Medical Biochemistry at a Glance

BEN GREENSTEIN MRPharmS PhD
Director of Endocrine Research
Lupus Research Unit
Rayne Institute
St Thomas's Hospital
London SE1 7EH

Honorary Lecturer in Pharmacology
United Medical and Dental Schools
St Thomas's Campus, London

ADAM GREENSTEIN BSc (Hons)
Medical student
University of Manchester,
Manchester M13 9PL

Blackwell
Science

Blackwell Science Ltd
Editorial Offices:
Osney Mead, Oxford OX2 0EL
25 John Street, London WC1N 2BL
23 Ainslie Place, Edinburgh EH3 6AJ
238 Main Street, Cambridge
Massachusetts 02142, USA
54 University Street, Carlton
Victoria 3053, Australia

Other Editorial Offices:
Arnette Blackwell SA
1, rue de Lille, 75007 Paris
France

Blackwell Wissenschafts-Verlag GmbH
Kurfürstendamm 57
10707 Berlin, Germany

Zehetnergasse 6
A-1140 Wien
Austria

First published 1996

Set by Setrite Typesetters, Hong Kong
Printed and bound in Great Britain
at the University Press, Cambridge

DISTRIBUTORS

Marston Book Services Ltd
PO Box 87
Oxford OX2 0DT
(*Orders*: Tel: 01865 791155
Fax: 01865 791927
Telex: 837515)

North America
Blackwell Science, Inc.
238 Main Street
Cambridge, MA 02142
(*Orders*: Tel: 800 215-1000
617 876-7000
Fax: 617 492-5263)

Australia
Blackwell Science Pty Ltd
54 University Street
Carlton, Victoria 3053
(*Orders*: Tel: 03 9347 0300
Fax: 03 9349 3016)

A catalogue record for this title
is available from the British Library

ISBN 0-86542-980-4 (BSL)
ISBN 0-86542-626-0 (IE)

Library of Congress
Cataloging-in-Publication Data

Greenstein, Ben, 1941–
Medical biochemistry at a glance/
Ben Greenstein, Adam Greenstein.
p. cm.
Includes bibliographical references
and index.
ISBN 0-86542-980-4
1 Biochemistry—Outlines, syllabi, etc.
2 Clinical biochemistry—Outlines, syllabi, etc.
I Greenstein, Adam. II Title.
[DNLM: 1 Biochemistry. QU 4 G815m 1996]
QP514.2.G76 1996
574. 19′2—dc20
DNLM/DLC
for Library of Congress
95-36168
CIP

Contents

Introduction 1

1 The eukaryotic cell 2

2 Membranes I 4

3 Membranes II 6

4 Membranes III 8

5 Intracellular receptors and receptor antagonists 10

6 Molecules I 12

7 Molecules II 14

8 DNA replication 16

9 DNA repair 18

10 Recombination 20

11 RNA 22

12 Transcription I 24

13 Transcription II 26

14 Errors of transcription 28

15 Protein synthesis I 30

16 Protein synthesis II 32

17 Protein synthesis III 34

18 Errors of translation 38

19 Collagen 40

20 Control of gene expression in prokaryotes 42

21 Control of gene expression in eukaryotes 44

22 Mechanisms of transcriptional control 46

23 Growth 48

24 Cancer 50

25 Genetic manipulation I 52

26 Genetic manipulation II 54

27 pH and buffers I 56

28 pH and buffers II 58

29 Chemical reactions I 60

30 Chemical reactions II 62

31 Enzymes I 64

32 Enzymes II 66

33 Enzymes III 68

34 Enzymes IV 70

35 Digestion: basic principles and cell types 72

36 Digestion of protein and carbohydrates 74

37 Digestion and absorption of lipids 76

38 The production of energy in the electron transfer chain 78

39 Glycolysis and gluconeogenesis 80

40 The citric acid cycle and mitochondrial carriers 82

41 Glycogen metabolism 84

42 Lipid metabolism I 86

43 Lipid metabolism II 88

44 Nucleotide synthesis 90

45 Breakdown of nucleotides 92

46 Catabolism of amino acids 94

47 Synthesis of amino acids 98

48 Integration of metabolism 100

49 Gas transport 102

50 Haemoglobin 104

51 Molecular chaperones 106

Abbreviations 108

Glossary 110

Sources 113

Index 115

Introduction

Medical Biochemistry at a Glance was envisaged, designed and written as a learning aid for undergraduate students. The book consists of a series of progressively organized, self-contained two-page spreads, in keeping with the general formula of the 'At a Glance' series. The emphasis has been on the diagrams, with supplementary text, and an attempt has been made to keep the diagrams as simple as possible, given the immense complexity of the subject. We hope the reader will find the diagrams rapidly and comfortably accessible, and that they will give the information, literally, at a glance.

Although the book has been called *Medical Biochemistry at a Glance*, it has been written not only for medical students, and we hope that students of Biology, including nurses, will find the book useful. We have related biochemical function to disease, but the emphasis has been on the basic information and mechanisms. There are branches of Biochemistry which are not covered here. Plant Biochemistry, for example, is a very important and rapidly growing subject, as is Clinical Biochemistry, which deals in great detail with the biochemistry underlying disease, and with the measurement of clinical biochemical parameters in health and disease. It is hoped nevertheless that students wishing to specialize in those areas will find this book most useful as an introduction to Biochemistry.

Biochemistry is a vast subject whose knowledge base expands daily at an almost unbelievable pace, and this book, therefore, modestly hopes to cover only the basic information which underpins the progress that is made, and which will be required to ensure a good examination result for the student. The book is not to be thought of as an alternative to the substantial volumes of Biochemistry which are available, and which deal with the subject in far more detail.

We have made every attempt to guarantee the clarity, accuracy and reliability of the information, and have had the contents read by a number of medical students, whose comments and suggestions greatly improved the clarity of the diagrams. In addition, the book has been read by several experts, who have contributed immeasurably to the confidence with which we have produced the book. In particular, we wish to thank Drs Kathleen Rowsell, Richard Gregory, Graham Wallace and Ms Carolyn Watson, who read the entire book and made invaluable comments and suggestions. We are grateful, also, to Ben Bromilow and to Drs Gavin Brooks, Brian Ellis and Hywel Thomas for reading several of the chapters. It goes without saying, however, that any remaining errors (none we hope) are the responsibility of the authors, and if the reader spots any we should like to know.

It remains only to thank the editorial staff of Blackwell Science, and particularly Stuart Taylor and Jonathan Rowley, for their patience and guidance in bringing this project to a fruitful conclusion.

Ben Greenstein
London, 1995
Adam Greenstein
Manchester, 1995

1 The eukaryotic cell

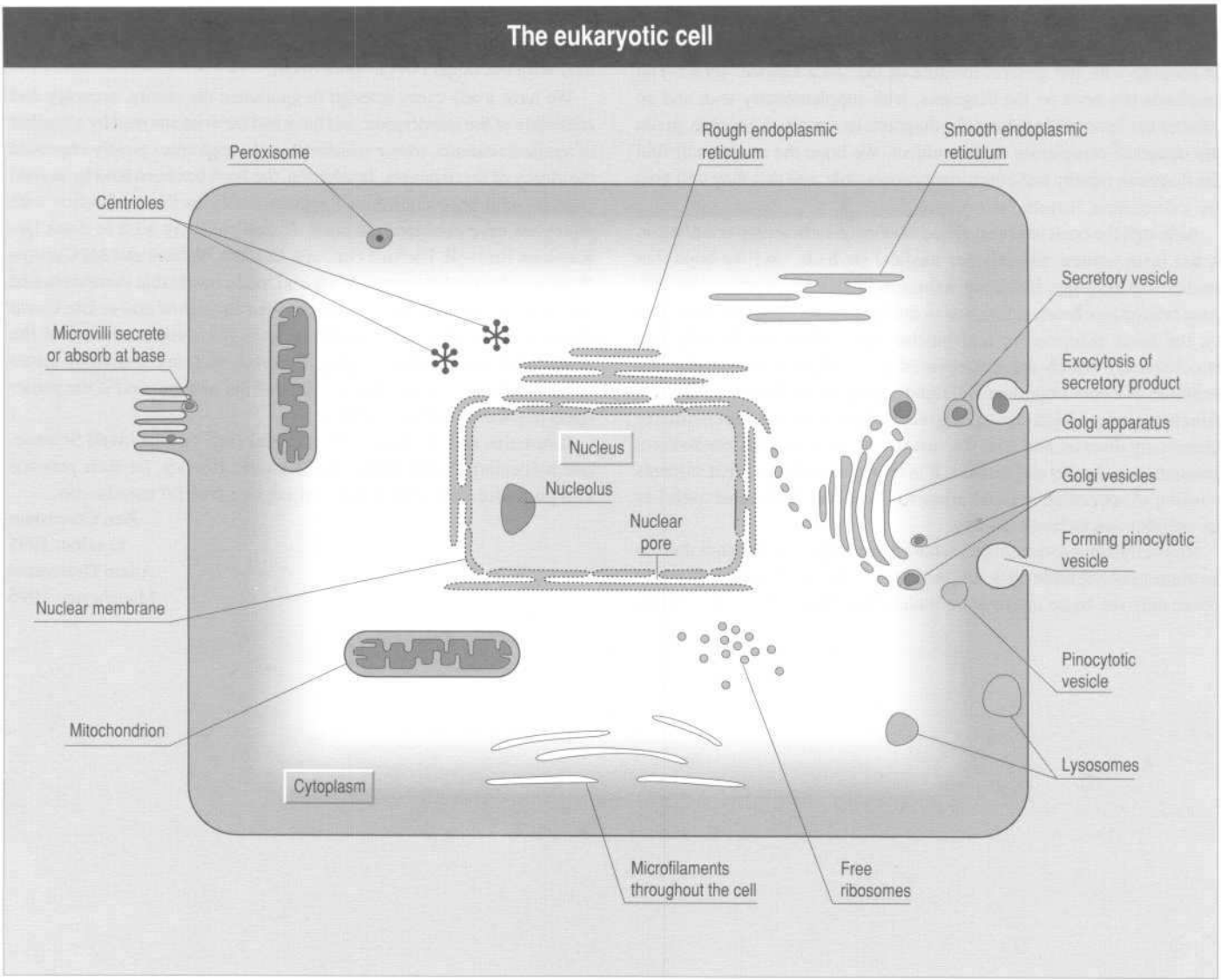

Fig. 1.1

INTRODUCTION

Eukaryotes have a membrane-bound **cell nucleus** with **chromosomes**, which are organized aggregations of genes. Algae, protozoa, fungi, slime moulds, plants and animals are eukaryotes. Prokaryotes are unicellular organisms, including bacteria, whose cells lack a nucleus and several other eukaryotic organelles.

Cell size

In animals, cells range in diameter from about 10 to 30 μm. Cells are usually microscopic to allow **diffusion**, a process whereby solutes distribute themselves in the available volume, and the rate of diffusion limits cell size. In most cells, no metabolically active region is more than 10–25 μm away from the cell surface. It would take days for amino acids, peptides and sugars to diffuse a couple of centimetres, but seconds to travel a few micrometres. Multicellular organisms have overcome this problem by developing a circulation.

INTERNAL ARCHITECTURE OF CELLS

The **cell nucleus** is bounded by two membranes: the inner, which defines it, and the outer, which is usually continuous with the cytoplasmic endoplasmic reticulum (ER). The nuclear membrane system is also called the **perinuclear envelope**. The space between inner and outer membranes is continuous with the lumen of the ER.

At points on the nuclear surface, the inner and outer membranes appear to fuse, creating **nuclear pores**, which may conduct materials between nucleus and cytoplasm. Most of the cell's deoxyribonucleic acid (DNA) occurs in the nucleus, as a DNA–protein complex called **chromatin**, which is organized into discrete, elongated bodies about 25 nm thick, called chromosomes. The DNA contains the cell's genetic information.

Inside the nucleus is the **nucleolus**, rich in ribonucleic acid (RNA). Within the nucleolus are one or more chromosomes, termed the **nucleolar organizer**, where ribosomal RNA (rRNA) is formed.

The non-nucleolar area of the nucleus is the nucleoplasm, within which occurs a small family of fibrous proteins termed **lamins**, which appear to bind DNA to the nuclear membrane.

Internal membrane systems of the cytoplasm

The ER is often the largest internal membrane system of the eukaryotic cell. Rough ER is studded with ribosomes, while smooth ER is not.

Smooth ER is the site of synthesis and metabolism of phospholipids and fatty acids. Smooth ER contains several enzymes that detoxify carcinogens or pesticides by rendering them water soluble and therefore more easily excreted. This may explain why some cells, such as liver hepatocytes, have more smooth ER than other cell types.

Rough ER is present in high abundance in cells that produce peptide hormones, e.g. insulin, and proteins, e.g. antibodies. Ribosomes bound to rough ER produce proteins, forming part of cell and organelle membranes. The rRNA–messenger RNA (mRNA) complex is attached to the ER, and usually passes the elongating peptide through a pore into the central lumen of the ER, where the chain aggregates prior to transportation elsewhere in the cell, or into the extracellular space.

The **Golgi apparatus** is a system of flattened vesicles and smooth membranes which transfers lipid precursors and carbohydrate to proteins to form lipoproteins and glycoproteins, respectively. The latter process, called glycosylation, is necessary for protein transport across the plasma membrane. The Golgi also produces much new cellular membrane in the form of vesicles, in which hormones, prohormones and some enzymes are packaged and exported from the cell. They also produce membrane for organelles such as peroxisomes and lysosomes.

Lysosomes are organelles with a single membrane, enclosing acid hydrolase enzymes in an acidic (pH 5) environment. The hydrolases degrade polymers such as DNA, RNA and protein into their monomeric units. The lysosomal membrane is impermeable to both large and small molecules, which are transported through it by specific mediators. Genetic lack of a lysosomal enzyme, β-*N*-hexosaminidase, which is important in the turnover of a membrane protein, G_{M2}, results in **Tay–Sachs disease** in which that protein accumulates in developing nerve cells and results in death by age about 5 years.

Peroxisomes are small organelles containing enzymes that use oxygen (O_2) to oxidize uric acid, D-amino acids and some 2-hydroxyacids, with the production of hydrogen peroxide (H_2O_2). H_2O_2 is converted in the peroxisome to water (H_2O) and O_2 by catalase. The peroxisome thus protects the cell from H_2O_2, a powerful oxidant. Peroxisomes also contain enzymes important in lipid metabolism. Peroxisomal enzymes vary from cell to cell, and with changes in cellular conditions. Absence of peroxisomes in brain, kidney, liver and skeletal muscle results in a rare autosomal recessive disease, **Zellweger syndrome**, which causes death within 6 months of birth.

Mitochondria are the energy powerhouses of the cell. They are large, being about 7 μm long and about 0.5–1.0 μm diameter, and may occupy up to 25% of the cytoplasm. They possess both inner and outer membranes. The outer membrane allows the passage of large molecules, up to 10 kDa, while the inner membrane is less permeable. The inner membrane has many infoldings or **cristae**, which protrude into the inner space or **matrix**. Respiratory enzymes protrude from the inner membrane into the matrix, as well as those which catalyse the production of adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and inorganic phosphate (Pi). The matrix is filled with the enzymes converting acetyl-coenzyme A (CoA) to carbon dioxide (CO_2). Many substances, such as ATP, ADP, citrate and phosphate, which need to move into and out of the mitochondria, cannot pass passively through the membrane and are transported by permease proteins which form channels for them.

In the matrix are several copies of a small DNA molecule that codes for several key mitochondrial membrane proteins. Most mitochondrial proteins, however, are produced by cytoplasmic ribosomes using mRNA originating in the cell nucleus.

The **cytoskeleton** is a network of filaments and microtubules, which maintains cellular morphology, transport, mitosis, meiosis and cell motility. The microtubules are composed of polymers of the protein tubulin. At least three mechanochemical proteins, kinesin, dynesin and myosin, which convert chemical into mechanical energy, occur in the cytoskeleton.

The **cytosol**, where many reactions occur, contains soluble constituents. (Note: the **ribosome** is dealt with more fully on pp. 32–33.)

2 Membranes I

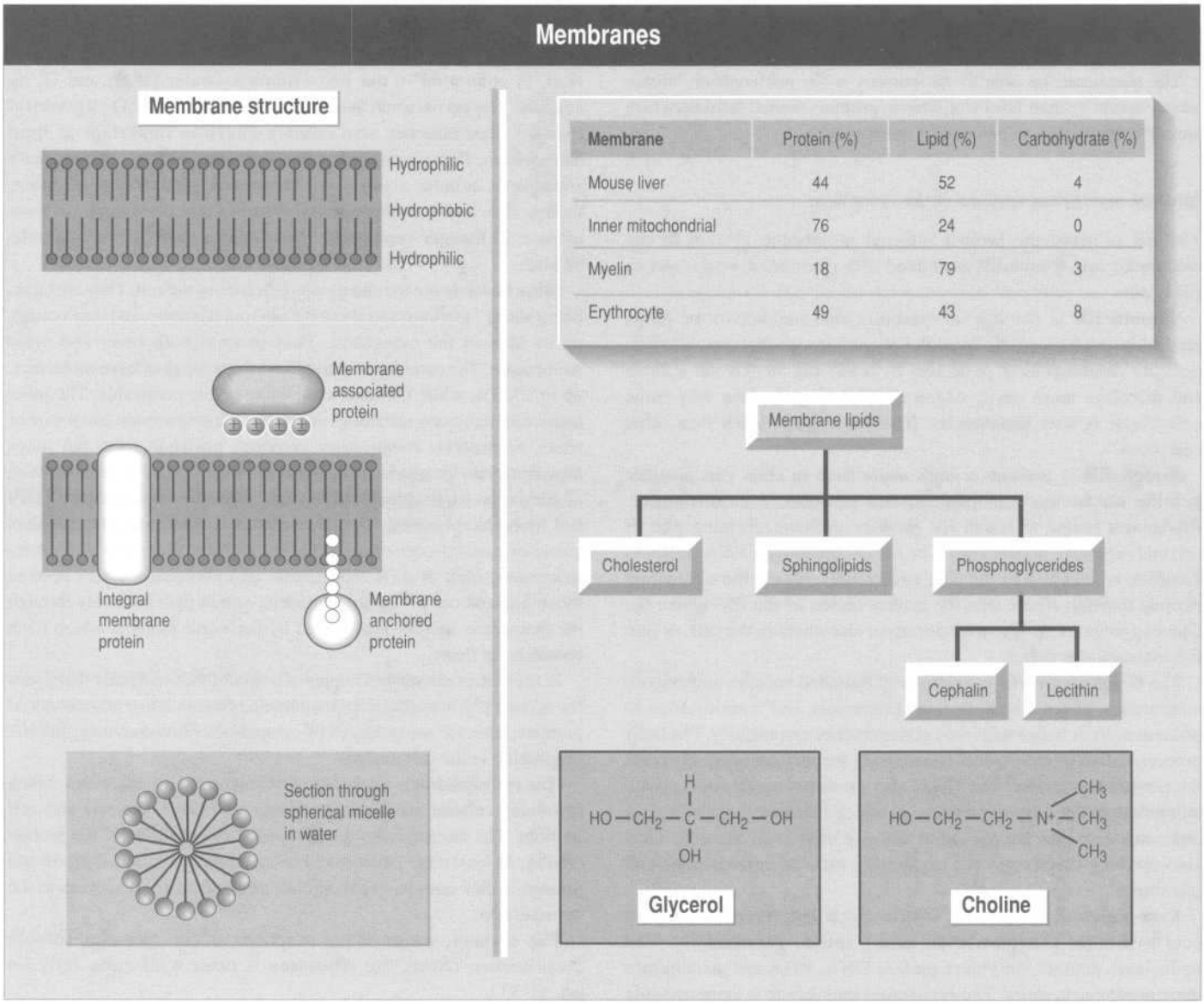

Fig. 2.1

LIPID MEMBRANES

Functions

Membranes: (i) define the shape of an organelle or cell; (ii) control the exchange of solutes, e.g. sodium (Na^+), potassium (K^+) and chloride (Cl^-) ions between interior and exterior; (iii) form the site of chemical reactions, e.g. oxidative phosphorylation on mitochondrial membranes; (iv) are a site for recognition of chemical messengers such as hormones and neurotransmitters, whose receptors may be situated on the membrane; (v) have a role in cell–cell recognition; and (vi) facilitate cellular locomotion.

Architecture

All biological membranes range from about 5 to 10 nm in thickness, and contain protein and lipid, the ratio of the two varying with the membrane source. Mammalian membranes are especially abundant in phospholipids and cholesterol. The phospholipid bilayer is the common structural unit. Phospholipids are **amphipathic**, i.e. possessing both hydrophilic and hydrophobic portions in the same molecule. Hydrophobic interactions between the fatty acyl chains of the lipid molecules produce a phospholipid bilayer, a sheet or leaflet possessing two layers of phospholipids whose polar heads face the H_2O, while the fatty acyl chains form the hydrophobic interior. When phospholipids

are shaken up with H_2O, they form spherical **micelles**, whose fatty acyl chains point away from the H_2O surface.

The lipid bilayer is coated on both sides with proteins, and according to the fluid mosaic model, the lipids themselves and some proteins move around in the plane of the bilayer.

