is brought to a focus by the condensing lens at the other pupil.

The field of illumination is largest in myopia and smallest in hypermetropia, and in all refractive states the size of the subject's pupil limits the field of illumination.

Retinoscope

The retinoscope can give an objective measure of the refractive state of an eye by the technique of retinoscopy. Light is shone into the patient's eye to illuminate the retina (illumination stage). An image of the illuminated retina is formed at the far point of the subject's eye (reflex stage).

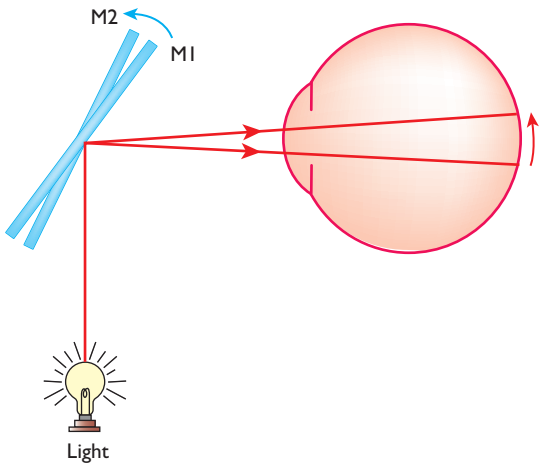


Fig. 10.3 Illumination stage of retinoscopy.
With permission from Neil Modi.

The image at the far point is viewed by the examiner by moving the light across the fundus and observing the behaviour of the reflex in the subject's pupil (projection stage).

When a plane mirror is used, light is moved across the fundus from A to B by rotating the mirror from M1 to M2 (Fig. 10.3). The light rays move in the same direction as the mirror.

An image A_1B_1 of the illuminated retina is formed at the patient's far point (Fig. 10.4). If the examiner is at the patient's far point, all the light enters the examiner's pupil and illumination is uniform. If the peephole is not at the far point of the patient's eye, some rays from the pupil will not enter the peephole and illumination of the pupil is incomplete.

If the patient is emmetropic (Fig. 10.4b), emerging light rays are parallel. If the patient is myopic (Fig. 10.4c), the emerging light rays are convergent. If the patient is hypermetropic (Fig. 10.4a), the emerging rays will be divergent.

These images are constructed using three rays:

1. A ray from point A on the retina, R, along the principal axis.
2. A ray from point B, parallel to the principal axis, up to the principal plane, P, of the eye. From here it is refracted to pass onward through the anterior principal focus, F_a, of the eye.
3. A ray from point B, which passes undeviated through the nodal point, N.

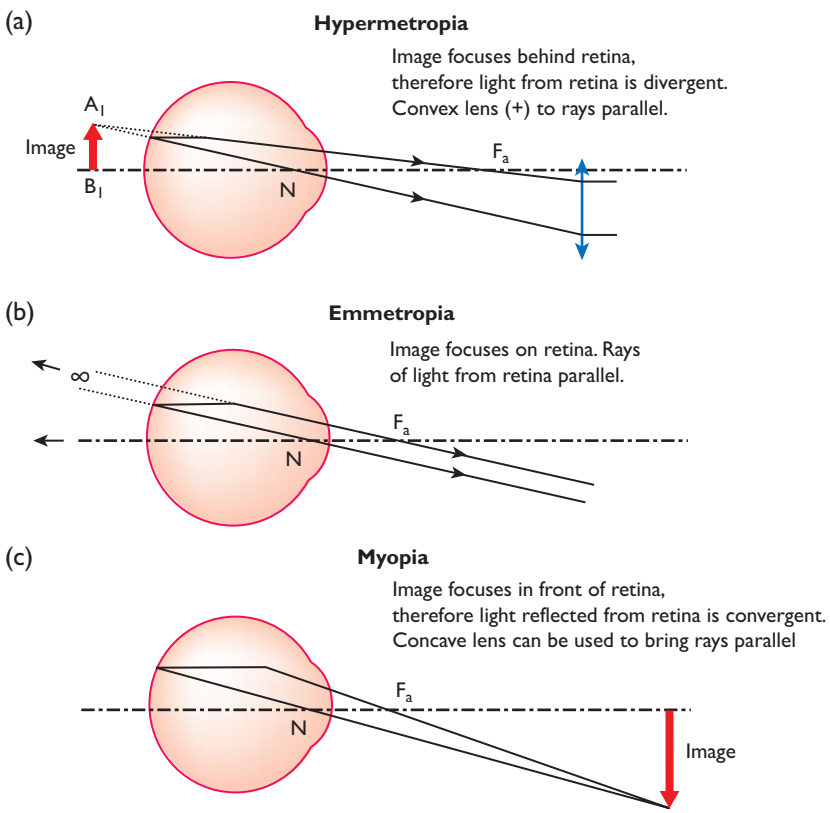


Fig. 10.4 Reflex stage of retinoscopy.
With permission from Neil Modi.

These diagrams can best be remembered based on the position of the far point, which is furthest from the principal plane in hypermetropia and vice versa in myopia.

