CRC REVIVALS

Handbook of Biochemistry and Molecular Biology

Proteins

3rd Edition - Volume I

Edited by
Gerald D. Fasman

CRC Press
Taylor & Francis Group

Handbook of Biochemistry and Molecular Biology

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Handbook of Biochemistry and Molecular Biology

3rd Edition

Proteins —Volume I

EDITOR

Gerald D. Fasman, Ph. D.

Rosenfield Professor of Biochemistry
Graduate Department of Biochemistry
Brandeis University
Waltham, Massachusetts

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

First published 1976 by CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Reissued 2018 by CRC Press

ISBN 13: 978-1-138-59688-7 (hbk)
ISBN 13: 978-0-429-48737-8 (ebk)

Visit the Taylor & Francis Web site at http://www.taylorandfrancis.com and the
CRC Press Web site at http://www.crcpress.com

Handbook of Biochemistry and Molecular Biology

3rd Edition

Proteins

Volume I

Editor
Gerald D. Fasman, Ph. D.
Rosenfield Professor of Biochemistry
Graduate Department of Biochemistry
Brandeis University
Waltham, Massachusetts

The following is a list of the four major sections of the *Handbook*, each consisting of one or more volumes

Proteins – Amino Acids, Peptides, Polypeptides, and Proteins

Nucleic, Acids – Purines, Pyrimidines, Nucleotides, Oligonucleotides, tRNA, DNA, RNA

Lipids, Carbohydrates, Steroids

Physical and Chemical Data, Miscellaneous – Ion Exchange, Chromatography, Buffers, Miscellaneous, e.g., Vitamins

PREFACE

The rapid pace at which new data is currently accumulated in science presents one of the significant problems of today – the problem of rapid retrieval of information. The fields of biochemistry and molecular biology are two areas in which the information explosion is manifest. Such data is of interest in the disciplines of medicine, modern biology, genetics, immunology, biophysics, etc., to name but a few related areas. It was this need which first prompted CRC Press, with Dr. Herbert A. Sober as Editor, to publish the first two editions of a modern *Handbook of Biochemistry*, which made available unique, in depth compilations of critically evaluated data to graduate students, post-doctoral fellows, and research workers in selected areas of biochemistry.

This third edition of the *Handbook* demonstrates the wealth of new information which has become available since 1970. The title has been changed to include molecular biology; as the fields of biochemistry and molecular biology exist today, it becomes more difficult to differentiate between them. As a result of this philosophy, this edition has been greatly expanded. Also, previous data has been revised and obsolete material has been eliminated. As before, however, all areas of interest have not been covered in this edition. Elementary data, readily available elsewhere, has not been included. We have attempted to stress the areas of today's principal research frontiers and consequently certain areas of important biochemical interest are relatively neglected, but hopefully not totally ignored.

This third edition is over double the size of the second edition. Tables used from the second edition without change are so marked, but their number is small. Most of the tables from the second edition have been extensively revised, and over half of the data is new material. In addition, a far more extensive index has been compiled to facilitate the use of the Handbook. To make more facile use of the Handbook because of the increased size, it has been divided into four sections. Each section will have one or more volumes. The four sections are titled:

Proteins – Amino Acids, Peptides, Polypeptides, and Proteins

Nucleic Acids – Purines, Pyrimidines, Nucleotides, Oligonucleotides, tRNA, DNA, RNA

Lipids, Carbohydrates, Steroids

Physical and Chemical Data, Miscellaneous – Ion Exchange, Chromatography, Buffers, Miscellaneous, e.g., Vitamins

By means of this division of the data, we can continuously update the *Handbook* by publishing new data as they become available.

The Editor wishes to thank the numerous contributors, Dr. Herbert A. Sober, who assisted the Editor generously, and the Advisory Board for their counsel and cooperation. Without their efforts this edition would not have been possible. Special acknowledgments are due to the editorial staff of CRC Press, Inc., particularly Ms. Susan Cubar Benovich, Ms. Sandy Pearlman, and Mrs. Gayle Tavens, for their perspicacity and invaluable assistance in the editing of the manuscript. The editor alone, however, is responsible for the scope and the organization of the tables.