The membrane proteins serve several functions. They may: (i) transport molecules through the membrane; (ii) act as receptors for chemical messengers such as hormones; (iii) make possible cell–cell interactions through their branching carbohydrate chains, which also make possible the recognition of antigens; and (iv) act as enzymes.

Proteins may be **integral**, being firmly bound to the membrane, or **associated**, being loosely or reversibly held to the membrane and able to be released by mild treatments. Integral proteins may be anchored, i.e. bonded covalently to the membrane by a link between the carboxy terminus of the protein and a membrane glycophospholipid (see below). Many integral proteins are insoluble in H_2O, and are embedded in the membrane and held to it by three main forces: (i) ionic interactions with the polar heads; (ii) hydrophobic interactions with the lipid interior; and (iii) specific interactions with cholesterol or other membrane molecules. Most integral proteins span the lipid bilayer and have polar regions at both ends of the protein.

MEMBRANE CHEMISTRY

Membranes are made up of protein, lipids and varying amounts of glycolipid and glycoprotein ('glyco-' implies a sugar moiety).

The three main lipids in eukaryotic membranes are **cholesterol**, **sphingolipids** and **phosphoglycerides**. Phosphoglycerides are the major membrane lipid components, and the two most abundant phosphoglycerides are **lecithin** (also called phosphatidylcholine) and **cephalin** (also called phosphatidylethanolamine). Sphingolipids are amphipathic molecules consisting of long-chain fatty acids with an amide linkage, which provides the polar head. Such a compound is termed a **ceramide**. **Glycosphingolipids** have a sugar moiety, either glucose or galactose, attached to the ceramide. Cerebrosides are examples of glycosphingolipids. **Galactocerebrosides** are cerebrosides found mainly in the central nervous system.

Cholesterol is a rigid molecule that intercalates among the phospholipids in the membrane. Its four-membered hydrophobic steroid ring interacts with the fatty acyl chains of membrane phospholipids. At 37°C, in eukaryotic cells, cholesterol limits the fluidity of the membrane. But, it also prevents the membrane from becoming less fluid at lower temperatures by preventing the chains from binding to each other. Membrane fluidity depends not only on the cholesterol content, but also on temperature and the lipid composition. Fluidity is promoted by shorter, unsaturated fatty acids. There is evidence that fluidity in membranes of certain cells may be influenced by diet.

THE ERYTHROCYTE MEMBRANE

The erythrocyte plasma membrane is relatively easily separated from other constituents. The lipid components are asymmetrically distributed across the membrane, in contrast to their symmetrical distribution in micelles. For example, cephalin occurs predominantly on the inner lipid layer. This asymmetry may be maintained by the transverse movement of phospholipids across the membrane, assisted by membrane proteins using metabolic energy to do so. An uncatalysed 'flip-flop' movement of sphingolipids and phosphoglycerides across the membrane is slow due to the tendency for the polar heads not to enter the hydrophobic bilayer, and may take days or weeks.

The erythrocyte membrane contains an integral glycoprotein, glycophorin, which contains 131 amino acids and spans the membrane, and another called **band 3**, because of its mobility on a polyacrylamide gel after electrophoresis. Band 3, a 900-amino acid protein, may have a role in the facilitated diffusion of hydrogen carbonate (HCO_3^-) and Cl^- across the membrane. It binds the cytosolic peripheral protein **ankyrin**, which in turn binds the protein **spectrin**. Spectrin and ankyrin are members of the erythrocyte cytoskeleton.

3 Membranes II

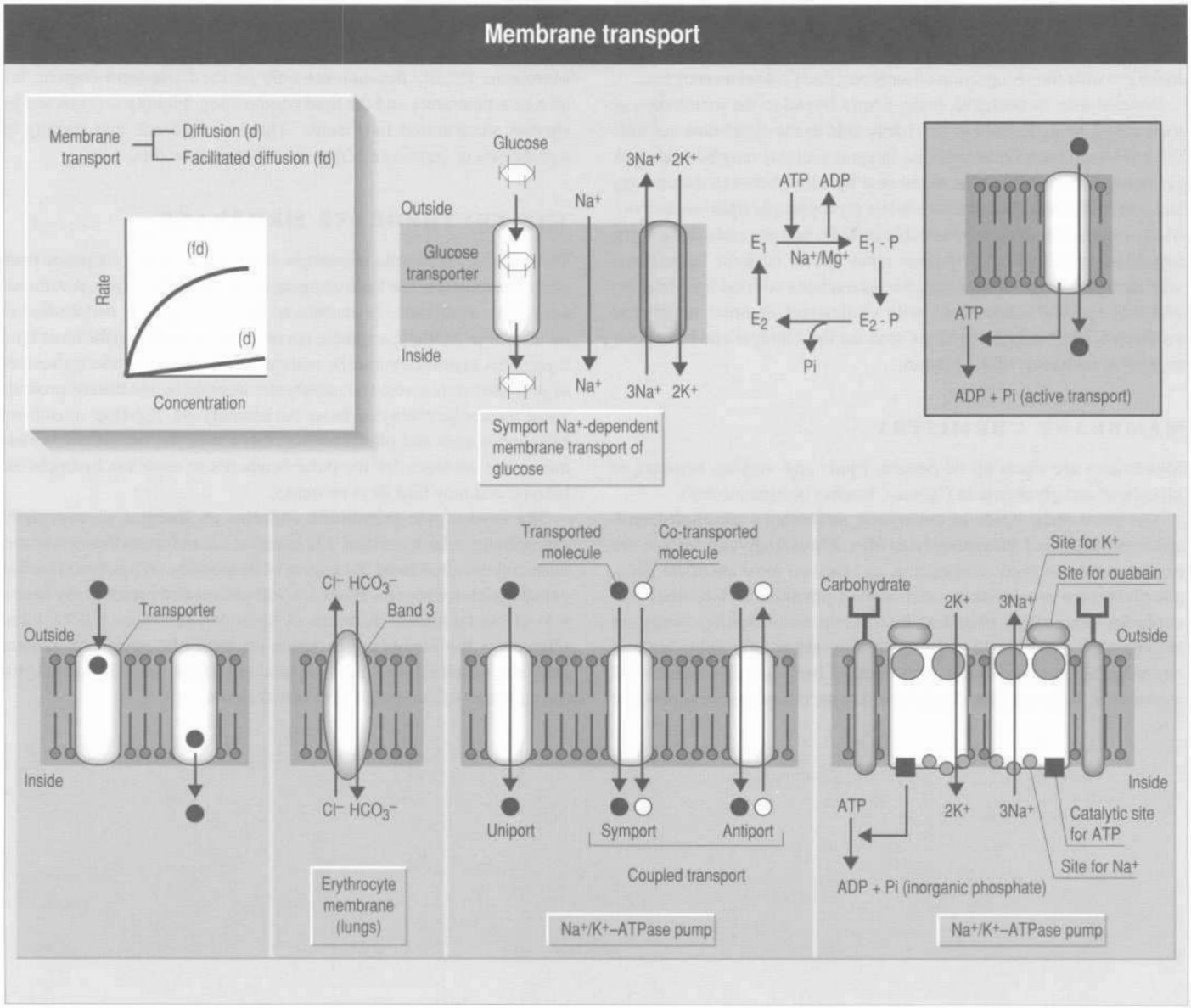

Fig. 3.1

MEMBRANE TRANSPORT

The membrane is selectively permeable to solutes in order to: (i) retain a barrier to the extracellular environment; (ii) ensure that essential molecules such as lipids, glucose and amino acids enter the cell, that these stay in the cell and that waste products leave the cell; and (iii) maintain ionic gradients across the membrane. Intracellular organelles may also have selectively permeable membranes. For example, the lysosome membrane maintains a concentration of hydrogen ions (H^+) 1000 to 10000 times greater than in the cytosol. Transport across the membrane may be **passive**, **facilitated** or **active**.

Passive transport is the movement of a molecule or ion down a concentration or electrochemical gradient. It may be by simple diffusion, which is how gases such as O_2 and CO_2 and how a simple molecule like ethanol cross the plasma membrane. In simple diffusion, a small molecule dissolved in the extracellular fluid dissolves in the membrane and then in the intracellular fluid. The process is non-specific, and the rate-limiting factor for entry of the molecule into the membrane is its hydrophobicity, i.e. its solubility in oil. The rate of diffusion through the phospholipid bilayer is proportional to the hydrophobicity. It is also proportional to the concentration gradient across the membrane.

Facilitated diffusion is the rapid movement of molecules across the membrane, with the help of specific membrane proteins called **permeases**. The process is specific, is faster than would be expected from diffusion alone and there is a maximum rate of transport.

Active transport is the movement of ions or molecules across the membrane against a concentration gradient, which utilizes energy in the form of ATP hydrolysis to do so. There are three main types of active ion transport: (i) the Na^+/K^+–adenosine triphosphatase (ATPase) pump, which transports Na^+ out and K^+ in; (ii) the calcium ion (Ca^{2+}) pump (also called the Ca^{2+}–ATPase pump), which drives Ca^{2+} out of the cell or from the cytosol into the sarcoplasmic reticulum; and (iii) the proton (H^+) pump. Ion gradients generated by active transport can be coupled to the active transport of molecules such as certain amino acids and sugars ('secondary' active transport).

Cotransport is the transport of an ion or molecule coupled to a cotransported ion. **Symport** is the simultaneous movement of both in the same direction, while **antiport** is simultaneous movement in opposite directions. If transport is not coupled to a cotransported ion, the process is termed **uniport**. Cotransport may occur during facilitated diffusion or during active transport. Glucose can be transported by a symport-facilitated diffusion, while Cl^- and HCO_3^- are transported across the erythrocyte membrane by an antiport-facilitated diffusion pump called band 3, which pumps Cl^- and HCO_3^- in opposite directions, the direction depending on the prevailing concentration gradient.

Active transport requires energy generated by ATP hydrolysis to ADP, coupled to the pumping of ions against their concentration gradient: ATP → ADP + Pi. Like facilitated diffusion, the transport is specific, has saturation kinetics and can be inhibited. An example is the primary active transport **Na^+/K^+–ATPase pump**. This is an antiporter enzyme system which requires Na^+, K^+ and magnesium ions (Mg^{2+}), and is present in virtually all animal cells, with especially high concentrations in excitable tissues such as nerve and muscle, and in cells actively involved in Na^+ movement across the plasma membrane, for example the kidney cortex and salivary glands.

The ATPase enzyme is an oligomer, composed of two α-subunits of 110 kDa, and two glycoprotein β-subunits of 55 kDa each. During ATP hydrolysis, the α-subunit is phosphorylated and dephosphorylated on a specific aspartate residue to form a β-aspartamyl phosphate. Phosphorylation requires Na^+ and Mg^{2+}, but not K^+, while dephosphorylation requires K^+ but not Na^+ or Mg^{2+}. The protein complex has been described in two conformations related to energy level, and the ATPase is thus referred to as an **E_1–E_2-type transporter**. The ATPase pump is inhibited by the cardiotonic glycosides, including **digoxin** and **ouabain**. Ouabain, due to its high H_2O solubility, has been extensively used to characterize the pump.

GLUCOSE TRANSPORT

Glucose transport provides an example of both facilitated diffusion and active transport, the former utilizing a uniport mechanism and the latter a symport mechanism. Glucose can be transported into erythrocytes by facilitated diffusion. The Michaelis constant (K_m) for glucose uptake into erythrocytes is about 1.5 mmol/l (i.e. at this concentration of glucose, about 50% of the available permease molecules will be bound to glucose molecules). Since the concentration of glucose in human blood is around 4–6 mmol/l, glucose uptake into erythrocytes will occur at virtually maximum rates. The permease is specific, since the L-isomer of glucose is not significantly transported into the erythrocyte. D-galactose and D-mannose are transported, but higher concentrations are required to half saturate the transport system. Once inside the cell, glucose is phosphorylated and can no longer leave the cell. The permease for glucose is also termed a D-hexose permease. It is an integral membrane protein of molecular weight 45 kDa.

Glucose can also be transported by a Na^+-dependent symport system found in plasma membranes of tissues, including the kidney tubule and intestinal epithelium. One molecule of glucose is moved against its concentration gradient, and one ion of Na^+ is moved down its concentration gradient by facilitated diffusion. But, the system is ultimately driven by the Na^+/K^+–ATPase pump. The symport is therefore a secondary active transport system. Amino acids are similarly transported.

THE CA^{2+} PUMP

The Ca^{2+} pump is an E_1–E_2-type active transport pump. It is an integral membrane protein, phosphorylated on an aspartamyl residue during Ca^{2+} transport. Two Ca^{2+} ions are transported for each molecule of ATP hydrolysed. In eukaryotic cells, Ca^{2+} binds to a calcium-binding protein called **calmodulin**, and the complex binds to the Ca^{2+} pump. Other Ca^{2+}-binding proteins include troponin C and parvalbumin.

4 Membranes III

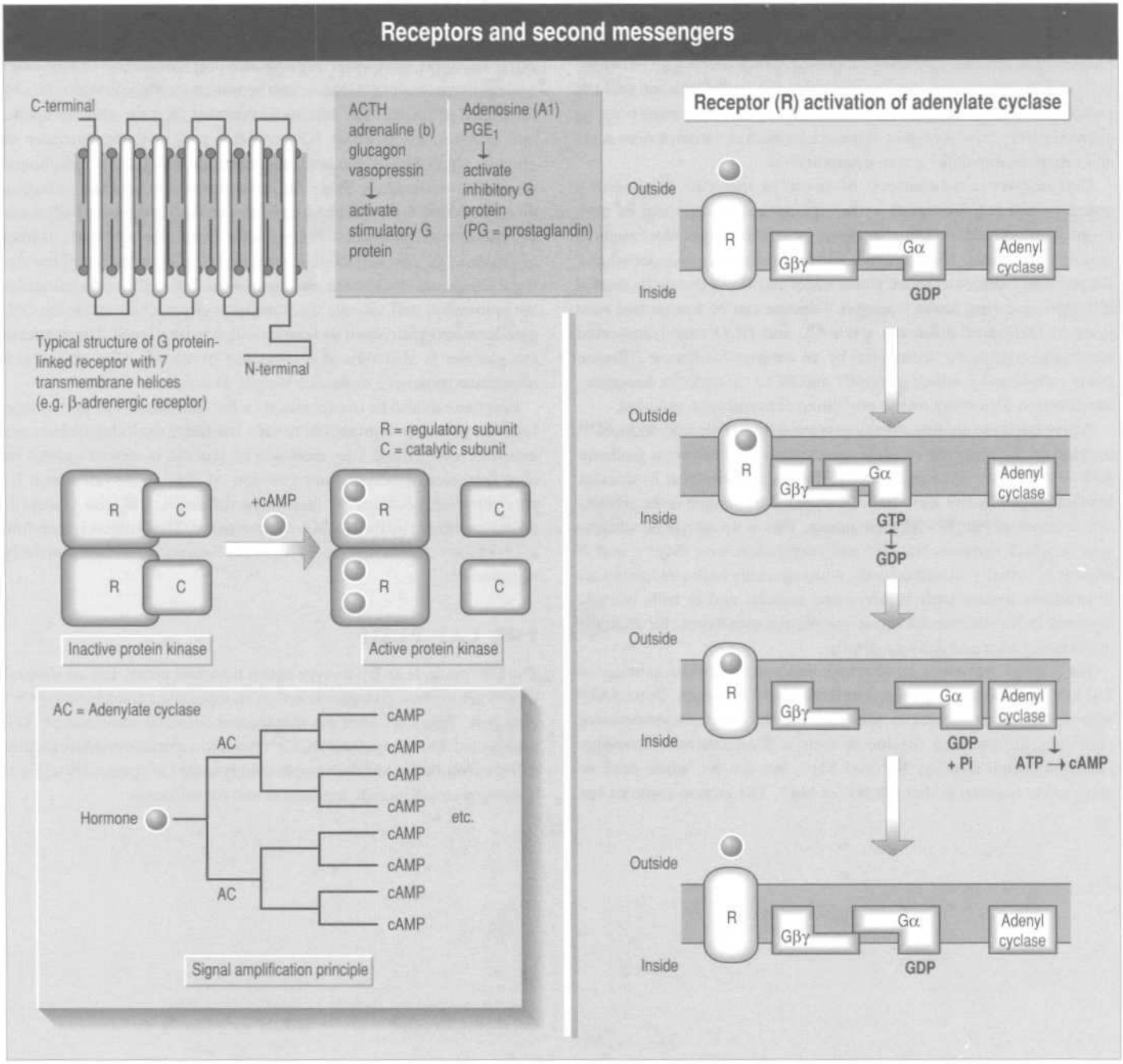

Fig. 4.1

CHEMICAL COMMUNICATION

Endocrine signalling occurs when cells or organs release chemicals or **hormones** which travel in the bloodstream to target cells or organs, which recognize them through specific **receptors**. Receptors are proteins that may be situated on the plasma membrane or inside the cell. The receptor has sites that recognize and bind the hormone, and the binding reaction produces a change in the receptor which allows it to pass on to the cell the message that the hormone has bound to it. Examples of hormones are adrenaline, insulin, the sex hormones and thyroid hormone. **Paracrine** signalling occurs when the secreting and target cells are adjacent or close to each other. **Autocrine** signalling

involves the release by a cell of a chemical that acts on the cell which released it. Growth factors are often paracrine or autocrine hormones. Sometimes, a chemical such as adrenaline, is both endocrine and paracrine.

Chemical signals or hormones acting on plasma membrane receptors are generally H_2O soluble (e.g. insulin, adrenaline) and produce relatively fast responses (seconds or minutes). Chemicals acting inside the cell are generally lipid soluble (e.g. cortisol, vitamin D), enabling them to pass easily through the plasma membrane. Their effects are slower in onset (hours or days) because their receptors alter gene expression and subsequent protein synthesis.

Receptors detect chemicals. They exhibit the properties of binding **selectivity**, high **affinity**, **reversibility** and **effector specificity** for the hormone that binds to them. An example of effector specificity is the action of adrenaline, which in liver cells causes glycogen breakdown and glucose release, while in neurones adrenaline may generate an electrical impulse.

Membrane receptors are integral proteins that span the membrane. On the extracellular surface is generally the N-terminal domain; inside the membrane is a helical hydrophobic ('H_2O-hating') domain; and the C-terminal domain extends into the cytosol. Membrane receptors may be linked to different signal transduction systems: (i) G proteins; (ii) ion channels; and (iii) enzymes.

A **G-protein-linked receptor**, adrenaline, binds to several different receptor subtypes, termed α_1, α_2, β_1 and β_2. For example, the β_2-adrenergic receptor recognizes both adrenaline and the neurotransmitter **noradrenaline**, as well as several artificial compounds, such as **isoprenaline**. The β_2-adrenergic receptor message transduction system has been relatively well characterized. The receptor itself has seven membrane-spanning helices, whose arrangement within the membrane may dictate the specificity with which the receptor binds the chemical. After the chemical has bound, the receptor interacts with other separate membrane components, which are the **G proteins** and the enzyme **adenylate cyclase**.

SECOND MESSENGER SYSTEM

Tightly bound to the cytoplasmic surface of the membrane is the integral **G protein**, so-called because it binds guanosine triphosphate (GTP) with high affinity. The G protein consists of three subunits, called α, β and γ, of molecular weight about 42, 35 and 10 kDa, respectively. The α-subunit can bind guanosine diphosphate (GDP) and GTP. The third protein integral to the membrane is the enzyme **adenylate cyclase**, which has an ATP-binding site on the cytoplasmic face of the membrane and, when the enzyme is activated, it converts ATP to cyclic adenosine monophosphate (cAMP).

In the absence of hormone, the G protein binds GDP, and adenylate cyclase is inactive. But, when the hormone binds to its site on the receptor, the receptor's conformation is changed and it binds the G protein. GDP dissociates, allowing GTP to bind instead. Consequently, the G protein dissociates into $G_{\beta,\gamma}$- and G_α-subunits. The G_α-subunit binds to adenylate cyclase, which is activated, and converts ATP to cAMP. cAMP is a so-called **second messenger**, relaying to the cell the fact that the hormone has bound to the receptor. Adenylate cyclase activation is rapidly terminated by hydrolysis of bound GTP to GDP, resulting in the resetting of the system for further stimulation.

Within the cell, cAMP initiates a cascade of protein phosphorylations by binding to a **protein kinase**. When cAMP binds to it, the kinase dissociates into two subunits, one regulatory and the other catalytic. The kinase is then able to phosphorylate other proteins by transferring the terminal phosphate group of ATP to serine, threonine or tyrosine residues of the substrate proteins. The end result of the cascade may be, for example, glycogen breakdown in liver or muscle, the hydrolysis of triacylglycerol to fatty acids and glycerol in adipose (fat) cells or the synthesis of steroid hormones in the adrenal cortex.

This type of cascade is an extremely efficient system for **amplifying** the signal, since the binding of a single adrenaline molecule results in the activation of adenylate cyclase molecules, and the generation of many molecules of cAMP.

Inhibitory G proteins also exist in the membrane, and these are activated to **inhibit** production of cAMP. They are activated by different receptors and hormones. These G proteins have $G_{\beta,\gamma}$-subunits, but a different G_α-subunit, called $G_{i\alpha}$, which binds GTP when activated, but which somehow inhibits adenylate cyclase. For example, adrenaline, through its β-receptor, stimulates cAMP production, while the neurotransmitter adenosine, through one of its receptors, the A_1-receptor, inhibits cAMP production. The adenosine A_2-receptor, however, stimulates cAMP production.

