Essentials of Chemical Biology

Structure and Dynamics of Biological Macromolecules

Andrew Miller
Department of Chemistry, Imperial College London, UK

Julian Tanner
Department of Biochemistry, University of Hong Kong

John Wiley & Sons, Ltd

Essentials of Chemical Biology

Essentials of Chemical Biology

Structure and Dynamics of Biological Macromolecules

Andrew Miller
Department of Chemistry, Imperial College London, UK

Julian Tanner
Department of Biochemistry, University of Hong Kong

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstrasse 12, 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 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Boulevard, Mississauga, Ontario, Canada L5R 4J3

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

Miller, Andrew, 1963–
Essentials of chemical biology : structure and dynamics of biological
macromolecules / Andrew Miller and Julian Tanner.
p. ; cm.
ISBN 978-0-470-84530-1 (cloth : alk. paper)
1. Biochemistry. 2. Molecular biology. 3. Macromolecules.
I. Tanner, Julian, 1976– II. Title.
[DNLM: 1. Biochemistry. 2. Chemistry. 3. Macromolecular Substances–chemistry.
4. Molecular Biology. QU 4 M6565e 2007]
QP514.2.M55 2007
612′.015–dc22

2007050369

British Library Cataloguing in Publication Data

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

ISBN 978-0-470-84530-1 (HB) 978-0-470-84531-8 (PB)

Typeset in 10.5/12.5pt Minion by Aptara Inc., New Delhi, India
Printed and bound by Grafos SpA., Barcelona, Spain
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.

To Izumi,

without whose love, patience, common sense
and great encouragement, this book
may never have been completed

Contents

Preface

Mapping the essentials of chemical biology

Chemical biology is a new, rapidly emerging branch of chemistry that represents all aspects of chemical endeavour, devoted to understanding the way biology works at the molecular level. Chemical biology is unashamedly inter-disciplinary, and chemical biology research is essentially problem driven and not discipline driven. Organic, physical, inorganic and analytical chemistry all contribute towards the chemical biology whole. 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. 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 graduate student and young 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 and appealing 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 measurable 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 4–6), followed by a consideration of dynamic behaviour and molecular interactions (Chapters 7–9), concluding with molecular evolution and thoughts on the origins of life, quintessentially from the chemistry point of view (Chapter 10). Armed with such essentials, we hope that readers will then be ready 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 hope 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 must leave their chemical principles behind them!

Andrew Miller
Julian Tanner

Glossary of physical terms

Chapter 1			
Potential energy	$\boldsymbol{V}$	J	[kg m^2 s^{-2}]
Electrical point charge	$\boldsymbol{q}_n$	C	
Vacuum permittivity	ε_0	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	$\boldsymbol{r}$	m	
(Electric) dipole moment	μ_n	C m	
Polarisability volume	$\boldsymbol{\alpha}'_n$	m^3(Å^3, cm^3)	
Ionisation energies	$\boldsymbol{I}_n$	J	[kg m^2 s^{-2}]
J is Joule; F is Farad; C is Coulomb	$\boldsymbol{\mu}_{ind}$ or $\boldsymbol{\mu}_{ind}$	C m	
Chapter 4			
(Time dependent) induced dipole moment			
(Time dependent) electronic polarisability	α ($\boldsymbol{\nu}_v$)	C m^2 V^{-1}	
(Oscillating) electric field (of light)	$\boldsymbol{?}$($\boldsymbol{\nu}_v$)	V m^{-1}	
Absorbance	$\boldsymbol{A}$ or $\boldsymbol{A}$ ($\boldsymbol{\lambda}$)	arbitrary units	
Optical density	$\boldsymbol{OD}$ ($\boldsymbol{\lambda}$)	arbitrary units	
Pathlength (optical)	$\boldsymbol{l}$	cm	
Extinction coefficient	$\varepsilon_{???}$ or gε ($\boldsymbol{\lambda}$)	l mol^{-1} cm^{-1}	
Biological macromolecular concentration	$\boldsymbol{c}_M$	mol l^{-1}	
Wavelength	λ	nm	
Molecular weight (of protein)	$\boldsymbol{M}_p$	D or kD	[g mol^{-1}]
Molecular weight (of nucleotide)	$\boldsymbol{M}_{nt}$	D or kD	[g mol^{-1}]
Concentration (of nucleotide)	$\boldsymbol{c}_{nt}$	mol l^{-1}	
Differential absorbance	$\boldsymbol{\Delta}$A($\boldsymbol{\lambda}$),	arbitrary units	

Essentials of Chemical Biology Andrew Miller and Julian Tanner

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)	$?_0$	V m^{-1}	
Equilibrium polarisability component	$\alpha_0(\nu_v)$	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}]
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_Fs^{-1}		
Rate of internal conversion (fluorescence)	k_{IC}s^{-1}		
Rate of intersystem crossing (fluorescence)	k_{IS}s^{-1}		
Rate of quenching (fluorescence)	k_qs^{-1}		
Fluorescence intensity (no **Q**)	I_{em} or F_0	arbitrary units	
Fluorescence intensity (in presence of **Q**)	F	arbitrary units	
FÖrster length	R_0m		
Interfluorophore distance	R_Fm		
Fluorescence quantum yield	ϕ_F		
Fluorescence quantum yield (of donor, **D**)	ϕ_D		
Refractive index	$?_R$		
X-ray absorption coefficient	μ_{ab} m^{-1}		
Incident intensity (of X-ray)	I_{x0}	arbitrary units	
Transmitted intensity (of X-rays)	I_x	arbitrary units	
*V is Volt (*J C^{-1}*); D is Daltons and kD kiloDaltons*	J	J s rad^{-1}	[kg rad s^{-1}]

