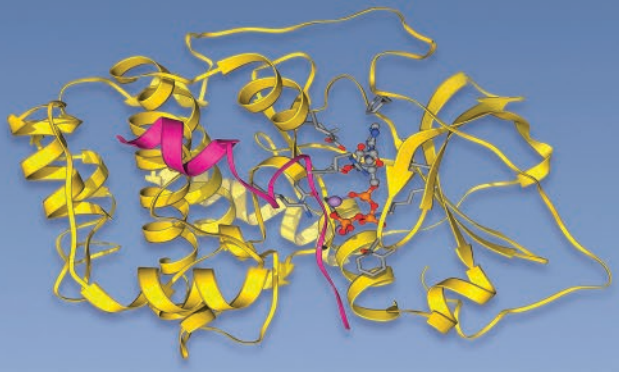
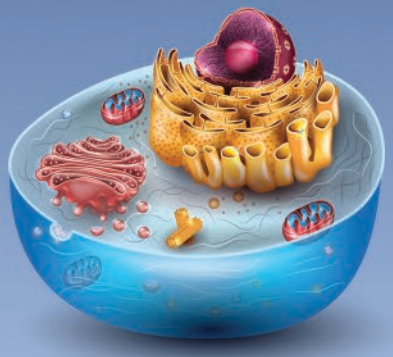
Second Edition

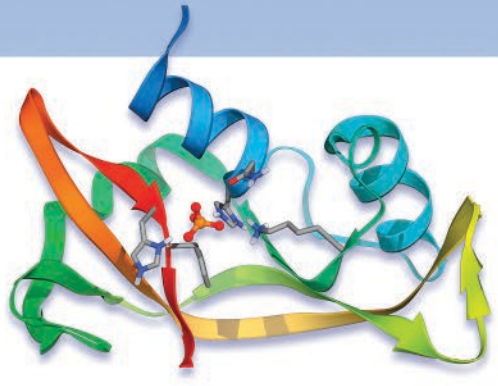
Essentials of Chemical Biology

Structures and Dynamics of Biological Macromolecules *In Vitro* & *In Vivo*





Andrew D. Miller | Julian A. Tanner



WILEY

Essentials of Chemical Biology

Essentials of Chemical Biology

Structures and Dynamics of Biological Macromolecules *In Vitro* and *In Vivo*

Andrew D. Miller
Department of Chemistry and Biochemistry, Mendel University, Brno, Czech Republic

Julian A. Tanner
School of Biomedical Sciences, University of Hong Kong

Second Edition

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:
Names: Miller, Andrew, author. | Tanner, Julian, author.
Title: Essentials of chemical biology : structures and dynamics of biological macromolecules *In Vitro* and *In Vivo* / Andrew D. Miller, Department of Chemistry and Biochemistry, Mendel University, Brno, Czech Republic, Julian A. Tanner, School of Biomedical Sciences, University of Hong Kong.
Description: Second edition. | Hoboken, New Jersey : John Wiley & Sons, Inc., [2024] | Includes bibliographical references and index.
Identifiers: LCCN 2022053778 | ISBN 9781119437970 (Paperback) | ISBN 9781119437963 (ePDF) | ISBN 9781119437840 (ePUB)
Subjects: LCSH: Biochemistry. | Molecular biology. | Macromolecules.
Classification: LCC QP514.2 .M55 2023 | DDC 612/.015--dc23/eng/20221215
LC record available at https://lccn.loc.gov/2022053778

Cover Images: © LAGUNA DESIGN/Getty Images; ksenvitaln/Shutterstock; Tefi/Shutterstock
Cover design: Wiley

Set in 10.5pt/12.5pt Minion Pro by Integra Software Services Pvt. Ltd, Pondicherry, India

To my father, who was the first to show me that chemistry does not have to be about disciplines but can also be about interesting problems waiting to be solved with all that the subject has to offer. In addition, thanks to my mother for her limitless encouragement in all I do or try to do. Thanks also to my children Nadia Nozomi, Tatiana Hikari and Samuel Kiyoshi for having grown up into such lovely people and for being such a primary inspiration for all that I do in my life. Thanks, too, to all my past and present students, co-researchers and collaborators who have made my professional life in chemical biology so enthralling, stimulating and such a great adventure. Finally, thanks to the Czech Ministry of Education, Youth and Sports (MŠMT) for the award of OPVVV Project FIT (Pharmacology, Immunotherapy, nanoToxicology) (CZ.02.1.01/0.0/0.0/15_003/0000495), with financial support from the European Regional Development Fund, which provided me with the necessary time and opportunity to bring into being this second edition.

Andrew D. Miller

To my wife, Ali, for all her love and support, her wonderful positivity and optimism that has encouraged me through all the ups and downs, plus her passion for student learning. She will always be an inspiration for how best to enable students to flourish, excel and enjoy learning. Thanks also to my children Alex, Nicky and Sophia for all their perceptive questions which always remind me of the deep curiosity essential for good science. Thanks, too, to my parents Andrew and Christine, and brother, Alastair, for creating around me an environment where learning and critical thinking were treasured from when I was young. Finally, thanks to all my students and teachers from whom I continue to learn so much.

Julian A. Tanner

Contents

Preface

Mapping the Latest Essentials of Chemical Biology

Since the first edition of this textbook, chemical biology has become established as a major branch of scientific activity devoted to understanding the way biology works at the molecular level. Chemical biology itself remains unashamedly multidisciplinary and chemical biology research is essentially problem driven and not discipline driven. Organic, physical, inorganic and analytical chemistry all contribute towards chemical biology, alongside newer emerging molecular disciplines. Some might say that chemical biology is just another way to rebadge biochemistry. However, such a comment misses the point. Biochemistry may have started as a discipline devoted to the study of individual biological macromolecules, but this discipline has been steadily evolving into increasingly descriptive, empirical studies of larger and larger macromolecular assemblies, structures and interacting molecular networks. The molecular increasingly gives ground to the cellular. In contrast, chemical biology is about chemistry-trained graduates and researchers taking a fundamental interest in the way biology works at the molecular level. Consequently, the focus is on the molecular and the quantitative, where molecular properties are investigated, studied and then gradually linked to macromolecular and cellular behaviour where possible. This is a fundamentally "bottom-up" approach to understanding biology, in keeping with the chemist's natural enthusiasm and appreciation for molecular structure and behaviour first and foremost.

This textbook has been produced with the third/fourth year undergraduate student and young (post) graduate school researcher in mind, namely those who have a solid background in chemical principles and are ready to apply and grow their chemical knowledge to suit a future degree or career interest in chemical biology. In preparing this textbook our objective has not been to try and cover everything currently seen as chemical biology, but instead to ask ourselves what topics and themes should be described as the essentials of chemical biology and how should these be presented in a way most appropriate for those of a chemical rather than a biological orientation. In doing this, we concluded that the true essentials of chemical biology are represented by the structure, characterisation and measureable behaviour of the main biological macromolecules and macromolecular lipid assemblies found in all cells of all organisms. We have also concluded that the activities of small molecules in biology for respiration and primary and secondary metabolism should not be included in the essentials of chemical biology, except where they feature as protein prosthetic groups or otherwise modify macromolecule behaviour. In our view, simple metabolism and metabolite interconversions are the stuff of biochemistry, whilst fascination with secondary metabolism, secondary metabolites and their interconversions has been the traditional preserve of bio-organic chemistry (a subset of organic chemistry).

Hence, in our textbook we begin with structure (Chapter 1) and synthesis (Chapters 2 and 3), then consider how structure is determined (Chapters 4, 5 and 6), followed by a consideration of dynamic behaviour and molecular interactions (Chapters 7, 8 and 9), culminating in molecular evolution and thoughts on the origins of life, quintessentially from the chemistry point of view (Chapter 10). New chapters (Chapters 11, 12, 13 and 14) focus on the ongoing transition of chemical biology research from reductionist studies at the molecular level to studies on molecules in cells *in vitro* then

molecules in organisms *in vivo*. This transition is neatly framed in terms of the contribution of chemical biology research to the creation and development of advanced therapeutics and diagnostics for the future management and treatment of chronic diseases. Finally, we conclude (Chapter 15) with the interface between chemical biology and synthetic biology. Armed with such essentials, we hope that readers will be empowered to think about and then tackle any problem of their chosen interest at the chemistry-biology and/or chemistry-medicine interfaces, after a little more detailed and specific reading, of course. Foremost, we remain hopeful that our textbook will provide a valuable tool for chemical biology students and researchers to open the door and step through into the extraordinary world of biology without feeling that they should have to leave their chemical principles behind them.