We invite comments and criticisms regarding format and selection of subject matter, as well as specific suggestions for new data (and their sources) which might be included in subsequent editions. We hope that errors and omissions in the data that appear in the Handbook will be brought to the attention of the Editor and the publisher.

Gerald D. Fasman
Editor
August 1975

PREFACE TO AMINO ACIDS, PEPTIDES, POLYPEPTIDES, AND PROTEINS, VOLUME I

The section of the *Handbook of Biochemistry and Molecular Biology* on Amino acids, Peptides, Polypeptides and Proteins is divided into three volumes. The first volume contains information relating to the naturally occurring amino acids, α, β-unsaturated amino acids, amino acid antagonists, and α-keto analogues of amino acids.

Data on ultraviolet absorption, fluorescence, optical rotatory dispersion and circular dichroism of amino acids and their derivatives are contained herein.

Relevant data on peptide synthesize is included: Amino acid derivatives, preparation of sequential polypeptides, solid state synthesize, and poly-α-amino acids.

The second and third volumes will contain material mainly on proteins.

Although the data, for which the editor alone is responsible, are far from complete, it is hoped these volumes will be of assistance to those working in the field of biochemistry and molecular biology.

Gerald D. Fasman
Editor
January 1976

THE EDITOR

Gerald D. Fasman, Ph.D., is the Rosenfield Professor of Biochemistry, Graduate Department of Chemistry, Brandeis University, Waltham, Massachusetts.

Dr. Fasman graduated from the University of Alberta in 1948 with a B.S. Honors Degree in Chemistry, and he received his Ph.D. in Organic Chemistry in 1952 from the California Institute of Technology, Pasadena, California. Dr. Fasman did postdoctoral studies at Cambridge University, England, Eidg. Technische Hochschule, Zurich, Switzerland, and the Weizmann Institute of Science, Rehovoth, Israel. Prior to moving to Brandeis University, he spent several years at the Children's Cancer Research Foundation at the Harvard Medical School. He has been an Established Investigator of the American Heart Association, a National Science Foundation Senior Postdoctoral Fellow in Japan, and recently was a John Simon Guggenheim Fellow.

Dr. Fasman is a member of the American Chemical Society, a Fellow of the American Association for the Advancement of Science, Sigma Xi, The Biophysical Society, American Society of Biological Chemists, The Chemical Society (London), the New York Academy of Science, and a Fellow of the American Institute of Chemists. He has published 180 research papers.

The Editor and CRC Press, Inc. would like to dedicate this third edition to the memory of Eva K. and Herbert A. Sober. Their pioneering work on the development of the Handbook is acknowledged with sincere appreciation.

TABLE OF CONTENTS

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Nomenclature

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

BIOCHEMICAL NOMENCLATURE

This synopsis of the recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature (CBN) was prepared by Waldo E. Cohn, Director, NAS-NRC Office of Biochemical Nomenclature (OBN, located at Biology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830), from whom reprints of the CBN publications listed below and on which the synopsis is based are available.

The synopsis is divided into three sections: Abbreviations, symbols, and trivial names. Each section contains material drawn from the documents (A1 to C1, inclusive) listed below, which deal with the subjects named.

Additions consonant with the CBN Recommendations have been made by OBN throughout the synopsis.

RULES AND RECOMMENDATIONS AFFECTING BIOCHEMICAL NOMENCLATURE AND PLACES OF PUBLICATION (AS OF FEBRUARY 1975)

I. IUPAC-IUB Commission on Biochemical Nomenclature

- A1. Abbreviations and Symbols [General; Section 5 replaced by A6]
- A2. Abbreviated Designation of Amino-acid Derivatives and Peptides (1965) [Revised 1971; Expands Section 2 of A1]
- A3. Synthetic Modifications of Natural Peptides (1966) [Revised 1972]
- A4. Synthetic Polypeptides (Polymerized Amino Acids) (1967) [Revised 1971]
- A5. A One-letter Notation for Amino-acid Sequences (1968)
- A6. Nucleic Acids, Polynucleotides, and their Constituents (1970)