> *Note*: cholera toxin, produced by the bacterium *Vibrio cholerae*, irreversibly activates G_α, so that it cannot hydrolyse GTP to GDP, resulting in continuously raised cAMP in the intestinal epithelium cell, causing a flood of Na^+ and H_2O into the intestinal lumen, massive diarrhoea and possible death from dehydration.

Other examples of second messengers are inositol trisphosphate (IP_3) and diacylglycerol (DAG), which are generated by activation of membrane receptors by hormones, for example through the action of adrenaline on α_1-receptors, or the neurotransmitter acetylcholine on muscarinic cholinergic receptors. The activated receptor binds to a membrane G protein, which stimulates the membrane-bound enzyme phospholipase C (PLC) to hydrolyse phosphatidylinositol-4,5-bisphosphate (PIP_2) to DAG and IP_3. IP_3 diffuses into the cytosol and binds to a receptor on the ER, which discharges free Ca^{2+} into the cytoplasm. The ions facilitate intracellular processes such as vesicle exocytosis or glycolysis.

5 Intracellular receptors and receptor antagonists

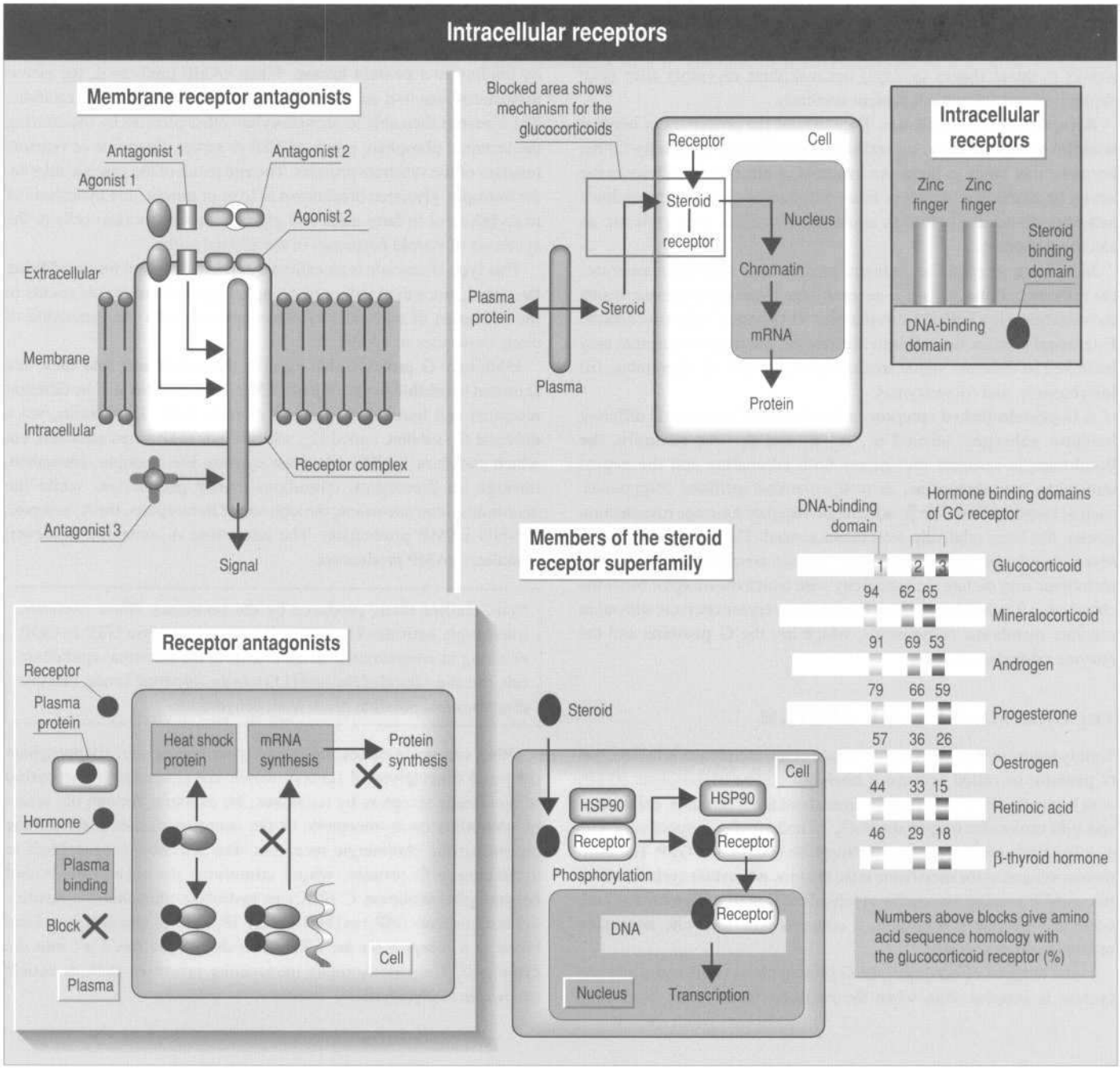

Fig. 5.1

INTRACELLULAR RECEPTORS

Intracellular receptor proteins bind to the lipophilic hormones which pass easily through the cell membrane. The intracellular receptors for **steroids**, **vitamin D** and **thyroid hormone** form part of a large so-called **receptor superfamily**, whose members bind to the nuclear chromatin and alter transcription.

Mechanism of receptor activation

The mechanism of intracellular receptor activation by hormone is not known, but there is evidence that in the inactive state, the receptor is bound by a **heat shock protein (HSP90)**. The number 90 refers to the size of the protein. The term 'heat shock' was given because these proteins were originally detected after cells were injured by heat shock,

although it is now known that they are expressed in untraumatized tissues. There is a family of HSPs associated with diseases, fever, ischaemia, ageing and the inflammatory process. Some HSPs are also molecular chaperones (see p. 107).

When the hormone binds to the receptor, the HSP90 dissociates from the receptor, and the receptor proteins form homodimers. The receptor–hormone complex binds to specific sites on the DNA, called **hormone response elements (HREs)**, upstream from initiation sites. Many of the deoxynucleotide sequences of the HREs are known. Receptors for the sex hormones, and for glucocorticoids such as cortisol, are associated with HSPs, although receptors for thyroid hormone and vitamin D are not, and these receptors appear able to bind to their HREs in the absence of the hormone. In all cases, the process of receptor activation involves phosphorylation of the receptor, although the exact mechanism is unclear.

Nature of the receptor

The members of the intracellular superfamily of receptors contain three main regions. The first is a DNA-binding domain (region 1) which consists of two 'zinc fingers', so-called because each binds an ion of zinc (Zn^{2+}). This region is rich in cysteine and basic amino acids. It is believed that the first zinc finger determines the specificity of the binding of the receptor to DNA, while the other stabilizes the receptor to its response element on the DNA. Regions 2 and 3 of the receptor determine the hormone specificity of the binding reaction. From the scheme shown in Fig. 5.1, it can be seen that region 1 is very highly conserved among members of the superfamily, while the hormone-binding regions show much less homology.

RECEPTOR ANTAGONISM

The specificity of the binding reaction between ligand and receptor creates many opportunities for designing drugs that will lessen or prevent the action of the ligand. Receptor-blocking drugs play a large part in therapy, and examples include: (i) the β-receptor blocking drugs, for example, **propranolol**, used to treat cardiovascular disorders; and (ii) the anticancer drug **tamoxifen**, which inhibits the binding of the sex hormone oestradiol to its intracellular receptor, and is used in certain forms of breast cancer.

Antagonists can block hormone effects by binding directly to the receptor, or through indirect means. In classical pharmacological terminology, the ligand that activates the receptor is termed an **agonist**, and the ligand that blocks the action of the agonist is termed an **antagonist**.

Membrane receptor antagonism

In the scheme shown in Fig. 5.1, agonist 1 binds to its site in order to elicit the response, while antagonist 1 binds to an allosteric site on the receptor to block the action of agonist 1. An example of such an agonist–antagonist pair is the neurotransmitter **glutamate**, which binds to its site in order to open ion channels, and the antagonist **2-amino-phosphonovalerate**, which binds to an allosteric site on the same receptor to which glutamate binds.

Two molecules of agonist 2 are required to bind in order to elicit a response, and antagonist 2 blocks by occupying the sites. An example of such an agonist is **acetylcholine**, which binds to two sites on the nicotinic receptor on skeletal muscle fibres. The antagonist **tubocurarine**, a muscle relaxant, occupies both sites, and blocks the action of acetylcholine.

Antagonist 3 does not directly antagonize the action of a ligand, but is able to penetrate the membrane and interact with the post-receptor mechanisms to block the action of the agonist. It may, for example, inhibit the passage of ions through a channel opened by the agonist. An example of antagonist 3 is the anticonvulsant drug **phenytoin**, which blocks the transmission of ions through the membrane.

Intracellular receptor antagonism

Intracellular hormone action can be antagonized by substances that interfere with the normal hormone–receptor interaction and post-receptor binding processes. The receptor itself may be blocked, or the post-receptor-mediated events, for example DNA binding, transcription or translation.

Dimerization block

Drugs have been developed that block the dimerization of the receptor once it has been activated by the hormone. An example of such a drug is the compound **ICI164384**, which appears to block the dimerization of the oestrogen receptor homodimers after they have been bound by oestradiol. (ICI is the company that developed the drug.) Failure to dimerize will compromise the ability of the receptor complex to bind to the HRE.

Transcriptional block

Transcriptional block is the mechanism whereby two important antagonists of steroid hormone action exert their effects. **Tamoxifen**, mentioned earlier, and the controversial substance **RU486** (developed by the company Roussel to block the actions of the hormone progesterone, and thus terminate pregnancy), both inhibit activation of transcription activation sites after they are bound by the receptor.

6 Molecules I

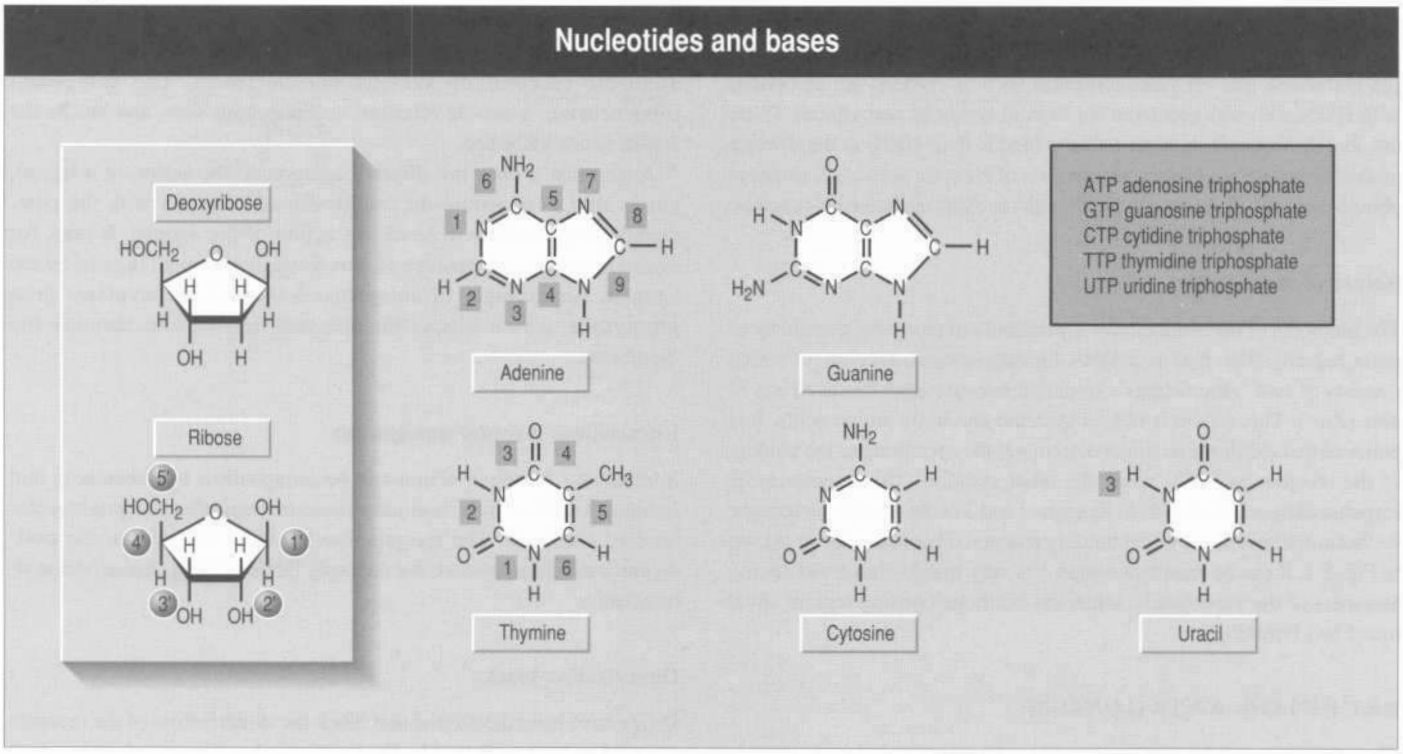

Fig. 6.1

CHEMICAL BASIS OF CELLS

All cells depend on chemical activity, which can conveniently be classified in terms of **function and structure**. Functionally, cells rely on chemicals to provide energy, to maintain electrochemical gradients across membranes and to provide the messengers which convey information. Structurally, cells require chemicals to provide the building blocks for the structures which have been described in previous chapters, and for growth and repair of these structures.

Chemical size

Chemicals can be classified broadly in terms of size, as small or large molecules. Small molecules consist, generally, of less than 50 atoms, with molecular weights mainly less than 1000, and have been referred to as **metabolic intermediates**. They include the inorganic ions such as Na^+, sugars such as glucose, amino acids, for example glycine, nitrogenous bases including the purines and pyrimidines, fatty acids and steroids.

Larger molecules are called **macromolecules**. In the cell, the macromolecules are the **polysaccharides**, **proteins** and the **nucleic acids**. Macromolecules are composed of several smaller molecules, linked chemically by covalent bonds. When the smaller molecules are of one type, the macromolecule is called a **polymer**, and the smaller molecules are termed **monomers**. Thus, proteins are polymers, composed of monomers called amino acids. Monomers are also commonly called **residues**. Nucleic acids are polymers of nucleotides, and polysaccharides are polymers of sugars. Some proteins contain polysaccharide moieties or groups covalently attached, when they are called **glycoproteins**.

NUCLEIC ACIDS

Nucleic acids are composed of nucleotides arranged in immense chains. Each nucleotide in turn is made of a base, either a purine or a pyrimidine, a pentose sugar (five-carbon sugar) and a phosphate group. The sugar is **ribose** or **deoxyribose**. When the sugar is ribose the nucleic acid is **RNA**. When the sugar is deoxyribose the nucleotide is **DNA**, and the nucleotides are deoxynucleotides. DNA, together with associated basic proteins called **histones**, makes up the **chromosome**. Humans have 46 chromosomes arranged in 23 pairs in the cell nucleus, and these contain the hereditary information of the body. The chromosomes have stretches of DNA, called **genes**, which code for specific proteins, as well as having associated sequences of deoxynucleotides which control the expression of the genes with which they are associated. A cell that contains two sets of the chromosomes is termed **diploid**, while a cell that contains only one set is termed **haploid**. Humans are therefore diploid organisms. The human germ cell, whether spermatozoon or unfertilized ovum, has only one set of 23 chromosomes, and is termed haploid. When the two combine during fertilization, the diploid status of the cell is restored.

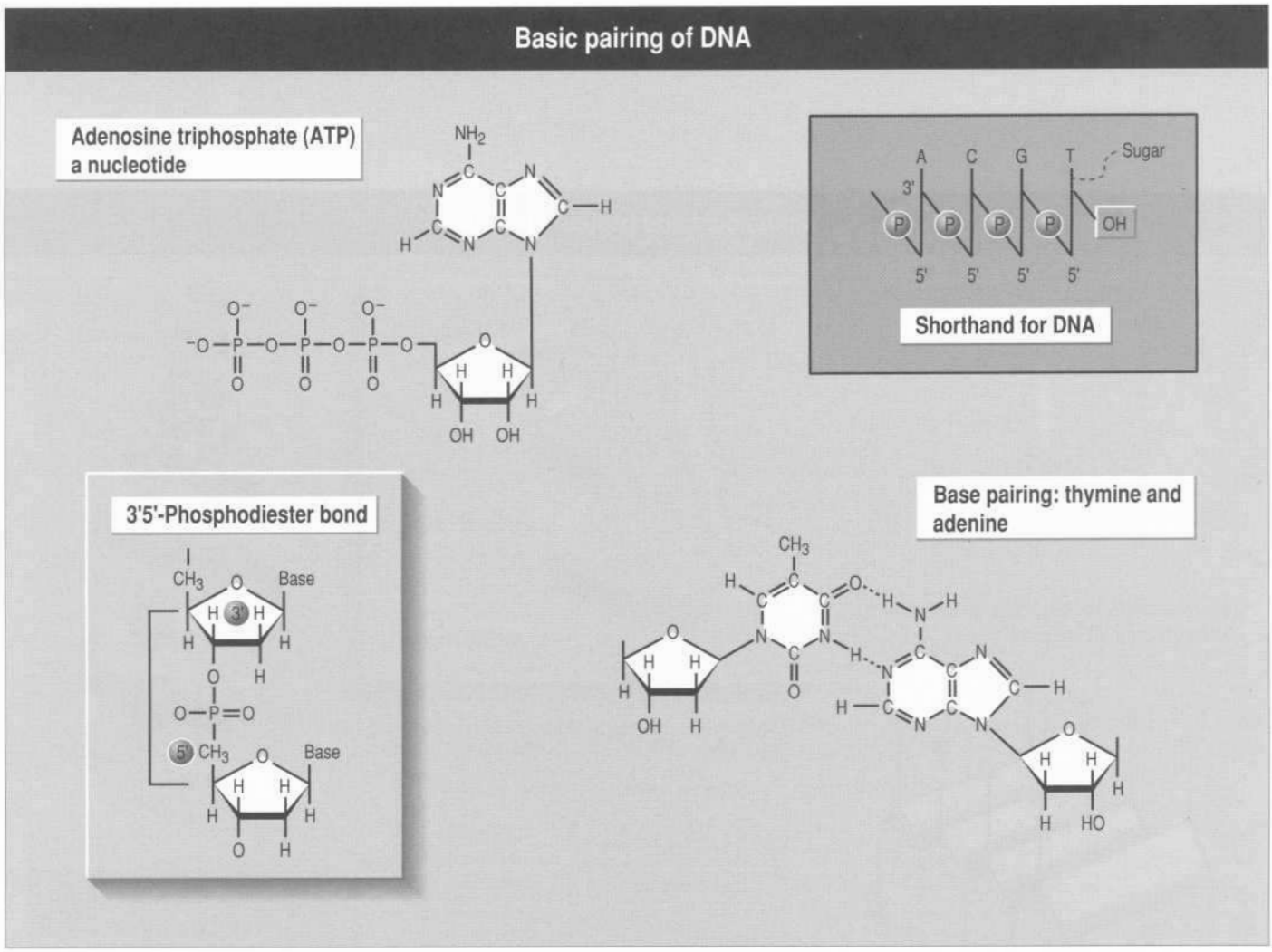

Fig. 6.2

Nucleotide structure

For DNA, the bases of the nucleotides are ring structures, either **pyrimidines** (thymine (T) or cytosine (C)) or **purines** (guanine (G) or adenine (A)). For RNA, the thymine is replaced by uracil (U). For both DNA and RNA, the nucleotides are linked (or 'condensed') through the sugars, where phosphate groups link the 3′ point of one sugar to the 5′ point of another, releasing a molecule of H_2O. These phosphate linkages are termed 3′5′-**phosphodiester bonds**. In this way, a linear chain of nucleotides is built up. Because there are four different nucleotides, and because the chains can be assembled in any order of nucleotides, there are 4^n different possible nucleic acids having n nucleotides. For example, a nucleic acid containing 15 nucleotides has 4^{15} possible combinations (more than 1×10^6). The chain is represented as a strand with the 3′ end having an unsubstituted hydroxyl (–OH) group on the left, and the phosphorylated 5′ end on the right. During synthesis of DNA (or RNA), nucleotides are added to the 3′ end.

Base pairing and the double helix

The DNA molecules in the cell consist of two complementary intertwined nucleotide strands. The strands are held together by what is called **base pairing**. When two strands of DNA come together, guanine will pair with cytosine, and adenine will pair with thymine through non-covalent hydrogen bonds. Thus, the base sequence of one strand will always be complementary to that of the other. Also, due to the way DNA is synthesized, the paired chains run in opposite directions; the 3′ end of one chain lying adjacent to the 5′ end of the other. Structurally, the chain forms what is termed a **double helix**. The double helix is a form of twisted or spiral right-handed staircase coiled round an imaginary central core, with about 10 bases every turn and the helix makes a complete turn every 3.4 nm. The double helix is held together not only by hydrogen bonding but also because the bases are planar and form stacks of base pairs held together by hydrophobic and van der Waal's forces (see Glossary, p. 112). The pairs can be made to separate simply by heating the DNA, when it 'melts' or denatures.