The examiner views the image (A_1B_1) of the illuminated retina (AB) from a distance of usually ⅔ m. A_0B_0 is the image of A_1B_1 on the examiner's retina. Fig. 10.5 shows this construct of light leaving the subject's retina and reaching the examiner's retina for hypermetropes, emmetropes, and myopes. The examiner here is not able to see an actual image of the retina as in ophthalmoscopy, but is able to see rays from A_1B_1 as an illuminated area of retina.

In hypermetropes (Fig. 10.5a), the image is behind the subject. Hence the luminous reflex in the observer's retina is seen as a 'with' movement. The converse is true in myopes (Fig. 10.5c), where the image is between the subject and examiner and hence appears inverted to the examiner. This produces an 'against' movement where the luminous reflex moves in the opposite direction to the movement of the illuminating light.

It must be remembered that the examiner is determining the refractive error from the distance at which he/she is looking through the peephole of the retinoscope. The diopteric equivalent of this working distance should be subtracted from the correcting lens to work out the patient's actual distance correction. If working at ⅔ m then the observer subtracts 1.5D from the correcting lens.

If a myope has myopia less than the diopteric value of the observer's working distance (i.e. for ⅔ m working distance, less than 1.5D) then a 'with' movement is seen; for example for a myope of –1.0D a 'with' movement will be seen and this can be brought to neutral by using a lens of +0.5D minus the working distance of –1.0D myopia. In this situation, the image is behind the examiner.

Clinically this is used to measure objective refraction by placing lenses of varying strengths in front of an eye until the point of reversal is noted. A correction is made for working distance (–1.5D for ⅔ m working distance and –1.0D for 1 m) to account for the additional necessary refractive power for light to form an image on the examiner's retina given that they are at a distance of ⅔ or 1 m.

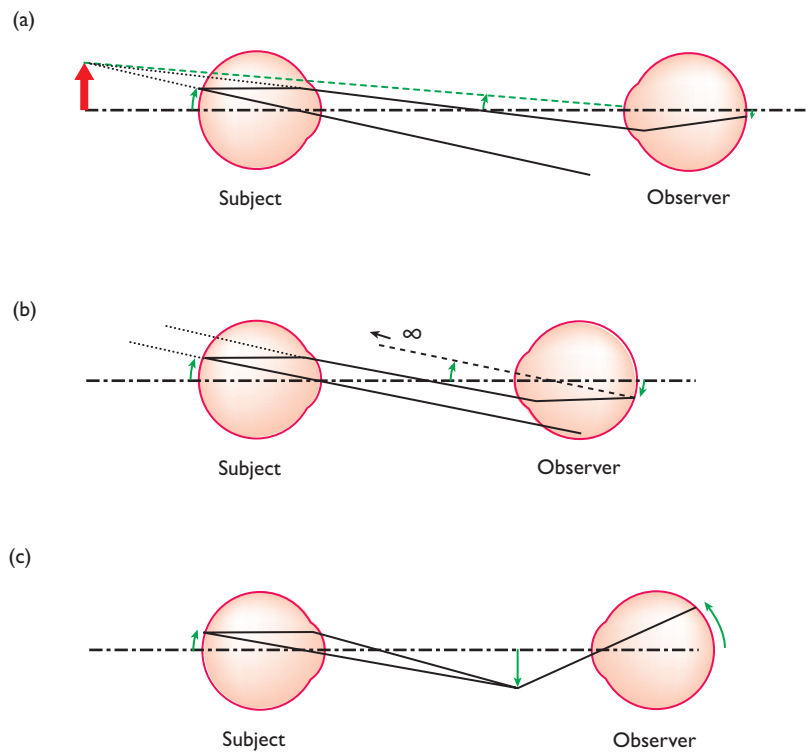


Fig. 10.5 Projection stage in (a) hypermetropia, (b) emmetropia, and (c) myopia greater than –1.5D (working distance).
With permission from Neil Modi.

Compound microscope

This provides a magnified view of a near object and in ophthalmology it is used to provide a magnified view of the eye (Fig. 10.6).

- It consists of two convex lenses (+), which are termed the objective and eyepiece lenses.
- The object to be studied must be placed outside the anterior focal point of the objective lens, F_o.
- This provides a real and magnified but inverted image some distance behind the objective lens.
- The eyepiece is positioned so that the image formed by the objective falls at or close to its principal focal plane, F_e.
- The eyepiece acts as a 'loupe' and magnifies further the image seen by the observer.
- The final image is vertically and horizontally inverted.

Slit lamps in clinical use have Porro prisms incorporated to obtain an image that is correctly orientated, both vertically and horizontally.

Slit lamp and fundoscopy

The slit lamp is a Nobel-prize-winning examination tool that is universally used in ophthalmology.

A slit lamp consists of two compound microscopes orientated at 13–14° from each other to provide a binocular view with stereopsis and an illumination system. Illumination is provided by an adjustable light source that can produce a slit of light of measurable dimension that rotates around an axis to give cross-sectional illumination of the eye.

The microscope and illumination are arranged such that the point on which the microscope is focused corresponds to the point at which light is focused.

Slit lamps have the ability to change their magnification (zoom). Given that the distance from the object to the lens is fixed, the power of the lens must be changed. Most simply this can be achieved by placing a lens between the microscope lenses, but this results in a change of image position. This problem can be remedied by using several mobile lens elements.

