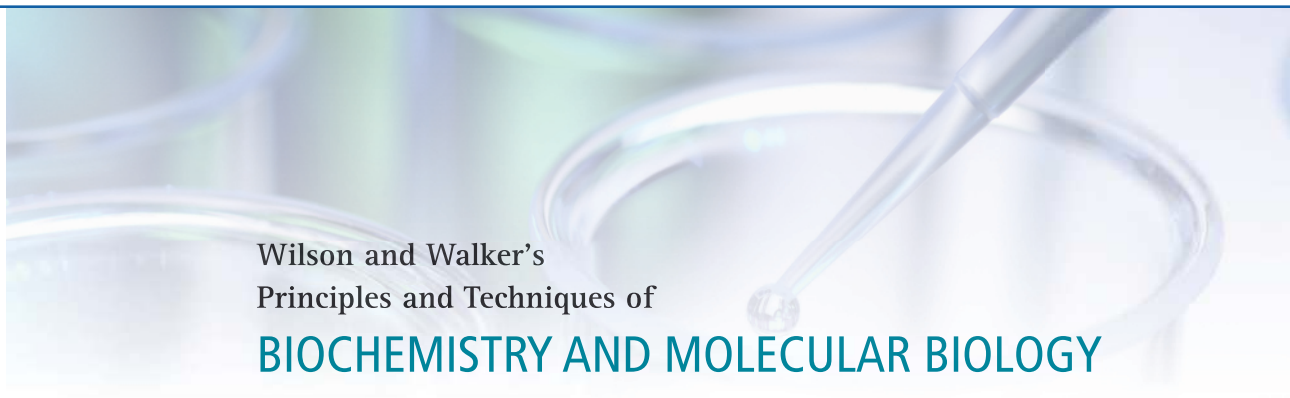


Wilson and Walker's
Principles and Techniques of

BIOCHEMISTRY AND MOLECULAR BIOLOGY

Bringing this best-selling textbook right up-to-date, the new edition uniquely integrates the theories and methods that drive the fields of biology, biotechnology and medicine, comprehensively covering both the techniques students will encounter in lab classes and those that underpin current key advances and discoveries. The contents have been updated to include both traditional and cutting-edge techniques most commonly used in current life science research. Emphasis is placed on understanding the theory behind the techniques, as well as analysis of the resulting data. New chapters cover proteomics, genomics, metabolomics and bioinformatics, as well as data analysis and visualisation. Using accessible language to describe concepts and methods, and with a wealth of new in-text worked examples to challenge students' understanding, this textbook provides an essential guide to the key techniques used in current bioscience research.

ASSOCIATE PROFESSOR ANDREAS HOFMANN is the Structural Chemistry Program Leader at Griffith University's Eskitis Institute and an Honorary Senior Research Fellow at the Faculty of Veterinary and Agricultural Sciences at The University of Melbourne. He is also Fellow of the Higher Education Academy (UK). Professor Hofmann's research lab focusses on the structure and function of proteins in infectious and neurodegenerative diseases, and also develops computational tools. He is the author of *Methods of Molecular Analysis in the Life Sciences* (2014) and *Essential Physical Chemistry* (2018).

DR SAMUEL CLOKIE is the Principal Clinical Bioinformatician at the Birmingham Women's and Children's Hospital, UK, where he leads the implementation of bioinformatic techniques to decipher next-generation sequencing data in clinical diagnostics. He is also Honorary Senior Research Fellow at the Institute of Cancer and Genomic Sciences at The University of Birmingham and before that was a research fellow at the National Institutes of Health in Bethesda, USA.

Wilson and Walker's
Principles and Techniques of

Biochemistry and Molecular Biology

Edited by

ANDREAS HOFMANN
Griffith University, Queensland

SAMUEL CLOKIE
Birmingham Women's and Children's Hospital

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781316614761

DOI: 10.1017/9781316677056

First published 2018

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

Library of Congress Cataloging-in-Publication data
Names: Wilson, Keith, 1936– editor. | Walker, John M., 1948– editor. | Hofmann, Andreas, editor. | Clokie, Samuel, editor.
Title: Wilson and Walker's principles and techniques of biochemistry and molecular biology / edited by Andreas Hofmann, Griffith University, Queensland, Samuel Clokie, Birmingham Women's and Children's Hospital.
Other titles: Principles and techniques of biochemistry and molecular biology.
Description: [2018 edition]. | Cambridge : Cambridge University Press, 2018. | Includes bibliographical
references and index.
Identifiers: LCCN 2017033452 | ISBN 9781107162273 (alk. paper)
Subjects: LCSH: Biochemistry – Textbooks. | Molecular biology – Textbooks.
Classification: LCC QP519.7 .P75 2018 | DDC 572.8–dc23
LC record available at https://lccn.loc.gov/2017033452

