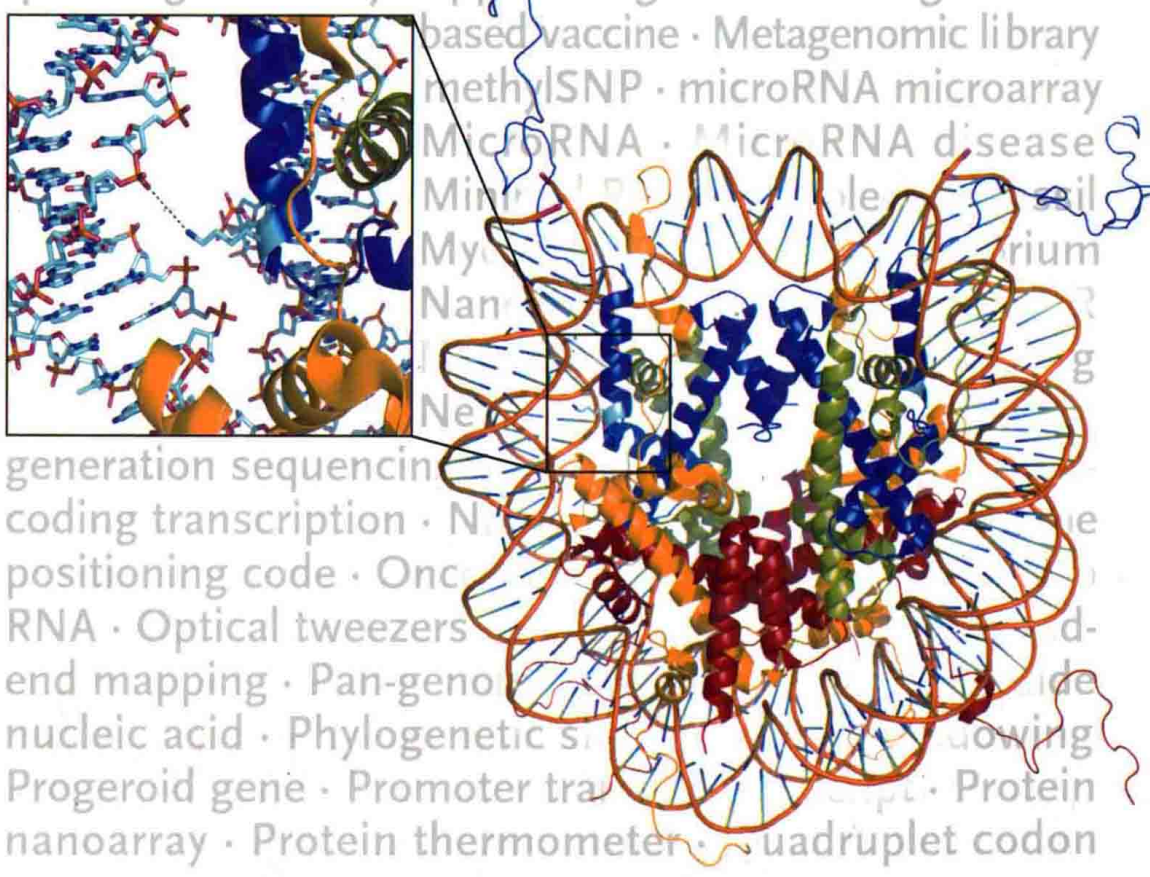
Günter Kahl

WILEY-BLACKWELL

The Dictionary of Genomics, Transcriptomics and Proteomics

Fourth, Greatly Enlarged Edition

Volume 2
L–Q



Günter Kahl

The Dictionary of Genomics, Transcriptomics and Proteomics

Fourth, Greatly Enlarged Edition

Volume 2: L–Q

WILEY-
BLACKWELL

WILEY-VCH Verlag GmbH & Co. KGaA

The Author

Günter Kahl
Johann Wolfgang Goethe-University
Molecular BioSciences, Biocenter
Max-von-Laue-Staße 9
60438 Frankfurt am Main
Germany

and

GenXPro GmbH
Frankfurt Innovation Center Biotechnology (FIZ)
Altenhöferallee 3
60438 Frankfurt am Main
Germany

Chinese Edition published in 2004

Cover Illustration
The Title Page shows a three-dimensional image of a nucleosome, in which the DNA double-helix is depicted as intertwined brownish ribbons (the phosphate backbone), from which the bases protrude (in green-blue colour). About 1.7 turns of DNA are wrapped around the histone core, where histone H2A comes in yellow, H2B in red, H3 in blue, and H4 in green colours. The H_2N-termini of the four histones emerge from the nucleosome as reels in the corresponding colour. The inset portrays the interaction between the phosphate backbone of the DNA with histone H3 K56 mediated by water molecules.