CLINICAL TIP

Methods of examination with a slit lamp

- Direct focal illumination—this is most commonly used and is appropriate in most situations. The slit-beam of light is focused on the part of the eye being examined.
- Diffuse illumination—the light is out of focus and diffusely illuminating the area being examined.
- Lateral illumination—the structure being examined is illuminated by light from the slit lamp that has been illuminated to one side of it.
- Transillumination—this is used to show areas of iris atrophy or the patency of an iridotomy. The illuminating and viewing systems are co-axial and light is shone through the pupil. Light is reflected from the retina and shows up defects in the iris, for example as area with red reflex rather than an iris silhouette in the normal eye.
- Specular reflection—the patient's gaze is directed straight ahead. The examiner will position the illuminating and viewing arms at equal angles from the direction in which the subject is gazing (effectively the 'normal'). Examining the light reflected off a surface in specular reflection is the best way of examining the corneal endothelium.
- Sclerotic scatter—light directed at the limbus illuminates the whole limbal area because the light from the slit beam is reflected internally within the cornea and is scattered all around the cornea.

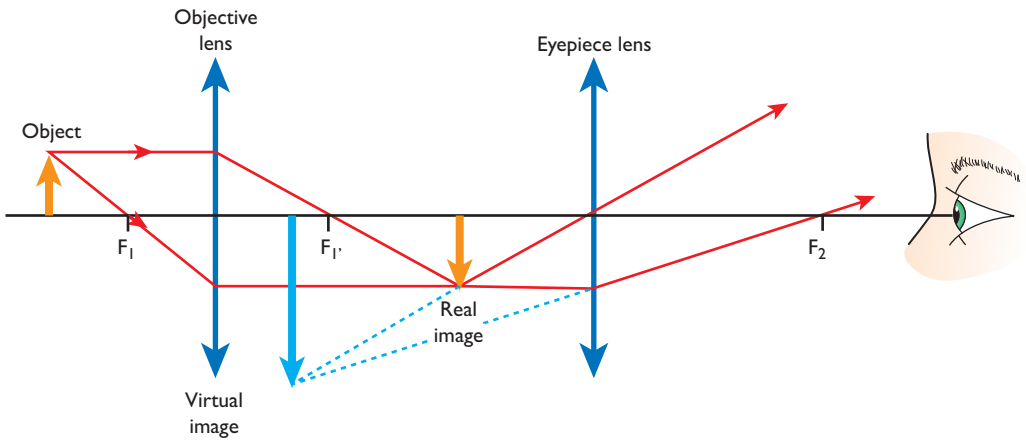


Fig. 10.6 The compound microscope.

With permission from Neil Modi.

There is a long working distance between the microscope and the patient's eye. This allows manoeuvres to be carried out such as applanation tonometry or a condensing lens (e.g. +90D or +78D). Fundus and drainage angle examination can be conducted using various lenses, for example:

- Hruby lens: this is a powerful plano-convex lens (–58.6D) that is held directly in front of the eye (Fig. 10.7). This allows an image to be formed within the focal length of the slit lamp by focusing the light emerging from the retina that otherwise comes out as parallel rays because of the refractive properties of the cornea and lens. A virtual, erect, and diminished image of the illuminated retina is formed.
- 90D and 78D lenses: these use the same principle as the indirect ophthalmoscope. The image is inverted vertically and horizontally, and reduced in size, but this is compensated for by the magnification of the slit lamp (Fig. 10.8).
- Panfundoscope contact lens (Rodenstock): this is a contact lens with a high convex power. It forms a real but inverted image of the fundus, which is located within the lens. A spherical glass element in the lens redirects the image towards the examiner. This lens gives a wide field of view, which means that the entire fundus can be examined as far as the equator without moving the lens (Fig. 10.9).
- Gonioscopy/three-mirror contact lens: the property of total internal reflection must be overcome to look at the drainage angles of the eye and the peripheral retina (Fig. 10.10). A contact lens is used with a higher refractive index than the eye, and the space between the lens and the eye is filled with saline (or viscous gel). This arrangement destroys the cornea/air refracting surface and allows light to pass from the drainage angle and peripheral retina, through the cornea, and on to the examiner (Fig. 10.11).

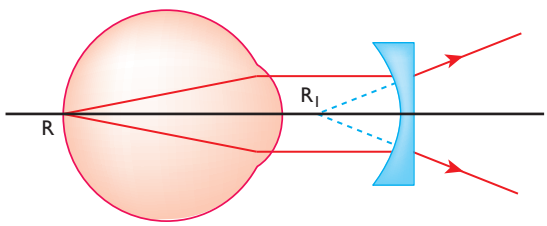


Fig. 10.7 A concave Hruby lens—the image lies in the focal range of the slit lamp.
With permission from Neil Modi.

Applanation tonometer

The applanation tonometer is used to measure IOP. The tonometer is applied to the cornea (which has been anaesthetized with topical anaesthetic). Because it has a standard area of contact of 3.06 mm in diameter, the force required is directly proportional to the IOP. This is because the effects of surface tension and rigidity of the cornea cancel each other out when the diameter is 3.06 mm.