Chapter 5

(Nuclear) angular momentum			
Gyromagnetic ratio	γ	rad s^{-1} T^{-1}	
Magnetic moment (z-axis)	μ_z	J T^{-1} or A m^2	
Nuclear magneton	μ_NJ T^{-1}		

Nuclear g-factor	g_I		
Charge (of an electron)	e C		
Mass (of proton)	$m_p$kg		
Applied magnetic field	B_zT or N m^{-1}A^{-1}		
Lamor (precession) frequency	ν_Ls^{-1}		
Planck's constant	h J s	[kg m^2 s^{-1}]	
	$h/2\pi$	J s rad^{-1}	[kg rad 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	$?_z(?)$		
Transverse magnetisation – coherence	$?_y(?)$		
Spectral line width (half peak intensity)	$\Delta\nu_{L,1/2}$	s^{-1}(Hz)	
(Electron) angular momentum	J_e	J s rad^{-1}	[kg rad s^{-1}]
Electron gyromagnetic ratio	γ_e	rad s^{-1} T^{-1}	
Electron magnetic moment	$\mu_z{}^e$ 7	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})	d_{hkl}	Å	

Chapter 6

Distance between lattice planes			
Scattering length	b_{X-ray}	cm	
Distance of resolvable separation - resolution	d_R	Å	
Planck's constant	h	J s	[kg m^2 s^{-1}]
Charge (of an electron)	e C		
Electrical potential difference (in Field Emission gun)	Φ	V or ?g?^{tx}	[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	Φ	m^{-1}	
Piezo electric bar changes in length	Δl_p	m	
(Piezo electric biomorph) displacement	Δx_p	m	

(Piezo electric) potential difference	$\boldsymbol{U}_{\mathrm{p}}$	V or ?g?$^{\mathrm{tx}}$	[kg m^2 s^{-2}C^{-1}]
(Piezo electric) coefficient	$\boldsymbol{d}_{31}$	m V^{-1}or C N^{-1}	[C s^2 m^{-1} kg^{-1}]
Tunnelling current	$\boldsymbol{I}_{\mathrm{T}}$A		
Van-der-Waals interactions (tip to surface)	$\boldsymbol{F}_{\mathrm{VDW}}(\boldsymbol{d}_{\mathrm{z}})$	N	[kg m s^{-2}]
Hamaker constant	$\boldsymbol{H}$	N m or J	[kg m^2 s^{-2}]
Distance (z-axis)	$\boldsymbol{d}_{\mathrm{z}}$	m	
Radius of tip above surface	$\boldsymbol{R}_{\mathrm{z}}$	m	
Surface-to-tip interaction forces	$\boldsymbol{F}_{\mathrm{ST}}$	N	[kg m s^{-2}]
Spring constant	$\boldsymbol{c}_{\mathrm{ST}}$	N m^{-1}	[kg s^{-2}]
Youngs Modulus	$\boldsymbol{E}_{\mathrm{M}}$	Pa or N m^{-2}	[kg m^{-1} s^{-2}]
Pa is Pascal (N m^{-2})	$\boldsymbol{V}_{\mathrm{h}}$	cm^{3}or m^3	
Chapter 7			
Hydrated volume			
Macromolecular molecular weight	$\boldsymbol{M}_{\mathrm{MM}}$	D or kD	[g mol^{-1}]
Avogadro's number	$\boldsymbol{N}_0$	mol^{-1}	
Macromolecular partial specific volume	$\boldsymbol{V}_{\mathrm{MM}}$	cm^3 g^{-1}	
Hydration level	$\boldsymbol{\Delta}$		
Coefficient of translational frictional force	$\boldsymbol{f}_{\mathrm{trans,sph}}$	kg s^{-1}or g s^{-1}	
Spherical macromolecular radius	$\boldsymbol{r}_{\mathrm{sph}}$	cm or m	
Viscosity	$\boldsymbol{\eta}$	Pa s or N m^{-2} s	[kg m^{-1} s^{-1}, g cm^{-1} s^{-1}]
Coefficient of rotational frictional force	$\boldsymbol{f}_{\mathrm{rot,sph}}$	kg m^2 s^{-1} or g cm^2 s^{-1}	
Spherical macromolecular volume	$\boldsymbol{V}_{\mathrm{sph}}$	m^3 or cm^3	
General coefficient of translational frictional force	$\boldsymbol{f}_{\mathrm{trans}}$	kg s^{-1}or g s^{-1}	
General coefficient of rotational frictional force	$\boldsymbol{f}_{\mathrm{rot}}$	kg m^2 s^{-1} or g cm^2 s^{-1}	
Macromolecular flux	$\boldsymbol{J}_{\mathrm{MM}}$ mol m^{-2} s^{-1}	kg m^{-2} s^{-1} or g cm^{-2} s^{-1}	
Macromolecular concentration	$\boldsymbol{C}_{\mathrm{MM}}$	kg m^{-3}or g cm^{-3} mol m^{-3}	
Average macromolecular velocity	$\langle \nu_{\mathrm{MM}} \rangle$	m s^{-1} or cm s^{-1}	
Macromolecular diffusion coefficient	$\boldsymbol{D}_{\mathrm{MM}}$	m^2 s^{-1} or cm^2 s^{-1}	
Boltzmann constant	$\boldsymbol{k}$	J K^{-1}	[kg m^2 s^{-2} K^{-1}]
Debye length	$\boldsymbol{r}_{\mathrm{D}}$	m	
Ionic strength	$\boldsymbol{?}$	mol m^{-3} or mol kg^{-1} M (mol l^{-1})	
Permittivity of medium	$\boldsymbol{\varepsilon}$	F m^{-1}or C^2 m^{-1} J^{-1}	[C^2 kg^{-1}m^{-3} s^2]
Association constant	$\boldsymbol{K}_{\mathrm{a}}$	M^{-1}	

