

Fundamentals of Medicinal Chemistry

Gareth Thomas

University of Portsmouth, UK



Fundamentals of Medicinal Chemistry

Fundamentals of Medicinal Chemistry

Gareth Thomas

University of Portsmouth, UK



Copyright © 2003 by John Wiley & Sons Ltd,
The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England
National 01243 779777
International (+44) 1243 779777
e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on <http://www.wiley.co.uk> or <http://www.wiley.com>

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London, UK W1P 9 HE, without the permission in writing of the publisher.

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158-0012, USA
Wiley-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany
John Wiley & Sons (Australia) Ltd, 33 Park Road, Milton,
Queensland 4064, Australia
John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 0512
John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books

Library of Congress Cataloging-in-Publication Data

Thomas, Gareth, Dr.

Fundamentals of medicinal chemistry / Gareth Thomas.

p. cm.

Includes bibliographical references and index.

ISBN 0-470-84306-3 (cloth : alk. paper) ISBN 0-470-84307-1 (paper : alk. paper)

1. Pharmaceutical chemistry. I. Title.

RS403.T446 2003

615'.19 — — dc21

2003014218

British Library Cataloguing in Publication Data

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

ISBN 0-470 84306 3 (Hardback)

ISBN 0-470 84307 1 (Paperback)

Typeset in 11/14pt Times by Kolam Information Services Pvt. Ltd, Pondicherry, India

Printed and bound in Great Britain by Antony Rowe Ltd

This book is printed on acid-free paper responsibly manufactured from sustainable forestry, in which at least two trees are planted for each one used for paper production.

Contents

Preface	xi
Acknowledgements	xii
Abbreviations/Acronyms	xiii
1 Biological Molecules	1
1.1 Introduction	1
1.2 Amino acids	1
1.2.1 Introduction	1
1.2.2 Structure	3
1.2.3 Nomenclature	3
1.3 Peptides and proteins	4
1.3.1 Structure	6
1.4 Carbohydrates	10
1.4.1 The structure of monosaccharides	11
1.4.2 The nomenclature of monosaccharides	14
1.4.3 Glycosides	15
1.4.4 Polysaccharides	17
1.4.5 The nomenclature of polysaccharides	18
1.4.6 Naturally occurring polysaccharides	18
1.5 Lipids	20
1.5.1 Introduction	20
1.5.2 Fatty acids	20
1.5.3 Acylglycerols (glycerides)	21
1.5.4 Steroids	21
1.5.5 Terpenes	22
1.5.6 Phospholipids	23
1.5.7 Glycolipids	25
1.6 Nucleic acids	26
1.6.1 Introduction	26
1.6.2 DNA, structure and replication	28
1.6.3 Genes and the human genome project	30
1.6.4 RNA, structure and transcription	31
1.6.5 Classification and function of RNA	33
1.7 Questions	34
2 An Introduction to Drugs and their Action	37
2.1 Introduction	37
2.2 What are drugs and why do we need new ones?	37
2.3 Drug discovery and design, a historical outline	39
2.4 Sources of drugs and lead compounds	43

2.4.1	Natural sources	43
2.4.2	Drug synthesis	45
2.4.3	Market forces and 'me-too drugs'	45
2.5	Classification of drugs	45
2.6	Routes of administration, the pharmaceutical phase	46
2.7	Introduction to drug action	49
2.7.1	The pharmacokinetic phase	49
	Absorption	49
	Distribution	50
	Metabolism	51
	Elimination	51
2.7.2	Bioavailability of a drug	53
2.7.3	The pharmacodynamic phase	53
2.8	Questions	55
3	An Introduction to Drug Discovery	57
3.1	Introduction	57
3.2	Stereochemistry and drug design	59
3.2.1	Structurally rigid groups	59
3.2.2	Conformation	60
3.2.3	Configuration	60
3.3	Solubility and drug design	61
3.3.1	The importance of water solubility	62
3.4	Solubility and drug structure	63
3.5	Salt formation	64
3.6	The incorporation of water solubilizing groups in a structure	65
3.6.1	The type of group	66
3.6.2	Reversibly and irreversibly attached groups	66
3.6.3	The position of the water solubilizing group	67
3.6.4	Methods of introduction	67
3.7	Questions	70
4	The SAR and QSAR Approaches to Drug Design	71
4.1	Structure–activity relationships (SARs)	71
4.2	Changing size and shape	73
4.3	Introduction of new substituents	73
4.3.1	The introduction of a group in an unsubstituted position	73
4.3.2	The introduction of a group by replacing an existing group	76
4.4	Quantitative structure–activity relationships (QSARs)	78
4.4.1	Lipophilicity	79
	Partition coefficients (P)	79
	Lipophilic substitution constants (π)	80
4.4.2	Electronic effects	82
	The Hammett constant (σ)	82
4.4.3	Steric effects	83
	The Taft steric parameter (E_s)	84
	Molar refractivity (MR)	84
	Other parameters	85

4.4.4	Hansch analysis	85
	Craig plots	88
4.5	The Topliss decision tree	89
4.6	Questions	92
5	Computer Aided Drug Design	95
5.1	Introduction	95
5.1.1	Molecular modelling methods	96
5.1.2	Computer graphics	98
5.2	Molecular mechanics	98
5.2.1	Creating a molecular model using molecular mechanics	102
5.3	Molecular dynamics	104
5.3.1	Conformational analysis	105
5.4	Quantum mechanics	105
5.5	Docking	109
5.6	Questions	110
6	Combinatorial Chemistry	113
6.1	Introduction	113
6.1.1	The design of combinatorial syntheses	115
6.1.2	The general techniques used in combinatorial synthesis	116
6.2	The solid support method	117
6.2.1	Parallel synthesis	118
6.2.2	Furka's mix and split technique	121
6.3	Encoding methods	123
6.3.1	Sequential chemical tagging methods	123
6.3.2	Still's binary code tag system	124
6.3.3	Computerized tagging	126
6.4	Combinatorial synthesis in solution	127
6.5	Screening and deconvolution	128
6.6	Questions	130
7	Selected Examples of Drug Action at some Common Target Areas	131
7.1	Introduction	131
7.2	Examples of drugs that disrupt cell membranes and walls	131
7.2.1	Antifungal agents	132
Azoles		133
Allylamines		134
Phenols		135
7.2.2	Antibacterial agents	135
Ionophoric antibiotic action		135
Cell wall synthesis inhibition		136
7.3	Drugs that target enzymes	138
7.3.1	Reversible inhibitors	139
7.3.2	Irreversible inhibition	140
7.3.3	Transition state inhibitors	142
7.4	Drugs that target receptors	144
7.4.1	Agonists	144
7.4.2	Antagonists	145
7.4.3	Partial agonists	147