Glossary of Physical and Chemical Terms

Equation glossary

Chapter 1			
Potential energy	V	J	[kg m^2 s^{-2}]
Electrical point charge	q_n	C	
Vacuum permittivity	ε_o	F m^{-1} or C^2 m^{-1} J^{-1}	[C^2 kg^{-1} m^{-3} s^2]
Permittivity of medium	ε	F m^{-1} or C^2 m^{-1} J^{-1}	[C^2 kg^{-1} m^{-3} s^2]
Distance between (charge/nuclear) centres	r	m	
(Electric) dipole moment	μ_n	C m	
Polarisability volume	α_n'	m^3 (Å^3, cm^3)	
Ionisation energies	I_n	J	[kg m^2 s^{-2}]
J is Joule; F is Farad; C is Coulomb			
Chapter 4			
(Time dependent) induced dipole moment	μ_{ind} or $\mu_{ind}(t)$	C m	
(Time dependent) electronic polarisability	$\alpha(\nu_v)$	C m^2 V^{-1}	
(Oscillating) electric field (of light)	$E(\nu_v)$	V m^{-1}	
Absorbance	A or $A(\lambda)$	arbitrary units	
Optical density	$OD(\lambda)$	arbitrary units	
Pathlength (optical)	l	cm	
Extinction coefficient	ε_{max} or $\varepsilon(\lambda)$	l mol^{-1} cm^{-1}	
Biological macromolecular concentration	c_M	mol l^{-1}	
Wavelength	λ	nm (m, Å)	
Molecular weight (of protein)	M_p	D or kD	[g mol^{-1}]
Molecular weight (of nucleotide)	M_{nt}	D or kD	[g mol^{-1}]
Concentration (of nucleotide)	c_{nt}	mol l^{-1}	
Differential absorbance	$\Delta A(\lambda)$	arbitrary units	
Differential molar extinction coefficient	$\Delta\varepsilon(\lambda)$	l mol^{-1} cm^{-1}	
Ellipticity	$\theta(\lambda)$	deg	
Molar ellipticity	$[\theta(\lambda)]$	deg l mol^{-1} cm^{-1}	
Vibrational frequency of light	ν_v	s^{-1}	
(Equilibrium) electric field (of light)	E_o	V m^{-1}	
Equilibrium polarisability component	$\alpha_o(\nu_v)$	C m^2 V^{-1}	
Nuclear oscillation component	$\alpha_R(\nu_R)$	C m^2 V^{-1}	

Frequency of vibrational modes (molecular)	ν_R	s^{-1}	
Frequency of emitted light	ν_{em}	s^{-1}	
Planck's constant	h	J s or N m s	[kg m^2 s^{-1}]
Reduced Planck's constant	$h/2\pi$ or $\hbar$	J s rad^{-1}	
Speed of light	c	m s^{-1}	
Radiative lifetime (fluorescence)	τ_R	s	
Radiative lifetime (phosphorescence)	$\tau_{R,\,Phor}$	s	
Rate of spontaneous emission (fluorescence)	k_F	s^{-1}	
Rate of internal conversion (fluorescence)	k_{IC}	s^{-1}	
Rate of intersystem crossing (fluorescence)	k_{IS}	s^{-1}	
Rate of quenching (fluorescence) (with **Q**)	k_q	M^{-1} s^{-1}	
Fluorescence intensity (no **Q**)	I_{em} or F_o	arbitrary units	
Fluorescence intensity (in presence of **Q**)	F	arbitrary units	
Förster length	R_o	m	
Interfluorophore distance	R_F	m	
Fluorescence quantum yield	ϕ_F		
Fluorescence quantum yield (of donor, **D**)	ϕ_D		
Refractive index	n_R		
X-ray absorption coefficient	μ_{ab}	m^{-1}	
Incident intensity (of X-ray)	I_{xo}	arbitrary units	
Transmitted intensity (of X-rays)	I_x	arbitrary units	

V is Volt (J C^{-1}); D is Daltons; kD is kiloDaltons

Chapter 5

(Nuclear) angular momentum	J	J s rad^{-1}	
Gyromagnetic ratio	γ	rad s^{-1} T^{-1}	
Magnetic moment (z-axis)	μ_z	J T^{-1} or A m^2	
Nuclear magneton	μ_N	J T^{-1}	
Nuclear *g*-factor	g_I		
Charge (of an electron or proton)	e	C	
Mass (of proton)	m_p	kg	
Externally applied magnetic field strength	B_z	T or N m^{-1} A^{-1}	
Larmor (precession) frequency	ν_L	s^{-1}	
Coupling constant	J	s^{-1} (Hz)	
Boltzmann constant	k	J K^{-1}	[kg m^2 s^{-2} K^{-1}]
(Absolute) temperature	T	K	
(Scalar) longitudinal relaxation time constant	T_1	s	
Transverse relaxation time constant	T_2	s	
Longitudinal magnetisation: polarisation	$M_z(t)$		
Transverse magnetisation: coherence	$M_y(t)$		
Spectral line width (half peak intensity)	$\Delta\nu_{L,½}$	s^{-1} (Hz)	
(Electron) angular momentum	J_e	J s rad^{-1}	
Electron gyromagnetic ratio	γ_e	rad s^{-1} T^{-1}	
Electron magnetic moment	μ_z^e	J T^{-1}	
Bohr magneton	μ_B	J T^{-1}	
g-factor	g_e		
Mass (of an electron)	m_e	kg	

rad is radians (2π); T is Tesla; A is ampere (C s^{-1})

Chapter 6

Distance between lattice planes	d_{hkl}	Å	
Scattering length	$b_{x\text{-}ray}$	cm	
Distance of resolvable separation			

resolution	d_R	Å	
Charge (of an electron)	e	C	
Electrical potential difference (in field emission gun)	Φ	V or J C^{-1}	[kg m^2 s^{-2} C^{-1}]
Mass (of an electron)	m_e	kg	
Maximum particle size	D	m	
Büttiker-Landauer tunnelling time	τ^{BL}	s	
Variable (z-axis) barrier dimension	s_z	m	
Barrier crossing constant	χ	m^{-1}	
Piezo electric bar changes in length	Δl_p	m	
(Piezo electric biomorph) displacement	Δx_p	m	
(Piezo electric) potential difference	U_p	V or J C^{-1}	[kg m^2 s^{-2} C^{-1}]
(Piezo electric) coefficient	d_{31}	m V^{-1} or C N^{-1}	[C s^2 m^{-1} kg^{-1}]
Tunnelling current	I_T	A	
Van der Waals interactions (tip to surface)	$F_{VDW}(d_z)$	N	[kg m s^{-2}]
Hamaker constant	H	N m or J	[kg m^2 s^{-2}]
Distance (z-axis)	d_z	m	
Radius of tip above surface	R_z	m	
Surface-to-tip interaction forces	F_{ST}	N	[kg m s^{-2}]
Spring constant	c_{ST}	N m^{-1}	[kg s^{-2}]
Young's modulus	E_M	Pa or N m^{-2}	[kg m^{-1} s^{-2}]

Pa is Pascal (N m^{-2})

Chapter 7

Hydrated volume	V_h	cm^3 or m^3	
Macromolecular molecular weight	M_{MM}	D or kD	[g mol^{-1}]
Avogadro's number	N_o	mol^{-1}	
Macromolecular partial specific volume	V_{MM}	cm^3 g^{-1}	
Hydration level	Δ		
Coefficient of translational frictional force	$f_{trans,\,sph}$	kg s^{-1} or g s^{-1}	
Spherical macromolecular radius	r_{sph}	cm or m	
Viscosity	η	Pa s or N s m^{-2}	[kg m^{-1} s^{-1}, g cm^{-1} s^{-1}]
Coefficient of rotational frictional force	$f_{rot,\,sph}$	kg m^2 s^{-1} or g cm^2 s^{-1}	
Spherical macromolecular volume	V_{sph}	m^3 or cm^3	
General coefficient of translational frictional force	f_{trans}	kg s^{-1} or g s^{-1}	
General coefficient of rotational frictional force	f_{rot}	kg m^2 s^{-1} or g cm^2 s^{-1}	
Macromolecular flux	J_{MM}	kg m^{-2} s^{-1} or g cm^{-2} s^{-1} mol m^{-2} s^{-1}	
Macromolecular concentration	C_{MM}	kg m^{-3} or g cm^{-3} mol l^{-1}	
Average macromolecular velocity	$\langle \nu_{MM} \rangle$	m s^{-1} or cm s^{-1}	
Macromolecular diffusion coefficient	D_{MM}	m^2 s^{-1} or cm^2 s^{-1}	
Debye length	r_D	m	
Ionic strength	I	mol m^{-3} or mol kg^{-1} M (mol l^{-1})	
Association constant	K_a	M^{-1}	
Dissociation constant	K_d	M	
Moles (of ligand) bound per mole (of receptor)	B	(Mol fraction)	
Total molar quantity (of ligand) bound (to receptor)	m_{RL}	mol	
Total molar quantity (of ligand) added	m_{Lo}	mol	
Total system volume	V_{tot}	m^3, dm^3 (l), cm^3	
Chemical potential of species *i*	μ_i	J mol^{-1}	[kg m^2 s^{-2} mol^{-1}]