With a larger contact area, corneal rigidity becomes more noticeable, whereas with smaller surface areas, the surface tension of the precorneal tear film causes errors. At a diameter of 3.06 mm, the surface tension and rigidity equal each other. With such a small diameter, the intraocular volume displaced is negligibly small and does not affect IOP.

To achieve a standard area of contact, the tonometer head consists of two alignment prisms with their bases in opposite directions. The examiner looks through the slit lamp to see the circle where the tonometer contacts the cornea, split into two hemicircles which are laterally displaced in opposite directions by the prisms. The applanation pressure is adjusted until the inner edges of the two hemicircles just touch. This means that the area of contact is 3.06 mm in diameter. When this happens, the intraocular pressure reading is proportional to the force applied and it can be read off a scale on the side of the tonometer.

High corneal astigmatism can lead to inaccuracy of 1 mm Hg per 4D of corneal astigmatism. The correct area of contact can be achieved if the measurement is made at 43° to the meridian of the lower corneal power. Clinically, the Goldmann tonometer has a white line in the horizontal meridian for relatively normal corneas without a great deal of astigmatism. There is also a red line at 43° to the horizontal. The

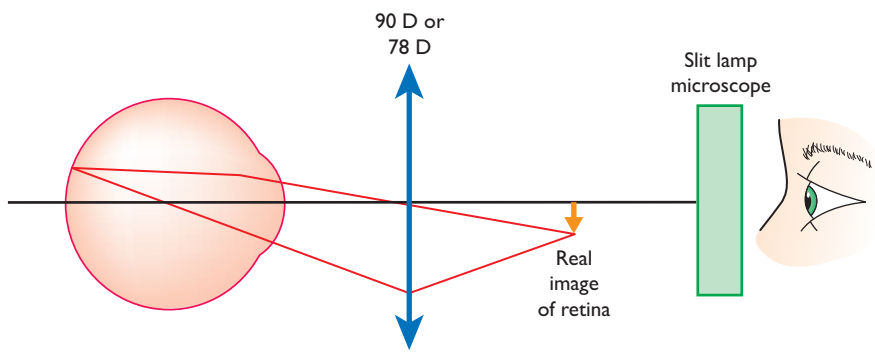


Fig. 10.8 Use of 90D (or other indirect lens) on slit lamp biomicroscope.
With permission from Neil Modi.

prism has markings on it for degrees. If the prism is mounted with the value of the minus (lower) cylinder at the red line, it will be correctly aligned to take the reading at 43° to the axis of the lower corneal power.

Pachymeter

Pachymetry is the measurement of corneal thickness. This can be either optical or ultrasound.

Optical pachymeters use Purkinje–Sanson images, which are formed by the anterior and posterior surfaces of the cornea. The pachymeter doubles the observer's image (Fig. 10.12).

The Jaeger pachymeter is commonly used. The image is horizontally split as it is seen through two glass plates. The lower one is split but the upper one may be tilted to displace the upper half of the image.

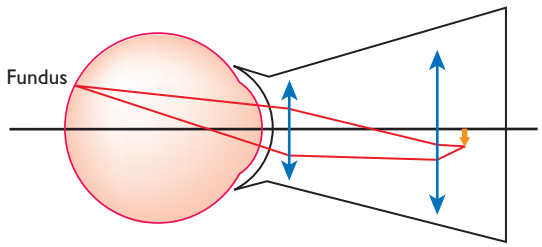


Fig. 10.9 Rodenstock panfundoscope contact lens.
With permission from Neil Modi.

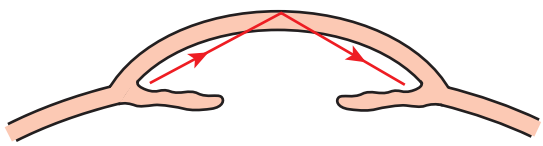


Fig. 10.10 Total internal reflection of light from the angle of the anterior chamber at the cornea.
With permission from Neil Modi.

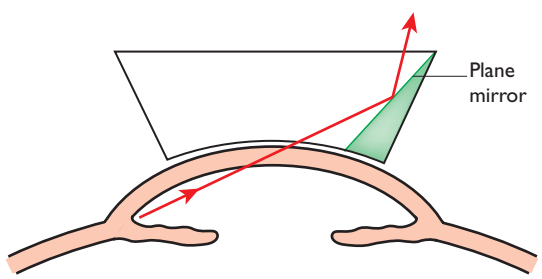


Fig. 10.11 Gonioscopy lens.
With permission from Neil Modi.

CLINICAL TIP

Ultrasound pachymetry is more commonly used in clinical practice as it is more accurate. This is invaluable for diagnosis of keratoconus, corneal and refractive surgery, and the diagnosis and treatment of glaucoma.

The observer aligns the anterior surface of the cornea in one half with the posterior surface in the other half. The degree of tilt of the glass plates required to do this is proportional to corneal thickness. A similar technique can be used to assess anterior chamber depth.

Investigation of corneal curvature

The anterior corneal surface is the main refracting surface of the eye and a small change here can cause a significant visual effect. There are various methods of examining the corneal surface curvature.