7.5	Drugs that target nucleic acids	147
7.5.1	Antimetabolites	147
7.5.2	Enzyme inhibitors	150
7.5.3	Intercalation agents	151
7.5.4	Alkylating agents	152
7.5.5	Antisense drugs	152
7.5.6	Chain cleaving agents	154
7.6	Antiviral drugs	155
7.6.1	Nucleic acid synthesis inhibitors	155
7.6.2	Host cell penetration inhibitors	156
7.6.3	Inhibitors of viral protein synthesis	157
7.7	Questions	157
8	Pharmacokinetics	159
8.1	Introduction to pharmacokinetics	159
8.1.1	General classification of pharmacokinetic properties	160
8.2	Pharmacokinetics and drug design	160
8.3	Pharmacokinetic models	161
8.4	Intravascular administration	162
8.4.1	Intravenous injection (IV bolus)	163
8.4.2	Clearance and its significance	166
8.4.3	Intravenous infusion	169
8.5	Extravascular administration	171
8.5.1	Single oral dose	174
8.5.2	The calculation of t_{\max} and C_{\max}	176
8.5.3	Repeated oral doses	176
8.6	The use of pharmacokinetics in drug design	177
8.7	Questions	178
9	Drug Metabolism	181
9.1	Introduction	181
9.1.1	The stereochemistry of drug metabolism	181
9.1.2	Biological factors affecting metabolism	183
9.1.3	Environmental factors affecting metabolism	184
9.1.4	Species and metabolism	184
9.2	Secondary pharmacological implications of metabolism	184
9.3	Sites of action	184
9.4	Phase I metabolic reactions	186
9.4.1	Oxidation	186
9.4.2	Reduction	186
9.4.3	Hydrolysis	189
9.4.4	Hydration	189
9.4.5	Other Phase I reactions	189
9.5	Phase II metabolic routes	190
9.6	Pharmacokinetics of metabolites	190
9.7	Drug metabolism and drug design	193
9.8	Prodrugs	195
9.8.1	Bioprecursor prodrugs	195
9.8.2	Carrier prodrugs	196

9.8.3 The design of prodrug systems for specific purposes	197
Improving absorption and transport through membranes	197
Improving patient acceptance	198
Slow release	198
Site specificity	198
Minimizing side effects	200
9.9 Questions	200
10 An Introduction to Lead and Analogue Syntheses	203
10.1 Introduction	203
10.2 Asymmetry in syntheses	205
10.2.1 The use of non-stereoselective reactions to produce stereospecific centres	206
10.2.2 The use of stereoselective reactions to produce stereospecific centres	207
10.2.3 General methods of asymmetric synthesis	208
Methods that use catalysts to obtain stereoselectivity	210
Methods that do not use catalysts to obtain stereoselectivity	211
10.3 Designing organic syntheses	214
10.3.1 An introduction to the disconnection approach	214
10.4 Questions	220
11 Drug Development and Production	223
11.1 Introduction	223
11.2 Chemical development	224
11.2.1 Chemical engineering issues	225
11.2.2 Chemical plant, health and safety considerations	226
11.2.3 Synthesis quality control	227
11.2.4 A case study	228
11.3 Pharmacological and toxicological testing	231
11.4 Drug metabolism and pharmacokinetics	234
11.5 Formulation development	235
11.6 Production and quality control	235
11.7 Patent protection	236
11.8 Regulation	237
11.9 Questions	238
Appendix	239
A.1 Sickle-cell anaemia	239
A.2 Bacteria	240
A.3 Cell membranes	244
A.4 Receptors	246
A.5 Transfer through membranes	249
A.6 Regression analysis	250
A.7 Enzymes	252

A.8 Prostaglandins	255
A.9 Cancer	256
A.10 Viruses	257
A.11 Blood–brain barrier	259
A.12 Enzyme structure and species	260
Answers to Questions	261
Selected Further Reading	273
Index	275

Preface

This book is written for second, and subsequent year undergraduates studying for degrees in medicinal chemistry, pharmaceutical chemistry, pharmacy, pharmacology and other related degrees. It is also intended for students whose degree courses contain a limited reference to medicinal chemistry. The text assumes that the reader has a knowledge of chemistry at level one of a university life sciences degree. The text discusses the fundamental chemical principles used for drug discovery and design. A knowledge of physiology and biology is advantageous but not essential. Appropriate relevant physiology and biology is outlined in the appendices.

Chapter 1 gives a brief review of the structures and nomenclature of the more common classes of naturally occurring compounds found in biological organisms. It is included for undergraduates who have little or no background knowledge of natural product chemistry. For students who have studied natural product chemistry it may be used as either a revision or a reference chapter. Chapter 2 attempts to give an overview of medicinal chemistry. The basic approaches used to discover and design drugs are outlined in Chapters 3–6 inclusive. Chapter 7 is intended to give the reader a taste of main line medicinal chemistry. It illustrates some of the strategies used, often within the approaches outlined in previous chapters, to design new drugs. For a more encyclopedic coverage of the discovery and design of drugs for specific conditions, the reader is referred to appropriate texts such as some of those given under *Medicinal Chemistry* in the *Selected Further Reading* section at the end of this book. Chapters 8 and 9 describe the pharmacokinetics and metabolism respectively of drugs and their effect on drug design. Chapter 10 attempts to give an introductory overview of an area that is one of the principal objectives of the medicinal chemist. For a more in depth discussion, the reader is referred to the many specialized texts that are available on organic synthesis. Drug development from the research stage to marketing the final product is briefly outlined in Chapter 11.

The approach to medicinal chemistry is kept as simple as possible. The text is supported by a set of questions at the end of each chapter. Answers, sometimes in the form of references to sections of the book, are listed separately. A list of recommended further reading, classified according to subject, is also included.

Gareth Thomas

Acknowledgements

I wish to thank all my colleagues, past and present without, whose help this book would have not been written. In particular, I would like to thank Dr. S Arkle for access to his lecture notes and Dr. J Wong for his unstinted help and advice, Dr. J Brown for again acting as a living pharmacology dictionary, Dr. P Cox for the molecular model diagrams and his patience in explaining to me the intricacies of molecular modelling and Mr. A Barrow and Dr. D Brimage for the library searches they conducted. I wish also to thank the following friends and colleagues for proof-reading chapters and supplying information: Dr. L Adams, Dr. C Alexander, Dr. L Banting, Dr. D Brown, Dr. S Campbell, Dr. B Carpenter, Mr. P Clark, Dr. P Howard, Dr. A Hunt, Mrs. W Jones, Dr. T Mason, Dr. J Mills, Dr. T Nevell, Dr. M Norris, Dr. J Smart, Professor D Thurston, Dr. G White and Mr. S Wills.