The graph was produced from pdb file 1KX5 with Pymol, and kindly provided by Heinz Neumann and Jason Chin, Division of Protein and Nucleic Acid Chemistry, Evolution and Synthesis of New Function, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.d-nb.de>.

Composition SNP Best-set Typesetter Ltd., Hong Kong
Printing betz-druck GmbH, Darmstadt
Bookbinding Litges & Dopf GmbH, Heppenheim

Printed in the Federal Republic of Germany
Printed on acid-free paper

ISBN: 978-3-527-32073-8

Related Titles

Singleton, P.

Dictionary of DNA and Genome Technology

2008
ISBN: 978-1-4051-5608-0

Dunn, M. J., Jorde, L. B., Little, P. F. R., Subramaniam, S. (eds.)

Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics

8 Volume Set

2006
ISBN: 978-0-470-84974-3

Kahl, G., Meksem, K. (eds.)

The Handbook of Plant Functional Genomics

Concepts and Protocols

2008
ISBN: 978-3-527-31885-8

Meksem, K., Kahl, G. (eds.)

The Handbook of Plant Genome Mapping

Genetic and Physical Mapping

402 pages with 77 figures and 11 tables
2005
Hardcover
ISBN: 978-3-527-31116-3

Günter Kahl

The Dictionary of Genomics, Transcriptomics and Proteomics

It is a pleasure to dedicate these three volumes
to my friends, colleagues, and foreign visitors of

in the Frankfurt Research Innovation Centre
Biotechnology in Frankfurt am Main (Germany).

Preface

Η γνώση των λέξεων οδηγεί στην γνώση των πραγμάτων
(The knowledge of words leads to a knowledge of things)
(Πλάτων, Platon, 427–347 B.C.)

The glamour and excitement of genetic engineering during the past three decades have given way to routine and almost trivial daily work in the laboratory. Many of the young researchers are nevertheless still fascinated by the precision of the various gene technologies and the surprising possibilities they offer, and even seasoned researchers are enchanted by this ever-growing field of molecular genetics. It is fair to say that gene technology has now infiltrated all areas of molecular biology, and massively contributed to the vast information in this field, that is accumulating at a more explosive rate than ever before. This phenomenal development forces to divide the field of gene technology into at least three subsections: **genomics**, **transcriptomics**, and **proteomics**, and this is exactly done in the present opus.

"This book contains a considerable volume of informations.
I deeply regret this, but unfortunately it was inevitable."
(Samuel Langhorne Clemens alias Mark Twain, 1835–1910)

With the immense growth of the three, and other related areas of molecular biology, the number of novel technologies, procedures and technical terms is soaring. So, the present three volumes contain a total of 12,000 different terms, many of them describing recent developments and brand-new technologies. It is therefore the most comprehensive collection of descriptions of molecular processes and techniques worldwide. Some of the terms and their multiple variants dominate. For example, traditional PCR and its numerous facets comprises some 190 entries, surpassed by the terminology around microarray with more than 250 entries, and well over 100 "omics" neologisms mess up both literature and daily language. The second ("next"), and emerging third ("next-next") generation sequencing technologies brought a burst in novel terminology, not to speak of the many other cutting-edge techniques that appear almost daily.

"Make everything as simple as possible, but not simpler!"
Albert Einstein (1879–1955)

This flood of terms and acronyms sometimes leaves the researcher a bit helpless, especially since a single term might mean different things, and many different terms may mean the same single thing. The present volumes aim at ordering this chaos a bit, and are not restricted to the omics trilogy (or even gene technology), but link to other related disciplines and describe relevant terms, if considered to be necessary or helpful for a

The Dictionary of Genomics, Transcriptomics and Proteomics, 4th Ed., Günter Kahl

ISBN: 978-3-527-32073-8

better understanding. Obviously, the growing number of proteins cannot be treated with in such a dictionary, especially since the peptides, proteins, and their isoforms will probably be in their millions. Therefore only a limited selection is portrayed. Another problem was, is, and will be, the extent of description. Some entries are described in some depth, some others only defined spartanically.