Concentration of species *i*	c_i	M (mol l^{-1})	
Molar gas constant	R	J K^{-1} mol^{-1}	[kg m^2 s^{-2} K^{-1} mol^{-1}]
Standard free energy change	ΔG^o	J mol^{-1}, kJ mol^{-1}	[kg m^2 s^{-2} mol^{-1}]
Standard enthalpy change	ΔH^o	J mol^{-1}, kJ mol^{-1}	[kg m^2 s^{-2} mol^{-1}]
Standard entropy change	ΔS^o	J mol^{-1} K^{-1}	[kg m^2 s^{-2} mol^{-1} K^{-1}]
(Exchangeable) heat energy	q	J	[kg m^2 s^{-2}]
(Fractional) change in enthalpy	dH	J	
Electric field	E_e	V m^{-1} or J C^{-1} m^{-1}	[kg m s^{-2} C^{-1}]
Electrophoretic velocity	ν_e	m s^{-1}	
Electrophoretic mobility	μ_e	m^2 V^{-1} s^{-1}	[C s kg^{-1}]
Apparent electophoretic mobility	μ_a	m^2 V^{-1} s^{-1}	[C s kg^{-1}]
EOF electophoretic mobility	μ_{EOF}	m^2 V^{-1} s^{-1}	[C s kg^{-1}]
Time to detector	t_e	s	
Effective length (of capillary)	l_e	m	
Total length (of apparatus)	L_e	m	
Applied potential difference	V_e	V or J C^{-1}	[kg m^2 s^{-2} C^{-1}]
Rate of association (complex formation) (***on***-rate)	k_{ass}	M^{-1} s^{-1}	
Rate of dissociation (complex) (***off***-rate)	k_{diss}	s^{-1}	
Resonant angle	Y_t	arc s	
Concentration dependent ***on***-rate (complex formation)	k_{on}	s^{-1}	

Chapter 8

Initial rate of (biocatalysis)	ν	M s^{-1}	[mol l^{-1} s^{-1}]
Initial substrate concentration	$[S]$	M	[mol l^{-1}]
Unimolecular rate constant for mechanism step **n**	k_n	s^{-1}	
Bimolecular rate constant for mechanism step **n**	k_n	M^{-1} s^{-1}	[l mol^{-1} s^{-1}]
Michaelis constant	K_m	M	[mol l^{-1}]
Equilibrium dissociation constant for **ES** complex	K_S	M	[mol l^{-1}]
Catalytic constant (when [**S**] $\gg K_m$)	k_{cat}	s^{-1}	
Maximum initial rate (when [**S**] $\gg K_m$)	V_{max}	M s^{-1}	[mol l^{-1} s^{-1}]
Specificity constant (when $K_m \gg$ [**S**])	k_{cat}/K_m	M^{-1} s^{-1}	[l mol^{-1} s^{-1}]
Inhibitor equilibrium dissociation constant	K_I	M	[mol l^{-1}]
Base equilibrium ionisation constant	K_d^B	M	[mol l^{-1}]
Acid equilibrium ionisation constant	K_d^A	M	[mol l^{-1}]
Saddle-point vibration frequency	ν_{TS}	s^{-1}	
Transition state forward decomposition rate constant	$k_C^{\ddagger}$	s^{-1}	
Quasi-equilibrium association constant	$K_C^{\ddagger}$	M^{-1}	
Microscopic rate constant	k_p	M^{-1} s^{-1}	
Partition function for molecular population **Z**	q^Z		
Transition state-ground state energy difference	E_o	J	[kg m^2 s^{-2}]
Standard free energy (of activation)	$\Delta G_o^{\ddagger}$	kJ mol^{-1}	[kg m^2 s^{-2} mol^{-1}]
Free energy (of activation) (from **E** and **S**)	$\Delta G_{ES}^{\ddagger}$	kJ mol^{-1}	[kg m^2 s^{-2} mol^{-1}]
Free energy (of activation) (from **ES** complex)	$\Delta G_T^{\ddagger}$	kJ mol^{-1}	[kg m^2 s^{-2} mol^{-1}]
Free energy (of association) of (**E** and **S**)	ΔG_S	kJ mol^{-1}	[kg m^2 s^{-2} mol^{-1}]
Rate constant for electron transfer	k_{ET}	s^{-1}	
Equilibrium association constant (for **D** and **A**)	$K_{a,DA}$	M^{-1}	[l mol^{-1}]
Edge to edge distance (between **D** and **A**)	R_{ET}	m	
Beta value	β_{ET}	m^{-1}	

Chapter 9			
Unitary charge of an ion	z		
Accelerating electrostatic potential	V_z	V or J C^{-1}	[kg m^2 s^{-2} C^{-1}]
Velocity of ion travel	ν_z	m s^{-1}	
Ion mass	m	D, kD (or amu)	
Time to detector	t_z	s	
Length (of field-free flight tube)	L_z	m	
Chapter 13			
Autocorrelation function decay constant	Γ	s^{-1}	
Nanomolecular diffusion constant	D_{NM}	m^2 s^{-1} or cm^2 s^{-1}	
Autocorrelation function wave vector	q	m^{-1} or cm^{-1}	
Zeta potential	ζ	V or J C^{-1}	[kg m^2 s^{-2} C^{-1}]
Chapter 14			
Concentration dependent water relaxivity rate	R_1	mM^{-1} s^{-1}	

Chemical glossary

Chapters 13 and 14

AtuFECT01	L-arginyl-2,3-L-diaminopropionic acid-*N*-palmityl-*N*-oleylamide trihydrochloride
CDAN	N^1-cholesteryloxycarbonyl-3,7-diazanonane-1,9-diamine
Chol	Cholesterol
DC-Chol	3β-[*N*-(*N*′,*N*′-dimethylaminoethane)carbamoyl] cholesterol
DLinDMA	1,2-dilinoleyloxy-3-dimethyl-aminopropane
DLin-KC2-DMA	2,2-dilinoleyl-4-(2-dimethyl-aminoethyl)-[1,3]-dioxolane
DLin-MC3-DMA (MC3)	(6Z, 9Z, 28Z, 31Z)-heptatriaconta-6, 9, 28, 31-tetra-en-19-yl 4-(dimethylamino)-butanoate
DODAG	*N*′, *N*′-dioctadecyl-*N*-4,8-diaza-10-aminodecanoyl glycine amide
DODMA	*N*-[1-(2,3-dioleyloxy) propyl]-*N,N*-dimethyl ammonium chloride
DOPC	1,2-dioleoyl-*sn*-glycero-3-phosphocholine or dioleoyl L-α-phosphatidylcholine
DOPE	1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine or dioleoyl L-α-phosphatidylethanolamine
DOPE-Rhoda	1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(lissamine rhodamine B sulfonyl) or dioleoyl L-α-phosphatidylethanolamine-*N*-(lissamine rhodamine B sulfonyl)
DPhyPE	l,2-diphytanoyl-*sn*-glycero-3-phosphoethanolamine or l,2-diphytanoyl L-α-phosphatidylethanolamine
DPPC	1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine or dipalmitoyl L-α-phosphatidylcholine
DSPC	1,2-distearoyl-*sn*-glycero-3-phosphocholine or distearoyl L-α-phosphatidylcholine
DSPE	1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine or distearoyl L-α-phosphatidylethanolamine
DS(14-yne)TAP	1,2-(distear-14-ynoyloxy)-3-(trimethylammonium) propane
folate-PEG^{2000}-DSPE	(folate-*N*-ω-polyethylene glycol 2000)-*N*-carboxy-1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine

Gd.DOTA.DSA	Gadolinium (III) 2-(4,7-bis-carboxymethyl-10-[(*N*,*N*-distearyl-amidomethyl)-*N*′-amidomethyl]-1,4,7,10-tetraazacyclododec-1-yl) acetic acid
GL-67	Lipid 67
PEG^{2000}-C-DMA	3-*N*-(ω-methoxy-polyethylene glycol 2000) carbamoyl-1,2-dimyristyloxypropylamine
PEG^{2000}-C-DMG	3-(ω-methoxy-polyethylene glycol 2000)carbamoyl-1,2-dimyristyl-*sn*-glycerol
PEG^{2000}-DSG	(ω-methoxy-polyethylene glycol 2000)-1,2-distearoyl-*sn*-glycerol
PEG^{2000}-DSPE	(ω-methoxy-polyethylene glycol 2000)-*N*-carboxy-1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine or (ω-methoxy-polyethylene glycol 2000)-*N*-carboxy-distearoyl L-α-phosphatidylethanolamine
PEG^{5000}-DMPE	PEG^{5000} conjugate of dimyristoyl-L-α-phosphatidylethanolamine

About the Companion Website

This book is accompanied by a companion website:

www.wiley.com/go/miller/essentialschembiol2

The website includes:
- List of Figures as PPTs
- List of Tables as PDFs

1

The Structures of Biological Macromolecules and Lipid Assemblies

1.1 General introduction

All living organisms are comprised of cells that may vary considerably in terms of size, shape and appearance. In complex multicellular organisms, many cells are organised into diverse, functional organs to perform a collective function (Figure 1.1). In spite of their wide morphological diversity, all cells of all living organisms, wherever they are located, are comprised of **proteins**, **carbohydrates**, **nucleic acids** and **lipid assemblies**. Together, these give a cell form and function. To know and understand the chemistry of these biological macromolecules is to comprehend not only the basic infrastructure of a cell but also of living organisms. In functional terms, macromolecular lipid assemblies provide for compartmentalisation in the form of membrane barriers that not only define the "outer limits" of each cell but also divide up the intracellular environment into different organelles or functional zones (Figure 1.2). Membrane barriers are fluidic and lack rigidity, so proteins provide a supporting and scaffolding function not only in the main fluid bulk of the cell, known as the **cytosol**, but also within organelles. Within the **nucleus**, proteins also provide a nucleic acid packaging function in order to restrain and constrain spectacular quantities of nucleic acids within the nuclear volume. Everywhere in any cell, proteins also perform other individualised functions in outer membranes (e.g. as pores or receptors), in organelle membranes (as selective transporters, redox acceptors or energy transducers), in the cytosol or organelle volumes (as enzyme catalysts, molecular chaperones or "communication and control" centres) and in the nucleus (as regulators and transcribers of the genetic code). The extraordinary variety of protein functions and the "work-horse" like nature of proteins in biology has made them endlessly fascinating to biochemists and now to chemical biologists alike.

Nucleic acids are found in two main classes, namely **deoxyribonucleic acid** (**DNA**) and **ribonucleic acid** (**RNA**). DNA is largely restricted to the nucleus and harbours genetic information that defines the composition and structure of cells and even the multicellular organisation of complex organisms, reaching even beyond that to influence organism behaviour as well. DNA molecules are partly segmented into **genes** that contain coding information for protein structures, but also into many other delineations associated with control over gene use. In fact, the level and sophistication of this control may well be the primary determinant of complexity in multicellular organisms, the more extensive and sophisticated the level of control, the more sophisticated and complex the multicellular organism. By contrast, RNA's most important role is in shuttling information from the nucleus to the cytosol. The primary function of RNA equates to the processing of genetic information from the DNA storage form into actual protein structures. RNA makes possible the central dogma of biology that ***genes code for proteins***. Finally, carbohydrates, if not stored in complex forms for primary metabolism, are known to decorate some intracellular proteins and attach to outer membrane proteins, forming a **glycocalyx** covering the surface of many cells,

Essentials of Chemical Biology: Structures and Dynamics of Biological Macromolecules In Vitro *and* In Vivo, Second Edition. Andrew D. Miller and Julian A. Tanner.

Companion Website: www.wiley.com/go/miller/essentialschembiol2

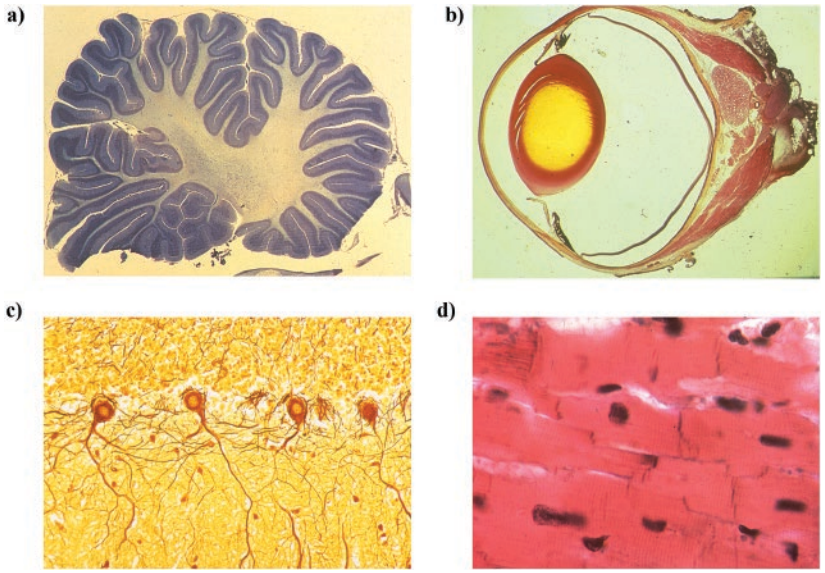


Figure 1.1 **Organs and cells.** (**a**) **Cross section of mammalian brain** showing the complex surface folds. There are an incalculable number of cells that make up the mammalian brain; (**b**) **Cross section of mammalian eye ball** in which the lens is made of proteins controlled in function by peripheral muscles. There is an enormous morphological and functional diversity between cells required for muscle control, light reception and signal transduction along the optic nerve; (**c**) **Cross section of mammalian neurological tissue** illustrating the **neuron cell bodies** with complex **axonal/dendritic processes** surrounded by support cells all of a wide range of size, shape, structure and function; (**d**) **Cross section of mammalian heart tissue** showing clusters of muscle fibres (single cell **myocytes**) that make up the heart wall. Myocytes are multinucleate with a very different shape, composition and function to neurological cells (all illustrations from Philip Harris Ltd).

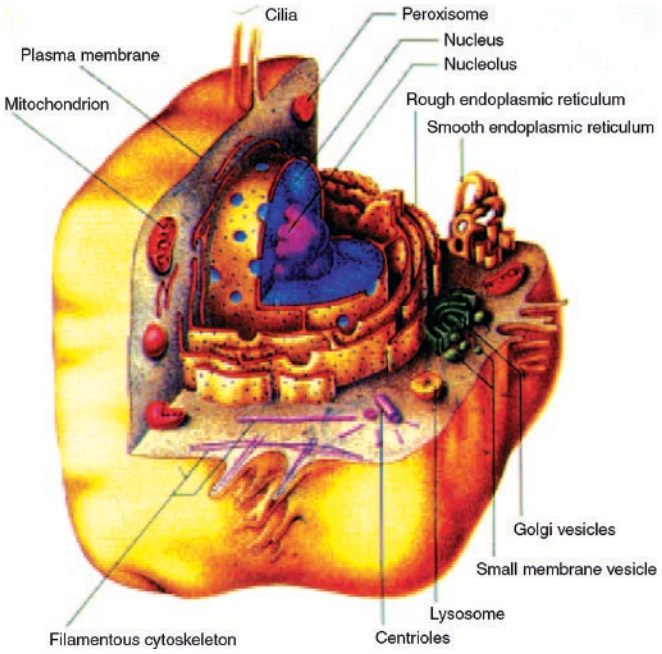


Figure 1.2 **General structure of a cell.** The main compartments (organelles) into which the interior is partitioned are illustrated. All cells of all organisms are constructed from the main biological macromolecules: **proteins**, **carbohydrates** and **nucleic acids**, together with macromolecular **lipid** structures that comprise the membranes (illustration from Philip Harris Ltd).

essential for communication between cells. In the plant and insect kingdoms, gigantic carbohydrate assemblies also provide the exo-skeleton framework to which cells are attached, giving form as well as function to plants and insects alike.

In all cases, proteins, carbohydrates and nucleic acids are polymers built from standard basis sets of molecular building blocks. In a similar way, lipid assemblies are built from a standard basis set of lipid building blocks associated through non-covalent bonds. What all biological macromolecules and macromolecular assemblies have in common is that they adopt defined three-dimensional (3D) structures that are the key to their functions (dynamics, binding and reactivity). Remarkably, these 3D structures are not only central to function but they are the result of **weak, non-covalent forces** of association acting together with stereoelectronic properties inherent within each class of polymer or macromolecular

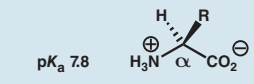
assembly. Without structure, function is hard to understand, although structure does not necessarily predict function. Therefore, the chemical biology reader needs to have a feel for the structures of proteins, carbohydrates, nucleic acids and lipid assemblies before embarking on any other part of this fascinating subject. Accordingly, the principles of structure are our main topic in this chapter, concluding with some explanation about critical, weak, non-covalent forces of association that are all so important in shaping and maintaining those structures.