7 Molecules II

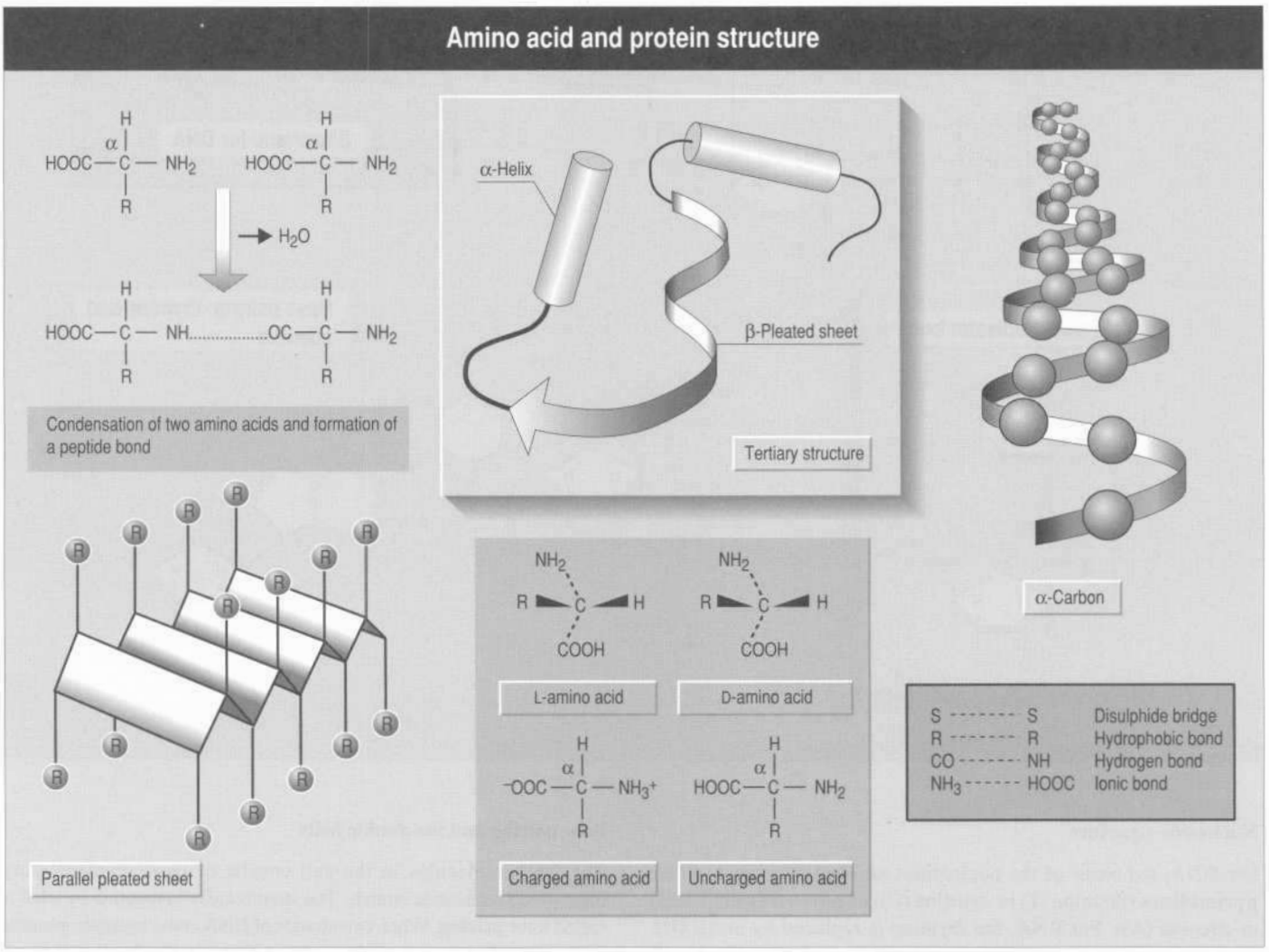

Fig. 7.1

PROTEINS

Proteins are polymers made up of monomers — amino acids. Proteins perform many functional roles inside and outside the cell, which also depends absolutely on proteins for its structure. Unlike DNA, proteins occur in many three-dimensional shapes, depending on their amino acid composition and arrangement. They may be fibrous (e.g. fibrin, collagen), globular (e.g. serum hormone-binding proteins), or antibodies.

Amino acids are so-called because all, with one exception, contain amino ($-NH_2$) groups. The exception is proline, which contains an imino (–NH–) group. They also contain acidic carboxyl (–COOH) groups. Chemically, amino acids have a common design: (i) a central α-carbon atom lying adjacent to the acidic carboxyl group is attached to (ii) the amino (or imino) group; (iii) a hydrogen atom; and (iv) to a variable side chain (R). At neutral pH, amino acids will be **ionized**, the charge depending on the relative numbers of amino and carboxyl groups. Lysine is positively charged, while aspartate is negatively charged. Some amino acids are more **hydrophobic** than others, depending on the hydrocarbon content of the R side chain. All amino acids, except glycine, have asymmetrical carbon atoms, and can therefore exist as **stereoisomers**. The two forms are mirror images, termed D and L, and proteins virtually always contain the L isomer.

Amino acids condense into polymers through the **peptide bond**, which joins the amino group of one amino acid to the carboxyl group of another, with the release of a molecule of H_2O. If the amino acid chain is less than 30 residues long, it is commonly termed a **peptide**, or **polypeptide**. Polypeptides can range in amino acid number from three (e.g. glutathione) to those which are made up of more than 1000 amino acids. Every polypeptide has a free amino group at one end, and a carboxyl group at the other. Proteins contain 20 different types of amino acids; therefore, a 200-amino acid polymer can have 20^{200} possible

structures. This offers enormous variability in protein structure and function.

Proteins have four structural levels:

1 primary — the linear sequence of amino acids and the S–S bonds;

2 secondary — protein folding into α-helix and pleated sheets;

3 tertiary — regional folding between pleated sheets and α-helix, determined by non-covalent bonds; and

4 quaternary — non-covalent binding of different polypeptide subunit chains into a single protein molecule (e.g. haemoglobin, Hb).

Polysaccharides are polymers of sugars (also termed saccharides). Examples of polysaccharides are starch (storage form of glucose in plant cells), cellulose (part of plant cell wall) and glycogen (storage form of glucose in animal cells). Linkages can occur several different ways, and polysaccharides can occur in many different branched forms.

8 DNA replication

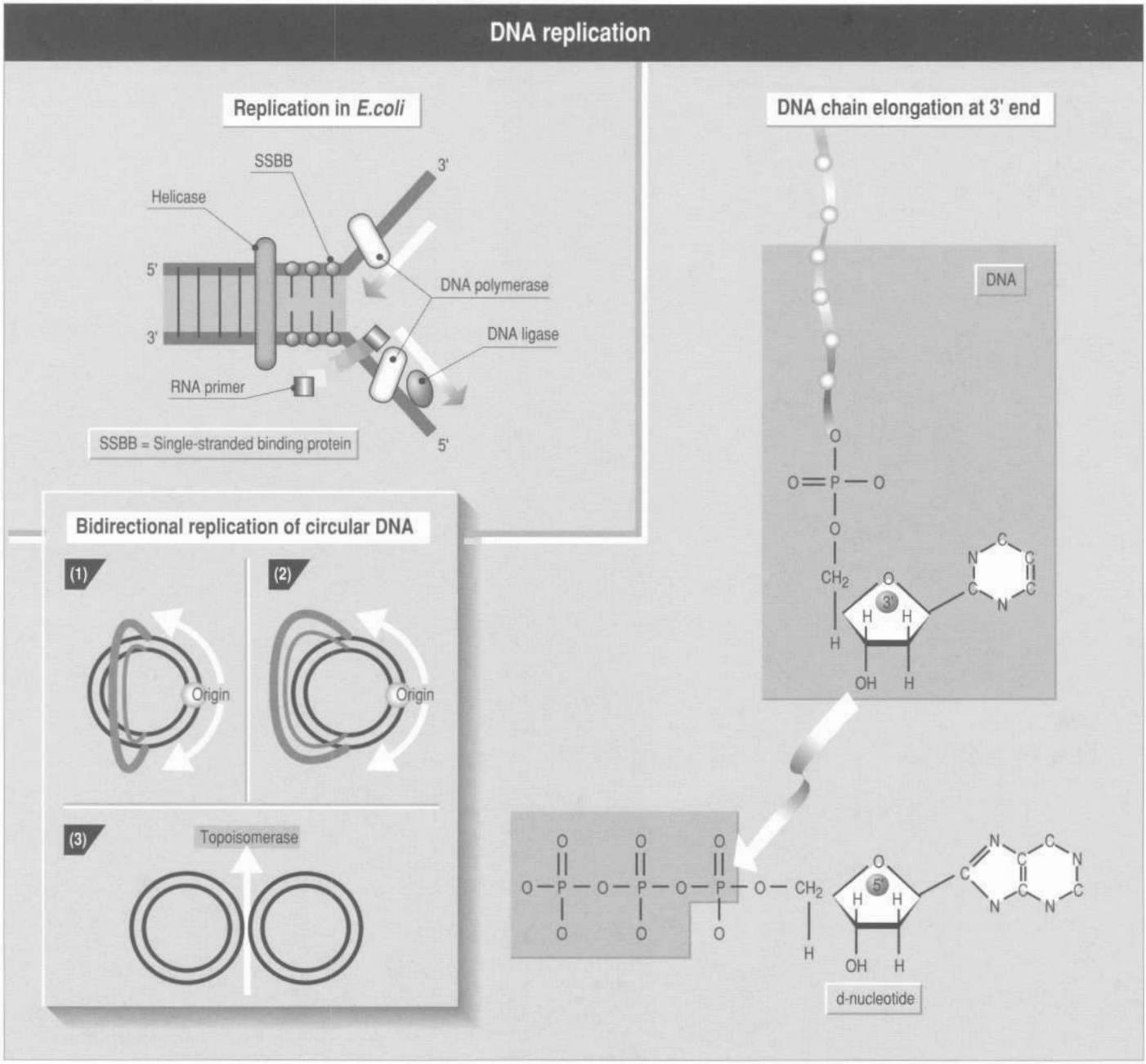

Fig. 8.1

DNA replication: (i) DNA chain synthesis; (ii) its initiation; (iii) termination; (iv) packaging with chromosomal proteins; (v) recombination; and (vi) DNA repair. A new strand of DNA combines with an old one. This is **semi-conservative** replication through **growing forks**. Replication is **bidirectional**, with both strands being copied. Each copied region is termed a replicon. Eventually, the replicons merge. In eukaryotes, the double helix is unwound simultaneously at several sites, and replicated bidirectionally, until eventually all replicated stretches join up to form the new complementary DNA (cDNA) strand.

Unwinding is initiated by an appropriate cellular signal. A **helicase** enzyme system binds to a specific nucleotide sequence and unwinds the DNA adjacent to it to create a **replication fork**. The DNA is unwound just ahead of the DNA polymerase moving up behind to replicate the exposed strands.

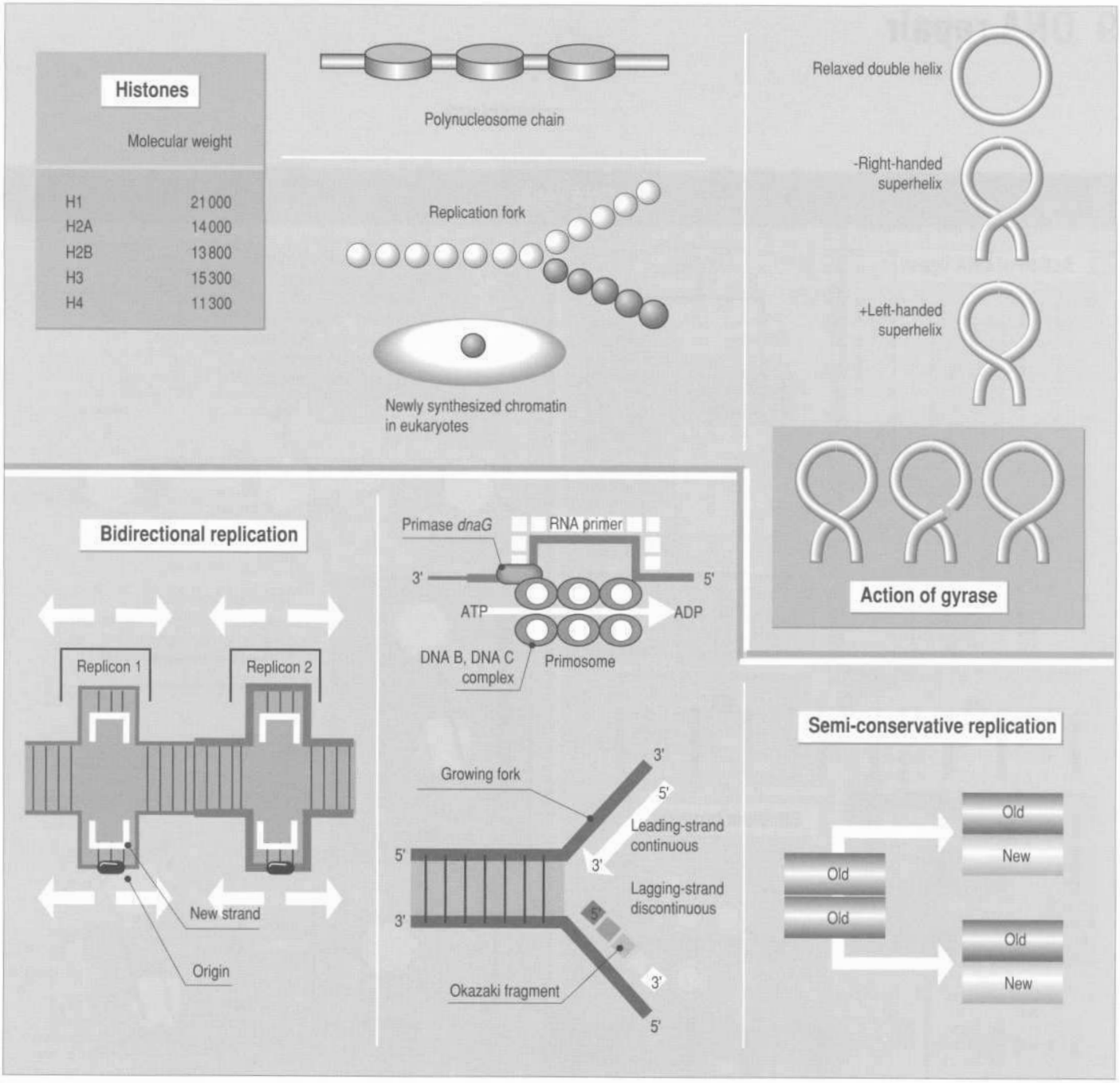

Fig. 8.2

REPLICATION

Replication is initiated by synthesis of a **primer** sequence of RNA at the starting point of replication on both unwound strands by **primase** enzyme activity of DNA polymerase α. (The primase gene is termed *dnaG*.) The polymerase begins elongation at the 3′ end of the primer. Elongation proceeds continuously along the 5′ → 3′ **leading strand**. On the so-called **lagging strand,** elongation must also run from 5′ → 3′, and runs in the opposite direction. Therefore, replication on the lagging strand is **discontinuous**, and has to stop periodically to wait for more of the strand to unwind. This creates segments of cDNA, called **Okazaki fragments**, about 100–200 deoxyribonucleotides long in eukaryotes. They are formed when proteins called **primosomes** attach themselves to the RNA–DNA complex and activate the reaction ready for the polymerase. After formation of the Okazaki fragment, the primers are removed and replaced by DNA, and the fragments are joined together by a **DNA ligase**. Errors in the matching of bases during elongation are detected by DNA polymerase δ, which 'proofreads' the bases.

Histones

In eukaryotes, DNA is associated with nucleoprotein histones. Their basic charge enables the histones to bind to the phosphate backbone of DNA. The **nucleosome** is a disc-shaped coil of DNA wound 1.5 times round a cluster of proteins including two molecules each of H4, H3, H2B and H2A. The **polynucleosome** consists of several nucleosomes joined by linker DNA. The **chromatin** formed from the DNA–protein complex is condensed into **chromosomes.**

9 DNA repair

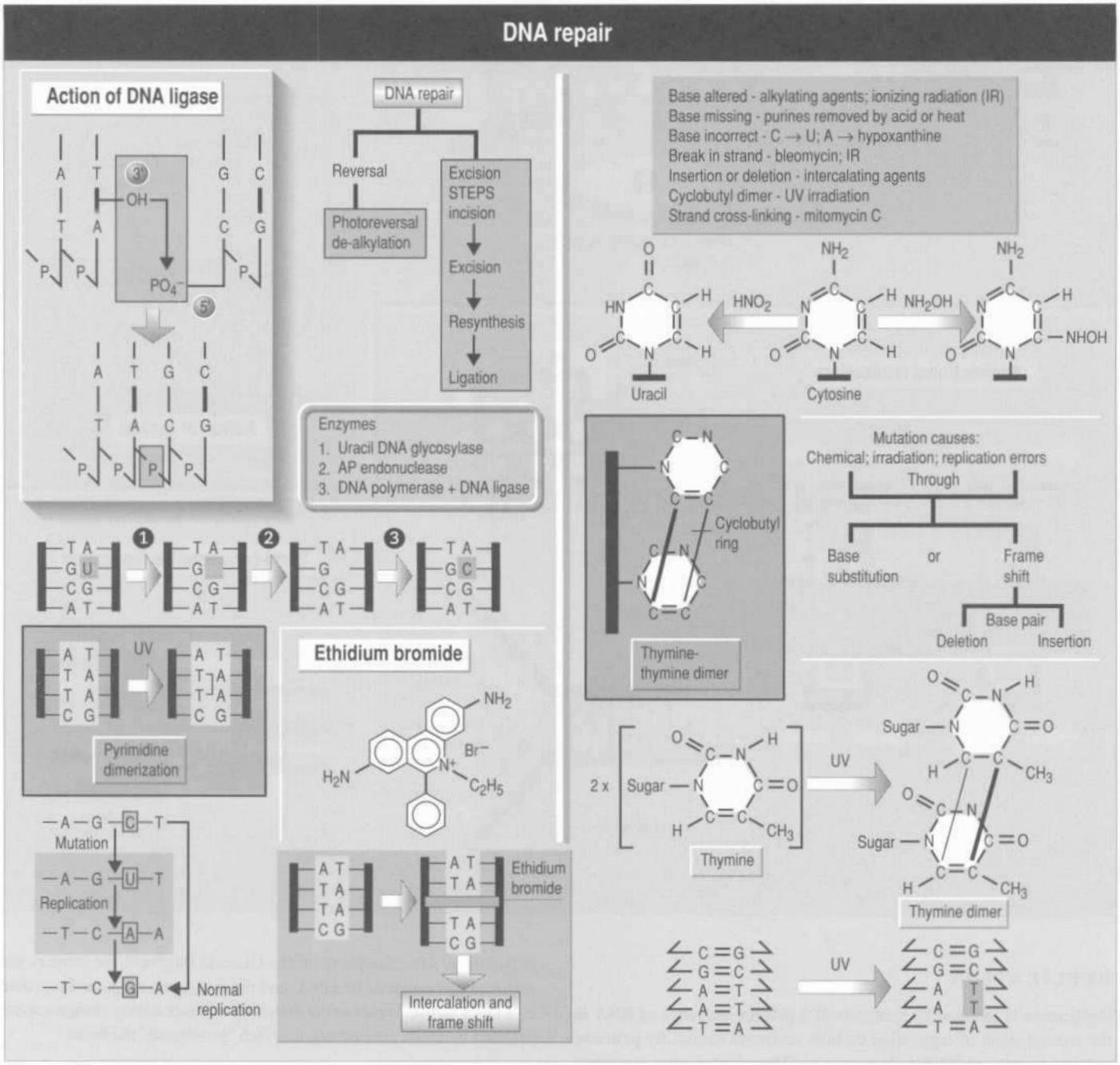

Fig. 9.1

DNA REPAIR

DNA repair is the maintenance of DNA stability through removal of errors in base sequence due to chemical or irradiation-induced damage, or through errors introduced during replication. Stability of DNA is essential for the preservation of genetic information. Survival of the species would be jeopardized by errors in rapidly dividing germ cells, and cellular function would not be possible in stable cells, such as brain or liver cells, which may not divide for years.

In *Escherichia coli*, DNA polymerase I introduces on average one incorrect base per 10^4 base pairs. Genes in *E. coli* are about 10^3 bases long. Therefore, errors would be introduced every tenth gene (10^{-1} errors (or mutations) per gene each generation). But, the actual rate of errors is much less in *E. coli*: around 10^{-5} errors per gene each generation. It is

thought that the rate of error production is similar in animal cells, which must also have a mechanism for repairing errors in coding regions of the genes.

Mutations can be defined as stable alterations in gene DNA structure. They may be **silent**, or **expressed** as phenotypic alterations. Mutations may be one of the following.