Placido's disc

This is a disc with concentric black and white rings. There is an aperture in the centre of the disc with a +2D sphere lens through which the examiner looks to see the reflection from the patient's cornea. The nature of the image shows any distortion in the cornea. The smaller the radius of curvature of the anterior corneal surface, the closer together the rings of Placido's disc will be as reflected by the cornea (much like on an OS map). When looking at an astigmatic cornea, the rings will therefore appear closer together in the steeper meridian.

Keratometer

Two types of keratometer are available: von Helmholtz and Javal–Schiotz keratometers. Both measure the radius of curvature at the central zone only, measuring around 3 mm. This optical power can be measured in millimetres (radius of curvature) or dioptres.

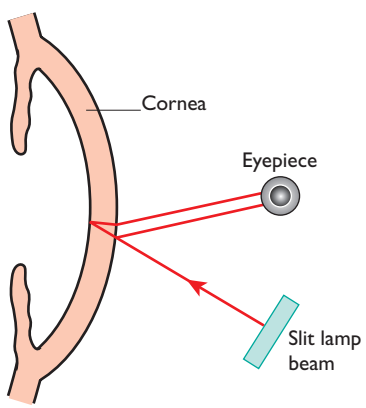


Fig. 10.12 An optical pachymeter.
With permission from Neil Modi.

Usually, the radius of curvature in the axial zone (central 4 mm) is around 7.8 mm or around 45D. The actual dioptric power depends on both the anterior and posterior corneal curvature and the refractive indices of the cornea and the media at both the anterior and posterior surface. The more peripheral cornea is flatter and non-spherical, and is not used for vision as it is obscured by the pupil.

The dioptric power (D) can be calculated by the following equation:

$$D = n_2 - n_1 / r = 1.3375 - 1 / r$$

where n_2 is the refractive index of tears, n_1 is the refractive index of air, and r is the radius of curvature.

The von Helmholtz keratometer

This has a fixed object size and the image size is adjusted. Light is shone onto the patient's retina and an image is formed. This image is reflected back though two glass plates (X and Y) which displace light as it passes through them and this causes the image to be doubled. As a result, two virtual images of I are formed, I' and I''. These images are viewed through a telescope. The distance between the two objects