Finally, I would like to thank my wife for her support whilst I was writing the text.

Abbreviations

A	Adenine
Abe	Abequose
ACE	Angiotensin-converting enzyme
ACh	Acetyl choline
ADME	Absorption, distribution, metabolism and elimination
ADR	Adverse drug reaction
Ala	Alanine
Arg	Arginine
Asp	Aspartate
ATP	Deoxyadenosine triphosphate
dATP	Adenosine triphosphate
AUC	Area under the curve
C	Cytosine
CNS	Central nervous system
CoA	Coenzyme A
CYP-450	Cytochrome P-450 family
Cys	Cysteine
d.e.	Diastereoisometric excess
DHF	Dihydrofolic acid
DHFR	Dihydrofolate reductase
DMPK	Drug metabolism and pharmacokinetics
DNA	Deoxyribonucleic acid
EC	Enzyme commission
e.e.	Enantiomeric excess
ELF	Effluent load factor
EMEA	European medicines evaluation agency
EPC	European Patent Convention
EPO	European Patent Office
E _s	Taft steric parameter

FDA	Food and drugs administration
FMO	Flavin monooxygenase
FGI	Functional group interconversion
Fmoc,	9-Fluorenylmethoxycarbonyl group
FdUMP	5-fluoro-2'-deoxyuridyline monophosphate
FUdRP	5-fluoro-2'-deoxyuridylic acid
G	Guanine
GABA	γ -Aminobutyric acid
GI	Gastrointestinal tract
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GSH	Glutathione
Hb,	Haemoglobin
HbS	Sickle cell haemoglobin
His	Histidine
HIV	Human immunodeficiency disease
hnRNA	Heterogeneous nuclear RNA
Ile	Isoleucine
IV	Intravenous injection
IM	Intramuscular injection
KDO	2-Keto-3-deoxyoctanoate
LDA	Lithium diisopropylamide
LDH	Lactose dehydrogenase
Leu	Leucine
Lys	Lysine
mACh	Muscarinic cholinergic receptor
MA(A)	Market authorisation (application)
MCA	Medical control agency
Met	Methionine
Moz	4-Methoxybenzyloxycarbonyl group
MR	Molar refractivity
mRNA	messenger RNA
nACh	Nicotinic cholinergic receptor
NAD ⁺	Nicotinamide adenine dinucleotide (oxidised form)
NADH	Nicotinamide adenine dinucleotide (reduced form)

NADP ⁺	Nicotinamide dinucleotide phosphate (oxidised form)
NADPH	Nicotinamide dinucleotide phosphate (reduced form)
NAG	β - <i>N</i> -Acetylglucosamine
NAM	β - <i>N</i> -Acetylmuramic acid
ONs	Sequence defined oligonucleotides
P-450	Cytochrome P-450 oxidases
PABA	<i>p</i> -Aminobenzoic acid
PCT	Patent Cooperation Treaty
PG	Prostaglandin
Phe	Phenylalanine
PO	Per oral (by mouth)
pre-mRNA	Premessenger RNA
Pro	Proline
ptRNA	Primary transcript RNA
QSAR	Quantitative structural-activity relationships
RNA	Ribonucleic acid
SAM	<i>S</i> -Adenosylmethionine
SAR	<i>See</i> Structural-activity relationships
Ser	Serine
SIN-1	3-Morpholino-sydnomine
T	Thymine
THF	Tetrahydrofolic acid
Thr	Threonine
dTMP	Deoxythymidylate-5'-monophosphate
tRNA	transfer RNA
Try	Tyrosine
U	Uracil
UDP	Uridine diphosphate
UDPGA	Uridine diphosphate glucuronic acid
dUMP	Deoxyuridylate-5'-monophosphate
UdRP	Deoxyuridylic acid
Val	Valine

1 Biological Molecules

1.1 Introduction

Chemical compounds and metallic ions are the basic building blocks of all biological structures and processes that are the basis of life as we know it. Some of these naturally occurring compounds and ions (**endogenous species**) are present only in very small amounts in specific regions of the body, whilst others, such as peptides, proteins, carbohydrates, lipids and nucleic acids, are found in all parts of the body. A basic knowledge of the nomenclature and structures of these more common endogenous classes of biological molecules is essential to understanding medicinal chemistry. This chapter introduces these topics in an attempt to provide for those readers who do not have this background knowledge.

The structures of biologically active molecules usually contain more than one type of functional group. This means that the properties of these molecules are a mixture of those of each of the functional groups present plus properties characteristic of the compound. The latter are frequently due to the interaction of adjacent functional groups and/or the influence of a functional group on the carbon–hydrogen skeleton of the compound. This often involves the electronic activation of C–H bonds by adjacent functional groups.

1.2 Amino acids

1.2.1 Introduction

Simple amino acids are the basic building blocks of proteins. Their structures contain both an amino group, usually a primary amine, and a carboxylic acid. The relative positions of these groups vary, but for most naturally occurring

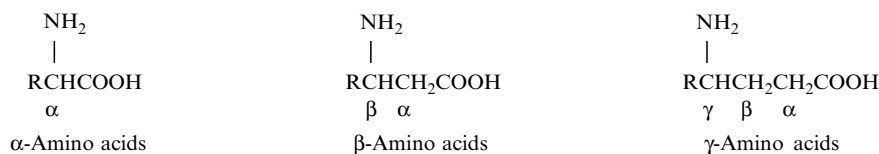




Figure 1.1 The general structural formulae of amino acids. Amino acids may be classified as α , β , γ ...etc. depending on the relative positions of the amine and carboxylic acid groups. α -Amino acids are the most common naturally occurring amino acids

compounds the amino group is attached to the same carbon as the carboxylic acid (Figure 1.1).

The structures of amino acids can also contain other functional groups besides the amine and carboxylic acid groups (Table 1.1). Methionine, for example, contains a sulphide group, whilst serine has a primary alcohol group.