1 **Frame shift** mutations, which result from:
   (a) deletion of a base pair or block of base pairs;
   (b) insertion of new base pairs.
2 **Base substitutions**, which result from:
   (a) **transversions**, which are the substitution of pyrimidine–purine base pairs by purine–pyrimidine base pairs;
   (b) **transitions**, which are the substitution of purine–pyrimidine base pairs by pyrimidine–purine base pairs.

Tests for potential carcinogens include tests for their mutagenic effects on bacteria. This implies that alterations to DNA lie at the root of both carcinogenesis and mutagenesis. One test is the **Ames test**, in which a strain of *Salmonella*, unable to synthesize histidine, is grown in a Petri dish in a medium that lacks histidine. Any bacteria that grow are natural mutants. The suspected mutagen is added, and many mutants may be formed — some of which can synthesize histidine — and they will appear as yet more visible colonies on the agar. Different strains have been identified, some of which respond to mutagens that cause base substitutions, while other strains respond to mutagens producing additions or deletions, i.e. frame shifts. This makes it possible to identify, tentatively, the mechanism of action of the mutagen/carcinogen.

Repair is effected by **proofreading**. The proofreading ability of DNA polymerases was first discovered in *E. coli*. The enzyme DNA polymerase 1 is believed to have $3' \rightarrow 5'$ exonuclease activity, which enables it to proofread newly added base sequences, excise non-matching bases and replace these with the correctly matching bases.

Repair of pyrimidine dimers

Pyrimidine dimers are formed after exposure of DNA to ultraviolet (UV) light, when adjacent pyrimidine residues on a DNA strand may become linked covalently. The distortion of DNA produced by the dimer is detected by a group of proteins expressed by the *uvrABC* genes; the proteins consist of a tetramer of two proteins, P_{uvrA}, and two proteins, P_{uvrB}. The exonuclease enzyme P_{uvrC} cuts the strands at two places: four nucleotides away from the dimer on the 3′ side, and eight nucleotides away on the 5′ side. The excised 12-residue piece of DNA is unwound by a helicase P_{uvrD}, and diffuses away. DNA polymerase 1 moves into the gap created, and uses the 3′ cut end as the primer and the intact complementary strand as the template to repair the cut strand. Finally, DNA ligase joins the 3′ end of the newly synthesized DNA and the original DNA.

Excision repair is also used by the cell to remove crosslinks formed by drugs used to treat cancer, such as cisplatin, mitomycin C and the nitrogen mustards.

> *Note*: *E. coli* contains an enzyme, DNA photolyase, which binds to the DNA region distorted by the dimer, and becomes photoactivated and splits the dimer.

Repair of deaminated cytosine

Cytosine may be deaminated to uracil, and this occurs spontaneously during the life of DNA. Since uracil can pair with adenine (U–A), the chemical change is potentially mutagenic. The presence of uracil on the DNA is recognized by uracil DNA glycosidase, which breaks the bond between uracil and deoxyribose by hydrolysis. The gap formed by removal of the pyrimidine is called an AP site (i.e. apurinic, containing no cytosine or thymine. The gap is recognized by the enzyme AP endonuclease, which cuts the DNA backbone adjacent to the missing base. DNA polymerase I cuts away the piece of deoxyribose phosphate and inserts cytosine opposite the intact guanine residue on the complementary strand, and DNA ligase seals the cut DNA strand.

Diseases associated with defects of DNA repair

These include hereditary retinoblastoma, Fanconi's anaemia and Xeroderma pigmentosum.

Xeroderma pigmentosum is the best understood. This rare disease is genetically transmitted as an autosomal recessive trait. Patients are highly sensitive to UV and sunlight, developing skin lesions soon after birth. The dermis atrophies, the eyelids scar and the cornea ulcerates. Freckles and skin ulcerations appear, followed by skin cancers.

The disease is caused by a defect of the exonuclease enzyme which nicks the DNA at the site of pyrimidine dimers, which are known to be caused by UV radiation. Skin fibroblasts from patients with Xeroderma pigmentosum have been shown to contain the deficient enzyme. Mutations in one of at least nine different genes can cause the disease. Although the incidence of the disease is low, about 1% of the population are carriers of at least one of the mutated genes.

10 Recombination

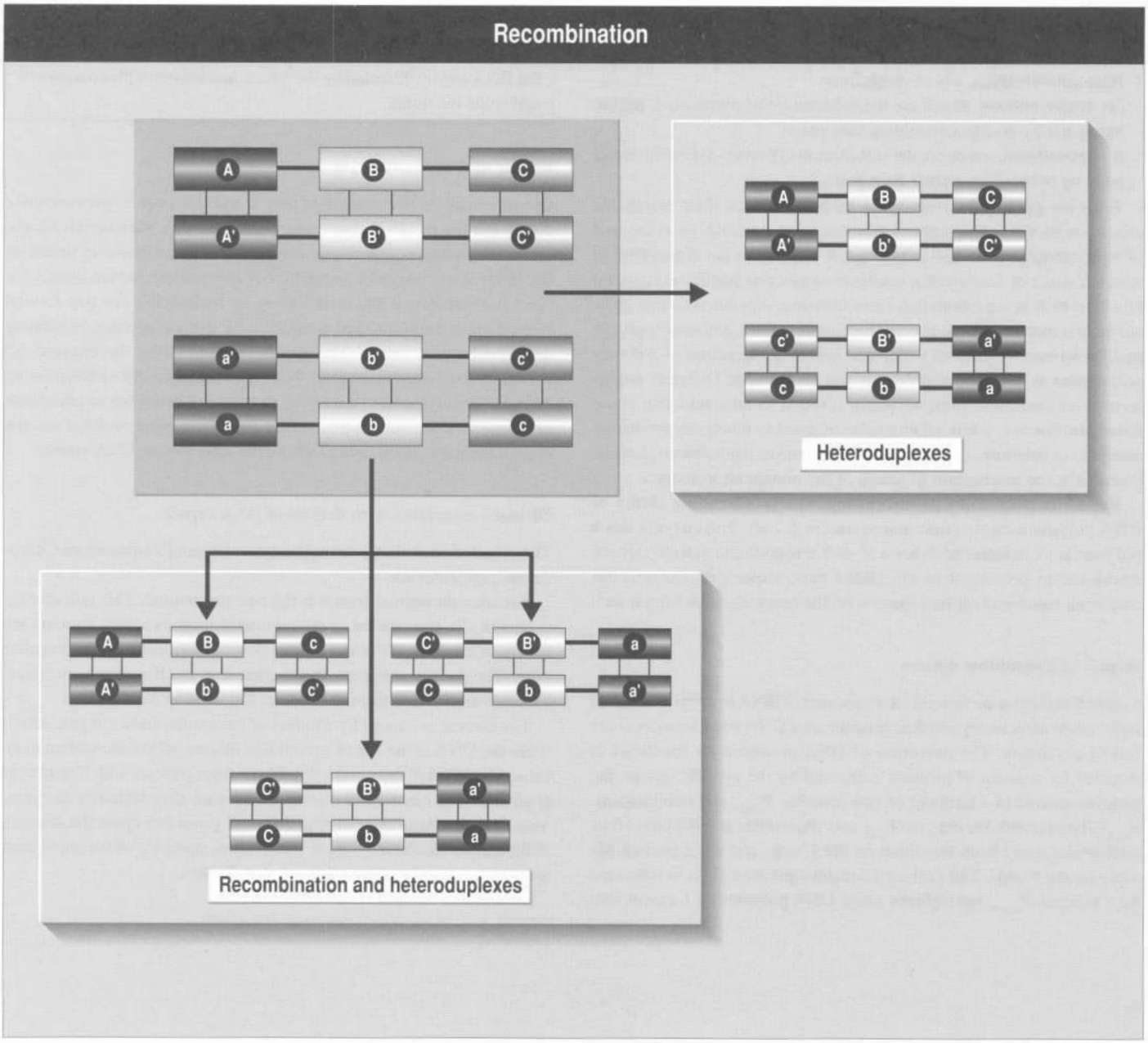

Fig. 10.1

INTRODUCTION

Recombination is the exchange or crossing over of blocks of genes by homologous chromosomes. Recombination takes place during meiosis in sexually reproducing organisms, and occurs in plants, animals and even in subcellular particles such as mitochondria and plasmids. The process involves: (i) the breaking of two homologous double-stranded DNA molecules; (ii) the exchange of both strands at the point of breakage; (iii) the exchange of genes; and (iv) the separation of the two altered strands of DNA. This process of formation of new DNA molecules is termed **general recombination**. When a gene is moved from one chromosome to another or from one part of the chromosome to another, this is called **transposition**, and does not require the same degree of homology as that for general recombination. The functions of rearranging genes are: (i) to repair damaged DNA; (ii) to generate new species of DNA, which is an important evolutionary event; and (iii) to regulate DNA expression.

Note: alternative forms of a gene, termed **alleles**, can be exchanged by general recombination, and can occur on the same chromosomal site.

MECHANISM OF RECOMBINATION

Initially, strands of DNA can join to each other by non-covalent base pairing to form **lap joints**, and these joints are made permanent by subsequent DNA synthesis.

ENZYMES OF RECOMBINATION

Single-stranded DNA for recombination is generated by a complex of proteins, P_{recB}, P_{recC} and P_{recD}, which are products of the *recB*, *recC* and *recD* genes, and which form a complex protein enzyme system. In *E. coli*, the complex: (i) recognizes a sequence — **5′-GCTGGTGG-3′** — the so-called *chi* sequence, and cuts the strand about four to six nucleotides away from the 3′ end of the *chi* sequence; and (ii) the complex unwinds the DNA strands. The process requires ATP hydrolysis.

In *E. coli*, the **recA protein** utilizes ATP to catalyse the assimilation of a single strand of DNA into a duplex. The protein: (i) binds to the single strand, and the protein–DNA filament thus formed binds to the duplex; (ii) partially unwinds and 'reads' the duplex for sequences complementary to those on the single strand; (iii) further unwinds the duplex where complementarity and pairing occurs; and (iv) the exchange process continues along the duplex, and these are the mechanics underlying the process of branch migration mentioned above. All these steps require the hydrolysis of ATP.

SOS RESPONSE

Normally, low levels of P_{recA} are present in *E. coli*, due to suppression of P_{recA} mRNA production by a repressor protein termed P_{lexA}. P_{lexA} suppresses the production of many repair proteins. When DNA is damaged in *E. coli*, the cell initiates what is called an SOS response, generating hundreds of new copies of recA protein: (i) damaged single-stranded DNA binds to existing recA protein; and (ii) the complex formed binds to lexA protein and splits its alanine–glycine bonds, rendering it inactive. Note that P_{recA} therefore acts not only as a recombinase but also as a protease because of its action on P_{lexA}. Another gene activated by inactivation of *lexA* is the *uvrA* gene, whose product, the *uvrA* nuclease, cuts away thymine dimers which are formed by UV radiation damage.

MOBILE GENETIC ELEMENTS (MGEs)

MGEs, also called **transposable elements** or **transposons**, mediate the large-scale rearrangements that cannot be effected by general recombination. **Plasmids** are important MGEs confined to prokaryotes. Plasmids are circular duplex DNA molecules which can replicate autonomously, and are therefore **replicons**. A replicon is any unit length of DNA which possesses its own origin of replication, and includes plasmids, bacterial chromosomes or regions of eukaryotic DNA.

Plasmids carry genes which express: (i) the generation of bacterial toxins; (ii) the metabolism of metabolites and other chemicals; and (iii) the inactivation of antibiotics. They are therefore tools of evolution and adaptation. Plasmids are also tools used by genetic engineers to introduce novel genes into cells. Engineers have termed particles such as plasmids, which are used to transfer genetic information between cells, **vectors**.

In recombination, plasmids may be F^+ factor plasmids or R^+ factor plasmids.

F^+ factor plasmid

The DNA sequences of F^+ factor plasmids ('F' stands for fertility) can be transferred between bacteria as follows.

1 A 'male' bacterium (F^+) fastens itself to a 'female' bacterium (F^-) by means of a **sex pilus** on its surface.

2 The pilus retracts to pull the two bacteria into close contact; one strand of the plasmid is cut and the DNA duplex unwinds.

3 The 5′ end of the cut strand passes into the 'female cell', where a complementary strand is synthesized and a circular duplex plasmid formed.

The recipient cell is now F^+, and the transferred plasmid has the genetic information required for production of a sex pilus.

The new plasmid can: (i) remain separate from the main bacterial chromosome; or (ii) become integrated into the chromosome, when the bacterium is now denoted as being a high frequency of recombination (hfr) cell. An hfr cell donates not just the F^+ factor plasmid during transfer, but its complete chromosomal database. An hfr cell can splice out the F factor from its chromosome, and thus reverts to the F^- state.

F^+ factor plasmids are not the only means of transferring genetic information between bacteria. This information can also be transferred by MGE called **bacteriophages**. A bacteriophage is a virus that infects bacteria.

R factor plasmids

These are so-called because they carry resistance to antibiotics. A bacterium can develop resistance to one or more antibiotics if it receives an R factor plasmid. The plasmid may contain a **resistance transfer factor** (**RTF**) as well as many so-called ***r* genes**. *r* Genes express enzymes that inactivate antibiotics such as tetracycline, streptomycin, chloramphenicol and sulphanilamide. Smaller plasmids, lacking an RTF region, will confer resistance to a single antibiotic when transferred.

11 RNA

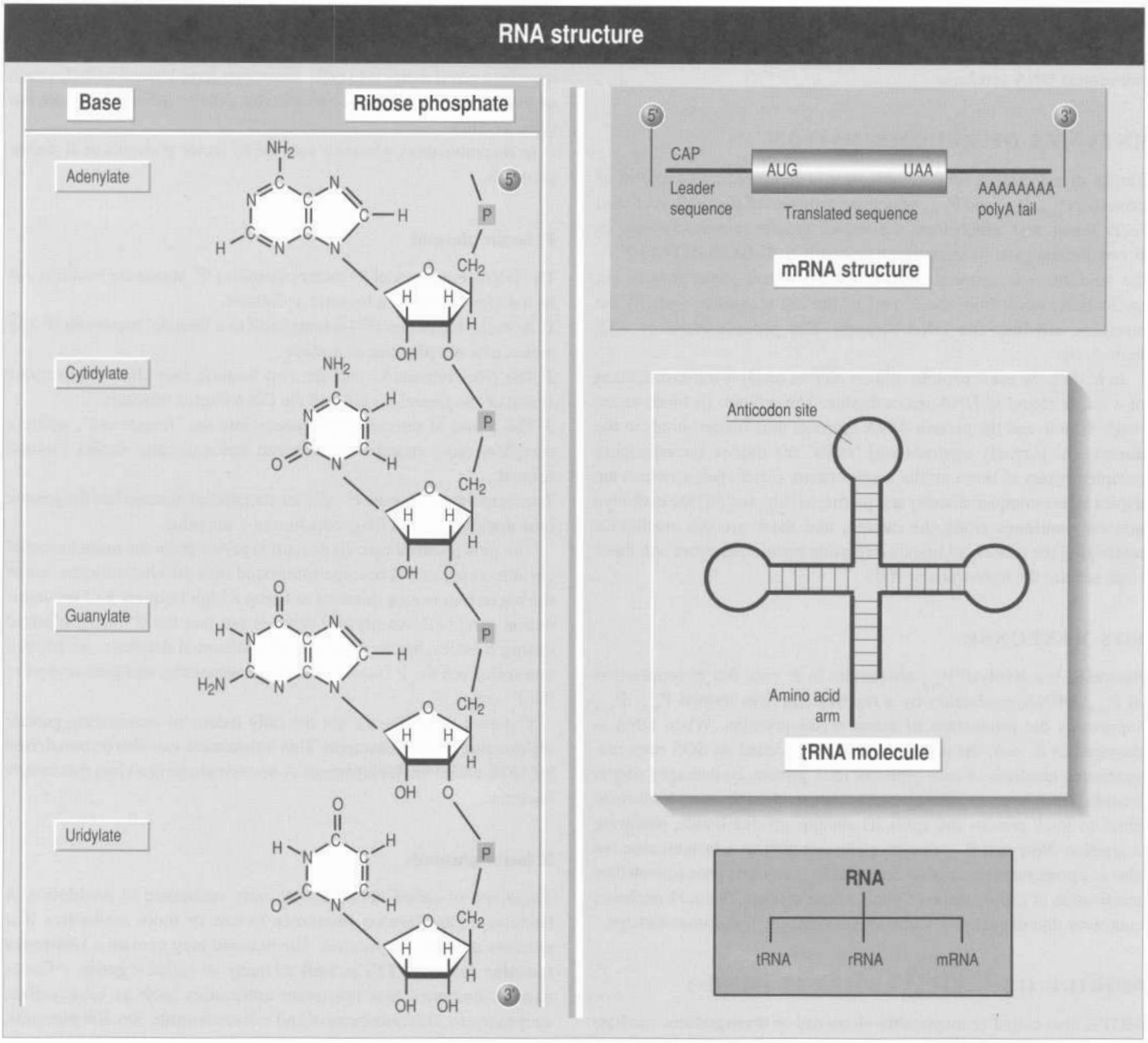

Fig. 11.1

RNA is an unbranched linear polymer whose chemical composition is similar to that of DNA, except that: (i) ribose is used instead of deoxyribose; (ii) adenine pairs with uracil (A–U) instead of with cytosine; and (iii) RNA exists as a single chain whereas DNA exists as a double helix. Two forms, transfer RNA (tRNA) and rRNA, may contain modified nucleosides, e.g. 2-methyladenosine, 5-methylcytosine, and there are several others. The base sequence of RNA reflects the base sequence of the DNA strand from which it was **transcribed**. In the case of mRNA, the information it carries enables the cell to synthesize, accurately, the proteins encoded by the genes.

RNA SPECIES

RNA may be classified in terms of its localization in the cell, its stability or its function. Thus, **mRNA** is the carrier of genetic information, **tRNA** is single stranded and carries the correct amino acid to the protein synthetic machinery and **rRNA** is part of ribosome structure.

mRNA

mRNA may be **monocistronic** or **polycistronic**. Eukaryotic mRNA is monocistronic, i.e. one mRNA molecule carries information about one protein only, whereas a prokaryotic mRNA molecule may be polycistronic, carrying information about more than one protein. mRNA is relatively unstable, being broken down rapidly after transcription has occurred. Some forms of mRNA may be stored in an inactive form until required, for example in amphibian ova or unfertilized eggs, where they remain 'silent' until fertilization. Structurally, mRNA differs from other forms of RNA because of its function as a template for **translation** into protein. At the 5′ end of the linear mRNA is a 'cap', made of a 7-methylated guanosine-5′-triphosphate. The cap prevents phosphatases and nucleases from metabolizing the 5′ end of the mRNA, and also serves as a recognition site for the ribosomal initiation factor (see p. 33). (Note, however, that 5-methylated cytidines and 6-methylated adenosines have been discovered further along the strand.) Following the cap is a **leader sequence**, and after this the **initiation codon** (see p. 32) or sequence, most commonly adenine–uracil–guanine (**AUG**). Following this is the coding region, whose message is ended by a **termination codon** which may be **UGA**, **UUA** or **UAG**. At the 3′ end is a non-translated **trailer sequence**, followed by a polyadenylated (polyA) tail, whose role is not known, but which may enhance stability.

rRNA

rRNA constitutes about 80% of total cellular RNA and is stable metabolically, mainly because of its association with the ribosomal proteins. The eukaryotic ribosome is made up of two subunits, the 40S subunit, and the 60S subunit. (Note: S is the Svedberg unit, which gives an indication of the relative size and molecular weight of a macromolecule.) The two subunits together contain around 75 proteins and four different species of rRNA, distinguished by size. The 40S subunit contains just over 50% of the protein and a molecule of 18S rRNA. The 60S subunit contains the rest of the protein, and a 5S, 5.8S and 28S rRNA. The 5S, 18S and 28S are synthesized in the nucleolus, and the 5S in the nucleoplasm.

rRNA molecules assume secondary structuring, forming base pairs within the molecule, allowing formation of helical regions. RNA may also form so-called hairpin loops. The conformations and nucleotide sequence of 5S, for example, are remarkably constant in evolutionary terms, being very similar in *E. coli* and humans. If the ribosome lacks 5S it cannot translate mRNA into protein. The large rRNA forms contain several modified nucleotides, particularly 2′-methylations on ribose, and several methylations of the bases, which may be necessary for processing RNA precursors.

> *Note*: methylation of bases in rRNA can confer resistance to antibiotics in bacteria. For example, *Staphylococcus aureus* can develop resistance to lincomycin and erythromycin if it acquires a plasmid that carries the gene coding for a methylase, which N6-methylates adenosine on 23S rRNA. This blocks the binding of erythromycin to the large bacterial ribosomal subunit.

rRNA may mediate translation by pairing with specific regions of mRNA, and by operating, together with the proteins, the hinging apparatus of the ribosome.

tRNA

tRNA constitutes around 15% of the total cellular RNA, and has several functions:

1 it binds specific amino acids, thereby raising their activity levels for formation of peptide bonds;

2 it carries the amino acid to the polyribosome; and

3 it accurately recognizes the correct codon in the mRNA, corresponding to the amino acid it carries.

One tRNA molecule carries one species only of amino acid, and although there are 20 amino acids, there are over 50 different species of tRNA in the cell. An amino acid may be able to bind to more than one species of tRNA, which are called **isoacceptors**.

tRNA has two active sites: (i) at the 3′ end –CCA–OH, to which the amino acid is enzymatically attached (if, for example, the tRNA carries arginine, it is called $tRNA^{Arg}$); and (ii) an **anticodon triplet**, which recognizes the complementary codon on the mRNA at the ribosomal complex.

