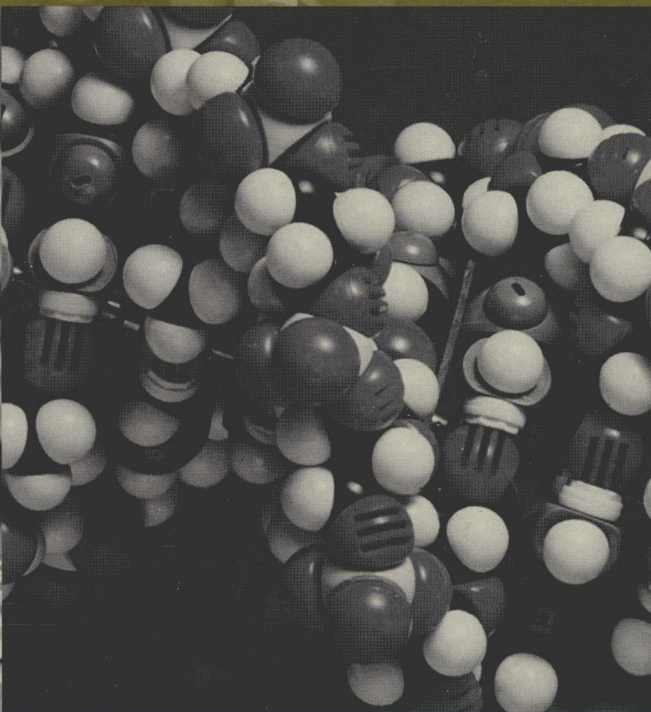


BIOMACROMOLECULES



*Introduction to
Structure, Function and Informatics*

C. STAN TSAI, Ph.D.

BIOMACROMOLECULES

Introduction to Structure,
Function and Informatics

C. STAN TSAI

Department of Chemistry, Carleton University



WILEY-LISS

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2007 by John Wiley & Sons, Inc. All rights reserved

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

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Tsai, C. Stan.

Biomacromolecules : introduction to structure, function, and informatics / C. Stan Tsai.
p. cm.

Includes bibliographical references and index.

ISBN-13: 978-0-471-71397-5

ISBN-10: 0-471-71397-X (cloth)

1. Macromolecules. 2. Biomolecules. I. Title.

QP801.P64T73 2006
572'.33—dc22

2006040639

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

BIOMACROMOLECULES

1977
1978
1979



THE WILEY BICENTENNIAL—KNOWLEDGE FOR GENERATIONS

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

WILLIAM J. PESCE
PRESIDENT AND CHIEF EXECUTIVE OFFICER

PETER BOOTH WILEY
CHAIRMAN OF THE BOARD

PREFACE

Biomacromolecules are fundamental structural and functional units of cells and therefore are at the very core of biochemical interest. They have always been the central topics of biochemical texts and literature. Various physicochemical and biochemical investigations greatly improve our knowledge of biomacromolecular structures and dynamics. The computational approach provides a new tool for structural and functional explorations of biomolecules. Sequence analyses and genetic recombination studies have contributed to our understanding of how biomacromolecules function at molecular and genetic levels. Recent years have witnessed an explosion in biological data that are derived primarily from studies of biomacromolecules. An application of information technology to organize, manage, distribute and analyze these biomacromolecular data has ushered in the new discipline of bioinformatics. There is an increased interest and sophistication in the study of biomacromolecules, as genomics and proteomics take central stage of biochemistry, molecular biology and bioinformatics. Slowly though, glycomics has now gained recognition. These developments give rise to a necessity for the comprehensive documentation and unified presentation of the structures, functions and informatics of biomacromolecules, for which this book is proposed to address.

Biomacromolecules, including nucleic acids (polynucleotides), proteins (polypeptides) and glycans (polysaccharides), are either briefly treated in the introductory biochemistry texts or extensively described in the advanced monographs concerning individual classes of compounds. In response to the renewed interest in biomacromolecules among various fields of biomedical sciences, a unified and comprehensive presentation of these topics is needed. The proposed textbook is aimed at bridging the gap between the introductory/elementary biochemistry course and advanced treatises on an individual class of biomacromolecules. The focus is on the integrated presentation of the structural, dynamic and informational biochemistry of nucleic acids, proteins and glycans, not separately, but as combined topics so that their similarities can be identified/acknowledged and differences compared/appreciated. The book intends to meet the demands of students who would like to broaden their biochemical knowledge beyond the introductory level and to prepare those who would like to venture into the advanced field of studies in genomics, proteomics and/or glycomics.