ISBN 978-1-107-16227-3 Hardback
ISBN 978-1-316-61476-1 Paperback

Additional resources for this publication at www.cambridge.org/hofmann

CONTENTS

FOREWORD TO THE EIGHTH EDITION

When the first edition of our book was published in 1975, biochemistry was emerging as a new discipline that had the potential to unify the previously divergent -ologies that constituted the biological sciences. It placed emphasis on the understanding of the structure, function and expression of individual proteins and their related genes, and relied on the application of analytical techniques such as electrophoresis, chromatography and various forms of spectrometry. In the ensuing 40 years the completion of the Human Genome Project has confirmed the central role of DNA in all of the activities of individual cells and the emergence of molecular biology as the means of understanding complex biological processes. This in turn has led to the new disciplines of bioinformatics, chemoinformatics, proteomics and metabolomics. The succeeding six editions of our book have attempted to reflect this evolution of biochemistry as a unifying discipline. All editions have placed emphasis on the experimental techniques that undergraduates can expect to encounter during the course of their university studies and to this end we have been grateful for the excellent feedback we have received from the users of the book.

The point has now been reached where we believed that it was appropriate for us to hand over the direction and academic balance of future editions to a new editorial team and to this end we are delighted that Andreas Hofmann and Samuel Clokie have agreed to take on the role. We wish them well and look forward to the continuing success of the book.

KEITH WILSON AND JOHN WALKER

PREFACE

It has been a tremendous honour being asked by Keith Wilson, John Walker and Cambridge University Press to take on the role of editorship for this eighth edition of *Principles and Techniques*. In designing the content, we extend the long and successful tradition of this text to introduce relevant methodologies in the biochemical sciences by uniquely integrating the theories and practices that drive the fields of biology, biotechnology and medicine. Methodologies have improved tremendously over the past ten years and new strategies and protocols that once were just applied by a few pioneering laboratories are now applied routinely, accompanied by ever more fluent boundaries between core disciplines – leading to what is now called the life sciences.

In this eighth edition, all core methodologies covered in the previous edition have been kept and appropriately updated and consolidated. New chapters have been added to address the requirements of today's students and scientists, who operate in a much broader area than did the typical biochemists one or two decades ago. The contents of this text are thus structured into six areas, all of which play pivotal roles in current research: basic principles, biochemistry and molecular biology, biophysical methods, information technology, -omics methods and chemical biology. Of course, due to space restraints that help to keep this text accessible, the addition of new material required us to consolidate and carefully select content to be kept from the previous edition. These decisions have been guided by the over-arching theme of this text, namely to present principles and techniques. Importantly, our experience as teachers has always been that many undergraduates are challenged by quantitative calculations based on these principles, and hence we have continued the tradition of this text by including relevant mathematical and numerical tools as well as examples.

Indeed, new chapters on data processing, visualisation and Python have example applications that can be accessed from the CUP website to aid understanding.

Sadly, two authors of previous editions have passed away: John Fyffe died shortly after the seventh edition was launched and Alastair Aitken passed away in 2014. Also, Keith Wilson, John Walker and Robert Burns decided to retire, and we thus invited several new authors to contribute to this eighth edition.

We are grateful to the authors and publishers who have granted us permission to reproduce and adapt their figures. We further wish to express our gratitude to Sonja Biberacher, Madeleine Dallaston, Emma Klepzig, Hannah Leeson and Megan Cross who have helped tremendously with preparing figures and photographs. The manuscript and figures for this book have been compiled entirely with open-source and academic software under Linux, and we would like to acknowledge the efforts of software developers and programmers who make their products freely available.

Finally, our sincere thanks also extend to Katrina Halliday and her team at Cambridge University Press whose invaluable support has made it possible to produce this new edition.

We welcome constructive comments from all students who use this book within their studies, and from teachers and academics who adopt the book to complement their courses.