Dissociation constant	$\boldsymbol{K}_d$	M	
Moles (of ligand) bound per mole (of receptor)	$\boldsymbol{B}$	(Mol fraction)	
Total molar quantity (of ligand) bound (to receptor)	$\boldsymbol{m}_{RL}$	mol	
Total molar quantity (of ligand) added	$\boldsymbol{m}_{L0}$	mol	
Total system volume	$\boldsymbol{V}_{tot}$	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	$\boldsymbol{c}_i$	M ($mol\ l^{-1}$)	
Molar gas constant	$\boldsymbol{R}$	$J\ K^{-1}\ mol^{-1}$	$[kg\ m^2\ s^{-2}\ K^{-1}\ mol^{-1}]$
Standard free energy change	$\boldsymbol{\Delta G}^0$	$J\ mol^{-1}$, $kJ\ mol^{-1}$	$[kg\ m^2\ s^{-2}\ mol^{-1}]$
Standard enthalpy change	$\boldsymbol{\Delta H}^0$	$J\ mol^{-1}$, $kJ\ mol^{-1}$	$[kg\ m^2\ s^{-2}\ mol^{-1}]$
Standard entropy change	$\boldsymbol{\Delta S}^0$	$J\ mol^{-1}K^{-1}$	$[kg\ m^2\ s^{-2}\ mol^{-1}K^{-1}]$
(Exchangeable) heat energy	$\boldsymbol{?}$	J	$[kg\ m^2\ s^{-2}]$
(Fractional) change in enthalpy	$\mathbf{d}\boldsymbol{H}$	J	
Electric field	$\boldsymbol{E}_e$	$V\ m^{-1}$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	$m^2\ V^{-1}\ s^{-1}$	$[C\ s\ kg^{-1}]$
EOFelectophoretic mobility	μ_{EOF}	$m^2\ V^{-1}\ s^{-1}$	$[C\ s\ kg^{-1}]$
Time to detector	$\boldsymbol{t}_e$	s	
Effective length (of capillary)	$\boldsymbol{l}_e$	m	
Total length (of apparatus)	$\boldsymbol{L}_e$	m	
Applied potential difference	$\boldsymbol{?}_e$	V or $J\ C^{-1}$	$[kg\ m^2\ s^{-2}\ C^{-1}]$
Rate of association (complex formation) (***on***-rate)	$\boldsymbol{k}_{ass}$	$M^{-1}s^{-1}$	
Rate of dissociation (complex) (***off***-rate)	$\boldsymbol{k}_{diss}$	s^{-1}	
Resonant angle	$\boldsymbol{Y}_t$	arc s	
Concentration dependent ***on***-rate (complex formation)	$\boldsymbol{k}_{on}$	s^{-1}	
Chapter 8	$\boldsymbol{\nu}$	$M\ s^{-1}$	$[mol\ l^{-1}\ s^{-1}]$
Initial rate of (biocatalysis)			
Initial substrate concentration	$[\mathbf{S}]$	M	$[mol\ l^{-1}]$
Unimolecular rate constant for mechanism step **n**	$\boldsymbol{k}_n$	s^{-1}	
Bimolecular rate constant for mechanism step **n**	$\boldsymbol{k}_n$	$M^{-1}\ s^{-1}$	$[l\ mol^{-1}\ s^{-1}]$
Michaelis constant	$\boldsymbol{K}_m$	M	$[mol\ l^{-1}]$
Equilibrium dissociation constant for **ES** complex	$\boldsymbol{K}_s$	M	$[mol\ l^{-1}]$

Catalytic rate constant (when [**S**] ≫ K_m)	k_{cat}	s^{-1}	
Maximum initial rate (when [**S**] ≫ K_m)	V_{max}	$M\ s^{-1}$	$[mol\ l^{-1}\ s^{-1}]$
Catalytic rate constant (whenK_m ≫ [**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 ionization constant	$K_d{}^B$	M	$[mol\ l^{-1}]$
Acid equilibrium ionization constant	$K_d{}^A$	M	$[mol\ l^{-1}]$
Saddle-point vibration frequency	$\nu_{??}$	s^{-1}	
Transition state forward decomposition rate constant	$k_c{}^{\ddagger}$	s^{-1}	
Quasi-equilibrium association constant	$K_c{}^{\ddagger}$.	M^{-1}	
Forward rate constant	k_p	$M^{-1}s^{-1}$	
Partition function for molecular population	$z\ q^z$		
Boltzmann constant	k	$J\ K^{-1}$	$[kg\ m^2\ s^{-2}\ K^{-1}]$
Transition state-ground state energy difference	E_0	J	$[kg\ m^2\ s^{-2}]$
Planck's constant	h	J s	$[kg\ m^2\ s^{-1}]$
Standard free energy (of activation)	$\Delta G_0{}^{\ddagger}$	$kJmol^{-1}$	$[kg\ m^2\ s^{-2}mol^{-1}]$
Free energy (of activation) (from **E** and **S**)	$\Delta G_{ES}{}^{\ddagger}$	$kJmol^{-1}$	$[kg\ m^2\ s^{-2}mol^{-1}]$
Free energy (of activation) (from **ES** complex)	$\Delta G_T{}^{\ddagger}$	$kJmol^{-1}$	$[kg\ m^2\ s^{-2}mol^{-1}]$
Free energy (of association) of (**E** and **S**)	$\Delta G_S\ kJmol^{-1}$	$[kg\ m^2\ s^{-2}mol^{-1}]$	
Molar gas constant	R	$J\ K^{-1}mol^{-1}$	$[kg\ m^2\ s^{-2}\ K^{-1}mol^{-1}]$
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	$?_z$	V or $J\ C^{-1}$	$[kg\ m^2\ s^{-2}\ C^{-1}]$
Velocity of ion travel	$\nu_?s^{-1}$		
Ion mass	**m**	D, kD (or a.m.u.)	
Time to detector	t_z	s	
Length (of field-free flight tube)	L_z	m	