1.2 Protein structures

1.2.1 Primary structure

Proteins are polymers formed primarily from the linear combination of 20 naturally occurring **L-α-amino acids** which are illustrated (see Table 1.1 and Figure 1.3). Almost all known protein structures are constructed from this fundamental set of 20 α-amino acid building blocks, which fall into two main classes, **hydrophobic** and **hydrophilic**, according to the nature of their **side chain** (Table 1.1). Protein architecture is intimately dependent upon having two such opposite sets of α-amino acid building blocks to call upon. Individual α-amino acid building blocks are joined together by a **peptide link** (Figure 1.4). When a small number (2–20) of amino acids are joined together by peptide links to form an unbranched chain then the result is known as an **oligopeptide** (Figure 1.5). However, peptide links can join together

Table 1.1 Structures and simple properties of all naturally occurring L-α-amino acids that are found in all proteins of all organisms. Included are the full name, the **three-letter code** name and the **one-letter code** name. Amino acids are grouped into those with **hydrophobic** side chains (**left panel**) and those with **hydrophilic** side chains (**right panel**). Where appropriate, measured functional group pK_a values are given.



R	Name	Abbrev.	R	Name	Abbrev.	pK_a
H—	Glycine	Gly (G)	HO	Tyrosine	Tyr (Y)	9.7
Me—	Alanine	Ala (A)	HO	Serine	Ser (S)	15
	Valine	Val (V)				
	Leucine	Leu (L)	HO H	Threonine	Thr (T)	15
			HS	Cysteine	Cys (C)	9.1
H	Isoleucine	Ile (I)	HO_2C	Aspartic acid	Asp (D)	4.0
	Phenylalanine	Phe (F)	H_2N O	Asparagine	Asn (N)	
HN	Tryptophan	Trp (W)	HO_2C	Glutamic acid	Glu (E)	4.5
			O H_2N	Glutamine	Gln (Q)	
S	Methionine	Met (M)	H_2N	Lysine	Lys (K)	10.4
			NH H_2N N H	Arginine	Arg (R)	12
N H CO_2H	Proline	Pro (P)	N HN	Histidine	His (H)	6.0

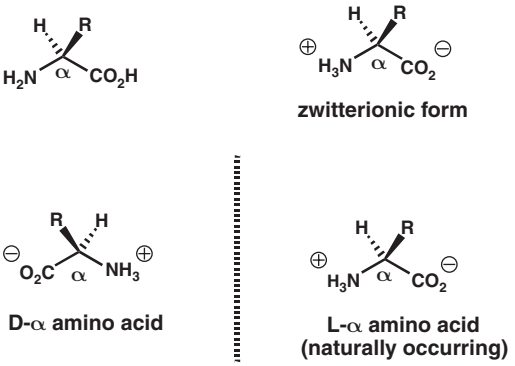


Figure 1.3 **Structures of α-amino acids.** L-α-amino acids are the preferred monomeric building blocks of proteins.

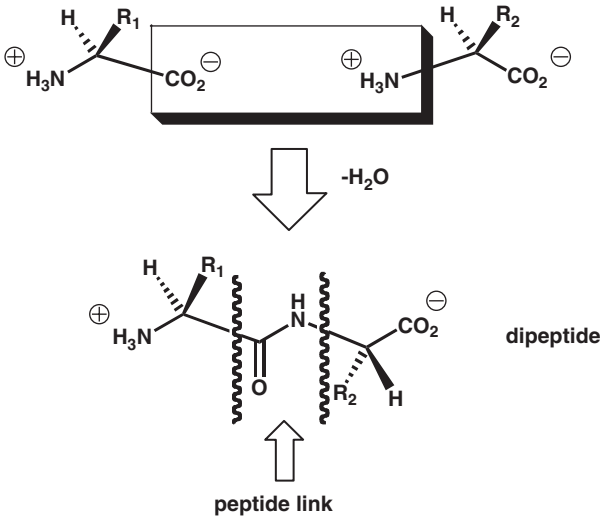


Figure 1.4 **Schematic of peptide link formation.**

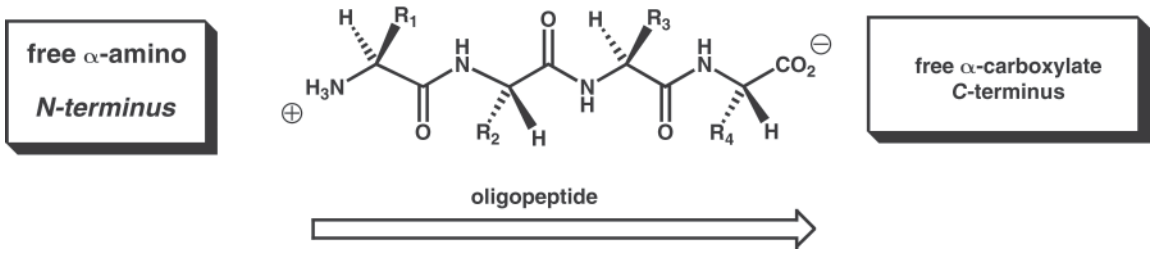


Figure 1.5 **General structure of tetrapeptide.** By convention the highest priority end is the free *N*-terminus and the lowest priority the free *C*-terminus, giving the backbone a directionality illustrated by the arrow. This convention applies for all peptides, polypeptides and proteins.

anything from 20 to 2000 amino acid residues in length to form substantial unbranched polymeric chains of L-α-amino acids. These are known as **polypeptides**. Within each polypeptide chain, the repeat unit $(\text{-N-C}_\alpha\text{-C(O)-})_n$, neglecting the α-amino acid side chains, is known as the **main chain** or **backbone**, whilst each constituent, linked α-amino acid building block is known as an **amino acid residue**. The order of amino acid residues, going from the free, uncombined α-amino terminal end (*N*-terminus) to the free, uncombined α-carboxyl terminus (*C*-terminus), is known as the **amino acid sequence**.

Quite clearly, each peptide link is in fact a simple secondary amide functional group, but with some unusual properties. In fact, the N, H, C and O atoms of a peptide link, together with each pair of flanking α-carbon atoms, actually

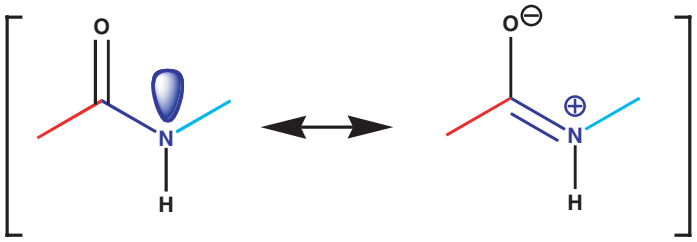


Figure 1.6 **Peptide link resonance structures.** These illustrate the partial double character in the C(O)-N bond (blue) sufficient to prevent free rotation, thereby restricting conformational freedom of peptide or polypeptide backbones.

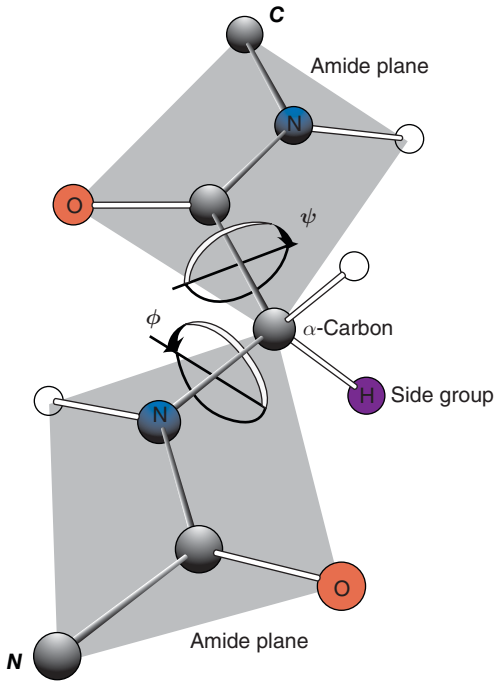


Figure 1.7 **Peptide link as virtual bond.** The C, O, N and H atoms act as a rigid coplanar unit, equivalent to a single bond (**virtual bond**) so that consecutive peptide links act as rigid coplanar units that pivot around individual C_α-atoms (from Voet *et al.*, 1999, Wiley, Figure 6.4).