Table 1.1 Examples of the names and structures of amino acids

Amino acid	Name	Symbol/letter	pI(25°)
$\text{CH}_3\text{CH}_2(\text{NH}_2)\text{COOH}$	Alanine	Ala A	6.0
$\text{NH}_2-\overset{\text{NH}}{\underset{\text{ }}{\text{C}}}-\text{NHCH}_2\text{CH}_2\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	Arginine	Arg R	10.8
$\text{NH}_2\text{COCH}_2\text{CH}(\text{NH}_2)\text{COOH}$	Asparagine	Asn N	5.4
$\text{HOOCCH}_2\text{CH}(\text{NH}_2)\text{COOH}$	Aspartic acid	Asp D	3.0
$\text{HOOCCH}_2\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	Glutamic acid	Glu E	3.2
$\text{H}_2\text{NCOCH}_2\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	Glutamine	Gln Q	5.7
$\text{CH}_2(\text{NH}_2)\text{COOH}$	Glycine	Gly G	6.0
 $\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	Histidine	His H	7.6
$\text{CH}_3\text{CH}_2\overset{\text{CH}_3}{\underset{ }{\text{C}}}\text{HCH}(\text{NH}_2)\text{COOH}$	Isoleucine	Ile I	6.0
$\text{CH}_3\overset{\text{CH}_3}{\underset{ }{\text{C}}}\text{HCH}_2\text{CH}(\text{NH}_2)\text{COOH}$	Leucine	Leu L	5.9
$\text{H}_2\text{NCH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	Lysine	Lys K	9.7
$\text{CH}_3\text{SCH}_2\text{CH}_2\text{CH}(\text{NH}_2)\text{COOH}$	Methionine	Met M	5.7
$\text{PhCH}_2\text{CH}(\text{NH}_2)\text{COOH}$	Phenylalanine	Phe F	5.5
 COOH	Proline	Pro P	6.3
$\text{CH}_2\text{OHCH}(\text{NH}_2)\text{COOH}$	Serine	Ser S	5.7
$\text{CH}_3\overset{\text{CH}_3}{\underset{ }{\text{C}}}\text{HCH}(\text{NH}_2)\text{COOH}$	Valine	Val V	6.0

The nature of the side chains of amino acids determines the hydrophobic (water hating) and hydrophilic (water loving) nature of the amino acid. Amino acids with hydrophobic side chains will be less soluble in water than those with hydrophilic side chains. The hydrophobic/hydrophilic nature of the side chains of amino acids has a considerable influence on the conformation adopted by a peptide or protein in aqueous solution. Furthermore, the hydrophobic/hydrophilic balance of the groups in a molecule will have a considerable effect on the ease of its passage through membranes (Appendix 5).

1.2.2 Structure

All solid amino acids exist as dipolar ions known as zwitterions (Figure 1.2(a)). In aqueous solution the structure of amino acids are dependent on the pH of the solution (Figure 1.2(b)). The pH at which an aqueous solution of an amino acid is electrically neutral is known as the **isoelectric point (pI)** of the amino acid (Table 1.1). Isoelectric point values vary with temperature. They are used in the design of electrophoretic and chromatographic analytical methods for amino acids.

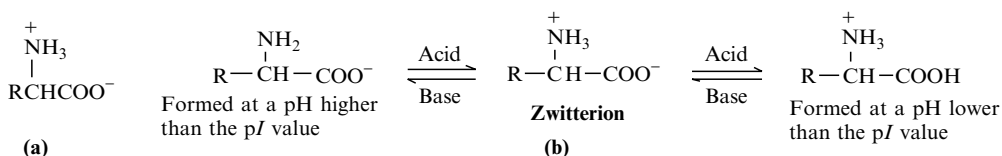


Figure 1.2 (a) The general structural formula of the zwitterions of amino acids. (b) The structures of amino acids in acidic and basic aqueous solutions

1.2.3 Nomenclature

Amino acids are normally known by their trivial names (Table 1.1). In peptide and protein structures their structures are indicated by either three letter groups or single letters (Table 1.1, and Figure 1.7). Amino acids such as ornithine and citrulline, which are not found in naturally occurring peptides and proteins, do not have an allocated three or single letter code (Figure 1.3).



Figure 1.3 Ornithine and citrulline

Most amino acids, with the notable exception of glycine, are optically active. Their configurations are usually indicated by the D/L system (Figure 1.4) rather than the R/S system. Most naturally occurring amino acids have an L configuration but there are some important exceptions. For example, some bacteria also possess D-amino acids. This is important in the development of some antibacterial drugs.

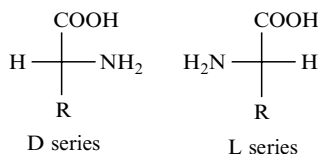


Figure 1.4 The D/L configurations of amino acids. Note that the carboxylic acid group must be drawn at the top and the R group at the bottom of the Fischer projection. Stereogenic centres in the R group do not affect the D/L assignment

1.3 Peptides and proteins

Peptides and proteins have a wide variety of roles in the human body (Table 1.2). They consist of amino acid residues linked together by **amide functional groups** (Figure 1.5(a)), which in peptides and proteins are referred to as **peptide links** (Figure 1.5(c)). The amide group has a rigid flat structure. The lone pair of its nitrogen atom is able to interact with the π electrons of the carbonyl group.

Table 1.2 Examples of some of the biological functions of proteins

Function	Notes
Structural	These proteins provide strength and elasticity to, for example, bone (collagen), hair (α -keratins) and connective tissue (elastin).
Enzymes	This is the largest class of proteins. Almost all steps in biological reactions are catalysed by enzymes.
Regulatory	These are proteins that control the physiological activity of other proteins. Insulin, for example, regulates glucose metabolism in mammals.
Transport	These transport specific compounds from one part of the body to another haemoglobin transports carbon dioxide too and oxygen from the lungs. Cell membranes contain proteins that are responsible for the transport of species from one side of the membrane to the other.
Storage	These provide a store of substances required by the body. For example, the protein ferritin acts as an iron store for the body.
Protective	These proteins that protect the body. Some form part of the bodies immune system defending the body against foreign molecules and bacteria. Others, such as the blood clotting agents thrombin and fibrinogen, prevent loss of blood when a blood vessel is damaged.