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**. These together give a cell form and function. To know and understand the chemistry of these biological macromolecules is to comprehend the basic infrastructure not only of a cell but also of living organisms. In functional terms, macromolecular lipid assemblies provide for compartmentalisation in the form of membrane barriers, which 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 (as pores or receptors for example), 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

Essentials of Chemical Biology Andrew Miller and Julian Tanner

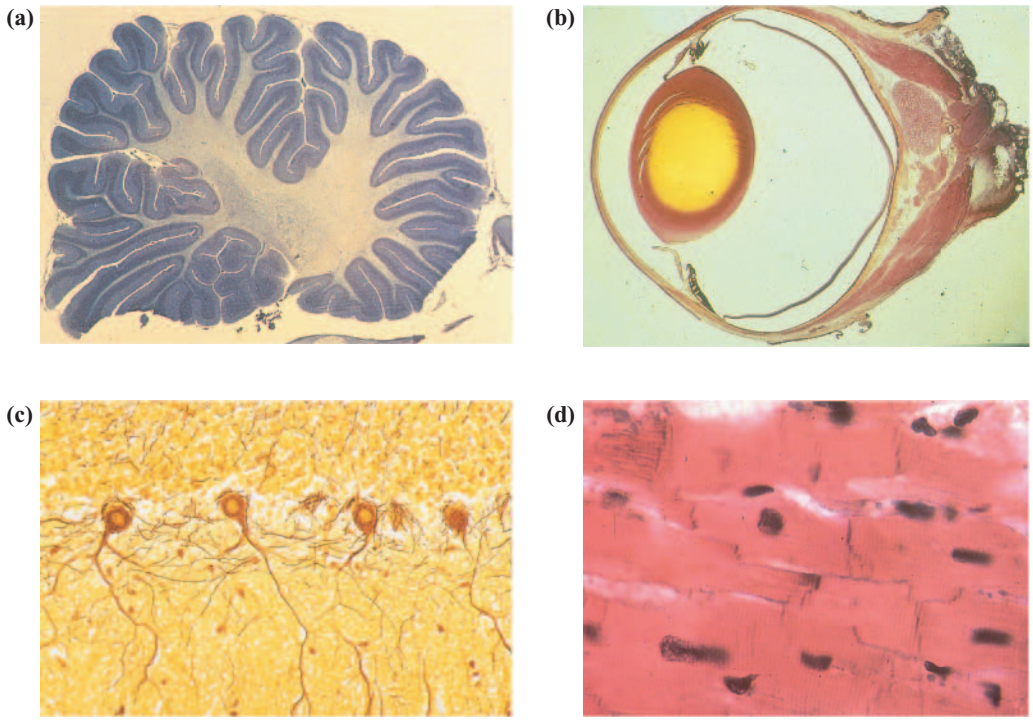


Figure 1.1 **Organs and Cells. (a) cross section of mammalian brain** showing the complex surface folds. There are an incalculable number of cells that make up the mammalian brain; **(b) cross section of mammalian eye ball** in which the lens is made of proteins controlled in function by peripheral muscles. 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, Weston Super Mare, UK).

of protein functions and the 'workhorse'-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 this 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

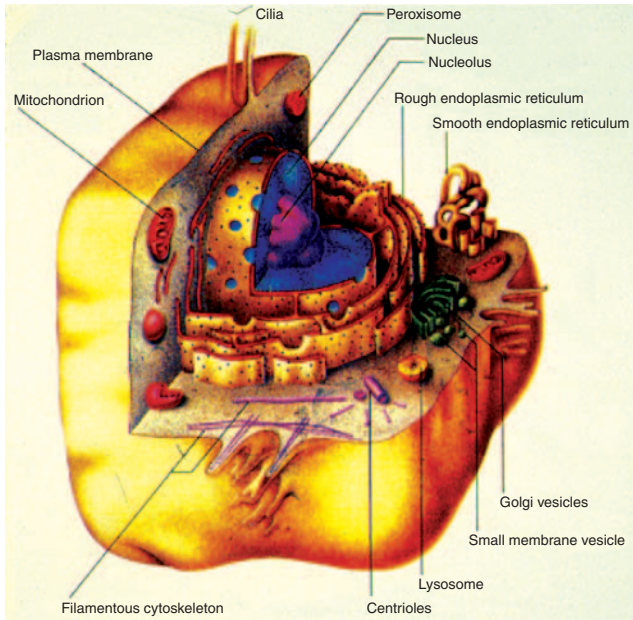


Figure 1.2 **General structure of a cell** showing the main compartments (organelles) into which the interior is partitioned. All cells of all organisms are constructed from the main biological macromolecules **proteins**, **carbohydrates**, and **nucleic acids**; together with macromolecular **lipid**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