Structurally, tRNA forms the so-called **clover leaf** secondary structure through base pairing; the anticodon active site is situated on one of the clover leaves, and the –CCA–OH site on the stem. There are several modified bases in tRNA, although their precise function is not known. They may be involved in tRNA–protein interactions and in enhancing tRNA stability.

RNA type	Site synthesized, S units*	Function
Messenger RNA (mRNA)	Nucleoplasm	Template for protein synthesis
Transfer RNA (tRNA)	Nucleoplasm 4S	Transfers amino acids to mRNA
Small nuclear RNA (snRNA)	Nucleoplasm	Chromatin regulatory and structural RNA
Heterogeneous nuclear RNA (hnRNA)	Nucleoplasm 30–100S	Precursors of other RNA
Ribosomal RNA (rRNA)	Nucleoplasm 5S	Part of ribosome structure
	Nucleolus 5.9–18S	
	Mitochondria 12–16S	
Mitochondrial mRNA (mt mRNA)	Mitochondria 9–40S	Template for protein synthesis
Mitochondrial tRNA (mt tRNA)	Mitochondria 3.2–4S	Transfers amino acids to mRNA
Small cytoplasmic RNA (scRNA)	Rough ER and cytosol 7S	Selects protein for export

* S, Svedberg units; coefficient of sedimentation.

12 Transcription I

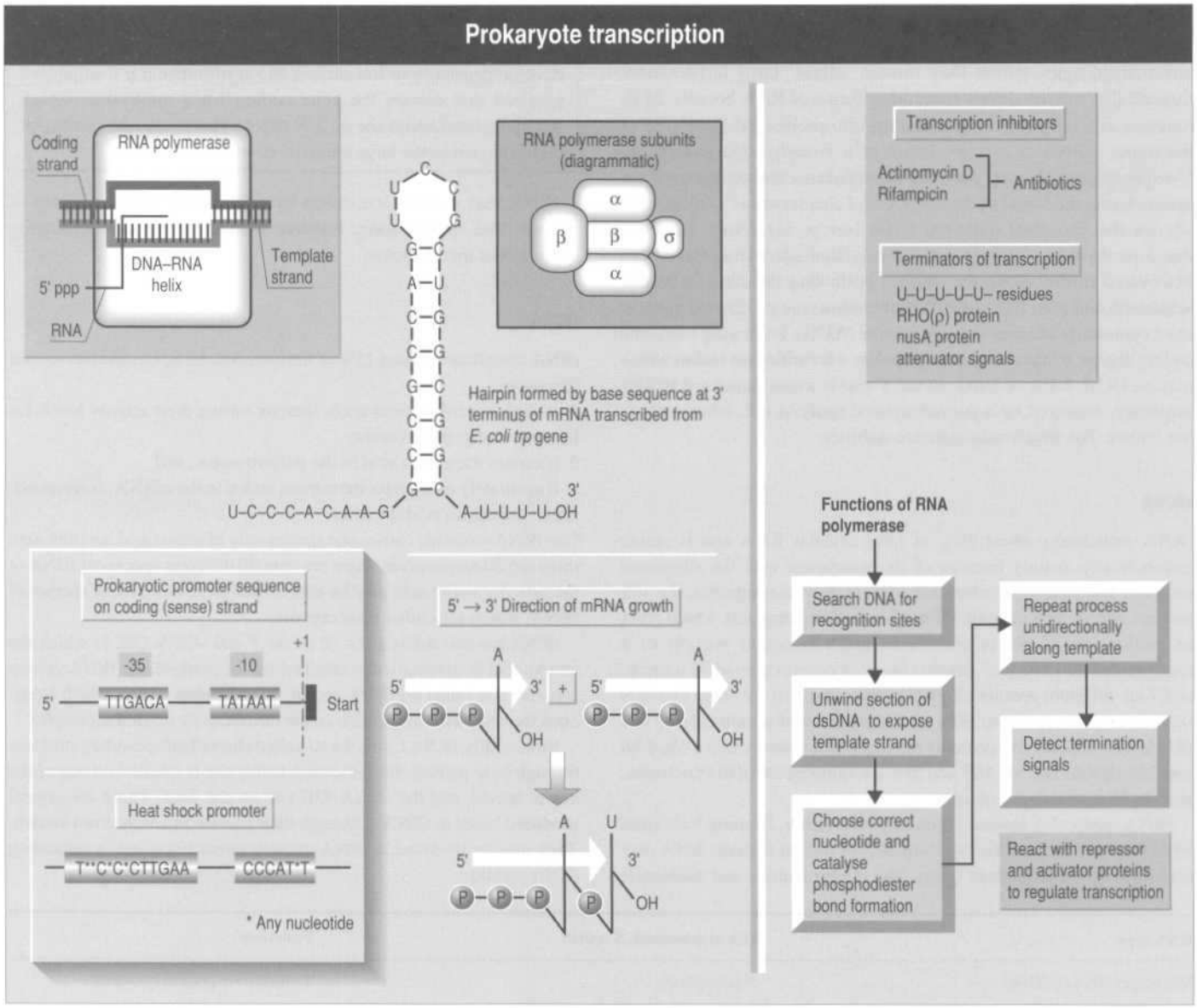

Fig. 12.1

RNA molecules are synthesized in the cell by RNA polymerases, which catalyse the formation of linear polynucleotides whose sequence is complementary to the DNA template. RNA is synthesized from the 5′ to the 3′ end. In prokaryotes, for example *E. coli*, from which much of the information about transcription has been obtained, RNA is synthesized in the cytoplasm and immediately used as a template for protein synthesis.

PROKARYOTES

In *E. coli*, transcription occurs in three phases: (i) initiation; (ii) elongation; and (iii) termination, and is catalysed by a single RNA polymerase, consisting of four subunits: $\alpha_2\beta\beta'\sigma$, and the complete enzyme is termed a **holoenzyme**. The subunits have different functions: σ identifies the promoter site, initiates transcription, and thereafter dissociates from the enzyme; β' binds to the DNA template; and β binds the nucleotides. The functions of the two α-subunits are not known. $\alpha_2\beta\beta'$ is called the core enzyme, and contains the catalytic site.

The σ-subunit quickly enables RNA polymerase to identify the promoter region, without the need for the double helix to be unwound; it also decreases the affinity of the enzyme for non-specific regions of the DNA molecule.

Promoter sites

Promoters are short sequences on the coding strand (as opposed to the template strand) upstream, i.e. on the 5′ side, of the **start site** of RNA synthesis. Two promoters have been identified in *E. coli*: the –35 and the –10 (Pribnow box) sequences, separated from each other, optimally, by 25 bases. If either of these sequences is altered by one base, the promoter loses most if not all activity.

Heat shock prompts *E. coli* to produce several HSPs, which help it to cope with the heat trauma. This occurs because the cell rapidly produces a different type of σ-unit called σ^{32}; σ^{32} recognizes a different set of promoters.

DNA unwinding and RNA synthesis

When RNA polymerase binds to the correct section of the double helix, it unwinds a 17-base-pair-long section of DNA, corresponding to 1.6 turns of the helix. Unlike DNA synthesis, RNA synthesis can begin without a primer. RNA synthesis proceeds from 5′ to 3′ end (similar to DNA synthesis), and at the 5′ end of the newly synthesized RNA is invariably found pppA or pppG (p, phosphate). At the free 3′ end (or terminus) is a free –OH group. The newly formed RNA is called **nascent RNA**; nascent means newborn, but not yet active. The complex consisting of the nascent RNA, the RNA polymerase and the portion of unwound DNA is sometimes called the **transcription bubble**. In *E. coli*, the speed of transcription is around 50 nucleotides incorporated into nascent RNA per second, and during this period the bubble moves 17 nm along the DNA. As the bubble moves along, unwinding the helix, so the helix is rewound at the same rate after the bubble has passed. Unlike DNA polymerase, RNA polymerase does not proofread or alter any errors (i.e. edit) of the newly synthesized RNA. Therefore, the error rate is greater than for DNA synthesis; DNA polymerase may make errors once every 10^{10} bases, whereas RNA polymerase may make errors once every 10^{5} bases.

Termination of transcription

Transcription is terminated when RNA polymerase encounters **stop signals** on the DNA template strand. For example, on the DNA template of the *trp* gene of *E. coli* is a **palindromic** length rich in GC bases, followed by one rich in AT bases. The term 'palindrome' refers to a sequence of bases that is identical when read from left to right, or from right to left, e.g. TAAT or GCCG. The RNA bases corresponding to this stretch of DNA spontaneously assume the so-called **hairpin** structure. After the hairpin, a series of U bases (uridine triphosphate, UTP) are added. Shortly after this addition: (i) RNA polymerase stops; (ii) the nascent RNA dissociates from the bubble; and (iii) the DNA reforms the double helix.

> *Note*: RNA polymerase activity is inhibited by the hairpin structure, and the series of U bases forms an especially weak pairing with the template strand. This facilitates dissociation of RNA from the DNA template.

There is in *E. coli*, a protein called **nusA** protein that terminates transcription. There are also **attenuator sites** on some *E. coli* genes that terminate transcription.

The **ρ (Rho) factor** is an enzyme, found in some phage particles, that hydrolyses ATP in the presence of single-stranded RNA, and which binds newly formed RNA. ρ acts as a terminator of transcription by using the energy from hydrolysis of ATP to move along the nascent RNA towards the bubble, and it pulls RNA away from the bubble when it reaches a certain signal in the RNA molecule itself. As with ρ-independent termination of transcription, the termination signal resides in the RNA and not on the DNA template.

Post-transcription processing

In prokaryotes, there is very little if any post-transcriptional processing of mRNA. But, rRNA and tRNA nascent chains are cleaved and some of their bases are modified.

1 One nascent chain may be cleaved by enzymes called **nucleases** to yield several species of tRNA and rRNA.

2 Another modification is the addition of terminal nucleotides at the 3′ end. For example, tRNA receives the sequence CCA at the 3′ end (see p. 23).

3 Molecules of rRNA may be altered by methylation of bases.

Inhibition of transcription

Transcription can be inhibited using drugs, for example some antibiotics.

Rifampicin reacts with the β-subunit of RNA polymerase to block the first phosphodiester bond between nucleotides. But, if transcription has started, rifampicin is ineffective. Bacteria can develop resistance to rifampicin by producing a mutated β-subunit, when they are termed *rif-r* mutants.

A molecule of **actinomycin D** intercalates between two GC pairs of double-helical DNA in the narrow groove of the helix. Actinomycin D does not bind to the RNA–DNA duplex, nor does it bind to single-stranded DNA or RNA.

13 Transcription II

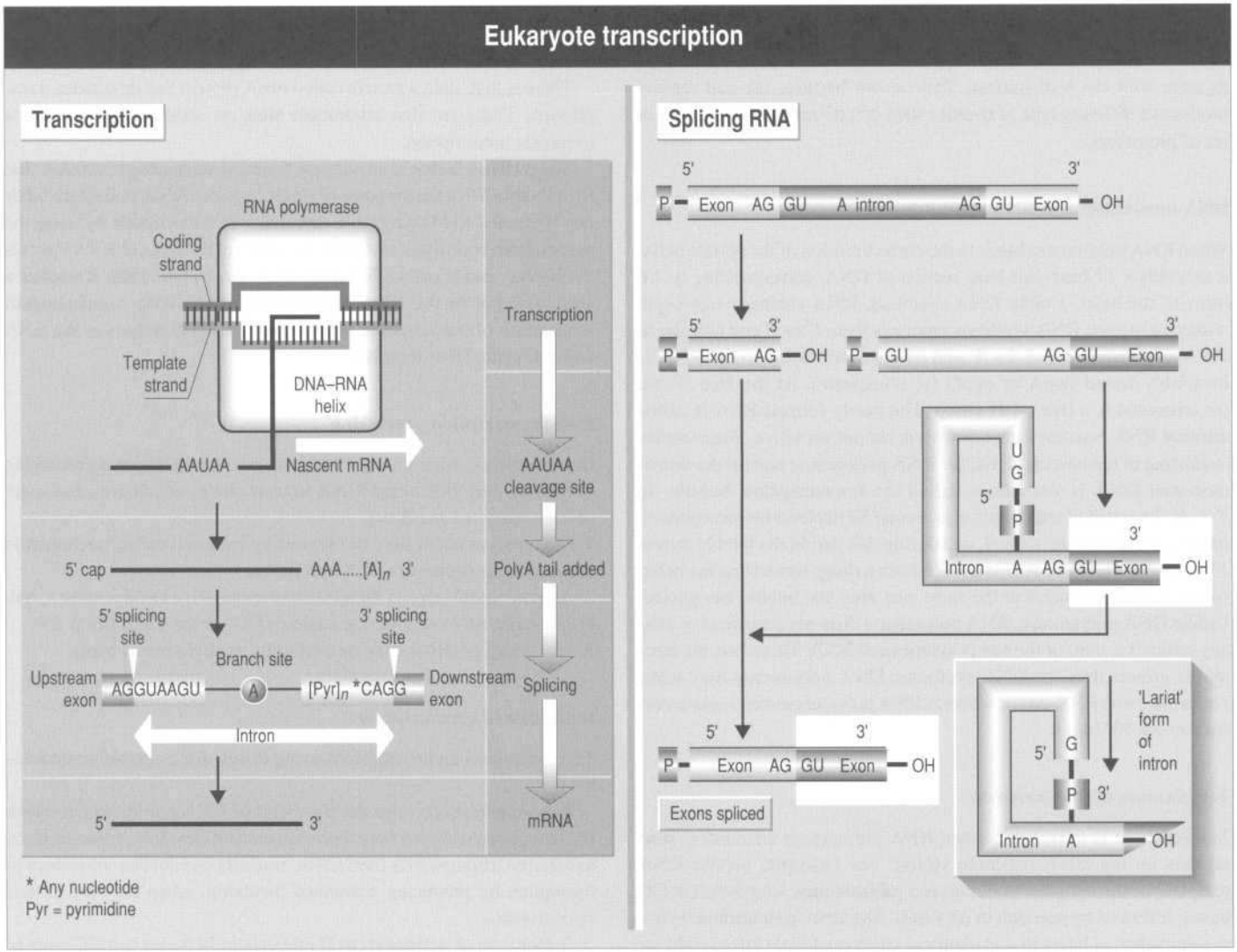

Fig. 13.1

EUKARYOTIC TRANSCRIPTION

Eukaryotic transcription occurs inside the nucleus, and the transcribed mRNA is translated outside the nucleus. In eukaryotes, there are three different types of RNA polymerase, and the nascent RNA formed is extensively processed before being utilized by the cell. In particular, the RNA is **spliced**, and the nature of the splicing depends on the function of the RNA. In eukaryotes, mRNA is derived from heterogeneous nuclear RNA (hnRNA; the primary transcript), which is extensively hydrolysed within the nucleus.

RNA polymerase

Transcription is initiated without the need for a primer, and mRNA is synthesized by addition of nucleoside triphosphates in the direction 5′ → 3′.

Species	Nuclear localization	RNA products
I	Nucleolus	**rRNA**: 5.8S, 18S, 28S
II	Nucleoplasm*	**mRNA**: hnRNA and mRNA precursors
III	Nucleoplasm	**rRNA**: 5S; **tRNA**

* Strongly inhibited by low doses of the fungal toxin α-amanitin. (III is blocked by higher doses of α-amanitin; I is insensitive.)

EUKARYOTIC PROMOTERS

As in prokaryotes, RNA polymerase recognizes promoter sequences upstream of the start site, but polymerases I, II and III recognize different start sites. RNA polymerase II requires three start sites: at –110 is the

CAAT box (CA–CAATC); at –40 is the **GC box** (GGGCGG); at –25 is the **TATA box**, also called the **Hogness box** (TATAAA). The CAAT box aids the binding of polymerase II to the DNA, and the TATA box guides polymerase II to the correct start site. The GC box is most often found on constitutive genes (continuously expressed), rather than on those which are, for example, developmentally regulated.

In addition to promoters, there are **enchancer sequences** which may be thousands of bases away from the start site, either upstream, downstream or both, and may be on either coding or template strands of the DNA. Enhancers on their own, however, are not promoters. Enhancers may confer **specificity** of cellular or organ responses to chemical stimuli. They have sites that bind sets of protein modifiers, both positive or negative in action. Steroid and thyroid hormones, for example, bind to intracellular receptors which in turn bind to enhancers, thus initiating transcription. Only those cells that normally contain the receptors and/or the enhancers will respond to the chemical stimulus.

TRANSCRIPTION FACTORS

Transcription factors are proteins that are required in order for RNA polymerase to recognize promoter sites. For example, genes containing GC boxes require a protein termed **spl**. A transcription factor called **CTF** binds to the CAAT box, and a protein, **B protein**, discovered in *Drosophila*, binds to the TATA box.

The **5′ cap** is added to the 5′ end of eukaryotic mRNA very soon after transcription is initiated. The cap has at least three functions:

1 it protects mRNA from enzyme attack;
2 it is important in subsequent splicing;
3 it enhances translation of the mRNA.

The 5′ cap contains an 'inverted' base, 7-methylguanylate, attached to methylated ribose units.

The **3′ polyA tail** of mRNA has an unknown function. It is not encoded by the gene and is added after cutting of the primary mRNA transcript at the cleavage signal AAUAAA, which is recognized by a specific **endonuclease enzyme**. The polyA tail is not required for transcription, and some species of mRNA do not have a polyA tail.

SPLICING

Splicing is the removal of part, if not most, of the newly synthesized mRNA precursors. Only the mRNA corresponding to DNA **exons** are needed for transcription. Exons are the regions within the eukaryotic genes that are expressed, and in the DNA are separated from each other by **introns**, which are non-coding regions of the DNA. mRNA corresponding to the DNA introns is cut away from the pre-mRNA by specific splicing enzymes. All eukaryotic systems described so far have introns that begin with **5′-GU** and end with **AG-3′**. Eukaryotic DNA contains many non-coding regions of repetitious DNA, which has been called 'junk DNA'.

The **site** of splicing is determined by the 5′ and 3′ **splice sites** and a region of the intron called the **branch site**. The branch site is a short sequence of ribonucleotides which may vary with the cell type and the species. All three splice sites, i.e. 5′, 3′ and the branch site, must be in order for correct splicing. Mutations of the sites can cause disease due to incorrect splicing.

The mechanism of splicing requires: (i) the cutting of the 5′ splice site; (ii) the cutting of the 3′ splice site; (iii) the release of the intron in 'lariat' form (so-called because it resembles the cowboys' lasso); and (iv) the joining of the two exons which are on either side of the spliced intron. The two exons are joined by means of two **transesterification** reactions, in which the free –OH group of one exon is linked to the free 5′ phosphate of the other. While these steps are being carried out, the whole assembly is held together by the splicosome.

The **splicosome** is a relatively large 60S complex made up of the mRNA precursor and three different species of small nuclear RNA (snRNA). Molecules of snRNA have been nicknamed 'snurps' (small cytoplasmic RNA has been nicknamed 'scurps'). Different snurps have different functions in splicing.

Snurp (snRNA)	Function in splicosome
U_1	Binds 5′ splice site
U_2	Binds branch site
U_3	Binds 3′ splice site
U_4–U_6	Assembles the splicosome

Catalytic RNA (self-splicing) was discovered in protozoa, in which rRNA can splice itself in the absence of any protein, i.e. enzymes.

Catalytic RNA is stable, and appears to function like an enzyme. It has been shown to catalyse the cleavage and joining together of other nucleotides. It is therefore both a ribonuclease and an RNA polymerase. Furthermore, it exhibits the features of enzymes, being highly selective for substrates, obeying saturation kinetics and the kinetics of competitive inhibition. Self-splicing RNA is important evolutionarily, since biochemical reactions of DNA and RNA could have taken place even before the evolution of protein enzymes involved in DNA and RNA synthesis.