ANDREAS HOFMANN AND SAMUEL CLOKIE

CONTRIBUTORS

PARISA AMANI
Structural Chemistry Program, Eskitis Institute for Drug Discovery,
Griffith University,
Nathan, Queensland, Australia

ANWAR R. BAYDOUN
School of Life Sciences, University of Hertfordshire,
Hatfield, Hertfordshire, UK

DAVID CAMP
School of Environment, Griffith University,
Nathan, Queensland, Australia

CINZIA CANTACESSI
Cambridge Veterinary School,
University of Cambridge,
Cambridge, UK

JEAN-BAPTISTE CAZIER
Centre for Computational Biology,
University of Birmingham,
Edgbaston, Birmingham, UK

SAMUEL CLOKIE
Bioinformatics Department,
West Midlands Regional Genetics Laboratory
Birmingham Women's and Children's NHS Foundation Trust,
Birmingham, UK

MEGAN CROSS
Structural Chemistry Program, Eskitis Institute for Drug Discovery,
Griffith University,
Nathan, Queensland, Australia

KATJA FISCHER
Scabies Laboratory, QIMR Berghofer,
Locked Bag 2000, Royal Brisbane Hospital,
Herston, Queensland, Australia

ROBIN B. GASSER
Faculty of Veterinary and Agricultural Sciences,
The University of Melbourne,
Parkville, Victoria, Australia

JOHN GRAINGER
Faculty of Biology, Medicine and Health,
University of Manchester,
Manchester, UK

STEPHEN E. HARDING
School of Biosciences,
University of Nottingham,
Sutton Bonington, UK

SONJA HESS
Proteome Exploration Laboratory, Beckman Institute,
California Institute of Technology,
Pasadena, USA

ANDREAS HOFMANN
Structural Chemistry Program, Eskitis Institute for Drug Discovery,
Griffith University,
Nathan, Queensland, Australia
and
Faculty of Veterinary and Agricultural Sciences,
The University of Melbourne,
Parkville, Victoria, Australia

JOANNE KONKEL
Faculty of Biology, Medicine and Health,
University of Manchester,
Manchester, UK

PASI K. KORHONEN
Faculty of Veterinary and Agricultural Sciences,
The University of Melbourne,
Parkville, Victoria, Australia

JOANNE MACDONALD
School of Science and Engineering,
University of the Sunshine Coast,
Maroochydore DC, Queensland, Australia
and

Department of Medicine,
Columbia University,
New York, NY, USA

JAMES I. MACRAE
Metabolomics Unit,
The Francis Crick Institute,
London, UK

KAY OHLENDIECK
Department of Biology,
National University of Ireland, Maynooth,
Co. Kildare, Ireland

STEPHEN W. PADDOCK
Howard Hughes Medical Institute,
Department of Molecular Biology,
University of Wisconsin,
Madison, Wisconsin, USA

ANNA V. PROTASIO
Wellcome Trust Sanger Institute,
University of Cambridge,
Hinxton, Cambridge, UK

RALPH RAPLEY
Department of Biosciences,
University of Hertfordshire,
Hatfield, Hertfordshire, UK

GILL RUMSBY
Clinical Biochemistry,
HSL Analytics LLP,
London, UK

ANNE SIMON
Université Claude Bernard Lyon 1,
Bâtiment Curien, Villeurbanne,
France
and
Laboratoire Chimie et Biologie des Membranes et des Nanoobjets,
Université de Bordeaux, Pessac,
France

ROBERT J. SLATER
Royal Society of Biology,
Charles Darwin House,
London, UK

TIM J. STEVENS
MRC Laboratory of Molecular Biology,
University of Cambridge,
Cambridge, UK

PAUL TAYLOR
School of Biological Sciences,
University of Edinburgh,
Edinburgh, Scotland, UK

MICHAEL WEISS
Phase I Pilot Consortium,
Children's Oncology Group,
Monrovia, CA, USA

0 Tables and Resources

Table 0.1 **Abbreviations**

ADP	adenosine 5′-diphosphate
AMP	adenosine 5′-monophosphate
ATP	adenosine 5′-triphosphate
bp	base pairs
cAMP	cyclic AMP
CAPS	*N*-cyclohexyl-3-aminopropanesulfonic acid
CHAPS	3-[(3-cholamidopropyl)dimethylamino]-1-propanesulfonic acid
cpm	counts per minute
CTP	cytidine triphosphate
DDT	2,2-bis-(*p*-chlorophenyl)-1,1,1-trichloroethane
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
e^-	electron
EDTA	ethylenediaminetetra-acetate
ELISA	enzyme-linked immunosorbent assay
FACS	fluorescence-activated cell sorting
FAD	flavin adenine dinucleotide (oxidised)
$FADH_2$	flavin adenine dinucleotide (reduced)
FMN	flavin mononucleotide (oxidised)
$FMNH_2$	flavin mononucleotide (reduced)
GC	gas chromatography
GTP	guanosine triphosphate
HAT	hypoxanthine, aminopterin, thymidine medium
HEPES	4(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid
HPLC	high-performance liquid chromatography
IMS	industrial methylated spirit
kb	kilobase pairs