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

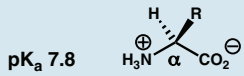
In all cases, proteins, carbohydrates and nucleic acids are polymers built from standard basis sets of molecular building blocks. In a similar way, lipid assemblies are built from a standard basis set of lipid building blocks associated through non-covalent bonds. What all biological macromolecules and macromolecular assemblies have in common is that they then adopt defined three dimensional structures that are the key to their functions (dynamics, binding and reactivity). Remarkably, these three dimensional 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 for Chapter 1, concluding with some explanation about those critical weak non-covalent forces of association that are all so important in shaping and maintaining these structures.

1.2 Protein structure

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 (Table 1.1) (Figure 1.3). Almost all known protein

Table 1.1 Structures of all **naturally occuring 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 Note that in Chapter 8, the term pK_a is replaced by pK_d^A or pK_d^B depending upon whether an acid or base dissociation is under consideration respectively.



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)	HO, H	Threonine	Thr (T)	15
	Leucine	Leu (L)	HS	Cysteine	Cys (C)	9.1
H	Isoleucine	lle (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
N H, CO_2H	Proline	Pro (P)	NH, H_2N, N H	Arginine	Arg (R)	12
			N, HN	Histidine	His (H)	6.0

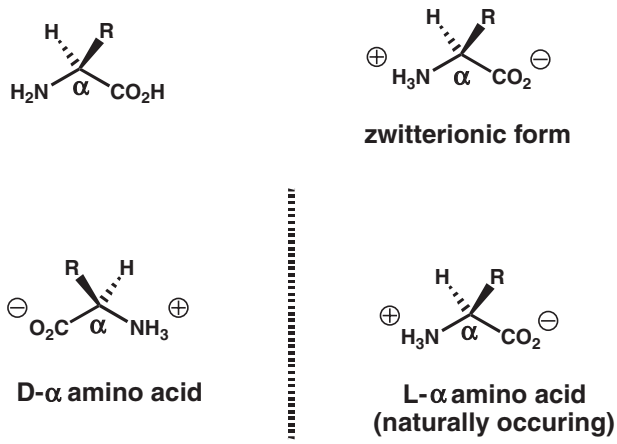


Figure 1.3 **Structures of α-amino acids** the monomeric building blocks of proteins.

structures are constructed from this fundamental set of 20 α-amino-acid building blocks. These building blocks fall into two main classes, **hydrophobic** and **hydrophilic**, according to the nature of their **side-chains** (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 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

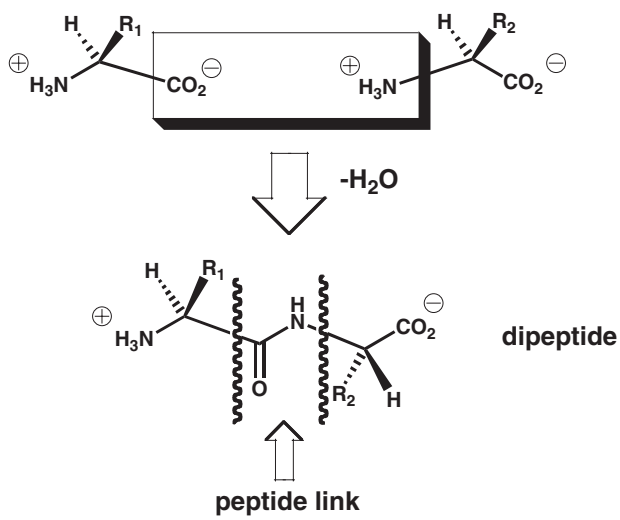


Figure 1.4 **Schematic of peptide link formation**, central virtual bond of polypeptide and protein backbones.

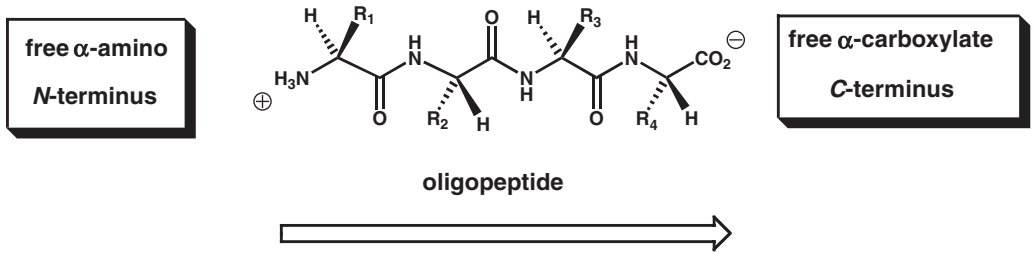


Figure 1.5 General structure of tetrapeptide; by convention the highest priority end is the free *N*-terminus and the lowest priority the free *C*-terminus giving the backbone a directionality illustrated by the arrow. This convention applies for all peptides, polypeptides and proteins.

each polypeptide chain, the repeat unit $(–N–C_\alpha–C(O)–)n$, 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 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 three dimensional structures of proteins. In effect, conformational restrictions imposed by lack of

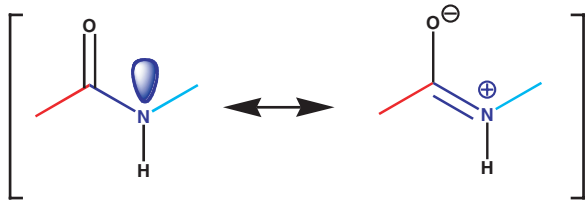


Figure 1.6 **Peptide link** resonance structures illustrating partial double character in the C(O)—N bond (blue) sufficient to prevent free rotation thereby restricting conformational freedom of peptide or polypeptide backbones.