"Going too far is as bad as not going far enough"

过犹不及

(Chinese proverb)

My prime appreciations go to my son Uwe Kahl, who took the tantalizing task to draw a multitude of figures and schemes from partly fragmented and absolutely insufficient samples. He did a great job! I also thank Achim Wilz for his patient introduction into the various facets and pitfalls of the computer world. Sigrid – as always – gave me all her support and the freedom needed for such a work.

These three books are dedicated to all people of GenXPro GmbH in the Frankfurt Innovation Center Biotechnology (FIZ, Frankfurt am Main, Germany). Since excellent science is nowadays also daily work in companies, I learned a lot and hopefully could give a tiny bit to the energetic people of this dynamic enterprise. I am in fact grateful for all the discussions, scientific turns and innovations, and all the adventures that accompany a young spin-off company: a very rewarding experience.

I appreciate the hospitality of various institutions in different countries, where I have been working on this opus over the last years, as there is The Research Institute for Bioresources (Kurashiki, Japan), the Department of Biology and Molecular Biology (University of California at Los Angeles, USA), the International Center for Agricultural Research in the Dry Areas (Aleppo, Syria), the Centro Agronomico Tropical (Turrialba, Costa Rica), the Iwate Biotechnology Research Institute (Kitakami, Japan), and the Pharma Center (University of Vienna, Austria).

Frankfurt am Main, September 2008 Günter Kahl

Instructions for Users

- All the entries are arranged in strict alphabetical order, letter by letter. For example, "mismatched primer" precedes "mismatch gene synthesis", and this is followed by "mismatch repair". Or, "photo-digoxygenin" precedes "photo-footprinting", which in turn precedes "photo-reactivation". In case an entry starts with, or contains a Roman, Greek or Arabic numeral, it has first to be translated into Latin script. A few examples illustrate the translation:

cI	: c-one
exonuclease VII	: exonuclease seven
exonuclease III	: exonuclease three
5′	: five prime
G 418	: G fourhundred and eighteen
λ	: lambda
P1	: p-one
ΦX 174	: phi X one-seven-four
Qβ	: q-beta
RP 4	: RP four

For help, the user may consult the Greek alphabet and the Roman numerals below.

- The main entry title, printed in bold type, is followed by synonyms in parentheses. Italicized letters in titles (and text) of entries indicate use of these letters for abbreviations.

- Cross referencing is either indicated by an arrow, or the words "see", "see also", and "compare".

- By using the cross-references as a road map between definitions, the reader will gain an appreciation of molecular biology as an integrated whole rather than a collection of fragments of isolated information.

- Organismal name: The formal Latin binomial names of organisms are italicized, whereas common names and derivatives of the Latin names are not.

- Etymology of the terms: Most biological terms originate from Greek or Latin language. Only the most common word roots are defined in this dictionary.

The Dictionary of Genomics, Transcriptomics and Proteomics, 4th Ed., Günter Kahl

ISBN: 978-3-527-32073-8

Greek Alphabet and Roman Numerals

Greek alphabet:

Capital	Lower case	Name	Capital	Lower case	Name
A	α	alpha	N	ν	nu
B	β	beta	Ξ	ξ	xi
Γ	γ	gamma	O	o	omicron
Δ	δ, δ	delta	Π	π	pi
E	ε	epsilon	P	ρ	rho
Z	ζ	zeta	Σ	σ, ς	sigma
H	η	eta	T	τ	tau
Θ	θ, ϑ	theta	Υ	υ	ypsilon
I	ι	iota	Φ	φ	phi
K	κ	kappa	X	χ	chi
Λ	λ	lambda	Ψ	ψ	psi
M	μ	mu	Ω	ω	omega

Roman numerals:

I	II	III	IV	V	VI	VII	VIII	IX	X
1	2	3	4	5	6	7	8	9	10
XX	XXX	XL	L	LX	LXX	LXXX	XC	IC	C
20	30	40	50	60	70	80	90	99	100
CC	CCC	CD	D	DC	DCC	DCCC	CM	XM	M
200	300	400	500	600	700	800	900	990	1000