14 Errors of transcription

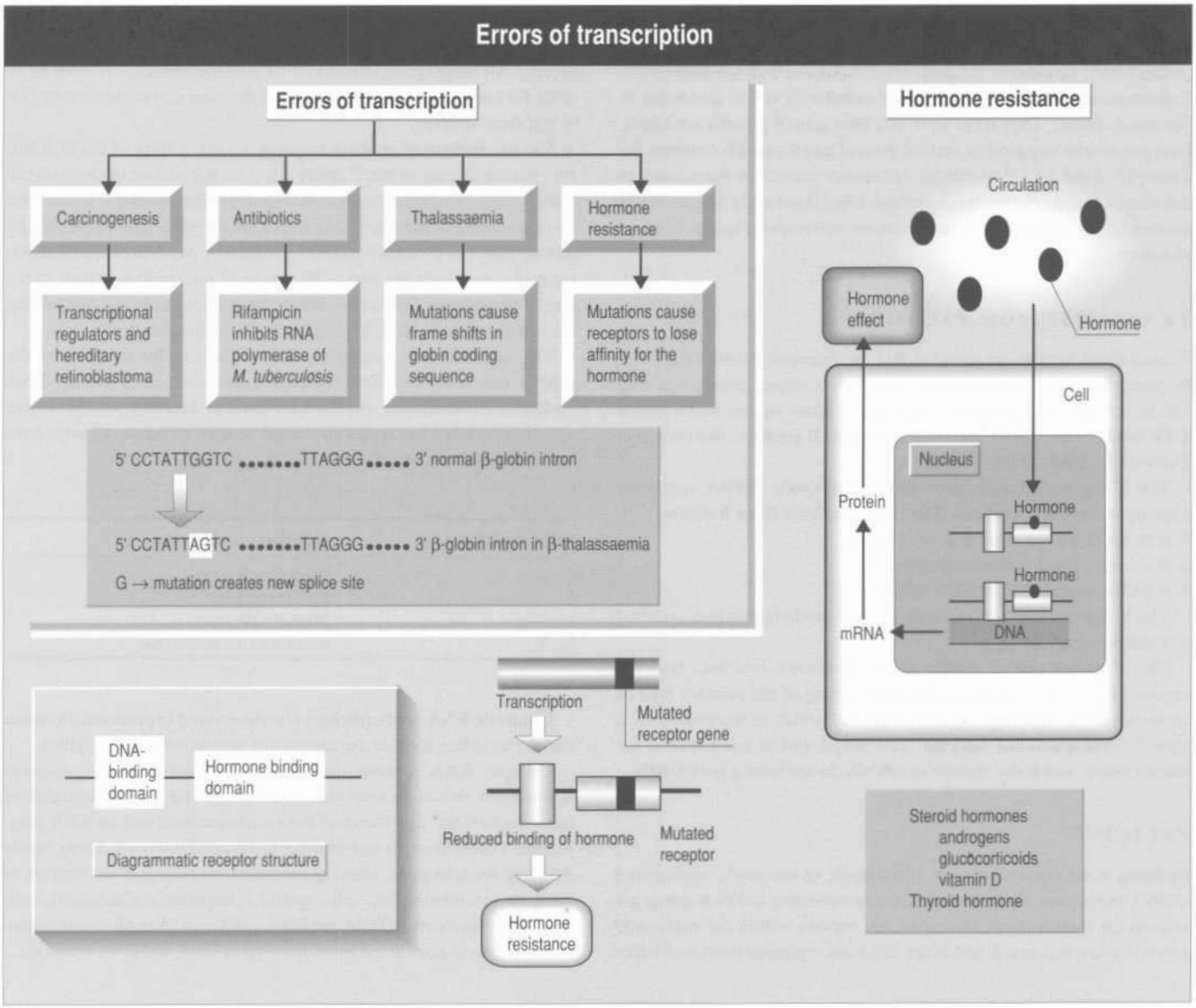

Fig. 14.1

INTRODUCTION

Errors of transcription can result in disease, and transcription is vulnerable to toxic agents such as α-amanitin, and to certain antibiotics. Inappropriate expression of transcription and growth factors can cause cancer. Mutations of the DNA can cause changes in the location of splice sites, with resultant defective functional and structural proteins. These mutations can result in defective receptor proteins for hormones. The vulnerability of transcription to drugs such as fungal toxins and to antibiotics means that these are useful both as research tools and as potential chemotherapeutic agents.

CANCER

The conversion of normal cells into malignant cells is mediated by, for example, viruses, mutagenic chemicals and ionizing radiations. Malignant cells divide out of control and may kill the host organism if not stopped. Malignant cells may start producing their own growth factors and reduce or abolish their susceptibility to negative growth factors.

Transforming growth factors (TGFs) are chemicals that can convert a normal cell to a malignant cell. For example, there is evidence that the hormone progesterone increases the expression of a growth factor called TGF-α, which may mediate the growth of breast cancer cells. In

another type of cancer, **hereditary retinoblastoma**, a rare cancer of the retina in infants, the cancer cells lack the ***Rb* gene**, which in normal cells may be a negative regulator of transcription.

ANTIBIOTICS

The antibiotic **rifampicin**, a semi-synthetic derivative of rifamycin, which was isolated from *Streptomyces*, blocks the formation of the first phosphodiester bond in RNA synthesis (see also p. 25). Chain elongation, however, is not affected. The bacterium that causes tuberculosis is *Mycobacterium tuberculosis*, and it is highly resistant to most antibiotics. But, it is susceptible to rifampicin, which is not as toxic to mammalian RNA polymerase. Rifampicin is used together with an antimetabolite, isoniazid, to treat the disease.

THALASSAEMIA

Thalassaemia (Cooley's anaemia) is an hereditary blood disease, especially prevalent in Africa, Asia and the Mediterranean countries, in which the globin part of the Hb molecule is altered, with decreased synthesis of the α- or β-chains. The geographical distribution of the disease coincides largely with that of malaria. Patients with the disease are anaemic because the Hb molecule cannot function normally. There are two types: (i) thalassaemia major, when the disease is inherited from both patients (i.e. there are two copies of the abnormal gene), and the patient is badly affected; and (ii) thalassaemia minor, when the defective gene is inherited from one parent, and the child has mild symptoms, or is free of symptoms. In thalassaemia major, affected individuals are anaemic and have swollen spleens and abnormalities of the bone marrow. They may require blood transfusions, which can result in an iron overload.

The disease can also be classified as **α-thalassaemia** or **β-thalassaemia**. α-Thalassaemia is a deficiency of the α-globin chains due to unequal crossover between adjacent α-alleles. β-Thalassaemia, which is rarer, can be due to one of a number of different mutations. For example, there is a G → A mutation of the β-globin coding region, resulting in a new splicing site; thus, a frame shift occurs. The promoter may be mutated, or the nascent mRNA prematurely released from the template strand or the splicing may be incorrect.

HORMONE RESISTANCE

Steroid and thyroid hormones, and vitamin D all act on their target cells by combining with intracellular receptors. The hormone–receptor complex binds to specific sequences upstream of transcription start sites and triggers transcription. The receptor protein has two important domains: one that binds the hormone and another that binds the DNA. This knowledge, and the knowledge of the structure of the genes that code for these receptors, has led to an understanding of the causes of several diseases characterized by a lack of responsiveness to these hormones.

For example, **androgen resistance** is the failure to respond to the male sex hormone testosterone, and to its powerful androgenic metabolite 5α-dihydrotestosterone (DHT). The disease may be caused by: (i) the complete or partial deletion of the gene that codes for the androgen receptor; (ii) splicing defects; (iii) premature termination codons; and (iv) amino acid substitutions resulting from mutated base substitutions. These mutations usually mean a loss of affinity of the hormone for the receptor. In the case of androgen resistance, most of the amino acid substitutions occur in the steroid-binding domain of the receptor.

Thyroid resistance, too, can be caused by mutations of the gene that codes for the thyroid receptor. These patients have retarded growth and lesions of bone, despite having high plasma levels of thyroid hormone. The tissues are not 'seeing' the hormone. Two distinct thyroid receptors, α and β, are encoded by two distinct genes, the *α*- and *β*-genes. In cases of generalized resistance to thyroid hormone, it has been discovered that most mutations occur within a tightly circumscribed region of the *β*-gene.

Glucocorticoid resistance can also be explained, in part, by mutations of the receptor, and this has important implications for therapy since glucocorticoids such as **prednisolone** are used extensively as anti-inflammatory agents in connective tissue diseases, and as immunosuppressants in autoimmune diseases. Thus, patients who are resistant to glucocorticoids will not respond to treatment with them. In some cases of glucocorticoid resistance, **point mutations** have been discovered, resulting in the substitution of single amino acids. This results in a reduced affinity for the hormone.

15 Protein synthesis I

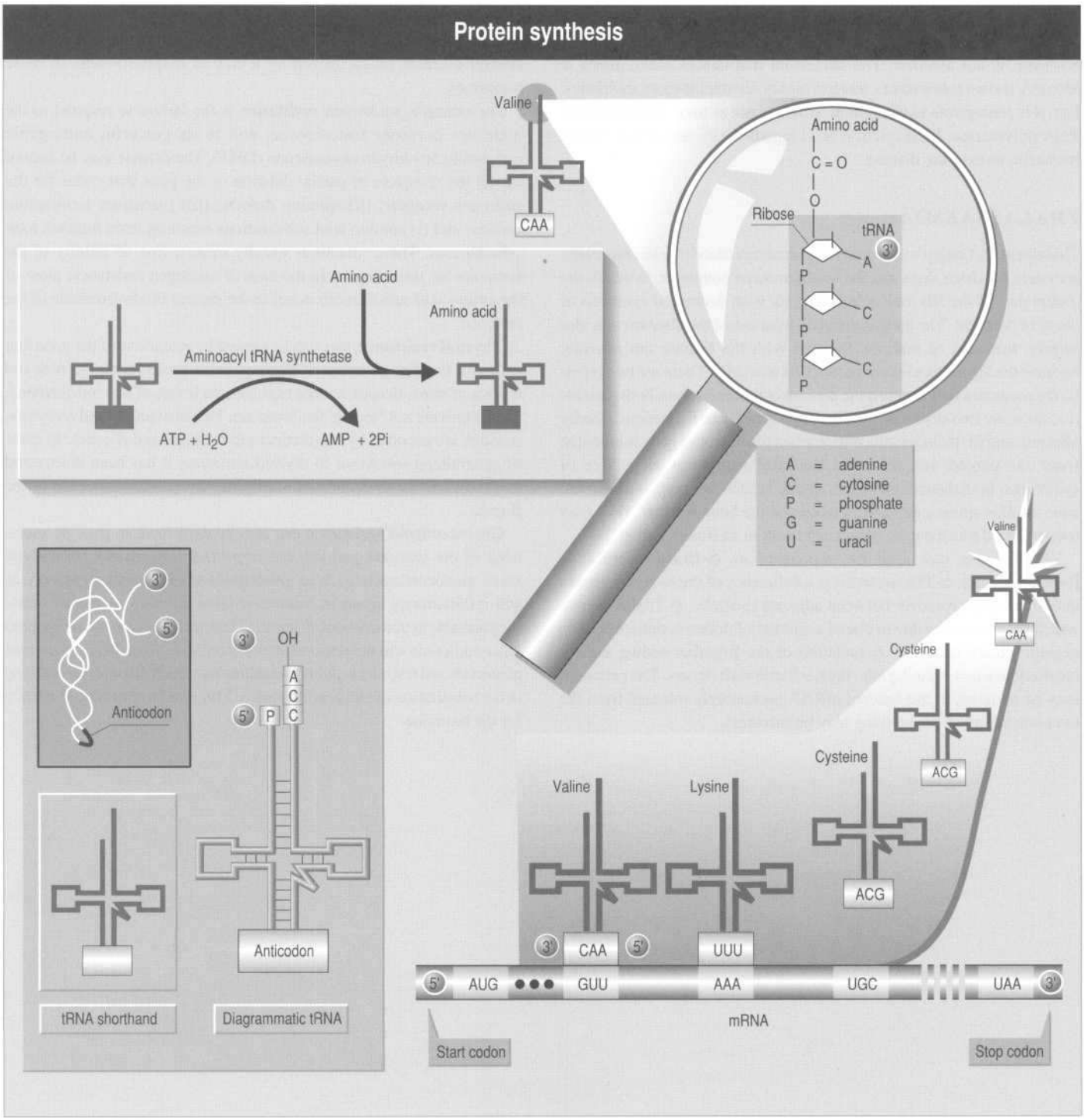

Fig. 15.1

INTRODUCTION

Proteins are synthesized by a process called **translation**. This is literally the translation of the sequence of bases in the DNA, and therefore in the corresponding mRNA, into a sequence of amino acids joined together to form the polypeptide chain, the primary structure of proteins. The primary

structure will determine, in turn, the secondary and tertiary structure, and therefore the function of the protein. Protein synthesis, as in the case of DNA and RNA synthesis, takes place in three stages: (i) initiation; (ii) elongation; and (iii) termination. Amino acids are activated through linkage to their tRNA molecules in a reaction catalysed by an enzyme, an **aminoacyl–tRNA synthetase** specific for the particular amino acid. In some cases, there may be more than one tRNA and enzyme serving a particular amino acid. Polypeptide chains are elongated in the direction amino to carboxyl group ($NH_3^+ \rightarrow COO^-$), by the polysome 'assembly line'. An initiator tRNA triggers synthesis by binding to a site on the ribosome; elongation begins with the binding of another tRNA to another site; and termination occurs when a protein release factor 'reads' a stop signal on the mRNA.

THE GENETIC CODE

At least 20 different amino acids occur in proteins, but there are only four bases in DNA and mRNA, and it was discovered that a sequence of three bases, called a **codon**, codes for each amino acid. The number of possible triplet codons derived from four bases is $4^3 = 64$. Experiments have determined that of these, 61 combinations code for amino acids. Although some amino acids, for example tryptophan, have only one codon (UGG), others, for example serine, have as many as six (see below). Nevertheless, the genetic code is specific and unambiguous; one codon codes for one amino acid only.

	Middle base of codon				
5′ end of codon	**U**	**C**	**A**	**G**	**3′ end of codon**
U	Phe	Ser	Tyr	Cys	U
U	Phe	Ser	Tyr	Cys	C
U	Leu	Ser	STOP	STOP	A
U	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
C	Leu	Pro	His	Arg	C
C	Leu	Pro	Gln	Arg	A
C	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
A	Ile	Thr	Asn	Ser	C
A	Ile	Thr	Lys	Arg	A
A	Met*	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
G	Val	Ala	Asp	Gly	C
G	Val	Ala	Glu	Gly	A
G	Val	Ala	Glu	Gly	G

* The codon for methionine (AUG) is a START codon for translation (see p. 30). For definitions of abbreviations see Abbreviations, p. 108.

Because there is more than one codon for amino acids, the code is referred to as **degenerate**. The code is also virtually **universal**, since substantially the same code has been found in all living organisms.

Codon	Mitochondrion	Normal meaning
CUA	Thr	Leu
AUA	Met	Ile
UGA	Trp	STOP

Mitochondria provide the only known exception to the rule of universality, since some codons which normally have one meaning are different in mitochondria.

The START signal for protein biosynthesis in eukaryotes is AUG, the codon for methionine. Methionine is therefore the first amino acid of the protein chain. The STOP signals, UGA, UAA and UAG, do not code for any amino acids, and may sometimes be referred to as **nonsense codons**.

THE 'WOBBLE' HYPOTHESIS

The 'wobble' hypothesis is an explanation for the fact that: (i) one tRNA can read more than one codon; and (ii) many of the codons can be read by more than one species of tRNA. Note, however, that one tRNA can carry only one species of amino acid; (i) and (ii) may be possible because the stringency requirements for base pairing normally stipulated by the Crick–Watson base-pairing rule do not apply, allowing base pairing between the third position of the codon on the mRNA reading $5' \rightarrow 3'$, and the first position of anticodon on the tRNA $5' \rightarrow 3'$. This makes the nucleotides literally wobble, changing the geometry of the codon–anticodon interaction, and allowing G–U base pairs to form. Wobble, and therefore the relaxation of stringency of base pairing between anticodon and codon in protein synthesis, may also be induced by the presence of modified nucleotides at or near the first position of the anticodon in some tRNA species. An important modified nucleotide for wobble is **inosinic acid** (I), which can form base pairs with A, C or U in the third position of mRNA codon. The relaxation of stringency of the base-pairing rule during wobble means that, theoretically, many of the different codons could be read by relatively few tRNA anticodons. In nature, however, this possibility has not necessarily been exploited, since most cells contain almost as many species of tRNA as there are of amino acids.

THE AMINOACYLATION REACTION

Amino acids are activated for protein synthesis, i.e. raised to a higher energy level for participation in protein synthesis, through coupling to tRNA in a reaction catalysed by a specific **aminoacyl–tRNA synthetase**. There are 20 different enzymes; one for each amino acid. The enzyme can: (i) recognize its specific amino acid–tRNA complex; and (ii) proofread the complex after it has bound it, and hydrolyse the incorrectly bound complex. It is not known with certainty how the enzyme proofreads the complex.

16 Protein synthesis II

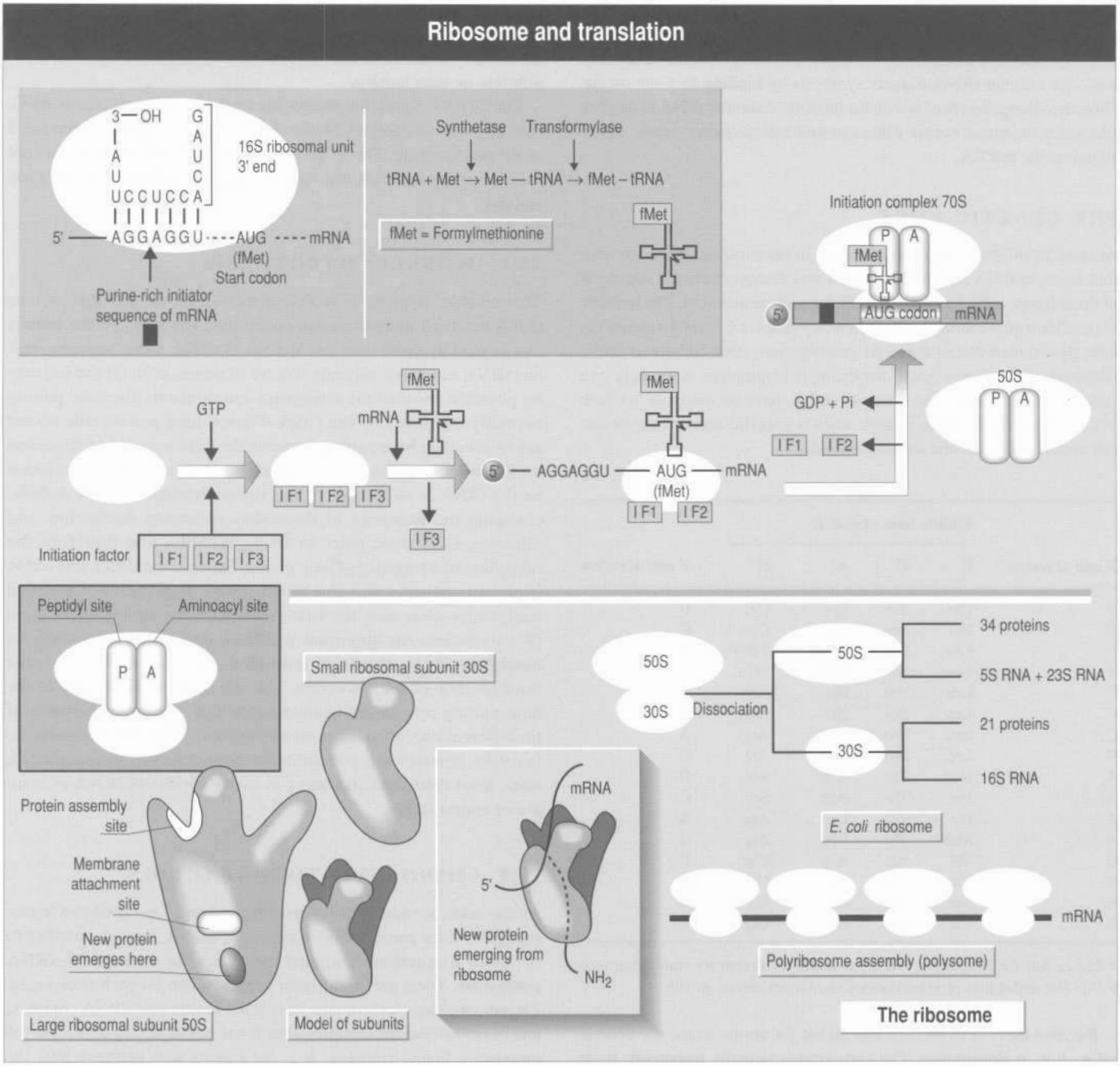

Fig. 16.1

INTRODUCTION

Protein synthesis is effected by the **ribosomes**, which are ribonucleoprotein particles. Ribosomes consist of two main subunits which fit together to form a structure that coordinates the assembly of proteins.

THE RIBOSOME

Bacterial ribosomes

In bacteria such as *E. coli*, whose ribosome has been extensively studied, the ribosome is an irregularly shaped ribonucleoprotein particle having a sedimentation coefficient of 70S, a diameter of about

20 nm and a mass of approximately 2700 kDa. The ribosome can be **dissociated** in the laboratory into a large 50S subunit and a smaller 30S subunit. These subunits can be dissociated further into the proteins and rRNA which make them up. The structure and function of the ribosome depends principally on the way the constituent rRNA molecules fold and associate with the proteins. Each bacterial cell has about 20 000 ribosomes, which constitute roughly 25% of the mass of the cell. Viewed under the electron microscope, ribosomes may be observed free in the cytoplasm.

Eukaryotic ribosomes

In mammalian cells, the structure and function of the ribosome is very similar to that of the *E. coli* ribosome, although its composition differs. The mammalian ribosome is 80S, having a mass of 42 kDa, and dissociates into 60S and 40S subunits. The 40S subunit has an 18S rRNA molecule and about 30 associated proteins, and the 40S subunit is made up of 5S, 5.8S and 28S rRNA, and about 45 associated proteins. Viewed under the electron microscope, ribosomes may be observed free in the cytoplasm or firmly attached to the ER. As a general rule, free ribosomes synthesize proteins for use in the cytoplasm, while membrane-bound ribosomes synthesize proteins for export from the cell, or for making membranes.