Table 0.1 **(cont.)**

M_r	relative molecular mass
MES	4-morpholine-ethanesulfonic acid
min	minute
MOPS	4-morpholine-propanesulfonic acid
NAD^+	nicotinamide adenine dinucleotide (oxidised)
NADH	nicotinamide adenine dinucleotide (reduced)
$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidised)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
PIPES	1,4-piperazinebis(ethanesulfonic acid)
ppb	parts per billion
ppm	parts per million
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
TAPS	*N*-[tris-(hydroxymethyl)]-3-aminopropanesulfonic acid
TRIS	2-amino-2-hydroxymethylpropane-1,3-diol
v/v	volume per volume
w/v	weight per volume

Table 0.2 **Biochemical constants**

Unit	Symbol	SI equivalent
Atomic mass unit	u	1.661×10^{-27} kg
Avogadro constant	L or N_A	6.022×10^{23} mol^{-1}
Faraday constant	F	9.648×10^4 C mol^{-1}
Planck constant	h	6.626×10^{-34} J s
Universal or molar gas constant	R	8.314 J K^{-1} mol^{-1}
Molar volume of an ideal gas at standard conditions[a]	V_m	22.41 dm^3 mol^{-1}
Velocity of light in a vacuum	c	2.997×10^8 m s^{-1}

Note: [a]Standard conditions: $p^{\emptyset}$ = 1 bar, θ_{normal} = 25 °C.

Table 0.3 **SI units: basic and derived units**

Quantity	SI unit	Symbol (basic SI units)	Definition of SI unit	Equivalent in SI units
Basic units				
Length	metre	m		
Mass	kilogram	kg		
Time	second	s		
Electric current	ampere	A		
Temperature	kelvin	K		
Luminous intensity	candela	cd		
Amount of substance	mole	mol		
Derived units				
Area	square metre	m^2		
Concentration	mole per cubic metre	mol m^{-3}		
Density	kilogram per cubic metre	kg m^{-3}		
Electric charge	coulomb	C	1 A s	1 J V^{-1}
Electric potential difference	volt	V	1 kg m^2 $s^{-3}A^{-1}$	1 J C^{-1}
Electric resistance	ohm	Ω	1 kg m^2 $s^{-3}A^{-2}$	1 V A^{-1}
Energy, work, heat	joule	J	1 kg m^2 s^{-2}	1 N m
Enzymatic activity	katal	katal	1 mol s^{-1}	60 mol min^{-1}
Force	newton	N	1 kg m s^{-2}	1 J m^{-1}
Frequency	hertz	Hz	1 s^{-1}	
Magnetic flux density	tesla	T	1 kg s^{-2} A^{-1}	1 V s m^{-2}
Molecular mass	atomic mass units (dalton)	u	1 u = 1 Da	
Molar mass	gram per mole	g mol^{-1}	1 g mol^{-1}	
Power, radiant flux	watt	W	1 kg m^2 s^{-3}	1 J s^{-1}
Pressure	pascal	Pa	1 kg m^1 s^{-2}	1 N m^{-2}
Volume	cubic metre	m^3		

Table 0.4 Conversion factors for non-SI units

Unit	Symbol	SI equivalent
Energy		
calorie	cal	4.184 J
erg	erg	10^{-7} J
electron volt	eV	1.602×10^{-19} J
Pressure		
atmosphere	atm	1.013×10^{5} Pa
bar	bar	10^{5} Pa
millimetres of mercury	mm Hg (Torr)	133.322 Pa
pounds per square inch	psi	6.895×10^{4} Pa
Temperature		
degree Celsius	°C	$\left[\frac{\theta}{1^{\circ}\mathrm{C}} + 273.15\right]$ K
degree Fahrenheit	°F	$\left[\left(\frac{t}{1^{\circ}\mathrm{F}} - 32\right) \times \frac{5}{9} + 273.15\right]$ K
Length		
ångstrøm	Å	10^{-10} m
inch	in	0.0254 m
Mass		
pound	lb	0.4536 kg