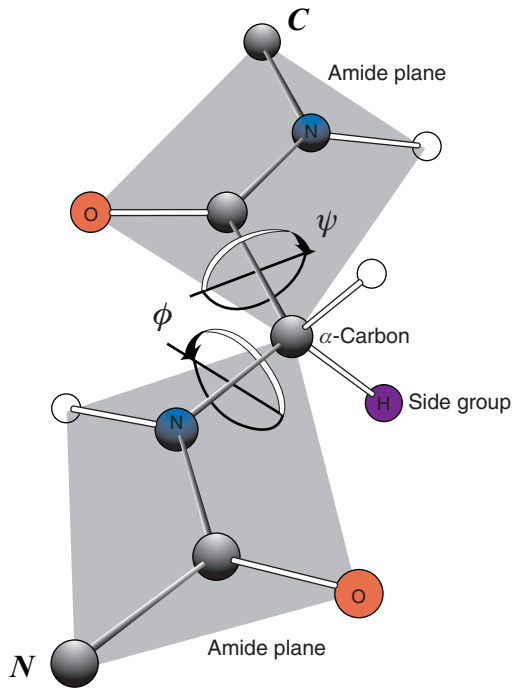


Figure 1.7 **Peptide link** C, O, N and H atoms act as a rigid coplanar units equivalent to a single bond (**virtual bond**) so that consecutive peptide links act as rigid coplanar units that pivot around individual C α-atoms (from Voet, Voet & Pratt, 1999 [Wiley], Fig. 6-4).

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 three dimensional 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 three dimensional structure but also actually determines this structure. In other words, all the necessary 'information' for the three dimensional structure of a protein is 'stored' and 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.

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 three dimensional structure. Secondary structures are essentially transient three dimensional structural elements that polypeptides may form in solution and that 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** (α_R)

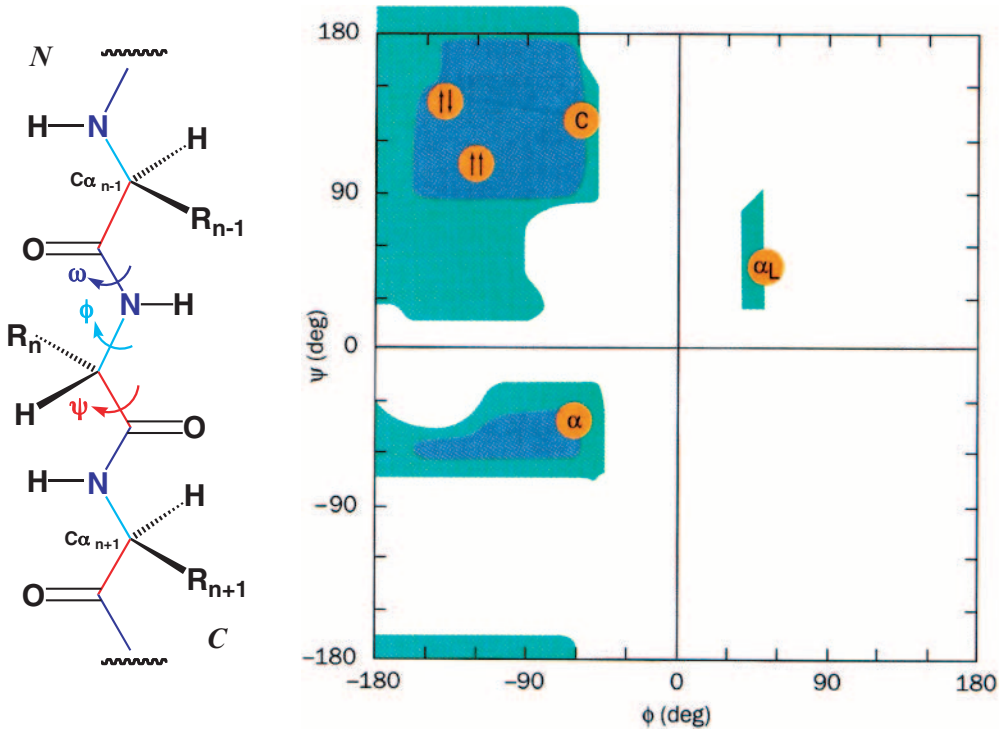


Figure 1.8 Amino acid residue side-chain interactions further restrict free rotation in peptide or polypeptide backbone. Rotational possibilities are defined by allowed values of dihedral angle ϕ subtended about N—C_α bond and ψ subtended about C_α—C(O) bond (left). Theoretically allowed angles are shown 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, P_{II} helix (see later). (Ramachandran plot from Voet, Voet & Pratt, 1999 [Wiley], Fig. 6-6).