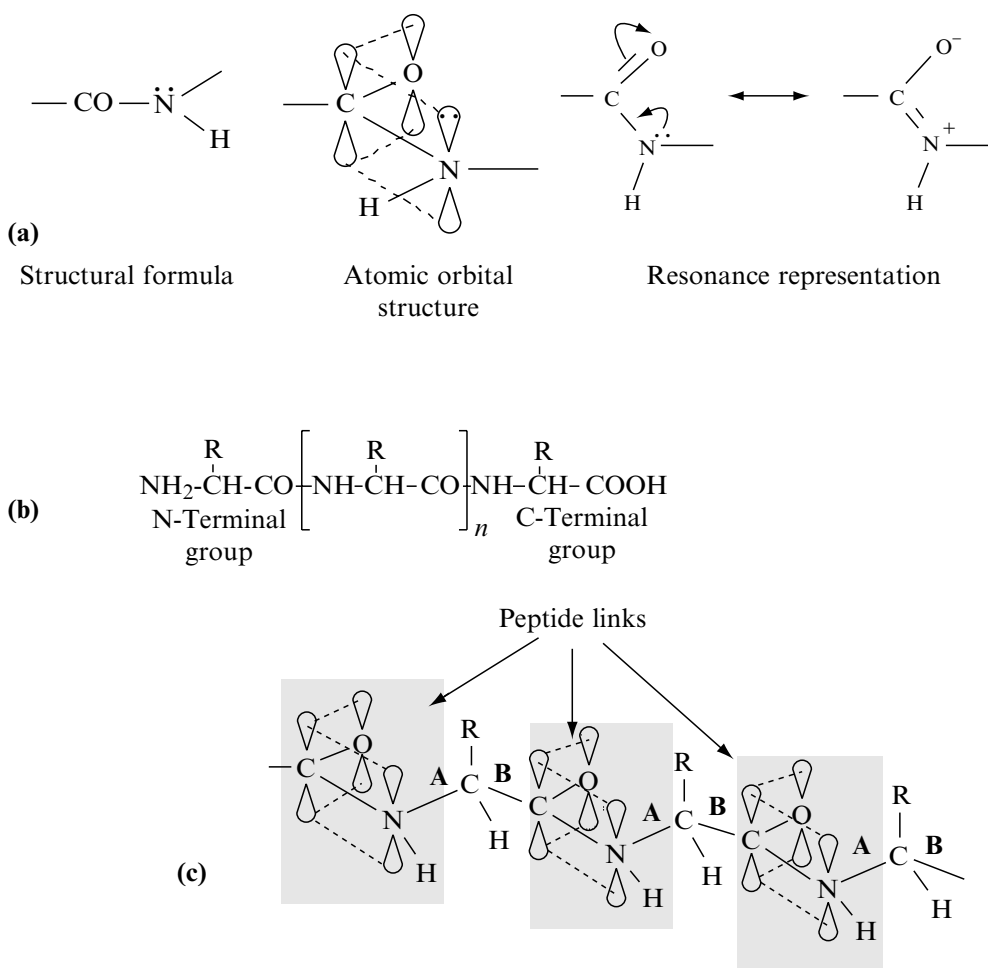


Figure 1.5 (a) The structure of the amide functional group. (b) The general structure of simple peptides. (c) The peptide link is planar and has a rigid conjugated structure. Changes in conformation can occur about bonds A and B. Adapted from G Thomas, *Chemistry for Pharmacy and the Life Sciences including Pharmacology and Biomedical Science*, 1996, published by Prentice Hall, a Pearson Education Company

This electron delocalization is explained by p orbital overlap and is usually shown by the use of resonance structures (Figure 1.5(a)).

The term **peptide** is normally used for compounds that contain small numbers of amino acid residues whilst the term **polypeptide** is loosely used for larger compounds with relative molecular mass (RMM) values greater than about 500 or more. **Proteins** are more complex polypeptides with RMM values usually greater than 2000. They are classified as **simple** when their structures contain

only amino acid residues and **conjugated** when other residues besides those of amino acids occur as integral parts of their structures. For example, haemoglobin is a conjugated protein because its structure contains a haem residue (Figure 1.6). These non-amino-acid residues are known as **prosthetic groups** when they are involved in the biological activity of the molecule. Conjugated proteins are classified according to the chemical nature of their non-amino-acid component. For example, glycoproteins contain a carbohydrate residue, haemoproteins a haem group and lipoproteins a lipid residue.

1.3.1 Structure

The structures of peptides and proteins are very varied. They basically consist of chains of amino acid residues (Figures 1.5(b), 1.5(c) and 1.7). These chains may be branched due to the presence of multi-basic or acidic amino acid residues in the chain (Figure 1.7(d)). In addition, bridges (cross links) may be formed between different sections of the same chain or different chains. Cysteine residues, for example, are responsible for the S–S bridges between the two peptide chains that form the structure of insulin (Figure 1.7(e)). The basic structure of peptides and proteins is twisted into a conformation (time dependent overall shape) characteristic of that peptide or protein. These conformations are dependent on both the nature of their biological environment as well as their chemical structures. The ability of peptides and proteins to carry out their biological functions is normally dependent on this conformation. Any changes to any part of the structure of a

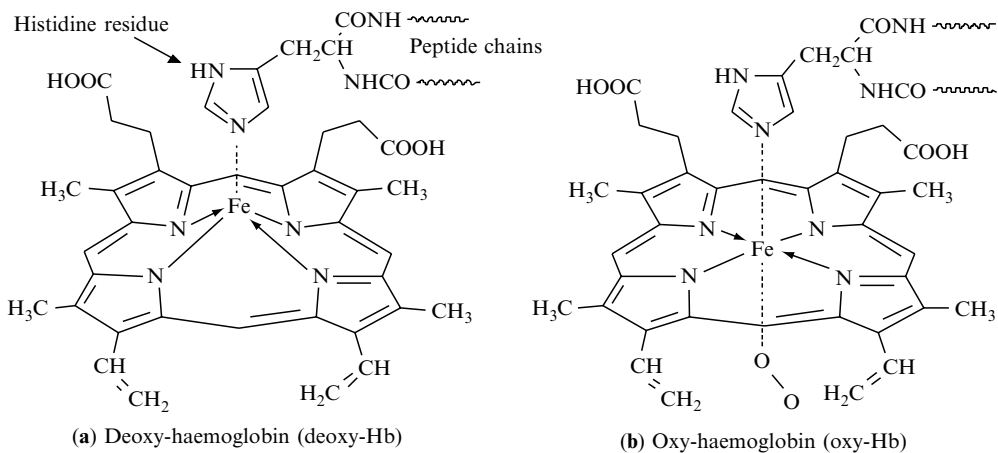


Figure 1.6 The structure of the haem residue in deoxy- and oxy-haemoglobins. In deoxy-Hb the bonding of the iron is pyramidal whilst in oxy-Hb it is octahedral