TRANSLATION

For protein synthesis, several ribosomes can bind simultaneoulsy to a mRNA molecule to form a polyribosome, or **polysome**. There may be up to one ribosome attached every eight nucleotides along the mRNA. The individual ribosomes of the polysome work independently of each other, and each produces a complete polypeptide chain. Polypeptide chains are synthesized in the direction $NH_2 \rightarrow COOH$, and the mRNA is read from 5′ to 3′. In eukaryotes, mRNA is exported from the nucleus to the cytoplasm, where protein synthesis occurs, whereas in prokaryotes such as *E. coli*, a mRNA molecule may be translated even while it is being transcribed.

Initiation in prokaryotes

Protein synthesis in prokaryotic cells is initiated by the combination, in the cytosol (see p. 30), of the free, smaller subunit of the ribosome with an initiator **tRNA** molecule that carries the amino acid methionine (Met). The initiator tRNA: (i) is linked with Met through a reaction catalysed by the corresponding aminoacyl–tRNA synthetase; and (ii) the Met attached to the tRNA is formylated by a transformylase ('formyl' is derived from formic acid, HCOOH). The initiator tRNA is therefore expressed as $tRNA_f$. The Met attached to $tRNA_f$ does not usually form part of the final polypeptide chain, but is removed after the protein has been made.

Note: Met attached to an intiator $tRNA_f$ can be formylated, but it cannot be formylated if attached to the species of tRNA that carries a molecule of Met destined for permanent inclusion in the protein. In this latter case, the tRNA that carries Met is designated $tRNA_{Met}$, or $tRNA_m$.

The free, smaller 30S subunit binds three initiation factors (IFs), IF1, IF2 and IF3. IF2 binds GTP, and also recognizes the Met–$tRNA_f$ complex. The reaction between IF2 and GTP also enables the mRNA molecule to be bound by the 30S subunit. IF3 dissociates as the 30S–Met–$tRNA_f$ complex is formed; GTP is hydrolysed as the larger 50S subunit joins the complex; and IF1 and IF2 dissociate. The Met–$tRNA_f$ molecule is located in the **P site** of the ribosome, and the **A site** is empty at the start of the elongation phase of protein synthesis. The resultant 70S complex is termed the **initiation complex.**

The Shine–Dalgarno sequence

The Met–$tRNA_f$ molecule has the anticodon UAC, which binds non-covalently to the initiation codon AUG on the mRNA molecule. AUG codes for Met. In *E. coli*, the site of initiation is situated at a purine-rich region of the mRNA upstream from the AUG start codon, and this purine-rich region is called the **Shine–Dalgarno** sequence (e.g. for *E. coli lacI*, the sequence is 5′-AGGAGG-3′). This sequence pairs with a complementary sequence very near the 3′ end of the 16S rRNA of the 30S subunit.

Therefore, protein synthesis will be initiated where the anticodon of the $tRNA_f$ binds the initiator AUG codon, and where the mRNA pairs with the complementary sequence at the 3′ end of the 16S rRNA molecule.

Initiation in eukaryotes

The mechanism of initiation of translation in eukaryotic cells is fundamentally the same as in prokaryotes, but with some differences.

1 The initiating tRNA is Met and **not** formylmethionine (fMet). The tRNA carrying Met is termed $tRNA_i$.

2 There are many more IFs (at least nine are known, and doubtless more will be discovered):

(a) eIF2 binds GTP and escorts tRNA to the 40S complex;

(b) cap-binding proteins bind to the mRNA 5′ cap, and eIF3 binds to the AUG start codon nearest to the cap, using energy provided by eIF4, which in turn derives its energy from ATP;

(c) Met–$tRNA_i$ binds to the start codon AUG and eIF5 causes eIF2 to hydrolyse GTP, which results in the release of eIF2 and eIF3 from the initiation complex;

(d) the 60S subunit is attached to form the complete initiation complex.

Note: in eukaryotes, there is one start codon only — AUG — and no Shine–Dalgarno purine-rich sequence. The 40S complex attaches itself to the mRNA at the 5′ end and uses energy (ATP see (2a) above) to move towards the 3′ end until it finds the AUG start signal.

17 Protein synthesis III

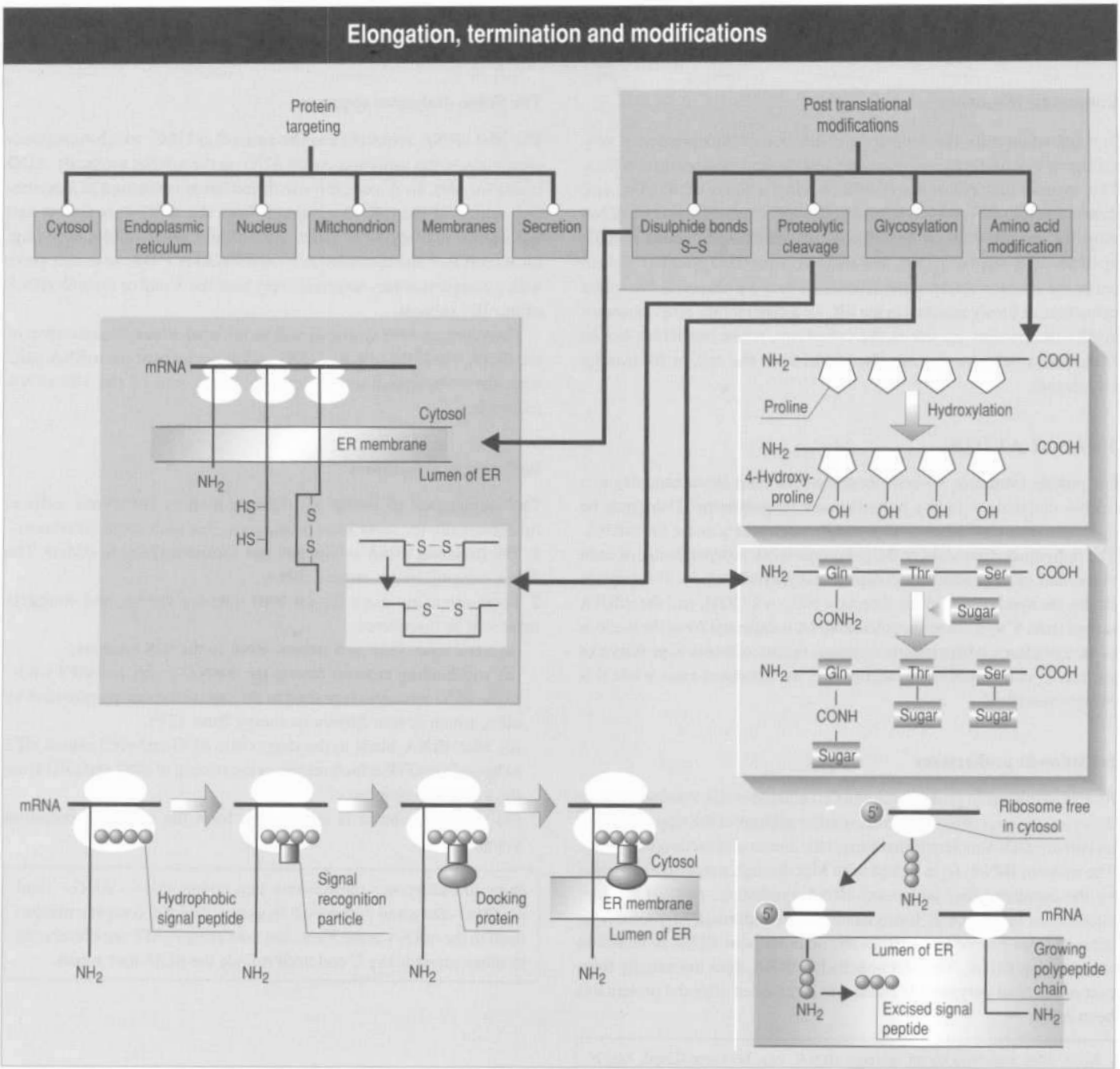

Fig. 17.1

INTRODUCTION

After the initiation of protein synthesis, the peptide chain is elongated, and when the appropriate signal is reached, elongation is terminated. The peptide is released from the ribosomal apparatus, and may be subject to post-translational modification. The protein will be modified depending on whether it is destined for other cellular organelles, such as mitochondria, the nucleus or lysosomes, or to structural units, such as the cell membrane. It may be modified for export. Errors in post-translational modification may result in disease.

ELONGATION

The amino acid to be added to the initial methionine is delivered, attached to its tRNA, to the A site of the ribosome by an **elongation factor**. It is critically important that the correct aminoacyl–tRNA is in the A site, because the cell cannot excise a 'wrong' amino acid once it has been added to the chain. The 'proofreading' is done by an elongation factor, **EF-TU** in prokaryotes. EF-TU binds GTP, and this enables it to bind the aminoacyl–tRNA and bring it to the A site. The amino acid cannot be added to the chain until EF-TU leaves the complex, and EF-TU cannot leave until it has hydrolysed its GTP to GDP. Therefore, there is time, while GTP is being hydrolysed, and while the EF-TU is leaving the complex, for an incorrect aminoacyl–tRNA to leave the complex. In eukaryotes, the proofreading elongation factor is **EF1**, whose subunit EF1α forms a complex with the aminoacyl–tRNA.

Elongation has two main steps, **peptide bond formation** and **translocation**. Peptide bond formation is catalysed by peptidyl transferase, which attaches the carbonyl atom of the P site aminoacyl–tRNA to the α-amino group of the amino acid of the aminoacyl–tRNA in the A site. The A site must now be emptied for the next aminoacyl–tRNA, and so the reading frame is shifted three bases along the mRNA until the next codon is at the A site. This movement is driven by another elongation factor, or **translocase**, which utilizes GTP hydrolysis for energy. In prokaryotes, the factor is EF-G, and in eukaryotes it is **EF2**. During the process, the tRNA–OH whose amino acid has been removed is shifted to a so-called 'exit site' on the ribosome, from where it is released back into the cytoplasm. In eukaryotes, a number of other elongation factors have been identified but their exact roles are not clear.

Termination

When a termination or STOP codon is reached, no aminoacyl–tRNA can be accepted in the A site. Instead, the codon is bound by a release factor–GTP complex **RF–GTP**. As a result, peptidyl transferase switches to being a hydrolase, adding H_2O to the carbonyl end of the peptide chain. RF hydrolyses its GTP to GDP, and undergoes a conformational change. These changes provide the energy to dissociate the elongation complex into the constituent mRNA, polypeptide chain and the ribosomal subunits. In prokaryotes, three release factors have been described: RF1, RF2 and RF3. RF1 and RF2 recognize different STOP codons, and RF3 potentiates the actions of RF1 and RF2.

Prokaryote–eukaryote protein synthesis summary

Feature	Prokaryotes	Eukaryotes
Large ribosome subunit	50S	60S
Small ribosome subunit	30S	40S
Whole ribosome	70S	80S
Large subunit rRNA	5S 23S	5S 5.8S 28S
Small subunit RNA	16S	18S
Large subunit protein numbers	34 proteins	50 proteins
Small subunit protein numbers	21 proteins	34 proteins
Initiation factors	IF1 IF2 IF3	eIF2 eIF3 eIF4a eIF4b eIF4c eIF5 eIF6 cap-binding protein
Initiating aminoacyl–tRNA	fMet–tRNA	Met–tRNA
Pre-start purine-rich sequence	Shine–Dalgarno	None
Elongation factors	EF-G (translocase)	EF1 EF2 (translocase)
Release factors	RF1 RF2 RF3	RF

Post-translational modifications

The amino acid sequence and conformational shape of a protein will determine its fate, whether it is to be targeted to a particular site, or to be a substrate for modifying enzymes. They will also determine its half-life.

Eukaryote ribosomes that produce lysosomal proteins, membrane proteins and proteins for export are bound to the ER. During protein synthesis in the cytosol, a **signal sequence** rich in hydrophobic residues, such as phenylalanine, is produced near the amino terminus. The sequence is recognized by a ribonucleoprotein termed **signal recognition particle (SRP)**, which binds to the ribosome, enabling it to bind to the surface of the ER at a 'docking' protein called **SRP receptor**. The ribosome interlocks with two ER membrane translocation proteins called **Riophorin I** and **II**, which drive the elongating peptide chain through the ER membrane into the lumen of the ER. Once inside the lumen, the signal sequence is excised.

Inside the lumen of the ER, several modifications may occur: (i) the protein may be crosslinked by disulphide bonds; (ii) part of the chain may be excised by proteolysis, for example the removal of an inactive portion of several prohormones, such as the conversion of proinsulin to insulin, for export; and (iii) proteins may be glycosylated. Glycosylation of proteins serves three main purposes: (i) changes their physical properties, for example solubility, size and stability; (ii) the carbohydrate addition is an important component of a membrane protein which has to recognize other proteins or cells; and (iii) enables the protein to be targeted to specific cellular sites.

In the lumen, oligosaccharides may be carried to the growing peptide chain by a lipid carrier, **dolichol phosphate**, which is located on the luminal surface of the ER membrane. The oligosaccharide attached to the dolichol phosphate becomes attached to an asparagine (Asn) residue on the peptide, in a reaction catalysed by a **glycosyltransferase**. Glycosylation may take place after the protein has moved through the ER to the **Golgi apparatus**, where it may also be packaged into vesicles for exocytosis if it is for export.

Proteins may be modified during translation (cotranslationally) through the alteration of amino acids. For example, in collagen synthesis, proline is hydroxylated to hydroxyproline.

TARGETING

To the mitochondrion

Proteins targeted to specific cellular sites need special signals. Most proteins destined for the mitochondrion are released into the cytoplasm by free ribosomes, and their signal will differ depending on whether they are targeted to the mitochondrial membranes or to the mitochondrial matrix. Once inside the mitochondrion, the targeting signal is usually cleaved from the protein.

To the nucleus

Most proteins destined for the cell nucleus have specific signal sequences rich in arginine and lysine sequences. The localization of histones to the nucleus appears to be mediated by a protein termed **nucleoplasmin**.

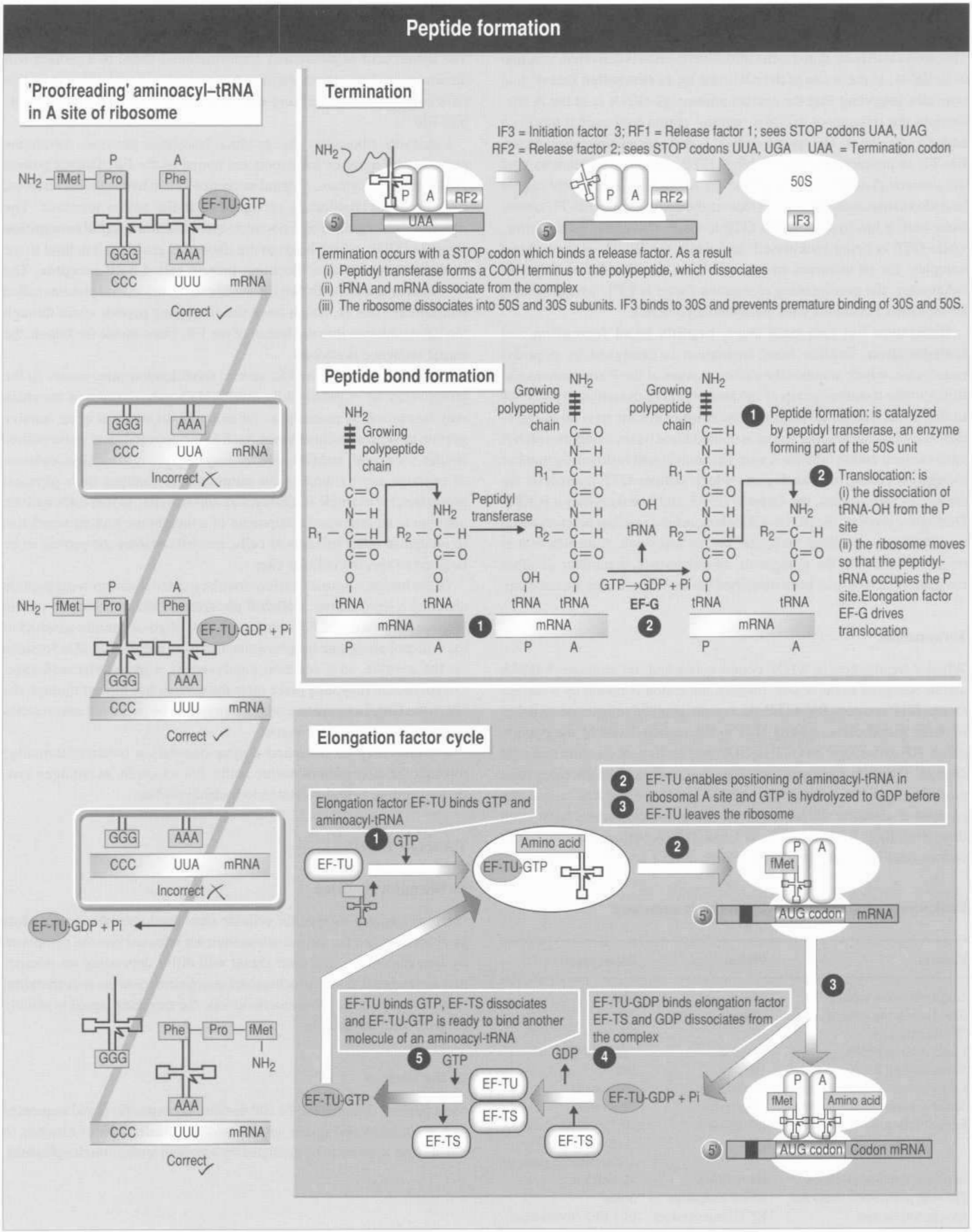
Peptide formation
'Proofreading' aminoacyl–tRNA in A site of ribosome
P
A
NH2 – fMet – Pro
Phe
EF-TU·GTP
GGG
AAA
CCC UUU mRNA
Correct
GGG AAA
CCC UUA mRNA
Incorrect
NH2 – fMet – Pro
Phe
EF-TU·GDP + Pi
GGG AAA
CCC UUU mRNA
Correct
GGG AAA
CCC UUA mRNA
Incorrect
EF-TU·GDP + Pi
GGG
Phe – Pro – fMet
NH2
AAA
CCC UUU mRNA
Correct
Termination
IF3 = Initiation factor 3; RF1 = Release factor 1; sees STOP codons UAA, UAG
RF2 = Release factor 2; sees STOP codons UUA, UGA UAA = Termination codon
NH2
P A RF2
5'
UAA
50S
IF3
Termination occurs with a STOP codon which binds a release factor. As a result
(i) Peptidyl transferase forms a COOH terminus to the polypeptide, which dissociates
(ii) tRNA and mRNA dissociate from the complex
(iii) The ribosome dissociates into 50S and 30S subunits. IF3 binds to 30S and prevents premature binding of 30S and 50S.
Peptide bond formation
NH2
Growing polypeptide chain
Peptidyl transferase
OH
tRNA
GTP→GDP + Pi
EF-G
mRNA
P A
1 Peptide formation: is catalyzed by peptidyl transferase, an enzyme forming part of the 50S unit
2 Translocation: is
(i) the dissociation of tRNA-OH from the P site
(ii) the ribosome moves so that the peptidyl-tRNA occupies the P site.Elongation factor EF-G drives translocation
Elongation factor cycle
Elongation factor EF-TU binds GTP and aminoacyl-tRNA
1 GTP
EF-TU
Amino acid
EF-TU·GTP
2 EF-TU enables positioning of aminoacyl-tRNA in ribosomal A site and GTP is hydrolyzed to GDP before
3 EF-TU leaves the ribosome
fMet
5'
AUG codon mRNA
EF-TU binds GTP, EF-TS dissociates and EF-TU-GTP is ready to bind another molecule of an aminoacyl-tRNA
EF-TU-GDP binds elongation factor EF-TS and GDP dissociates from the complex
5 GTP
GDP
4
EF-TU
EF-TS
EF-TU·GTP
EF-TU·GDP + Pi
Amino acid
AUG codon Codon mRNA

Fig. 17.2

PATHOPHYSIOLOGY OF PROTEIN MODIFICATION

Several diseases and abnormal protein status situations arise from errors in the post-translational modification of proteins, including errors in the targeting process.

Familial hyperproinsulinaemia is an autosomal dominant state in which the individual has approximately the same amount of proinsulin and insulin in the circulation. Those with the condition may exhibit none of the symptoms of diabetes, and glucose metabolism is apparently normal, despite the presence of very high levels of proinsulin in the blood. The cause of the condition is not known with certainty, but may involve point mutations in the proinsulin molecule, which prevents the action of the protease enzymes which splice proinsulin.

I-cell disease is one of a number of related disorders arising from aberrations in the targeting of lysosomal enzymes. Other diseases are **mucolipidosis II** and **III**. The lesion is a deficiency in the enzyme which catalyses the transfer of *N*-acetylglucosamine phosphate to polysaccharide moieties of proteins which are targeted to the lysosome. These proteins are secreted into the bloodstream, and are found in high concentrations in a number of body fluids. The defect is usually present at birth and manifested by skeletal abnormalities, facial coarsening of features and psychomotor retardation. Patients usually die before 7–8 years of age.