Table 0.5 Common unit prefixes associated with quantitative terms

Multiple	Prefix	Symbol	Multiple	Prefix	Symbol
10^{24}	yotta	Y	10^{-1}	deci	d
10^{21}	zetta	Z	10^{-2}	centi	c
10^{18}	exa	E	10^{-3}	milli	m
10^{15}	peta	P	10^{-6}	micro	μ
10^{12}	tera	T	10^{-9}	nano	n
10^{9}	giga	G	10^{-12}	pico	p
10^{6}	mega	M	10^{-15}	femto	f
10^{3}	kilo	k	10^{-18}	atto	a
10^{2}	hecto	h	10^{-21}	zepto	z
10^{1}	deca	da	10^{-24}	yocto	y

Table 0.6 Interconversion of non-SI and SI units of volume

Non-SI unit		Non-SI subunit		SI subunit		SI unit
1 litre (l)	=	10^3 ml	=	1 dm^3	=	10^{-3} m^3
1 millilitre (ml)	=	1 ml	=	1 cm^3	=	10^{-6} m^3
1 microlitre (µl)	=	10^{-3} ml	=	1 mm^3	=	10^{-9} m^3
1 nanolitre (nl)	=	10^{-6} ml	=	1 nm^3	=	10^{-12} m^3

Table 0.7 Interconversion of mol, mmol and µmol in different volumes to give different concentrations

Molar (M)		Millimolar (mM)		Micromolar (µM)
1 mol dm^{-3}	=	10^3 mmol dm^{-3}	=	10^6 µmol dm^{-3}
=		=		=
1 mmol cm^{-3}	=	10^3 µmol cm^{-3}	=	10^6 nmol cm^{-3}
=		=		=
1 µmol mm^{-3}	=	10^3 nmol mm^{-3}	=	10^6 pmol mm^{-3}

Table 0.8 pK_a values of some acids and bases that are commonly used as buffer solutions

Acid or base	pK_a
Acetic acid	4.8
Barbituric acid	4.0
Carbonic acid	6.1, 10.2
CAPS[a]	10.4
Citric acid	3.1, 4.8, 6.4
Glycylglycine	3.1, 8.1
HEPES[a]	7.5
MES[a]	6.1
MOPS[a]	7.1
Phosphoric acid	2.0, 7.1, 12.3
Phthalic acid	2.8, 5.5
PIPES[a]	6.8
Succinic acid	4.2, 5.6
TAPS[a]	8.4
Tartaric acid	3.0, 4.2
TRIS[a]	8.1

Note: [a]See list of abbreviations (Table 0.1).

Table 0.9 **Abbreviations for amino acids and average molecular masses, free and within peptides (residue mass)**

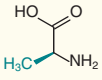
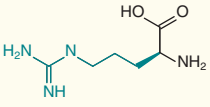
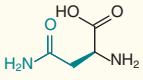
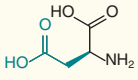
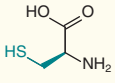
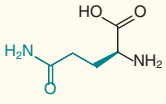
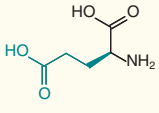
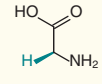
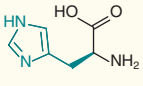
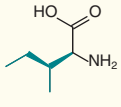
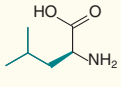
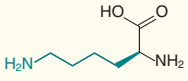
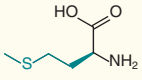
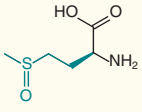
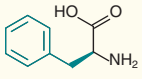
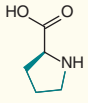
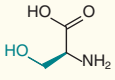
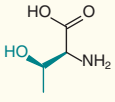
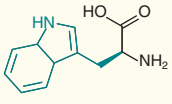
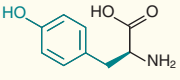
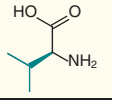
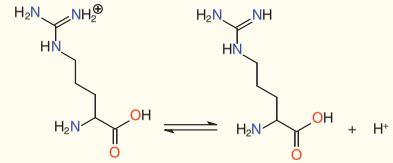
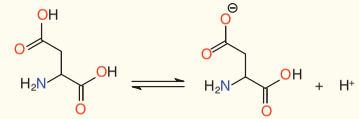
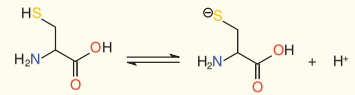
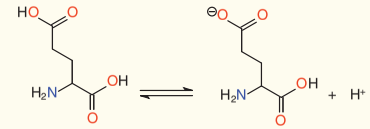
Amino acid	Three-letter code	One-letter code	Molecular mass *M*	Residue mass *M*-*M*(H_2O)		Side Chain
			Nominal mass (Da)	Average mass (Da)	Monoisotopic mass (Da)	
Alanine	Ala	A	89	71.08	71.037114	
Arginine	Arg	R	174	156.19	156.10111	
Asparagine	Asn	N	132	114.10	114.04293	
Aspartic acid	Asp	D	133	115.09	115.02694	
Asparagine or aspartic acid	Asx	B				
Cysteine	Cys	C	121	103.15	103.00919	
Glutamine	Gln	Q	146	128.13	128.05858	
Glutamic acid	Glu	E	147	129.12	129.04259	
Glutamine or glutamic acid	Glx	Z				
Glycine	Gly	G	75	57.05	57.021464	
Histidine	His	H	155	137.14	137.05891	
Isoleucine	Ile	I	131	113.16	113.08406	
Leucine	Leu	L	131	113.16	113.08406	