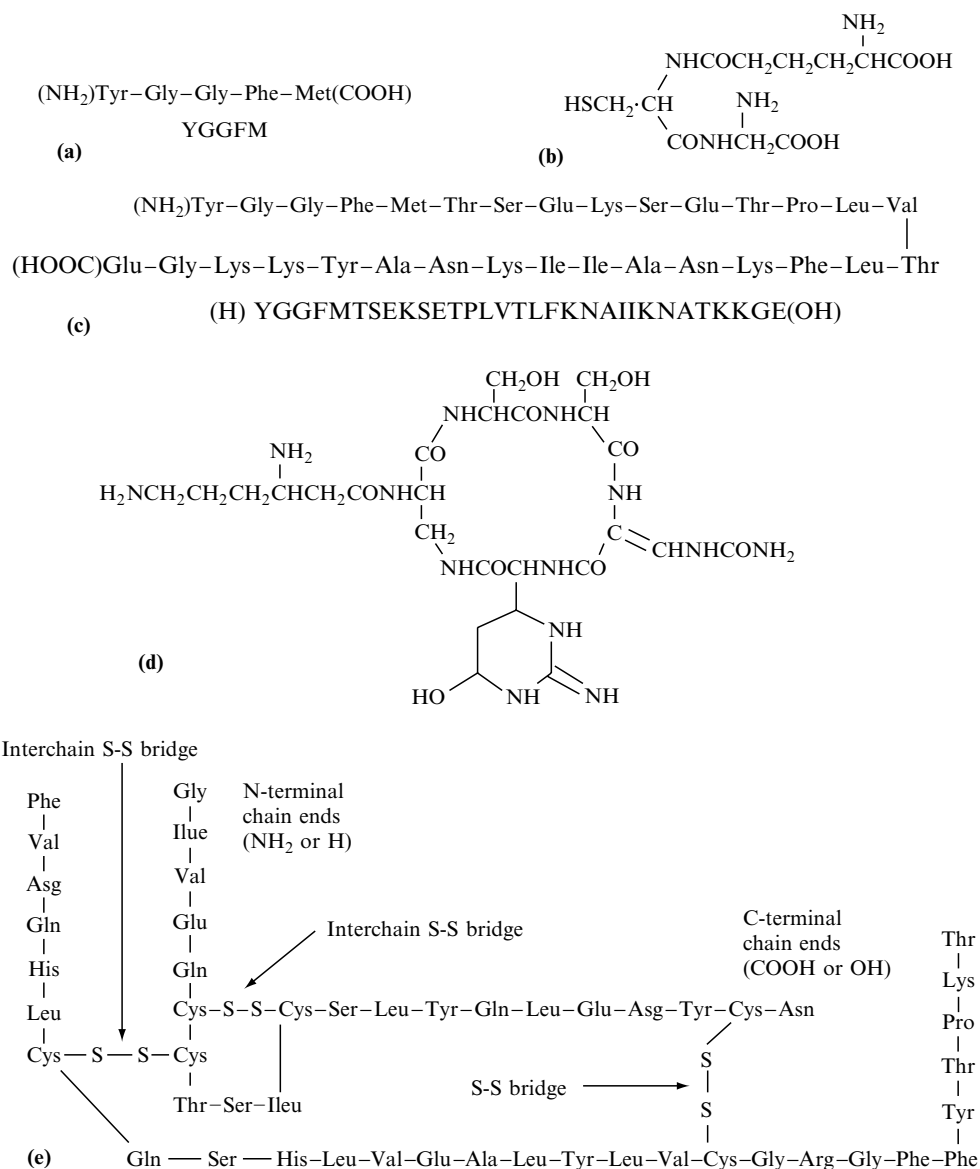


Figure 1.7 Representations of the primary structures of peptides. Two systems of abbreviations are used to represent primary structures. The single letter system is used for computer programs. In both systems the N-terminus of the peptide chain is usually drawn on the left-hand side of the structure. (a) Met-enkephalin. This pentapeptide occurs in human brain tissue. (b) Glutathione, an important constituent of all cells, where it is involved in a number of biological processes. (c) β -Endorphin. This endogenous peptide has opiate activity and is believed to be produced in the body to counter pain. (d) Viomycin, a polypeptide antibiotic produced by *Streptomyces griseovorticillatus* var. *tuberciculus*. The presence of the dibasic 2,3-diaminopropanoic acid residue produces the chain branching. (e) Insulin, the hormone that is responsible for controlling glucose metabolism

peptide or protein will either change or destroy the compound's biological activity. For example, sickle-cell anaemia (Appendix 1) is caused by the replacement of a glutamine residue by a valine residue structure of haemoglobin.

Proteins are often referred to as **globular** and **fibrous proteins** according to their conformation. Globular proteins are usually soluble in water, whilst fibrous proteins are usually insoluble. The complex nature of their structures has resulted in the use of a sub-classification, sometimes referred to as **the order of protein structures**. This classification divides the structure into primary, secondary, tertiary and quaternary orders of structures.

The **primary protein structure** of peptides and proteins is the sequence of amino acid residues in the molecule (Figure. 1.7).

Secondary protein structures are the local regular and random conformations assumed by sections of the peptide chains found in the structures of peptides and proteins. The main regular conformations found in the secondary structures of proteins are the α -helix, the β -pleated sheet and the triple helix (Figure 1.8). These and other random conformations are believed to be mainly due to intramolecular hydrogen bonding between different sections of the peptide chain.

The **tertiary protein structure** is the overall shape of the molecule. Tertiary structures are often formed by the peptide chain folding back on itself. These folded structures are stabilized by S-S bridges, hydrogen bonding, salt bridges (Figure 1.9(a)) and van der Waals' forces within the peptide chain and also with molecules in the peptide's environment. They are also influenced by hydrophobic interactions between the peptide chain and its environment. Hydrophobic interaction is thought to be mainly responsible for the folded shape of the β -peptide chain of human haemoglobin (Figure 1.9(b)). In this structure the hydrophilic groups of the peptide chain are on the outer surface of the folded structure.

Quaternary protein structures are the three dimensional protein structures formed by the noncovalent associations of a number of individual peptides and polypeptide molecules. These individual peptide and polypeptide molecules are known as subunits. They may or may not be the same. Haemoglobin, for example, consists of four subunits, two α - and two β -units held together by hydrogen bonds and salt bridges.