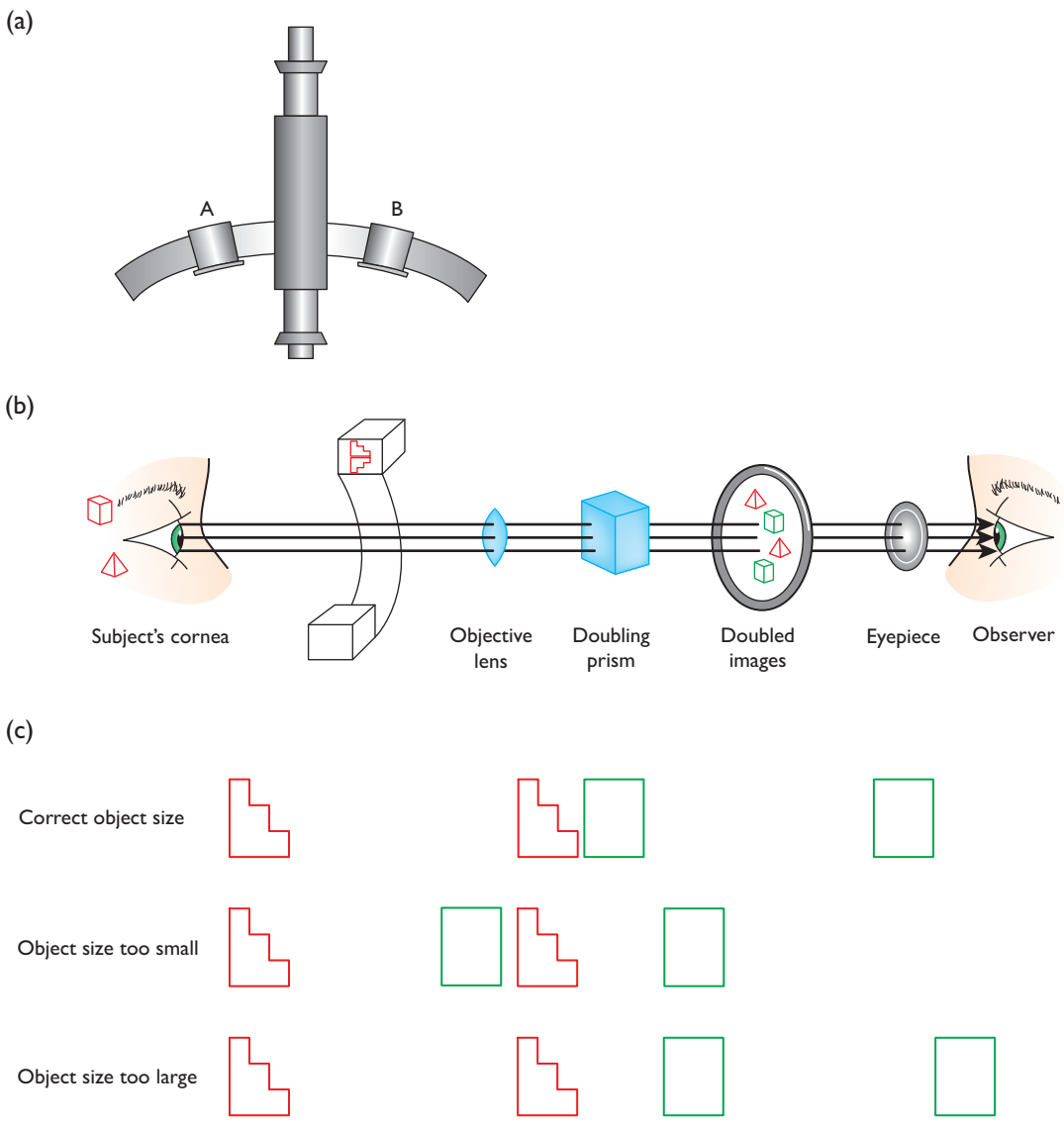


Fig. 10.13 The Javal–Schiotz keratometer (images of mires).
With permission from Neil Modi.

is adjusted by inclining the glass plates. When the edges of the two objects meet, the distance between their centres equals the diameter of the object, *l*. The corneal diameter can be calculated from this because the object is of a fixed size.

The Javal–Schiotz keratometer

This uses an object of variable size that consists of two mires, A and B, mounted on curved side-arms which protrude from either side of a viewing telescope (Fig. 10.13a). Each mire has a light source of different colours. One is step shaped and one is a rectangle. The space between the mires is used to measure the corneal curvature (Fig. 10.13b). The telescope and therefore the mires can be rotated in any direction so that readings can be made in any meridian. The images are doubled and the distance between the mires is adjusted until the doubled images touch.

CHAPTER 11

Basic biostatistical and epidemiological terms

This chapter provides a brief outline of basic biostatistical and epidemiological terms and it is important to practise calculations (sensitivity and specificity, negative and positive predictive values, odds ratio, and relative risk) to complement the learning from this section.

Basic descriptive statistics

Statistics involves the methods of collecting, summarizing, analysing, and drawing conclusions from data. This allows for data interpretation and extraction of the most useful information.

Data can be described according to the average and the spread around the average.

The average

There are various ways the average can be calculated (Table 11.1):

- The arithmetic mean: this is calculated by adding up all the values and dividing this sum by the number of values in the set.
- The median: data are arranged starting with the smallest value and ending with the largest value; the median is the middle value of the ordered data set. The median is similar to the mean if the data are symmetrical, less than the mean if the data are skewed to the right, and greater than the mean if the data are skewed to the left.
- The mode: this is the value that appears most often in the data set. The mode is rarely used as a measure.

The spread

The range is the difference between the largest and smallest observations in the data set, whereas spread describes a trend within the range. The range can be derived from percentiles. For example, the value within a data set where 1% of the observations are below it and 99% are above it is known as the 1st percentile. The 50th percentile would be the median.

Similarly, ranges could be derived by deciles and quartiles.

Percentiles can be used to measure spread by excluding those values within a data set that are at extreme ends of the data set (outliers). The interquartile range is the difference between the 25th and 75th percentiles, and contains 50% of the observations. Often the range is used that contains the central 95% of observations, excluding 2.5% above the upper limit and 2.5% below the lower limit. This interval is often referred to as the normal range.

Table 11.1 Advantages and disadvantage of averages

Types of average	Advantages	Disadvantages
Mean	• Uses all the data values • Algebraically defined • Mathematically manageable • Known sampling distribution	• Distorted by outliers • Distorted by skewed data
Median	• Not distorted by outliers • Not distorted by skewed data	• Ignores most of the information • Not algebraically defined • Complicated sampling distribution
Mode	• Easily determined for categorical data	• Ignores most of the information • Not algebraically defined • Unknown sampling distribution

Spread can also be calculated by measuring variation around the arithmetic mean. Units of variance are measured as the square of the unit.

The standard deviation is the square root of the variance. It is an average of the deviations from the mean and is measured in the same units as the data.

Variance and standard deviation change little as the sample size is increased.

The standard error (SE) is a measure of the uncertainty of a sample statistic, e.g. the mean. The standard error of the mean (SEM) is calculated by dividing the standard deviation by the square root of the number of observations and decreases as sample size increases.

The normal (Gaussian) distribution

The characteristics of normal distribution (Fig. 11.1) are:

- bell-shaped curve (unimodal)
- symmetrical about its mean
- shifted to the right if the mean is increased and to the left if the mean is decreased (constant variance)
- flattened if the variance is increased, more peaked if the variance is decreased (constant mean)
- equal mean and median.

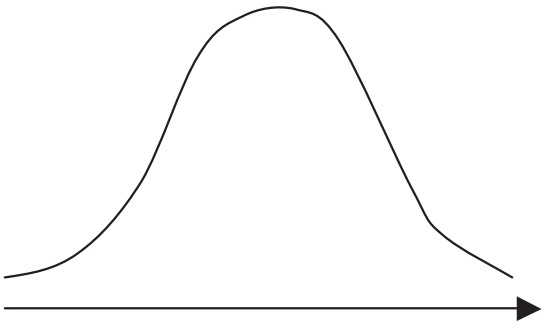


Fig. 11.1 An example of a normal distribution curve.