Table 0.9 **(cont.)**

Amino acid	Three-letter code	One-letter code	Molecular mass *M*	Residue mass *M*-*M*(H_2O)		Side Chain
			Nominal mass (Da)	Average mass (Da)	Monoisotopic mass (Da)	
Lysine	Lys	K	146	128.17	128.09496	
Methionine	Met	M	149	131.20	131.04048	
Met-sulfoxide[a]	MetSO	MSO	165	147.20	147.03540	
Phenylalanine	Phe	F	165	147.18	147.06841	
Proline	Pro	P	115	97.12	97.052764	
Serine	Ser	S	105	87.08	87.032029	
Threonine	Thr	T	119	101.11	101.04768	
Tryptophan	Trp	W	204	186.21	186.07931	
Tyrosine	Tyr	Y	181	163.18	163.06333	
Valine	Val	V	117	99.13	99.068414	

Notes: Although some amino acids are similar in mass, they can be distinguished by modern high-resolution mass spectrometry. [a]This is a frequently found modification in mass spectrometric investigation of proteins and peptides.

Table 0.10 **Ionisable groups found in amino acids**

Amino acid	pK_{a1} α-carboxyl group	pK_{a2} α-ammonium ion	pK_{a3} side-chain group	pI
Alanine	2.3	9.7	-	6.0
Arginine	2.2	9.0	12.5 	10.8
Asparagine	2.0	8.8	-	5.41
Aspartic acid	1.9	9.6	3.7 	2.8
Cysteine	2.0	8.2	8.4 	5.1
Glutamine	2.2	9.1	-	5.7
Glutamic acid	2.2	9.7	4.3 	3.2

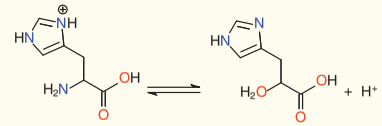
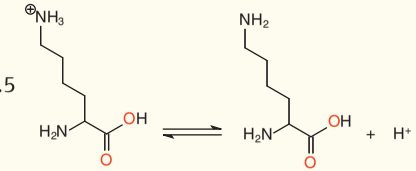
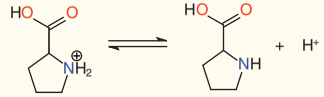
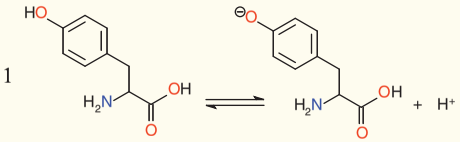
Glycine	2.3	9.6	-	6.0
Histidine	1.8	9.2	6.0 	7.6
Isoleucine	2.4	9.6	-	6.0
Leucine	2.4	9.6	-	6.0
Lysine	2.2	9.0	10.5 	9.7
Methionine	2.3	9.2	-	5.7
Phenylalanine	1.8	9.1	-	5.5
Proline	2.0	10.6 	-	6.3
Serine	2.2	9.2	-	5.7
Threonine	2.1	9.1	-	5.6
Tryptophan	2.8	9.4	-	5.9
Tyrosine	2.2	9.1	10.1 	5.7
Valine	2.3	9.6	-	6.0

Table 0.11 **The genetic code: triplet codons and their corresponding amino acids**

	U		C		A		G		
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	A
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G