The structures of peptides and proteins usually contain numerous amino and carboxylic acid groups. Consequently, water soluble proteins in aqueous solution can form differently charged structures and zwitterions depending on the pH of the solution (see 1.2.2). The pH at which the latter occurs is known as the isoelectric point (pI) of the protein (Table 1.3). The nature of the charge on the structures of peptides and proteins has a considerable effect on their solubility

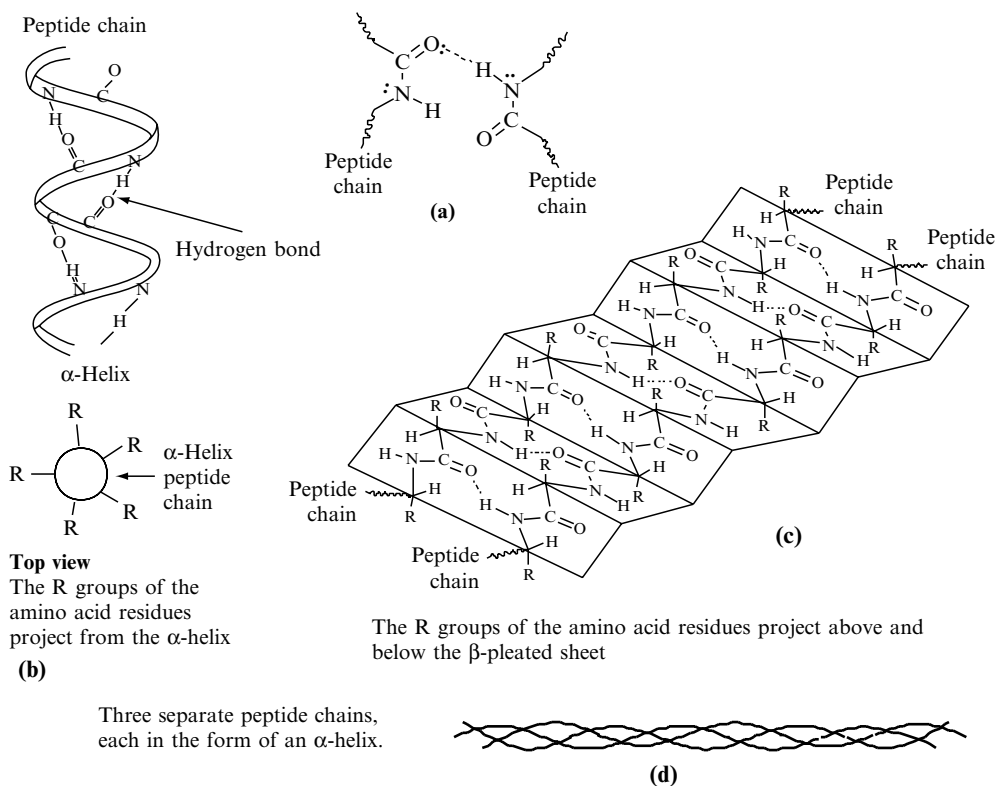


Figure 1.8 The secondary structures of proteins. (a) Hydrogen bonding between peptide links. The conjugated lone pair of the amide nitrogen atom is not available to form hydrogen bonds. (b) The α -helix. The peptide chain is largely held in this shape by intramolecular hydrogen bonds. (c) β -Pleated sheets are formed by hydrogen bonding between neighbouring peptide chains. Antiparallel β -sheets (shown) have the peptide chains running in opposite directions. Parallel β -sheet (not shown) have the peptide chains running in the same direction. Silk fibroin has a high proportion of antiparallel β -pleated sheets. (d) The triple helix in which the three peptide chains are largely held together by hydrogen bonding. For example, the basis of the structure of the fibrous protein collagen which occurs in skin, teeth and bones, consists of three chains of the polypeptide tropocollagen in the form of a triple helix. This forms a cable like structure known as a *protofibril*. Reproduced from G Thomas, *Chemistry for Pharmacy and the Life Sciences including Pharmacology and Biomedical Science*, 1996, by permission of Prentice Hall, a Pearson Education Company

and biological activity. For example, the water solubility of a protein is usually at a minimum at its isoelectric point whilst the charge on a protein may affect the ease of transport of a protein through a plasma membrane (see Appendix 5). It is also important in electrophoretic and chromatographic methods of protein analysis.

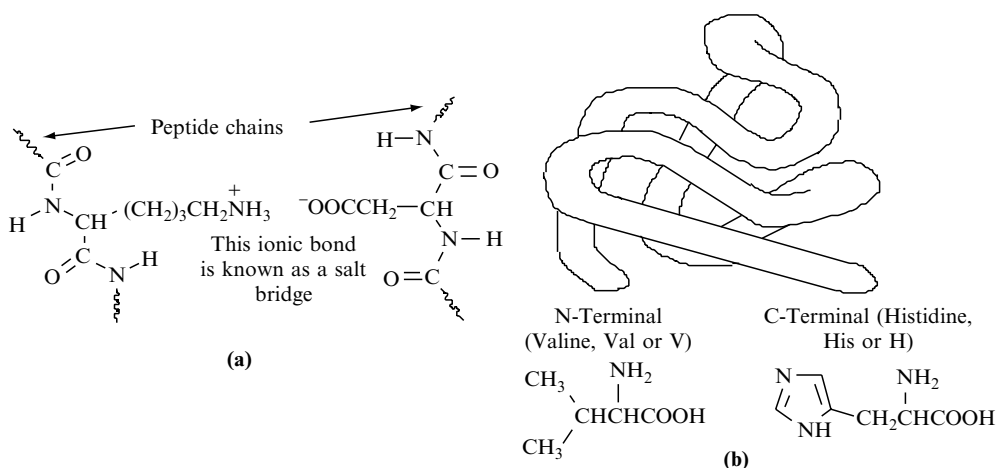


Figure 1.9 (a) A salt bridge. This is essentially an ionic bond. (b) The folded structure of a β -haemoglobin polypeptide chain. Reproduced from G Thomas, *Chemistry for Pharmacy and the Life Sciences including Pharmacology and Biomedical Science*, 1996, by permission of Prentice Hall, a Pearson Education Company

Table 1.3 Examples of the pI values of proteins (various sources)

Protein	pI (25°)	Protein	pI (25°)	Protein	pI (25°)
Cytochrome c (horse)	10.6	γ -Globulin (human)	6.6	Lysozyme (hen)	11.0
Fibrinogen (human)	5.5	Histone (bovine)	10.8	Ribonuclease A (bovine)	9.4
Haemoglobin A (human)	7.1	Insulin (human)	5.4	Serum albumin (human)	4.9

1.4 Carbohydrates

Carbohydrates, or sugars as they are commonly known, are classified as monosaccharides, oligosaccharides and polysaccharides. **Monosaccharides** are either polyhydroxyaldehydes (**aldoses**) or polyhydroxyketones (**ketoses**), which are not converted to any simpler polyhydroxyaldehydes and polyhydroxyketones respectively under aqueous hydrolysis conditions. Carbohydrates also include compounds such as glucosamine (Figure 1.10), whose structures contain amino groups as well as hydroxy groups. These compounds are known as **amino sugars**. However, not all polyhydroxyaldehydes and ketones are classified as carbohydrates.