The Dictionary of Genomics, Transcriptomics and Proteomics, 4th Ed., Günter Kahl

ISBN: 978-3-527-32073-8

Abbreviations and Symbols

a	– atto (10^{-18})
A	– adenine or adenosine, absorbance
Å	– Ångstrom unit (1 Å = 0.1 nm)
~	– approximately
≅	– approximately equals
A/D	– analog-to-digital
aa	– amino acid
Ab	– antibody
Ag	– antigen
Ap	– ampicillin
ATP	– adenosine triphosphate
B	– any nucleo*b*ase (A,C,G,or T)
BAC	– bacterial artificial chromosome
Bis	– *N, N'*-methylenebisacrylamide
BLAST	– basic local alignment search tool
bp	– base pair(s)
Bq	– Becquerel
BSA	– bovine serum albumin
c	– centi (10^{-2})
C	– cytosine or cytidine
^{14}C	– radioactive carbon
°C	– centigrade (degrees Celsius)
Ca	– Calcium
CBB	– Coomassie Brilliant Blue
CCD	– charge-coupled device
cDNA	– complementary DNA
CE	– capillary electrophoresis
CGE	– capillary gel electrophoresis
Ci	– Curie
cm	– centimeter(s)
Cm	– chloramphenicol
CO_2	– carbon dioxide
cpm	– counts per minute
CTAB	– cetyltrimethylammonium bromide
Cy	– cyanine
D, Da	– Dalton
DAF	– DNA amplification fingerprinting
dATP	– deoxyadenosine triphosphate
dCTP	– deoxycytosine triphosphate

The Dictionary of Genomics, Transcriptomics and Proteomics, 4th Ed., Günter Kahl

ISBN: 978-3-527-32073-8

ddNTP – 2′, 3′-dideoxynucleotide triphosphate
DGGE – denaturing gradient gel electrophoresis
dGTP – deoxyguanosine triphosphate
DMF – *N, N′*-dimethylformamide
DMSO – dimethyl sulfoxide
DMT, DMTr – dimethyloxytrityl
DNA – deoxyribonucleic acid
DNase – deoxyribonuclease
dNTP – deoxynucleotide triphosphate
ds – double-stranded
dT – deoxythymidine
DTT – dithiothreitol, Cleland's reagent
dTTP – deoxythymidine triphosphate
dUTP – deoxyuridine triphosphate
EC – enzyme classification number
ECL – enhanced chemiluminescence
E. coli –*Escherichia coli*
EDTA – ethylenediaminetetraacetic acid
EGTA – ethylene glycol-bis (β-aminoethylether) *N,N,N′,N′*-tetraacetic acid
e.g. – for example
ELISA – enzyme-linked immunosorbent assay
ESI – electrospray ionization
ESI-MS – electrospray ionization mass spectrometry
EST – expressed sequence tag
EtBr – ethidium bromide
EtOH – ethanol
f – femto (10^{-15})
Fab – antigen-binding region of an antibody
FACS – fluorescence-activated cell sorter
FIGE – field inversion get electrophoresis
FITC – fluorescein isothiocyanate
fmol – femto mol
5′ – carbon atom 5 of deoxyribose
g – gram(s) or gravity
G – guanine or guanidine, giga (10^{9})
Gb – gigabase
GC – gas chromatography
GFP – green fluorescent protein
Gm – gentamycin
GMO – genetically modified organism
> – greater than
h – hour(s)
HAC – human artificial chromosome
^{3}H – tritium, radioactive hydrogen
HCl – hydrochloric acid

HEPES – N-(2–hydroxyethyl) piperazine-N′-(2–ethanesulfonic acid)
HGP – human genome project
HIV – human immunodeficiency virus
HPCE – high-performance capillary electrophoresis
HPLC – high pressure liquid chromatography
HRP – horseradish peroxidase
HTE – high Tris-EDTA buffer
H_2O – water
H_2O_2 – hydrogen peroxide
HTML – hypertext mark-up language
HVR – hypervariable region
i.e. – that is
IEF – isoelectric focusing
Ig – immunoglobulin
IP – intellectual property
IVS – intervening sequence, intron
k – kilo (10^3)
kb – kilobase(s)
KB – kilobyte
kbp – kilobase pairs
kD (kDa) – kilo Dalton
kg – kilogram(s)
Km – kanamycin
l – liter(s)
< – less than
LC – liquid chromatography
LiCl – lithium chloride
LIF – laser-induced fluorescence
LTE – low Tris-EDTA buffer
mAb – monoclonal antibody
MALDI-MS – matrix-assisted laser desorption/ionization-mass-spectrometry
m – meter(s) or milli (10^{-3})
µ – micro (10^{-6})
µg – microgram(s)
µl – microliter(s)
M – molar or mega (10^6)
Mb (Mbp) – megabase pairs
MB – megabyte
MCS – multiple cloning site
Mg – magnesium
mg – milligram(s)
$MgCl_2$ – magnesium chloride
$MgSO_4$ – magnesium sulfate
min – minute(s)
ml – milliliter(s)