form a rigid, coplanar unit that behaves almost like a single bond, owing to restricted rotation about the N-C(O) bond caused by nitrogen atom lone pair resonance and the build up of N-C(O) double bond character (Figure 1.6). For this reason, the peptide link and flanking C_α-atoms together are sometimes referred to as a **virtual bond**. We might say that the C_α-atom of each amino acid residue in a polypeptide chain belongs simultaneously to two such virtual bonds (Figure 1.7). The spatial relationship between each C_α-linked pair of virtual bonds is then defined using the conformational angles ϕ and ψ, which are the main chain dihedral angles subtended about the N(H)-C_α and C_α-C(O) σ-bonds respectively of each amino acid residue (Figures 1.7 and 1.8). Only certain combinations of ϕ and ψ are now allowed owing to steric congestion between the side chains of adjacent amino acid residues (Figure 1.8). Consequently, the overall conformation of a given polypeptide chain is also very restricted, with direct consequences for the 3D structures of proteins. In effect, conformational restrictions imposed by lack of free-rotation in the peptide link and the natures of each peptide-linked amino acid residue, place substantial restrictions upon the conformational freedom of a given polypeptide and hence the range of possible 3D structures that may be formed by any given polypeptide polymer. In fact, the primary structure amino acid sequence of a protein not only influences the 3D structure but also actually determines that structure. In other words, all the necessary "information" for the 3D structure of a protein is "stored" and is available within the primary structure. This is the basis for self-assembly in biology and explains why proteins can be such excellent platforms or "workbenches" for the development of defined functions and the evolution of living organisms.

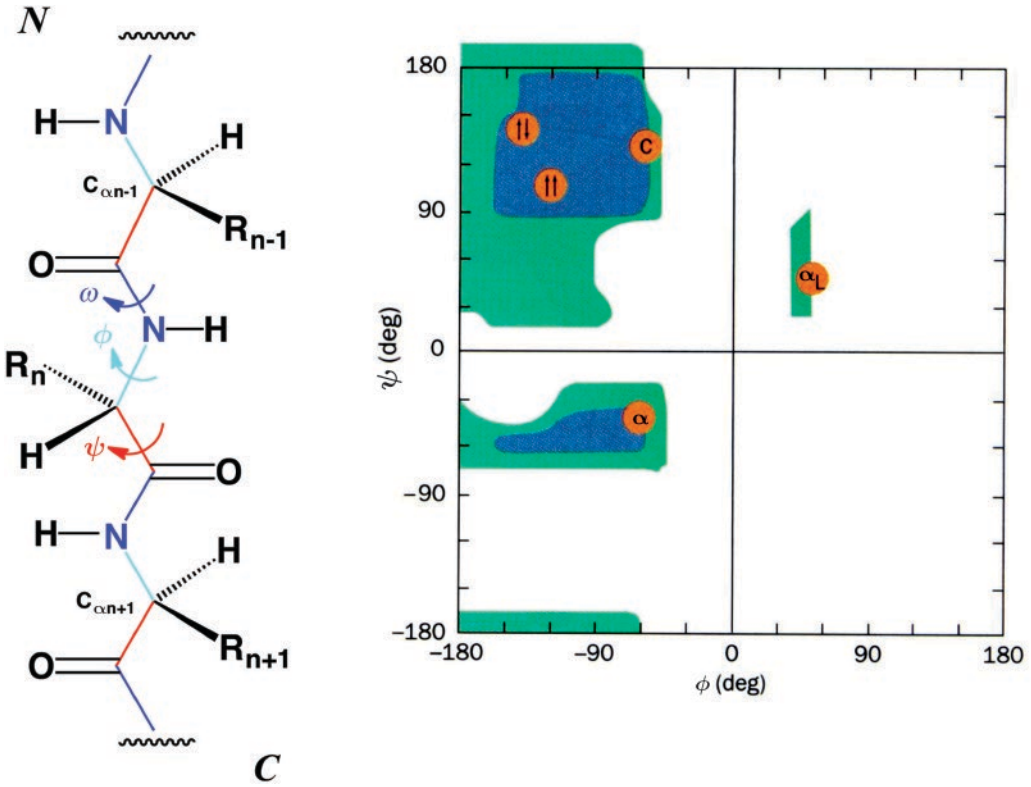


Figure 1.8 **Conformational restrictions in oligopeptide or polypeptide backbones.** Amino acid residue side-chain interactions further restrict free rotation in oligopeptide or polypeptide backbones. Rotational possibilities are defined by allowed values of dihedral angle ϕ subtended about N-C_α bond and ψ subtended about C_α-C(O) bond (left).Theoretically allowed angles are shown in **Ramachandran plot** (right), together with positions of actual angles found in real protein secondary structures: α: right-handed **α-helix**; α_L: **left-handed α-helix**; ↑↑: **parallel β-sheet**; ↑↓: **anti-parallel β-sheet**; **C**: **collagen** or **P_{II} helix** (see Figures 1.30 and 1.31). (Ramachandran plot from Voet *et al.*, 1999, Wiley, Figure 6.6).

1.2.2 Repetitive secondary structure

If primary structure is amino acid residue sequence, then **secondary structure** represents the first major steps towards a functional 3D structure. Secondary structures are essentially transient 3D structural elements that polypeptides may form in solution and which can interlock or dock together for stability. Polypeptides are capable of forming remarkably beautiful helical structures that are known as the right-handed **α-helix** and the right-handed **3_{10}-helix**. The term "right-handed" refers to the way in which the polypeptide main chain traces out the path of a right-handed corkscrew (incidentally, a **left handed α-helix** is possible, but is unknown in natural proteins so far). The α-helix can be a surprisingly sturdy, robust and regular structural feature (Figures 1.9 and 1.10). Typically, α-helices are comprised of up to 35 amino acid residues in length and are very stereo-regular; the ϕ and ψ conformational angles of each amino acid residue in the α-helix are both about -60° in all cases (Figure 1.11). Helices are held together by a regular network of non-covalent hydrogen bonds (see Section 1.6) formed between the peptide bond C = O and N-H groups of neighbouring amino acid residues (Figure 1.12). There are 3.6 amino acid residues per turn, with the result that the hydrogen bonds are formed between the C = O group lone pairs (hydrogen bond acceptors) of n-th residues and the N-H groups (hydrogen bond donors) of $(n + 4)$-th residues. The closed loop formed by one of these hydrogen bonds and the intervening stretch of polypeptide main chain contains 13 atoms (Figure 1.12). Hence, the α-helix has also been christened a **3.6_{13}-helix**. By contrast, the 3_{10}-helix (or **α_{II}-helix**) is effectively a smaller and slightly distorted version of the α-helix but with only three amino acid residues per turn and ten atoms involved in the intervening stretch of polypeptide main chain (Figures 1.13 and 1.14). Hydrogen bonds are therefore formed between the C = O group lone pairs of n-th residues and the N-H groups (hydrogen bond donors) of $(n + 3)$-th residues; ϕ and ψ conformational angles are approximately -60° and -30° respectively (Figure 1.15).

Sheet-like structures are the main alternative to helices. The origin of these structures can be found in the behaviour of polypeptide chains when they are fully extended into their **β-strand** conformations. A β-strand has a "pleated" appearance, with the peptide bonds orientated perpendicular to the main chain and with amino acid residue side chains alternating above and below (Figure 1.16). Both ϕ and ψ conformational angles are near 180°, but are typically between -120–150° and +120–150° respectively (Figure 1.17). All **β-strand** conformations are unstable alone, but may be stabilised

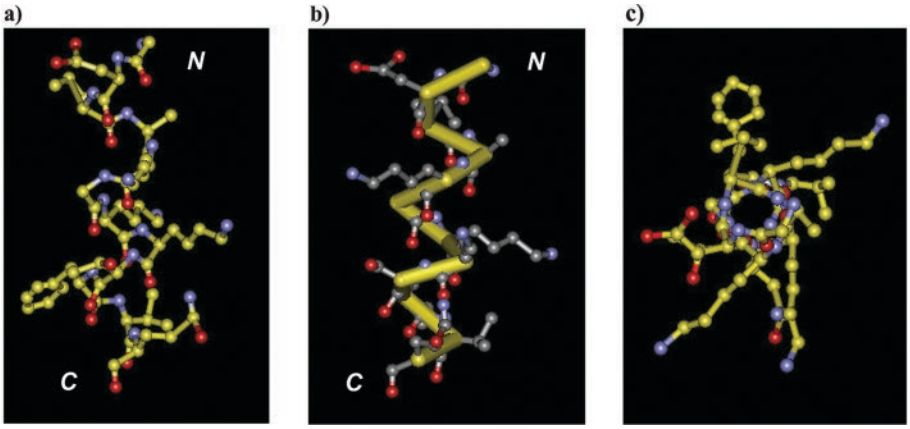


Figure 1.9 **Depictions of an α-helix.** The illustrated α-helix derives from **triose phosphate isomerase** (chicken muscle) (pdb: **1tim**). (**a**) **Ball and stick representation** (side view) of atoms and bonds shown with carbon (**yellow**), nitrogen (**blue**) and oxygen (**red**); (**b**) **CA stick display** of α-carbon backbone, atoms and bonds of amino acid side chains are rendered in **ball and stick representations** with carbon (**grey**), nitrogen (**blue**) and oxygen (**red**); (**c**) **Ball and stick representation** (top view) of atoms and bonds with labels as per (a).