and the right-handed $\mathbf{3_{10}}$ **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, the left-handed α helix, α_L, is possible but is unknown in natural proteins so far). The α_R helix can be a surprisingly sturdy, robust and regular structural feature (Figures 1.9 and 1.10). Typically, α_R 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_R 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 the nth residues and the N—H groups (hydrogen bond donors) of the $(n + 4)$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_R helix has also been christened a $\mathbf{3.6_{13}}$ **helix**. By contrast, the 3_{10} helix (or α_{II} **helix**) is effectively

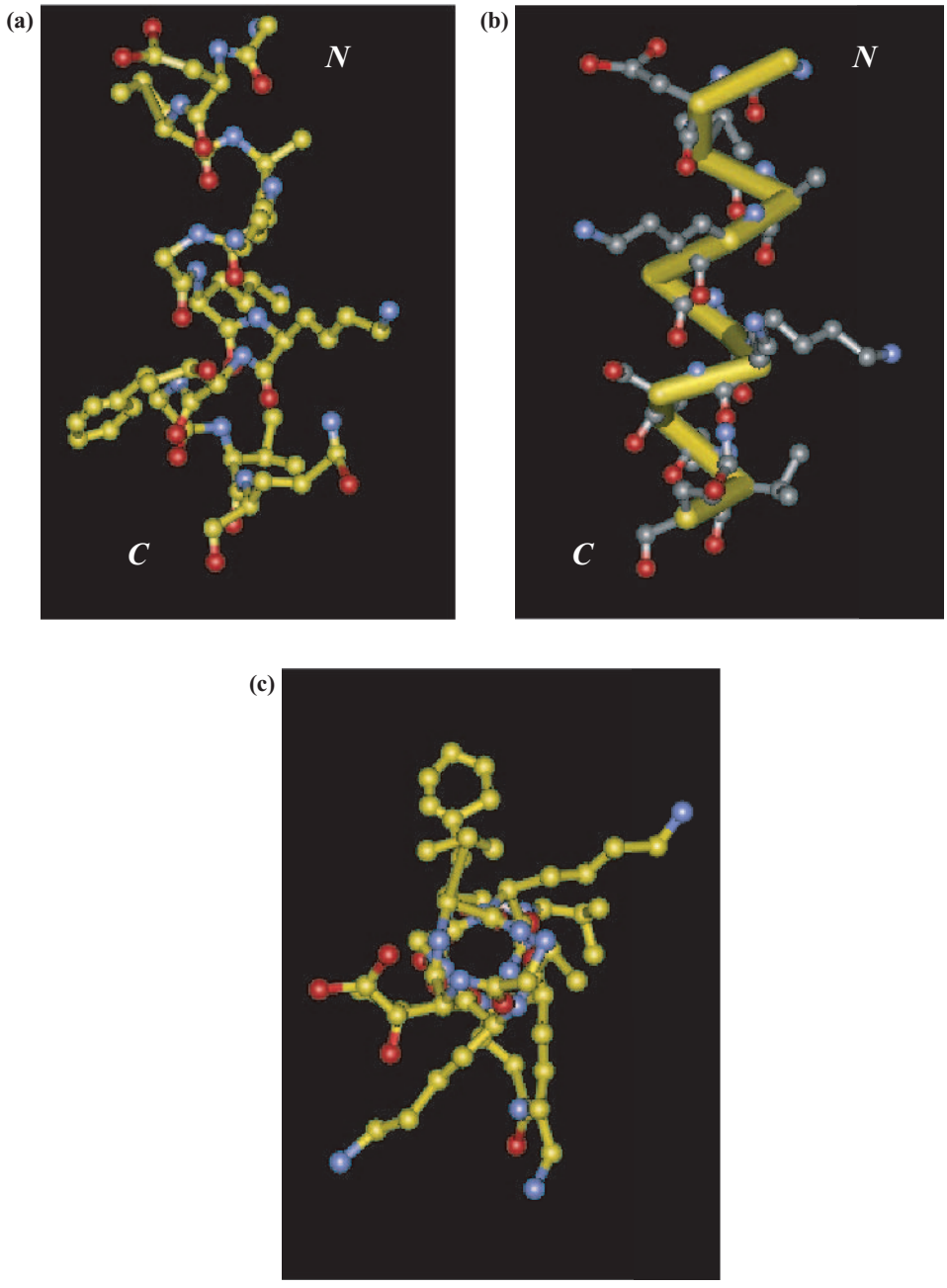


Figure 1.9 **Various depictions of an α-helix** from triose phosphate isomerase (chicken muscle) (pdb: **1tim**). (a) **ball and stick representation** (side view) of atoms and bonds shown with carbon (**yellow**), nitrogen (**blue**) and oxygen (**red**); (**b**) **CA stick display** of α-carbon backbone, 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).