- B1. (Nomenclature of Vitamins, Coenzymes, and Related Compounds)
  - a. Miscellaneous [A, B's, C, D's, tocols, niacins; see B2 and B3]
  - b. Quinones with Isoprenoid Side-chains: E, K, Q [Revised 1973]
  - c. Folic Acid and Related Compounds
  - d. Corrinoids: B-12's [Revised 1973]
- B2. Vitamins B-6 and Related Compounds [Revised 1973]
- B3. Tocopherols (1973)

- C1. Nomenclature of Lipids (1967) [Amended 1970; see also II, 2]
- C2. Nomenclature of α-Amino Acids (1974) [See also II, 5]

- D1. Conformation of Polypeptide Chains (1970) [See also III, 2]

- E1. Enzyme Nomenclature (1972)[a] [Elsevier (in paperback); Replaces 1965 edition.]
- E2. Multiple Forms of Enzymes (1971) [Chapter 3 of E1]
- E3. Nomenclature of Iron-sulfur Proteins (1973) [Chapter 6.5 of E1]
- E4. Nomenclature of Peptide Hormones (1974)

II. Documents Jointly Authored by CBN and CNOC [See III]

1. Nomenclature of Cyclitols (1968) [Revised 1973]
2. Nomenclature of Steroids (1968) [Amended 1971; Revised 1972]
3. Nomenclature of Carbohydrates-I (1969)
4. Nomenclature of Carotenoids (1972) [Revised 1975]
5. Nomenclature of α-Amino Acids (1974) [Listed under I, C2 in the following table]

III. IUPAC Commission on the Nomenclature of Organic Chemistry (CNOC)

1. Section A (Hydrocarbons), Section B (Heterocyclics): *J. Am. Chem. Soc.*, 82, 5545;[a] Section C (Groups containing N, Hal, S, Se/Te): *Pure Appl. Chem.*, 11, Nos. 1–2[a] [A, B, and C Revised 1969:[a] Butterworth's, London (1971)]
2. Section E (Stereochemistry):[b] *J. Org. Chem.*, 35, 2489 (1970); *Biochim. Biophys. Acta*, 208, 1 (1970); *Eur. J. Biochem.*, 18, 151 (1970) [See also I, D1]

[a] No reprints available from OBN; order from publisher.
[b] Reprints available from OBN (in addition to all in IA to ID and II).

RULES AND RECOMMENDATIONS AFFECTING BIOCHEMICAL NOMENCLATURE AND PLACES OF PUBLICATION (AS OF FEBRUARY 1975)(continued)

IV. Physiochemical Quantities and Units (IUPAC)[a] *J. Am. Chem. Soc.*, 82, 5517 (1960) [Revised 1970: *Pure Appl. Chem.*, 21, 1 (1970)]

V. Nomenclature of Inorganic Chemistry (IUPAC) *J. Am. Chem. Soc.*, 82, 5523[a] [Revised 1971: *Pure Appl. Chem.*, 28, No. 1 (1971)][a]

VI. Drugs and Related Compounds or Preparations
   1. U.S. Adopted Names (USAN) No. 10 (1972) and Supplement [U.S. Pharmacopeial Convention, Inc., 12601 Twinbrook Parkway, Rockville, Md.]
   2. International Nonproprietary Names (INN) [WHO, Geneva]

CBN RECOMMENDATIONS APPEAR IN THE FOLLOWING PLACES[a]