I have been teaching Biomacromolecules ever since the course was introduced into our program in 1968, on-and-off (and mostly on) before and after my retirement. Since the inception, the subject matter has undergone amazing transformations; from generally descriptive to molecular details, from mainly structure/function to informatics. The field has grown to encompass a large volume of information that any attempt to cover even the most superficial aspects of the topics in a single text is practically impossible if not a daunting task.

This book is written for students who have taken elementary/introductory biochemistry and would like to take further courses in biochemistry related to special topics in nucleic acids, proteins, and/or polysaccharides. Thus it is designed for students who are familiar with the general aspects of biochemistry and would like to further their

knowledge or for those who contemplate to pursue the field of studies related to biomacromolecules. It serves as an intermediate textbook in biochemistry, molecular biology and bioinformatics. Its content follows the organization of general/introductory biochemistry so that the continuity of biochemical curriculum is preserved; however the focus is on the macromolecular biochemistry. The book is unique in that it treats nucleic acids, proteins and glycans jointly as biomacromolecules and describes their structures, dynamics and informatics together.

Following introductory topics on biomacromolecules (Chapters 1–3), the elements of biomacromolecular structures (Chapters 4–6) and their studies (Chapters 7–9) are presented. The functions of biomacromolecules are discussed in terms of their interactions (Chapter 10), catalyses (Chapter 11) and metabolisms, including genetic transmission and applications (Chapters 12, 13). Biomacromolecular informatics (Chapter 14), namely genomics (Chapter 15), proteomics (Chapter 16) and glycomics (Chapter 16), are introduced. Chapter 18 describes biomacromolecular evolution. Each chapter presents a proper background in structures, dynamics or informatics of biomacromolecules, providing the context for further studies, which is supplemented by a list of references. In the areas where the speed of change and growth is high, a book cannot be either all-inclusive or entirely current. It is especially difficult for an introductory textbook of this nature to cover the topic materials up-to-date and comprehensively. Students are urged to consult the reference materials (literature cited and Web sites) for further understanding.

Balanced approaches include some general descriptions, which have been treated in general biochemistry texts, otherwise serving as introductory to advanced presentations, but however, not dwelling too deeply on the topics of specialized interest. Some materials that are commonly available in general texts are not repeatedly described here so that others may be considered. For example, discussion on the stereochemistry of monosaccharides (Chapter 3), spectral recordings (Chapter 7) and detailed descriptions on physiological functions or transformations of biomacromolecules, are either omitted or briefly mentioned. However, classical approaches of general interest in the study of biomacromolecules are presented, since they may serve as the background knowledge for advancing current understanding. Solid phase synthesis (Chapter 8) used in the fabrication of biochips (Chapter 14), and chemical modification of enzymes (Chapter 11) applied to the design of affinity/activity-based probes (Chapter 16), are some of examples. The choices of materials presented in this text are derived from many years of the author's teaching experience. The author alone is responsible for inadequate and erroneous presentations that may occur and readers' suggestions are very much appreciated.

I would like to thank all authors whose published works have contributed to a better understanding of biomacromolecular biochemistry and formed the resource materials of this textbook. The public accessibility of all the sequences and three-dimensional structures of biomacromolecules has greatly facilitated the advancement of our knowledge for the structure, function and informatics of biomacromolecules. The efforts of all the developers, contributors and managers of many outstanding Web sites of biomacromolecules are most appreciated. The writing of this text would not have been possible without the contributions and generosity of these investigators, authors and developers. My wife, Alice, has been most instrumental in helping me complete this text, which I would like to present to her as a gift on our tetracontyl anniversary. It is my pleasure to state that the realization of this text goes to former Editor, Luna Han and Editorial program coordinator, Kristin Hauser. They have patiently urged me to initiate the project prior to their departure for their new posts. I am grateful to the John Wiley staffs, Ian Collins, Thomas Moore, Dean Gonzalez, and Danielle Lacourciere for their timely help to assist the transformation of this manuscript to be publishable and to oversee the completion of this project.