mm – millimeter(s)
mM – millimolar
mmol – millimole
mol – mole
M_r – relative molecular mass (no dimension)
mRNA – messenger RNA
MS – mass spectrometry
MS/MS – tandem mass spectrometry
mtDNA – mitochondrial DNA
MW – molecular weight
m/z – mass-to-charge ratio
n – number or nano (10^{-9})
NaCl – sodium chloride
Na_2EDTA – disodium-EDTA
NC – nitrocellulose
ng – nanogram(s)
NH_4Cl – ammonium chloride
NH_4OAc – ammonium acetate
nm – nanometer(s)
NMR – nuclear magnetic resonance
nt – nucleotide
OD – optical density
ODN – oligodeoxynucleotide
OH – hydroxy
oligo – oligonucleotide(s)
ORF – open reading frame
ORN – oligoribonucleotide
P – phosphorus
p – pico (10^{-12})
P_i – inorganic phosphorus
^{32}P – radioactive phosphorus
PAGE – polycrylamide gel electrophoresis
PBS – phsphate buffered saline
PCR – polymerase chain reaction
PEG – polyethylene glycol
Petabyte (PB) – 10^{15} bytes
PFGE – pulsed field gel electrophoresis
pfu – plaque forming unit
pg – picogram(s)
pH – logarithm of reciprocal of hydrogen (H) ion concentration
pI – isoelectric point
PMS – phenazine methosulfate
PMSF – phenylmethylsulfonyl fluoride
PNA – peptide nucleic acid
pp – page(s)

ppm – parts per million
PSD – post-source decay
PTFE – polytetrafluoroethylene
PVDF – polyvinylidene difluoride
PVP – polyvinyl pyrolidone
RAPD – random amplified polymorphic DNA
RFL – restriction fragment length
RFLP(s) – restriction fragment length polymorphism(s)
RIA – radioimmunoassay
RNA – ribonucleic acid
RNase – ribonuclease
RP – reversed phase
rpm – revolutions per minute
rRNA – ribosomal RNA
RT – room temperature (also reverse transcriptase)
RT-PCR – reverse transcriptase PCR
^{35}S – radioactive sulfur
SAGE – serial analysis of gene expression
SD – standard deviation
SDS – sodium dodecyl sulfate, lauryl sulfate
SE (SEM) – standard error (standard error of the mean)
sec – second(s)
Σ – sum of
Sm – streptomycin
S/N – signal-to-noise ratio
SNP – single nucleotide polymorphism
ss – single-stranded
SSC – sodium chloride sodium citrate (saline sodium citrate)
SSCP – single-strand conformation polymorphism
ssDNA – single-stranded DNA
SSO – sequence-specific oligonucleotide
SSP – sequence-specific probe
SSPE – sodium chloride-sodium phosphate-EDTA
STR – short tandem repeat
T – thymine or thymidine, tera (10^{12})
$\tau_{1/2}$ – half-life
TAE – Tris-acetate-EDTA
TBE – Tris-borate-EDTA
TBS – Tris-buffered saline
Tc – tetracycline
TCA – trichloroacetic acid
TE – Tris-EDTA-buffer
TEMED – N, N, N′, N′-tetramethylethylene diamine
Terabyte (TB) – 10^{12} bytes
3′ – carbon atom 3 of deoxyribose

TLC	– thin-layer chromatography
T_m	– melting temperature
TOF	– time of flight
Tp	– trimethoprim
Tris	– tris (hydroxymethyl) aminomethane
tRNA	– transfer RNA
U	– unit(s)
U	– uracil or uridine
URL	– uniform resource locator
UV	– ultraviolet
V	– voltage, volt(s)
VNTR	– variable number of tandem repeats
vol	– volume
v/v	– volume/volume
w/v	– weight/volume
www	– world wide web
$\bar{X}$	– mean
χ^2	– chi squared
YAC	– yeast artificial chromosome
yr	– year(s)

Contents

The Dictionary of Genomics, Transcriptomics and Proteomics, 4th Ed., Günter Kahl

ISBN: 978-3-527-32073-8