	Arch. Biochem. Biophys.	*Biochem. J.*	*Biochemistry*	*Biochim. Biophys. Acta*	*Eur. J. Biochem.*	*J. Biol. Chem.*	*Pure Appl. Chem.*[b]	*Biochimie (Bull. Soc.)*[c]	*Molek. Biol.*[d]	*Z. Phys. Chem.*[e]
A1[f]	136,1	101,1	5,1445		1,259	241,527		50,3	1,872	348,245
A2(Revised)	150,1(R)	126,773(R)	11,1726(R)	263,205(R)	27,201(R)	247,977(R)	40,(R)	49,121*	2,282*	348,256*
A3(Revised)	121,6*	104,17*	6,362*	133,1*	1,379*	242,555*	31,649(R)	49,325*	2,466*	348,262*
A4(Revised)[g]	151,597(R)	127,753(R)	11,942(R)	278,211(R)	26,301(R)	247,323(R)	33,439(R)	51,205*	5,492(R)	349,1013*
A5	125(3),i	113,1	7,2703	168,6	5,151	243,3557	31,641	50,1577	3,473	350,793
A6[h]	145,425	120,449	9,4022	247,1	15,203	245,5171	40,		6,167	351,1055
B1*	118,505	102,15		107,1(a–c)	2,1	241,2987		49,331		348,266
B1b(Revised)	165.1(R)	147,15(R)		387,397(R)	53,15(R)		38,439			
Bld(Revised)	161(2),iii(R)	147,1(R)	13,1555(R)		45,7(R)					
B2(Revised)	162,1(R)	137,417(R)	13,1056(R)	354,155(R)	40,325(R)	245,4229*	33,447(R)			351,1165*
B3(Revised)	165,6(R)	147,11(R)			46,217(R)					
C1[f]	123,409	105,897	6,3287	152,1	2,127	242,4845		50,1363	2,784	350,279
Amendments		116(5)		202,404	12,1	245,1511				
C2			14,449		53,1					
D1[i]	145,405	121,577	9,3471	229,1	17,193	245,6489			7,289	
E2	147,1	126,769	10,4825	258,1	24,1	246,6127		54,123		353,852
E3	160,355	135,5	12,3582	310,295	35,1	248,5907				
E4		151,1	14,2559			250,3215				
II,I(Revised)	128,269*	112,17*		165,1*	5,1*	243,5809*	37,285(R)	51,3*		350,523*
II,2[f]	136,13	113,5	8,2227	164,453	10,1		31,285(R)	51,819		351,663
Amendments	147,4	127,613	10,4994	248,387	25,2					
II,3		125,673	10,3983	244,223	21,455	247,613				
II,4		127,741	10,4827	286,217	25,397	247,2633				
Amendments		151,507	14,1803							

[a]Reprints available from OBN.
[b]No reprints available from OBN; order from publisher.
[c]In French.
[d]In Russian.
[e]In German.
[f]Also in other journals.
[g]Also in *Biopolymers*, 11, 321.
[h]*J. Mol. Biol.*, 55, 299.
[i]*J. Mol. Biol.*, 52, 1.

*First, unrevised version.
(R) = revised version.

ABBREVIATIONS

Abbreviations are distinguished from **symbols** as follows (taken from Reference A1):

a. **Symbols,** for monomeric units in macromolecules, are used to make up abbreviated structural formulas (e.g., Gly-Val-Thr for the tripeptide glycylvalylthreonine) and can be made fairly systematic.

b. **Abbreviations** for semi-systematic or trivial names (e.g., ATP for adenosine triphosphate; FAD for flavinadenine dinucleotide) are generally formed of three or four capital letters, chosen for brevity rather than for system. It is the indiscriminate coining and use of such abbreviations that has aroused objections to the use of abbreviations in general.

[Abbreviations are thus distinguished from symbols in that they (a) are for semi-systematic or trivial names, (b) are brief rather than systematic, (c) are usually formed from three or four capital letters, and (d) are not used – as are symbols – as units of larger structures. ATP, FAD, etc., are abbreviations. Gly, Ser, Ado, Glc, etc., are symbols (as are Na, K, Ca, O, S, etc.); they are sometimes useful as abbreviations in figures, tables, etc., where space is limited, but are usually not permitted in text. The use of abbreviations is permitted when necessary but is never required.]

1. Nucleotides (N = A, C, G, I, O, T, U, X, ψ – see Symbols)

NMP	Nucleoside 5′-phosphate
NDP	Nucleoside 5′-di(or pyro)phosphate
NTP	Nucleoside 5′-triphosphate

Prefix d indicates deoxy.