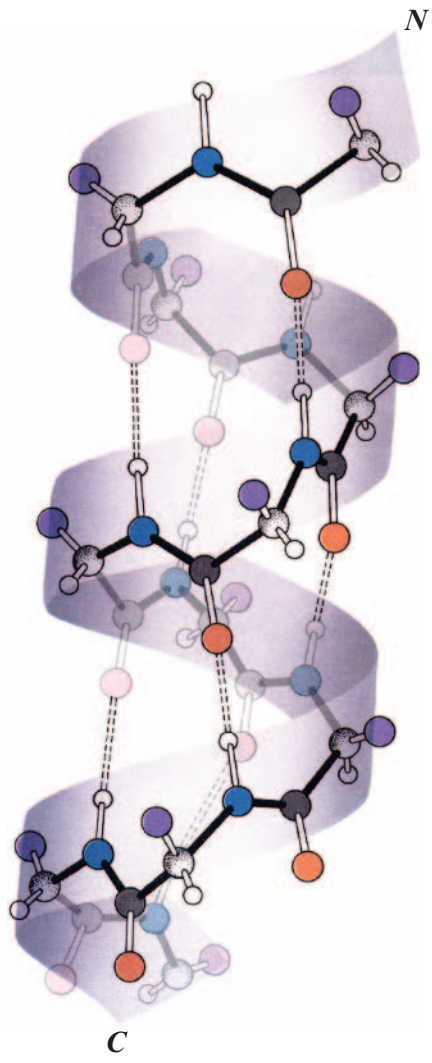


Figure 1.10 **Cartoon rendition of α-helix** in which right-hand helix path is illustrated as a ribbon over which a **ball and stick representation** of the **α-carbon backbone** is drawn using the code hydrogen (**white**), carbon (**grey**), nitrogen (**blue**) oxygen (**red**) and side chain atom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

a smaller and slightly distorted version of the α_R helix but with only three amino-acid residues per turn and 10 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 nth 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).

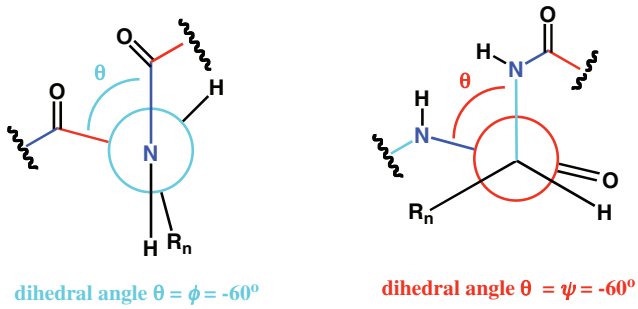


Figure 1.11 Newman projections involving the N—C_α bond and C_α—C(O) bond of the α-helix to demonstrate the consequences of highly regular dihedral angles ϕ and ψ respectively. Peptide backbone bonds are colour coded in the same way as in **Figs. 1.6** and **1.8**.

Sheetlike 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 and −150° and +120 and +150° respectively (Figure 1.17). All **β-strand** conformations are unstable alone, but may be stabilised 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 **antiparallel** (β) or **parallel** (β'), depending upon whether the β strands are orientated in opposite directions or the same direction with respect to each other (Figures 1.19 and 1.20). The hydrogen bonds that link β strands together are formed between

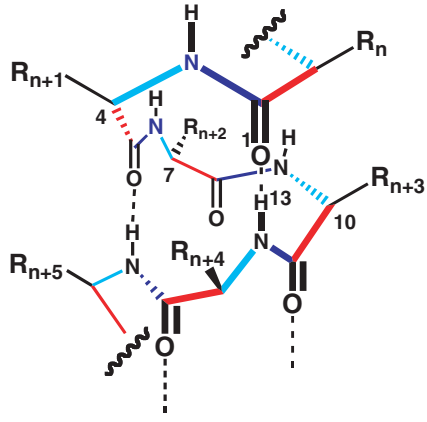


Figure 1.12 **Stereo-defined structure** of first turn of an α-helix to demonstrate the atom separation between N—H hydrogen bond donors and C=O hydrogen bond acceptors. The C=O acceptor of each n-th residue forms a hydrogen bond link with the N—H bond donor of the $(n + 4)$-th residue defining an atom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

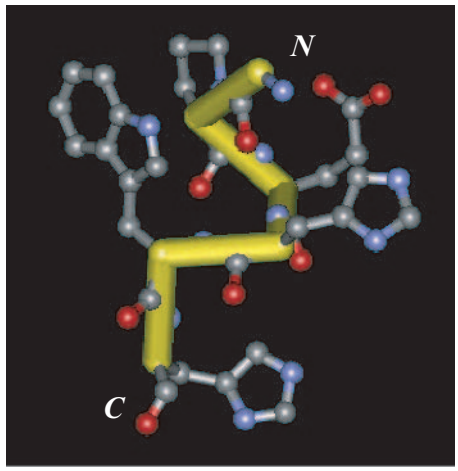


Figure 1.13 Depiction of 3_{10}-**helix** (turn) from triose phosphate isomerase (chicken muscle) (pdb: **1tim**). **CA stick display** of α-carbon backbone, atoms and bonds of amino acid side chains are rendered in ball and stick representations with carbon (**grey**), nitrogen (**blue**) and oxygen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

the same functional groups as in helices. In antiparallel (β) β sheets, hydrogen bonds are alternately spaced close together then wide apart; in parallel (β') β sheets they are evenly spaced throughout (Figure 1.20).

1.2.3 Non-repetitive secondary structure

Helices and sheetlike 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

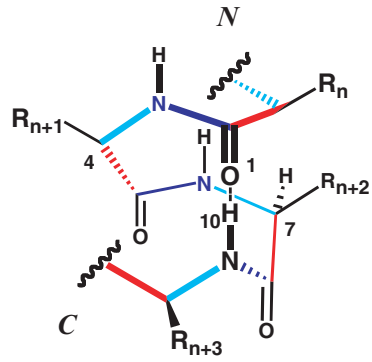


Figure 1.14 **Stereo-defined structure** of first turn of a 3_{10}-helix to demonstrate the atom separation between N—H hydrogen bond donors and C=O hydrogen bond acceptors. The C=O acceptor of each n-th residue forms a hydrogen bond link with the N—H bond donor of the $(n+3)$-th residue defining an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