Retired politicians and celebrities write personal memoirs. Retired entrepreneurs and investors write financial guides. Retired engineers and professionals write how-to or do-it-yourself manuals. Why cannot retired academics write text or reference books of their specialized fields? After many years of teaching and research experience, we certainly have lots to write about. I have taken up this project after my mandatory retirement, not without skeptics. However, I am relieved that I have made it.

C. Stan Tsai (stan@tsai-info.com)
Ottawa, Ontario, Canada

ABBREVIATIONS IN REPETITIVE USE

Some abbreviations that appear in the literature, but are not repeatedly used in this text, are mentioned but not listed here.

1D	one-dimensional
2D	two-dimensional
2DE	two-dimensional electrophoresis
2D-PAGE	two-dimensional polyacrylamide gel electrophoresis
3D	three-dimensional
7TM	seven transmembrane
Ψ(I)	hydrophobic (interactions)
A	adenosine/adenylate/adenine
A or Ala	alanine
A or Gal	galactose
AA	Allo A-H
aa-tRNA	3'-O-aminoacyl-tRNA, aminoacyl-tRNA
AB	<i>ab initio</i> prediction
ABP	activity-based probe
AC	accession number
AC	adenylyl cyclase
ACR	ancient conserved region
AD	transcription activation domain
ADEPT	antibody-directed enzyme prodrug therapy
AD(T)P	adenosine di(tri)phosphate
AFBP	affinity-based probe
AGE	advanced glycated end-product
AIF	apoptosis inducing factor
AISMAG	An Interactive Server-side Molecular Image Generator
ALU	arithmetic logic unit
AN or GalNAc	<i>N</i> -acetylgalactosamine
ANS	8-anilino-1-naphthalene sulfonate
AP	apurinic/apyrimidinic
Apaf	apoptotic protease activating factor
APC	anaphase-promoting complex
aRS	aminoacyl-tRNA synthetase
AS	active site/s
ASGPR	asialoglycoprotein receptor
BCM	Baylor College of Medicine
BD	DNA-binding domain
BER	base excision repair
BFGF	basic fibroblast growth factor
bgl	blood group locus (loci)

BHA	benzyhydrilamine
BIND	biomolecular interaction network database
BIOS	basic input/output system
BLAST	Basic Local Alignment Search Tool
BMCD	Biological Macromolecular Crystallization Database
BNL	Brookhaven National Laboratories
Boc	<i>tert</i> -butoxycarbonyl
bp	base pair(s)
Bzl	benzyl
C	cytosine
C or Cys	cysteine
CA	correspondence analysis
CAD or CID	collisionally-activated dissociation or collisionally-induced dissociation
cAMP	3',5'-cyclic adenosine monophosphate
CAPRI	Critical Assessment of Predicted Interactions
CASP	Critical Assessment of Techniques for Protein Structure Prediction
CASPER	Computer Aided Spectrum Evaluation of Regular Polysaccharides
CATH	Class, Architecture, Topology, Homology
CAZymes	carbohydrate active enzymes
CBP	CREB binding protein
CBS	Center for Biological Sequence Analysis
cc	comment
CCDC	Cambridge Crystallographic Data Centre
cccDNA	covalently closed circular DNA
ccDNA	closed circular DNA
CCSD	Complex Carbohydrate Structure Database
CD	circular dichroism (spectroscopy)
CD	cluster of differentiation
Cdb	cyclin destruction box
CDG	congenital disorders of glycosylation
cDNA	complementary DNA
CDR	complementarity determining region
CDS, cds	coding sequence
CE	cyanoethyl
CERMAV	Centre de Recherches sur les Macromolécules Végétales
CFG	Consortium for Functional Glycomics
CG	cancer gene/s
cGMP	3',5'-cyclic guanosine monophosphate
CICR	Ca ²⁺ -induced Ca ²⁺ -release
CID	collision-induced dissociation
cM	centiMorgan
CM	comparative (homology) modeling
CMD	congenital muscular dystrophy
CMR	carbon-13 magnetic resonance (spectroscopy)
CNS	central nervous system
coagF	coagulation factor
Con A	concanavlin A
COSY	J correlated spectroscopy
CP	chlorophenyl