Other types of asymmetrical distribution

An asymmetrical distribution is positively (right) skewed if the longer tail is to the right and negatively (left) skewed if the tail is to the left. The mean will be shifted towards the longest tail. The mode will continue to be the highest peak on the curve and the median will fall between the mean and the mode. Examples of asymmetrical distributions include the *t*-distribution and chi-squared distribution. Both of these are characterized by degrees of freedom.

Clinical study design

Experimental studies involve the researcher assessing the effect or outcome following an intervention. An example of this may be a laboratory experiment or clinical trial to assess therapy. Such trials are normally prospective.

Observational studies are those in which the researcher does nothing to affect the outcome but just observes what happens. Examples include cohort and case-control studies. They can be prospective or retrospective.

Longitudinal studies follow a sample of patients/controls over time. This is normally prospective but sometimes can be retrospective.

Types of study include:

- **Cross-sectional:** cross-sectional studies are observational and are therefore carried out at a single time point. An example may be a survey or census. They can be used to measure the point prevalence of a condition:

  point prevalence = number with the disease at a given time point / total number studied at same time point

- **Cohort:** this takes a group of individuals who should be representative of the normal population and follows them in time. It is therefore observational. The aim usually is to study whether exposure to a particular risk factor will affect the incidence of a disease. The relative risk (RR) can be calculated and indicates the increased or decreased risk of disease associated with the risk factor. If RR = 1 then the risk is the same for the exposed group as the unexposed group.

Case-control studies

These are observational, retrospective studies. They compare a group of patients with a disease outcome (cases) with a group of normal individuals without a disease outcome (controls). This assesses whether exposure of any factor occurred more or less frequently in the cases than the controls.

HELPFUL HINT

Advantages and disadvantages of cohort studies

Advantages	Disadvantages
• Time sequence of events can be assessed	• Long follow-up, therefore expensive
• A wide range of disease outcomes can be assessed	• In rare diseases large sample sizes are needed
• Incidence/risk is measured	• Patients may leave the study due to migration or loss to follow-up
• A wide range of risk factors can be assessed	• Consistency can be reduced with the length of the study
• Changes in exposure of time can be measured	• Disease outcomes can change over time
• Reduce recall and selection bias versus a case-control study	

HELPFUL HINT

Advantages and disadvantages of case-control studies

Advantages	Disadvantages
• Quick to perform • Less costly • Good for rare disease • Wide range of risk factors assessed • Not lost to follow-up	• Recall bias—cases have varying ability to remember details about history • Selection bias • Not suitable when exposure to the risk factor is rare

The odds ratio can be calculated:

odds ratio = odds of being a case in exposed group / odds of being a case in unexposed group

Clinical trial

This is any planned experimental study involving evaluation of a new treatment on outcome in humans. Clinical trials are longitudinal and prospective.

Phase I/II trials are small clinical studies that evaluate the effect and safety. Phase III is a larger study to evaluate the effect of the new treatment.

Treatment is always compared to a control, which may be standard treatment or no treatment (placebo).

The primary endpoint is usually the treatment efficacy and the secondary endpoints may relate to side effects.

Patients are randomized to either the treatment group or the control group. The idea is to create similar groups and avoid allocation bias. These studies are randomized controlled trials (RCTs).

Randomization is completed by random allocation, which can be done by a computer-generated list of random numbers and can be fine tuned. Stratified randomization controls for factors such as age to provide equal spread between two groups. Restricted randomization ensures equal group numbers and cluster randomization allocates 'clusters' of people rather than individuals.

To avoid assessment bias patients and doctors can be blinded (double-blind trial). If only the patient is blinded it is a single-blinded trial.

Clinical trials have to pass through an ethics committee and informed patient consent is obtained from the patient or guardian.

The term 'intention to treat' (ITT) describes the aim to keep all patients on whom information is analysed in their original group, whether they followed the treatment regime or not. This should avoid bias.

Bias

Bias can occur at any stage of a research process. It describes a difference between the results from the study completed and the true outcome.

HELPFUL HINT

The main types of bias

- Funding bias: reporting findings in the direction favoured by the sponsor.
- Publication bias: reporting of only topical results or positive reports.
- Selection bias: patients in the study are not representative of the population that the results will be applied to.
- Information bias: during data collection measurements on exposure and/or disease outcome are incorrectly recorded, for example reporter bias occurs when patients give answers in the direction they feel the researcher is interested in and observer bias occurs when an observer under-reports or over-reports a variable.

Confounding factors

A confounding factor is an exposure factor that is related to the outcome and to one or more other exposure factors. This will result in a false association or missing a real association. Study design and analysis are important to reduce the effects of confounding factors.

Null hypothesis

Studies are usually designed to test against a null hypothesis which assumes no effect in the population. If the null hypothesis is not true (i.e. an effect is found) then the alternative hypothesis is found to be true.

Studies only take a sample of the population; it is therefore always possible that the rejection or acceptance of the null hypothesis was indeed wrong.

95% confidence interval

This is the range of values within which we are 95% confident that the true population parameter lies.

P value

The *P* value is the probability of obtaining our results if the null hypothesis is true. Normally the *P* value is set at 0.05; therefore if a value less than 0.05 occurs the null hypothesis can be rejected and the results are significant at the 5% level. If the *P* value is equal to or greater than 0.05; the null hypothesis is accepted, i.e. there is not enough evidence to reject the null hypothesis or the null hypothesis is true. The selected *P* value should be chosen before a study is commenced and is known as the significance level.