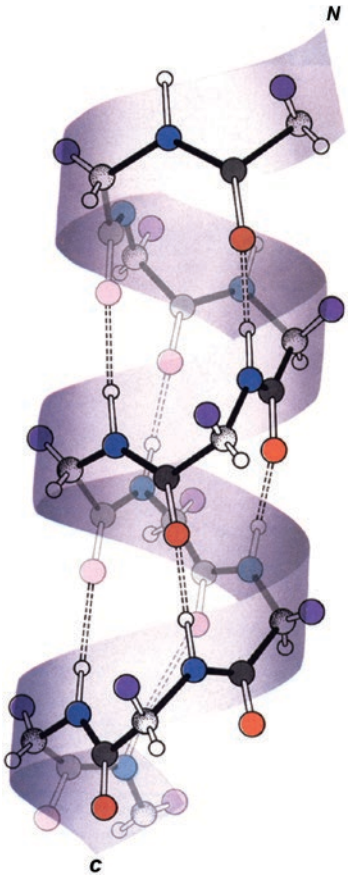


Figure 1.10 **Cartoon rendition of α-helix.** Here the right-hand helix path is illustrated as a **ribbon** over which a **ball and stick representation** of the α-carbon backbone is drawn using the code hydrogen (**white**), carbon (**grey**), nitrogen (**blue**) oxygen (**red**) and side chain atom (**purple**), in order to illustrate general **hydrogen-bonding** patterns in the helix (illustration from Voet *et al.*, 1999, Wiley, Figures 6–8).

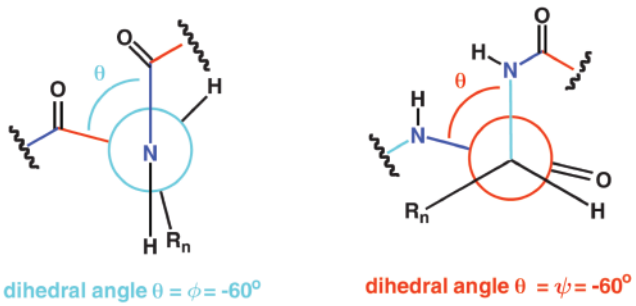


Figure 1.11 **Newman projections for α-helix.** These projections involve N-C_α bonds and C_α-C(O) bonds of the α-helix to demonstrate the consequences of highly regular dihedral angles ϕ and ψ respectively. Peptide backbone bonds are colour coded in the same way as Figures 1.6 and 1.8.

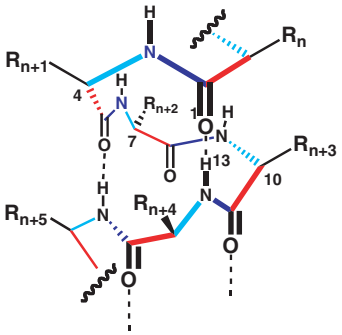


Figure 1.12 **Stereo-defined structure of first turn of an α-helix.** This demonstrates the atom separation between N-H hydrogen bond donors and C = O hydrogen bond acceptors. The C = O acceptor of each *n*-th residue forms a hydrogen bond link with the N-H bond donor of the (*n* + 4)-th residue defining an atom separation of 13 between acceptor O-atom and donor H-atom. Peptide backbone bonds are colour coded in the same way as Figures 1.6 and 1.8.

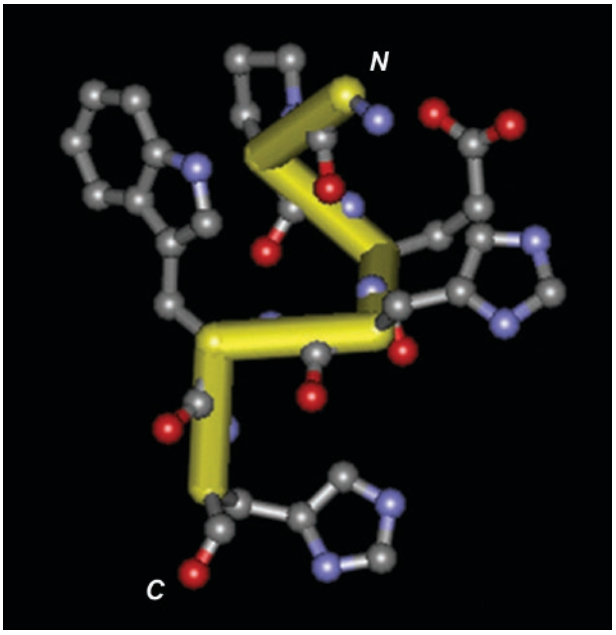


Figure 1.13 **Depiction of 3_{10}-helix (turn).** The illustrated 3_{10}-helix derives from triose phosphate isomerase (chicken muscle) (pdb: **1tim**). **CA stick display** of α-carbon backbone, atoms and bonds of amino acid side chains are rendered in **ball and stick representation** with carbon (**grey**), nitrogen (**blue**) and oxygen (**red**).

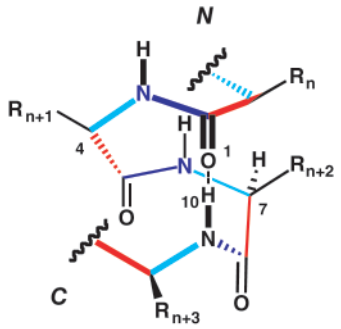


Figure 1.14 **Stereo-defined structure of an a 3_{10}-helix**. This structural representation demonstrates. This structural representation demonstrates the atom separation between N-H hydrogen bond donors and C = O hydrogen bond acceptors. The C = O acceptor of each *n*-th residue forms a hydrogen bond link with the N-H bond donor of the (*n* + 3)-th residue defining an atom separation of 10 between acceptor O-atom and donor H-atom. Peptide backbone bonds are colour coded in the same way as Figures 1.6 and 1.8.

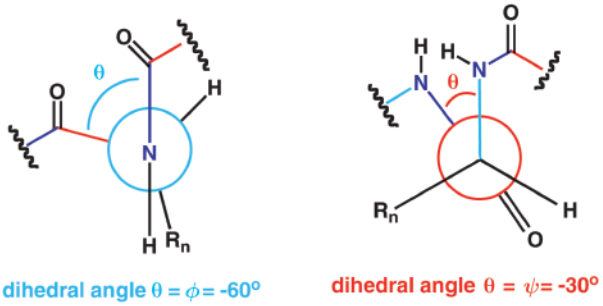


Figure 1.15 **Newman projections for a 3_{10}-helix**. These projections involve N-C_α bond and C_α-C(O) bonds of the 3_{10}-helix to illustrate the result of highly regular dihedral angles ϕ and ψ respectively. Peptide backbone bonds are colour coded in the same way as Figures 1.6 and 1.8. A tighter turn relates to smaller value of ψ.