CPG	controlled pore glass
CPK	Corey–Pauling–Koltun
CPS	carbamyl phosphate synthetase
CPU	central processor unit
CRD	carbohydrate-recognition domain
CRE(B)	cAMP-response element (binding protein)
CSA	Catalytic Site Atlas
CSM	common structures of monosaccharides
CSS	Carbohydrate Structure Suite
CT	carbamoyltransferase
CT	consensus trees/supertrees
CU	control unit
Cyt/cyt	cytochrome
D	dihydrouridine
D or Asp	aspartic acid
Da	daltons
(d)A	(deoxy)adenosine
DAG	diacylglycerol
DB	database
DBMS	database management system
(d)C	(deoxy)cytidine/cytidylate
dc	distance calculation
DCA	discriminant correspondent analysis
DCC	dicyclohexyl carbodiimide
DCP	dichlorophenyl
DD	death domain
DDBJ	DNA Data Bank of Japan
DD-PCR	differential display PCR
DE	description/s
DED	death effector domain
(d)G	(deoxy)guanosine/guanylate
DH	dehydrogenase
DIGE	difference in-gel electrophoresis
diS	disaccharides
DKFZ	German Cancer Research Center
DM	distance matrix
DMC	dichloromethane
DMTr	dimethoxy trityl
DNA	deoxyribonucleic acid(s)
DNP	dinitrophenyl
DNS	dansyl
dNTP	deoxyribonucleotide-5'-triphosphate
Dol	dolichol
dosDNA	defined ordered DNA sequences
DPO	dolichol phosphate-oligosaccharide
DP	degree of polymerization
DP	dolichol pyrophosphate
DR	database cross-reference/S
dsDNA	double-stranded DNA
DSS	sodium 2,2-methyl-2-silapentane-5-sulfonate

DT	date of entry
DTT	dithiothreitol
E or Glu	glutamic acid
EA	enzyme-substrate
EBI	European Bioinformatics Institute
EC	Enzyme Commission
EcorL	<i>Erythrina corallodendron</i> lectin
EDTA	ethylenediamine tetraacetate
EF	electrofocusing
EFF/FF	empirical force field/force field
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EGP	epidermal growth factor
Eif	eukaryotic initiation factor
EMBL	European Molecular Biology Laboratory
EMBnet	European Molecular Biology network
EMin	energy minimization
EP	enzyme-product
EP	eukaryotic primase
EPD	eukaryotic promoter database
EPL	expressed protein ligation
ER	endoplasmic reticulum
ERK	extracellular signal regulated protein kinase
ESI	electrospray ionization
ESP	electrostatic potential
EST	expressed sequence tag
ExPASy	Expert Protein Analysis System
F or Phe	phenylalanine
FAB	fast atom bombardment
FAD	flavin adenine dinucleotide
FaPy	formamidopyrimidine
FB(M)	fragment-based (method)
FF	force field
FG	functional genomics
fibF	fibrinolytic factor
FID	free induction decay
FITC	fluorescein isothiocyanate
FMN/FAD	flavin mononucleotide/flavin adenine dinucleotide
Fmoc	fluorenyl-9-methyloxycarbonyl
FP	fingerprint/s
FP	fluorescent probe
FR	fold recognition
FSSP	fold tree, domain dictionary, sequence neighbors, structure superposition
FT	feature/s
FT	feature table
FT	Fourier transform
FTP	file transfer protocol
G	general
G	guanosine/guanine