Type I and type II errors

Type I error occurs when the null hypothesis is rejected when it is true and it is concluded that there is an effect when actually there is none.

Type II error occurs when the null hypothesis is accepted when it is false and it is concluded that there is no evidence of an effect when one exists.

HELPFUL HINT

Factors that affect the power of a study

- Sample size: power increases with increasing sample size.
- Variability of the observations: power increases with decreasing variability.
- Effect of interest: power is greater with a larger effect.
- Significance level: power is greater if the significance level is larger.

Power

The power is the probability of rejecting the null hypothesis when it is false. It is usually expressed as a percentage and is the chance of detecting a statistically significant real treatment effect.

Sample size calculation

Sample sizes need to be big enough to detect an existing effect and small enough to allow for expense and time. To calculate the optimal sample size for a study, the power, significance level, variability (standard deviation), and smallest effect of interest must be defined first. Published studies or a pilot study can be used to help with this.

Statistical tests

Parametric versus non-parametric

Hypothesis tests based on normally distributed data are known as parametric tests. The *t*-test is a parametric test. If the data are not normally distributed or the sample is too small a non-parametric test should be used. In non-parametric tests the observations are ordered from low to high and a rank is assigned for each. Statistical tests are performed on the rank not the data values. Non-parametric tests include the Wilcoxon signed rank test (equivalent to a paired *t*-test) and the Mann-Whitney U-test (equivalent to an unpaired *t*-test).

HELPFUL HINT

Non-parametric tests

- Wilcoxon signed rank test: compares paired observations.
- Wilcoxon rank sum (two-sample) test: compares distributions of two independent groups of observations.
- Mann–Whitney U-test: similar to the Wilcoxon rank sum test. Produces the same *P* values.

HELPFUL HINT

Parametric tests

- Paired *t*-test: tests the null hypothesis that the mean of a set of differences of paired observations is equal to zero.
- Unpaired *t*-test: tests the null hypothesis that two means from independent groups are equal.
- Chi-squared (×2) test: tests the null hypothesis that there is no association between the factors that define a contingency table. Used on frequency data and to test differences in proportions. It uses actual values and not percentages. A value of zero means there is no difference between observed and expected frequencies. The larger the value the greater the differences and the less likely the null hypothesis is true. Fisher's exact test is used to calculate *P* values.
- Wald test: tests the significance of a parameter in a regression model. Its square follows the chi-squared distribution.
- McNemar's test: compares proportions in two related groups using a chi-squared test statistic.

HELPFUL HINT

Key statistical terms

Sensitivity is the proportion of individuals with the disease who are correctly diagnosed by the test.

Specificity is the proportion of individuals without the disease who are correctly identified by a diagnostic test.

False positive is an individual who is normal but who is incorrectly diagnosed as having the disease.

False negative is an individual who has the disease but is incorrectly diagnosed as normal.

Positive predictive value is the proportion of individuals with a positive diagnostic test result who have the disease.

Negative predictive value is the proportion of individuals with a negative test result who do not have the disease.

Likelihood ratio is the ratio of the chances of getting a particular test result in those who have or do not have the disease.

Table 11.2 Methods for selecting an appropriate statistical test

Categorical	> 2 categories	Chi-squared test
	2 categories	1 group = z test
		2 group = paired = McNemar's unpaired = chi square
Numerical	1 group = t-test	
	2 groups = paired or unpaired t-test or Wilcoxon rank sum test	
	> 2 groups = ANOVA	

ANOVA: analysis of variance. With permission from Louise Bye.

Choice of test

The choice of test depends on study design, sample size, the variable tested, and the distribution of the data. Table 11.2 summarizes a possible method for selecting a statistical test.

Commonly used statistical terms

As emphasized, it is important to practise calculations of the terms below to complement understanding of these areas.

Correlation and regression

Correlation measures the degree of association between two variables. A linear relationship between *x* and *y* occurs when a straight line can be drawn through the points. The Pearson correlation coefficient can be calculated to describe how close the observations are to a linear relationship. Spearman's rank correlation coefficient is a non-parametric equivalent to Pearson's correlation coefficient and can be used for the same reasons as other non-parametric tests as well as when a measure of association between two variables is required when the relationship is definitely non-linear.

Regression is used if it is thought that *y* is dependent on *x* and a change in *y* is attributed to a change in *x*. This would be univariate regression. If there are two or more *x* values the regression is termed multivariate. To analyse multivariate regression the ANOVA (analysis of variance) can be used in which the response is compared between groups of individuals (two or more treatment groups).

Basic epidemiological terms

- Incidence: measures the number of new cases of disease per unit population per unit time.
- Prevalence: measures the total number of cases (old and new) at a particular time.
- Primary prevention: designed to prevent the occurrence of disease, for example health education.
- Secondary prevention: aims to detect and treat disease before symptoms develop, for example blood pressure control.
- Tertiary prevention: prevents the consequences of disease after it has developed, for example taking aspirin following a transient ischaemic attack.

Further reading

Petrie A and Sabin C (2009). *Medical Statistics at a Glance*. Third Edition. Wiley-Blackwell.