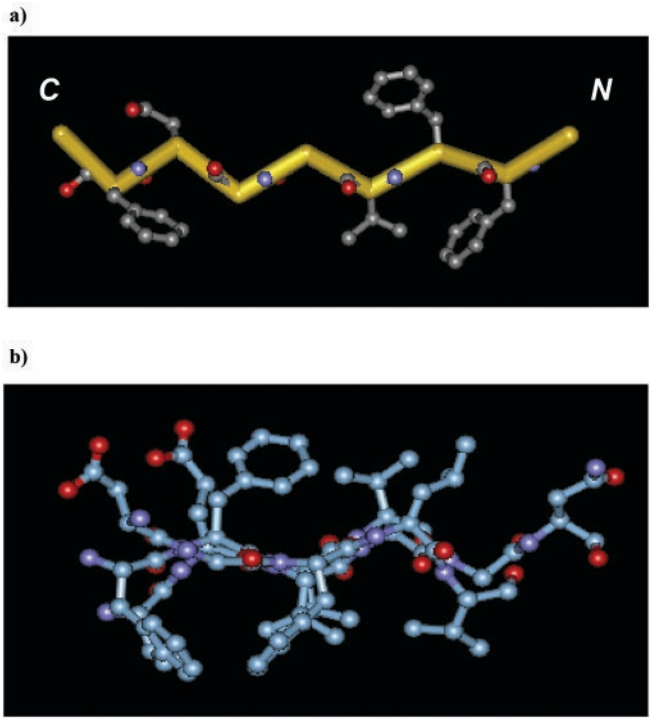


Figure 1.16 **β-Strand and β-sheet**. The illustrated β-strand and β-sheet derive from. from triose phosphate isomerase (chicken muscle) (pdb: **1tim**). (**a**) **CA stick display** of α-carbon backbone (side view), atoms and bonds of amino acid side chains are rendered in **ball and stick representation** with carbon (**grey**), nitrogen (**blue**) and oxygen (**red**). **β-strand** is shown to illustrate "zig-zag" extended conformation; (**b**) **Ball and stick representation** of **β-sheet** (side view) is shown with carbon (**grey**), nitrogen (**blue**) and oxygen (**red**) to illustrate "zig-zag" pleating and to show regular arrangement of amino acid residue side chains in close juxtaposition.

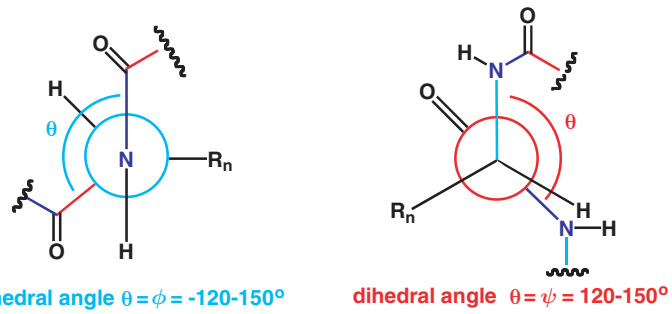


Figure 1.17 **Newman projections for β-strand.** These projections involve N-C_α bond and C_α-C(O) bonds of a β-strand to demonstrate the consequences of highly regular dihedral angles ϕ and ψ respectively and extending the conformation. Peptide backbone bonds are colour coded in the same way as Figures 1.6 and 1.8.

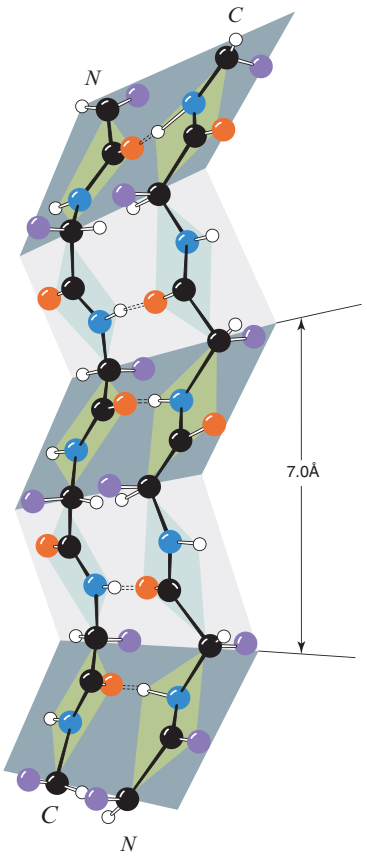


Figure 1.18 **Cartoon rendition of a β-sheet.** This rendition illustrates the pleating as a sequence of intersecting planes over which a **ball and stick representation** of the α-carbon backbone is drawn using the code hydrogen (**white**), carbon (**black**), nitrogen (**blue**) oxygen (**red**) and side-chain atom (**purple**), in order to illustrate general positioning of side chains above and below the sheet (illustration from Voet *et al.*, 1999, Wiley, Figures 6–10).

by the formation of non-covalent hydrogen bonds between strands, thereby resulting in a **β-sheet** (Figures 1.16 and 1.18). Such β-sheets may either be **anti-parallel** (β) or **parallel** (β′) depending upon whether the β-strands are orientated in the opposite or same direction with respect to each other (Figures 1.19 and 1.20). The hydrogen bonds that link β-strands together are formed between the same functional groups as in helices. In anti-parallel β-sheets, hydrogen bonds are alternately spaced close together then wide apart; in parallel β-sheets they are evenly spaced throughout (Figure 1.20).

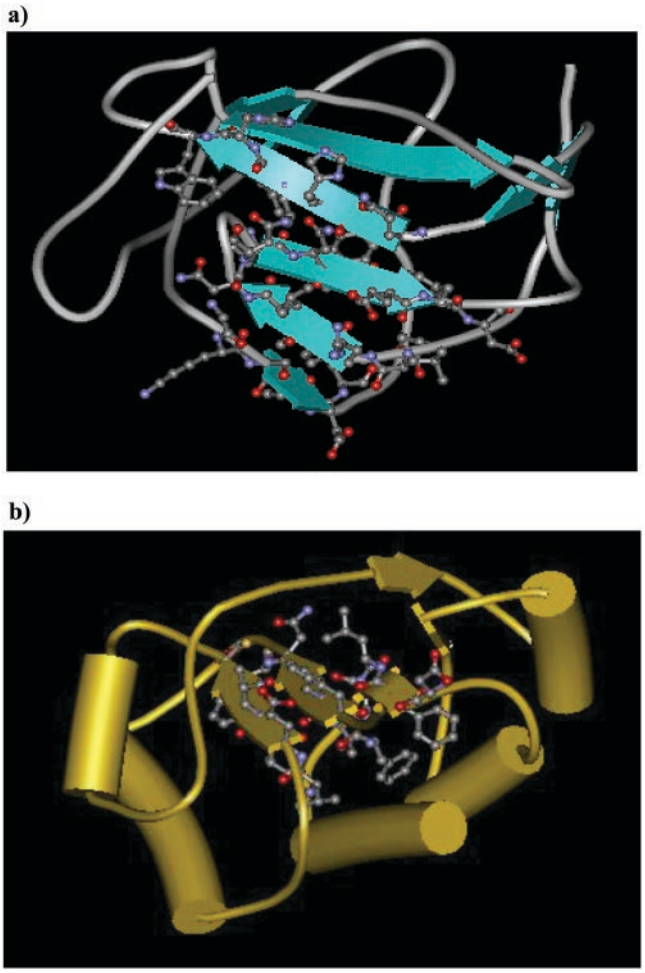


Figure 1.19 **Depiction of β-sheet structures from indicated proteins.** (**a**) **Schematic display structure** (see Section 1.2.5; each flat arrow is a β-strand with arrow head equal to *C*-terminus of each strand: cylinders are α-helices; remainder represent loops and turns) of **anti-parallel β-sheet segment** of **carbonic anhydrase I** (human erythrocyte) (pdb: **2cab**), atoms and bonds of amino acid side-chains are rendered in **ball and stick representation** with carbon (**grey**), nitrogen (**blue**) and oxygen (**red**); (**b**) **Schematic display structure** of **parallel β-sheet segment** of triose phosphate isomerase (chicken muscle) (pdb: **1tim**), atoms and bonds of side-chains are rendered as in (a).

1.2.3 Non-repetitive secondary structure

Helices and sheet-like structures are linked and/or held together by turns and loops in a given polypeptide main chain. **Tight turns** in the main chain (also known as **β-bends** or **β-turns**) are very common. These typically involve four amino acid residue units held together by a non-covalent hydrogen bond between C = O group lone pairs of the n-th residue and the N-H group of the $(n + 3)$-th residue. Given variations in the possible ϕ and ψ angles of the amino acid residues involved, there are at least six possible variants. However, these are usually divisible into just two main classes **Type I** and **Type II** that differ primarily in the conformation of the peptide link between the second and third residues of the turn (Figures 1.21 and 1.22). **Loops** in the main chain are also very common, but interactions between amino acid residue side chains provide stability rather than peptide link-associated **hydrogen bonding**. Consequently, the path mapped out by the main chain in forming a loop is a good deal less regular than that found in a tight turn (Figure 1.23). Occasionally, **disulfide bridges** in polypeptides replace and/or supplement peptide links. These bridges are formed between the thiol-functional groups of two different cysteine (Cys, C) residues separated by at least two other amino acid residues from each other in the amino acid sequence of a polypeptide. These may be thought of as the polypeptide equivalent of a "tie bar" or some other such reinforcing device. Both right and left-handed spiral forms are known and a series of conformational angles $(\chi_1, \chi_2, \chi_3, \chi_2', \chi_1')$ define the state of each given disulfide bridge (Figures 1.24 and 1.25).