Monosaccharides are classified according to the total number of carbon atoms in their structure. For example, an aldohexose is a monosaccharide that contains a total of six carbon atoms including that of the aldehyde in its structure. Similarly, a ketopentose has five carbons in its structure including the one in the keto group. **Oligosaccharides** are carbohydrates that yield from two to about nine monosaccharide molecules when one molecule of the oligosaccharide is hydrolysed. Small oligosaccharides are often classified according to the number of monosaccharide residues contained in their structures. For example, disaccharides and trisaccharides contain two and three monosaccharide residues respectively whilst **polysaccharides** yield larger numbers of monosaccharide molecules per polysaccharide molecule on hydrolysis. All types of carbohydrate occur widely in the human body. They exhibit a wide variety of biological functions but in particular act as major energy sources for the body.

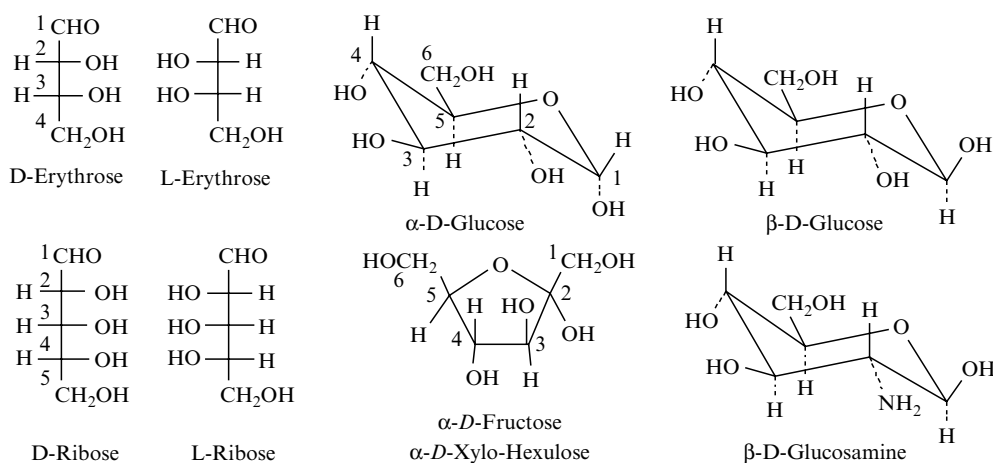


Figure 1.10 Examples of the cyclic and straight chain structures of monosaccharides. The carbon of the carbonyl group has the lowest locant

1.4.1 The structure of monosaccharides

Monosaccharides can exist as either straight chain or cyclic structures (Figure 1.10). Those with five or more carbon atoms usually assume either a five (**furanose**) or six (**pyranose**) membered ring structure. These cyclic structures are formed by an internal nucleophilic addition between a suitably positioned hydroxy group in the molecule and the carbonyl group (Figure 1.11). It results in the formation of the corresponding cyclic hemiacetal or hemiketal. The rings of these cyclic products exist in their normal conformations. For example, six

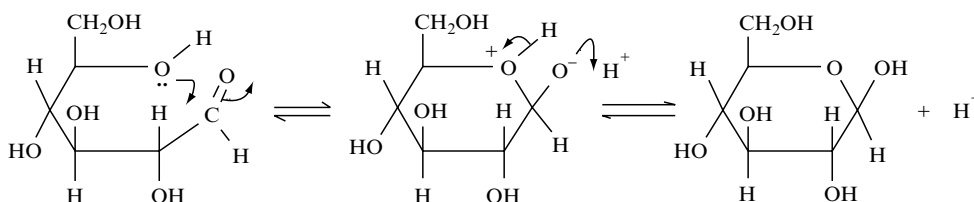


Figure 1.11 The cyclization of the *straight chain* form of glucose to form the β -hemiacetal cyclic form of the molecule

membered rings usually occur as chair conformations whilst five membered rings exist as envelope conformations.

This internal nucleophilic addition introduces a new chiral centre into the molecule. The carbon of the new centre is known as the **anomeric carbon** and the two new stereoisomers formed are referred to as **anomers**. The isomer where the new hydroxy group and the CH_2OH are on opposite sides of the plane of the ring is known as the alpha (α) anomer. Conversely, the isomer with the new hydroxy group and terminal CH_2OH on the same side of the plane of the ring is known as the beta (β) anomer (Figure 1.12).

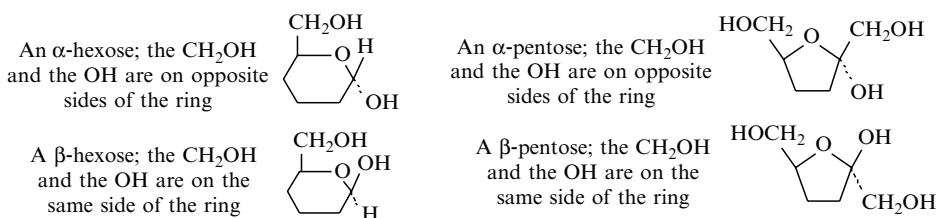


Figure 1.12 The α - and β -anomers of monosaccharides drawn using the Haworth convention. In this convention solid lines represent bonds above the plane of the ring whilst dotted lines are used to indicate bonds below the plane of the ring. Reproduced from G Thomas, *Chemistry for Pharmacy and the Life Sciences including Pharmacology and Biomedical Science*, 1996, by permission of Prentice Hall, a Pearson Education Company

In many cases pure α - and β -anomers may be obtained by using appropriate isolation techniques. For example, crystallization of D-glucose from ethanol yields α -D-glucose $[\alpha]_{\text{D}} +112.2^\circ$ whilst crystallization from aqueous ethanol produces β -D-glucose $[\alpha]_{\text{D}} +18.7^\circ$. In the solid state these forms are stable and do not interconvert. However, in aqueous solution these cyclic structures can form equilibrium mixtures with the corresponding straight chain form (Figure 1.13). The change in optical rotation due to the conversion of either the pure α - or pure β -anomer of a monosaccharide into an equilibrium mixture of both forms in aqueous solution is known as **mutarotation** (Figure 1.13).