G or Glc	glucose
G or Gly	glycine
GA	genetic algorithms
GA	wheat germ agglutinin
Gaba	γ -aminobutyric acid
GAP	GTPase activating proteins
Gase	glycosidase
GAT	glutamine amidotransferase
Gbp	(giga)-base pairs
GBP	glycan binding proteins
GC	gas chromatography
GD(T)P	guanosine di(tri)phosphate
GE	gel electrophoresis
GE	gene expression
GEF	guanine nucleotide exchange
GIF	graphical interchange format
Gla	γ -carboxyglutamate
Glc	D-Glucose
GlcNacT	GlcNAc transferase
GN or GlcNAc	N-acetylglucosamine
GN	gene name
GO	gene ontology
GOA	Gene Ontology Annotation
GOR	Garnier, Osguthorpe and Robson
GP	genome project(s)
gp	grid point
GP	glycogen phosphorylase
GP _a	phospho-phosphorylase a
GP _b	dephospho-phosphorylase b
GPCR	G-protein-coupled receptor
GPI	glycosylphosphatidylinositol
GR	glutathione reductase
GRE	glucocorticoid response element
GSS	genome survey sequence
GT	glycosylation pathways
GT	glycosyltransferase(s)
GTO	Gaussian type orbitals
GUI	graphical user interface
H or His	histidine
HB	hydrogen bond(s)
hCG	human chorionic gonadotropin
hd	helical domain
HDV	hepatitis delta virus
HEMPAS	hereditary erythroblastic multinuclearity with positive acidified-serum lysis test
HGP	Human Genome Project
HisF	histidine biosynthetic enzyme
HIV	human immunodeficiency virus
HM	histogram matching
HMM	hidden Markov model

hnRNA	heterogeneous nuclear RNA
hnRNP	heterogeneous nuclear ribonucleoprotein
HPLC	high performance (high-pressure) chromatography
HrPAGE	high-resolution polyacrylamide gel electrophoresis
HSE	heat shock element
HSP	heat-shock protein
HSP	high-scoring pair
HTML	HyperText Markup Language
HTPC	high-throughput protein crystallography
HTS	high-throughput screening
hs	heparan sulfate
HTTP	HyperText Transfer Protocol
Hyl (Hk)	hydroxylysine
Hx	hypoxanthine
Hyp (hP)	hydroxyproline
I	invariant method
I or Ile	isoleucine
ICAT	isotope-coded affinity tag
ID	Identifier
IEF	isoelectric focusing
Ig	immunoglobulin
IgG	immunoglobulin G
IGOT	isotope-coded glycosylation-site-specific tagging
IL	interleukin
IMAC	immobilized metal affinity chromatography
IMP	inosine monophosphate
InDel(s)	insertion(s) and/or deletion(s)
INSDC	International Nucleotide Sequence Database Collaboration
insR	insulin receptor
IP	Internet Protocol
IP	inositol phosphate
IPCR	immuno-polymerase chain reaction
IPG	immobilized pH gradient
IPI	International Protein Index
IPL	intein-mediated protein ligation
IP ₃ R	inositol triphosphate receptor
IR	infrared (spectroscopy)
IR	instruction register
IS	insertion sequence
ISOC	Internet society
ISP	Internet service provider
ISREC	Swiss Institute for Experimental Cancer Research
IUBMB	International Union of Biochemistry and Molecular Biology
IUPAC	International Union of Pure and Applied Chemistry
IVS	intervening sequence(s), intron(s)
JPEG	joint photographic experts group
K or Lys	lysine
Kb	thousand base-pairs
Kbp	kilo base-pairs
KcaM	KEGG Carbohydrate Matcher

kDa	Kilo Daltons
KEGG	Kyoto Encyclopedia of Genes and Genomes
KF-Pol I	Klenow fragment of DNA polymerase I
KW	keyword
L or Leu	leucine
LAN	local area network
LC	liquid chromatography
LCD	liquid crystal display
LDH	lactate dehydrogenase
LFA	<i>Limax flavus</i> agglutinin
LINE	long interspersed nuclear element
LINUCS	linear notation for unique description of carbohydrate sequence
LOL	<i>Lathyrus ochrus</i> lectin
M or Man	mannose
M or Met	methionine
MA	microarray
MAA	<i>Maackia amurensis</i> agglutinin
MAG	myelin-associated glycoprotein
MALDI	matrix-assisted laser desorption/ionization
MAP	mitogen activated protein
MAP-KKK	MAP kinase-kinase-kinase
MAR	memory address register
MAS	maskless array synthesizer
Mb	million base-pairs
Mbp	(mega)-base pairs
MBP	mannose binding protein
MBR	memory buffer register
MC	Monte Carlo method
MC-SYM	macromolecular conformations by symbolic programming
MD	mutation data
MEK	MAP (mitogen activated protein) kinase-ERK kinase
MeNPOC	methylnitropiperonyloxycarbonyl
MHC	major histocompatibility complex
MI	metastable ion
MIME	multipurpose Internet mail extensions
MIP	molecularly imprinted polymer
MIPS	Munich Info Center for Protein Sequences
miRNA	microRNA
ML	maximum likelihood
MLCK	myosin light chain kinase
MM	molecular mechanics
MMDB	molecular modeling database
MMTr	monomethoxy trityl
MO	molecular orbital(s)
MolD	molecular dynamics
momoS	monosaccharides
MPB	mannose binding protein
MPP	mitochondrial processing peptidase
mPu/Py	methylpurin/pyrimidine
MRE	metal response element

mRNA	messenger RNA
MRW	mean residue weight of monomer
mp	matching point
MS	mass spectrometry
ms	mean-square
MS-MS	tandem mass spectrometry
MSP	maximal-scoring segment pair/s
m.u.	mass unit/s
MW	molecular weight
m/z	mass-to-charge ratio
N or Asn	asparagine
N or Neu	neuramic acid
NAC	nascent polypeptide-associated complex
NAD(P) ⁺	nicotinamide adenine dinucleotide (phosphate)
NAD(P)H	reduced nicotinamide adenine dinucleotide (phosphate)
NAPPA	nucleic acid programmable protein array
NAT	natural antisense transcript
NBRF	National Biomedical Research Foundation
NCBI	National Center for Biotechnology Information
NCS	noncrystallographic symmetry
ncRNA	noncoding RNA
NDB	Nucleic Acid Database
NDP	nucleoside diphosphates
NER	nucleotide excision repair
NeuNAc or Sia	N-acetylneuramic acid or sialic acid
NGF(R)	nerve growth factor (receptor)
NIH	National Institute of Health
NJ	neighbor joining method
NK	natural killer
NLM	National Library of Medicine
NMD	nonsense-mediated mRNA decay
NMR	nuclear magnetic resonance (spectroscopy)
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect and exchange spectroscopy
NOS	nitric oxide synthase
NPG	nucleotide phosphoglycose
NR	non-reducing
NR	non-redundant
nrPTK	non-receptor protein tyrosine kinase
nt	nucleotide(s)
Nvoc	6-nitroveratryloxycarbonyl
NW	network
OC	organism classification
oligoS	oligosaccharides
ORD	optical rotatory dispersion (spectroscopy)
ORF	open reading frame
OUT	operational taxonomic unit
OS	operating system
OS	organism species
OST	oligosaccharyltransferase

OUT	operational taxonomic unit/s
P	parsimony
P or Pro	proline
PABA	poly(A) binding protein
PAGE	polyacrylamide gel electrophoresis
PAM	phenylacetamidomethyl
PAM	point accepted mutation
PC	personal computer
PCA	principle component analysis
PCD	programmed cell death
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PD	pyrimidine dimer
PDB	Protein Data Bank
PDC	pyruvate dehydrogenase complex
PDGF	platelet derived growth factor
PDP	pyruvate dehydrogenase phosphatase
PEG	polyethylene glycol
PEP	primer extension preamplification
Perl	Practical Extraction and Report Language
PEST	glutamic acid, serine and threonine
PEG	polyethylene glycol
PFGE	pulsed field gel electrophoresis
PEP	primer extension preamplification
PFK	phosphofructokinase
PGGF	platelet derived growth factor
PHP	<i>Pyrococcus horikoshii</i>
PHYLP	Phylogenetic Inference Package
PIC	phenyl isocyanate
PID	phosphotyrosine interaction domain
PIR	Protein Information Resource
PITC	phenyl isothiocyanate
pk	psudoknot
PKA	protein kinase A
PKC	protein kinase C
PL	phospholipase
PLC	phospholipase C
PLD	phospholipase D
PLP	pyridoxal-5'-phosphate
PMR	proton magnetic resonance (spectroscopy)
³¹ PMR	phosphorus magnetic resonance spectroscopy
PMSF	phenylmethylsulfonyl fluoride
PMW	position weight matrix
PNGF	peptide-N-glycosidase F
Pol	DNA polymerase
poly(DA)	poly(deoxyadenylate)
poly(dG-dC)	poly(deoxyguanidylate-deoxycytidylate)
poly(U)	poly(uridylate)
ppm	parts per million
PPP	point-to-point protocol