2. Coenzymes, vitamins

CoA(or CoASH)	Coenzyme A
CoASAc	Acetyl Coenzyme A
DPN[a]	Diphosphopyridine nucleotide
FAD	Flavin-adenine dinucleotide
FMN	Riboflavin 5′-phosphate
GSH	Glutathione
GSSG	Oxidized glutathione
NAD[b]	Nicotinamide-adenine dinucleotide (cozymase, Coenzyme I, diphosphopyridine nucleotide)
NADP[b]	Nicotinamide-adenine dinucleotide phosphate (Coenzyme II, triphosphopyridine nucleotide)
NMN	Nicotinamide mononucleotide
TPN[c]	Triphosphopyridine nucleotide

3. Miscellaneous

ACTH	Adrenocorticotropin, adrenocorticotropic hormone, or corticotropin
CM-cellulose	*O*-(Carboxymethyl)cellulose
DEAE-cellulose	*O*-(Diethylaminoethyl)cellulose
DDT	1,1,1-Trichloro-2,2-bis(*p*-chlorophenyl)ethane
EDTA	Ethylenediaminetetraacetate
Hb,HbCO,HbO_2	Hemoglobin, carbon monoxide hemoglobin, oxyhemoglobin
P_i	Inorganic orthophosphate

[a]Replaced by NAD (also DPN^+ by NAD^+, DPNH by NADH).
[b]Generic term; oxidized and reduced forms are NAD^+, NADH ($NADP^+$, NADPH).
[c]Replaced by NADP (also TPN^+ by $NADP^+$, TPNH by NADPH).

PP_i	Inorganic pyrophosphate
TEAE-cellulose	*O*-(Triethylaminoethyl)cellulose
Tris	Tris(hydroxymethyl)aminomethan (2-amino-2-hydroxymethylpropane-1,3-diol)

4. Nucleic Acids

DNA, RNA	Deoxyribonucleic acid, ribonucleic acid (or -nucleate)
hnRNA	Heterogeneous RNA
mtDNA	Mitochondrial DNA
cRNA	Complementary RNA
mRNA	Messenger RNA
nRNA	Nuclear RNA
rRNA	Ribosomal RNA
tRNA	Transfer RNA (generic term; sRNA should not be used for this or any other purpose)
$tRNA^{Ala}$	Alanine tRNA; $tRNA_1^{Ala}$, $tRNA_2^{Ala}$: isoacceptor alanine tRNA's
AA-tRNA	Aminoacyl-tRNA; aminoacylated tRNA; "charged" tRNA (generic term)
Ala-tRNA or $Ala\text{-}tRNA^{Ala}$	Alanyl-tRNA
$tRNA^{Met}$	Methionine tRNA (not enzymically formylatable)
$tRNA^{fMet}$ or $tRNA_f^{Met}$	Methionine tRNA, enzymically formylatable to . . .
fMet-tRNA	Formylmethionyl-tRNA (small f, to distinguish from fluorine F)

SYMBOLS

Symbols are distinguished from abbreviations in that they are designed to represent specific parts of larger molecules, just as the symbols for the elements are used in depicting molecules, and are thus rather systematic in construction and use. Symbols are not designed to be used as abbreviations and should not be used as such in text, but they may often serve this purpose when space is limited (as in a figure or table). Symbols are always written with a single capital letter, all subsequent letters being lower-case (e.g., Ca, Cl, Me, Ac, Gly, Rib, Ado), regardless of their position in a sequence, a sentence, or as a superscript or subscript.

Some abbreviations expressed in symbols (see also Section II F below), as examples of the use of symbols:

Dimethylsulfoxide	Me_2SO [a]
Tetranitromethane	$(NO_2)_4C$ [b]
Guanidine hydrochloride	Gdn · HCl [c]
Guanidinium chloride	GdmCl
Cetyltrimethylammonium bromide	$CtMe_3NBr$ [d]
Ethyl methanesulfonate	$MeSO_3Et$
Methylnitronitrosoguanidine	MeN_2O_3Gdn
-nitrosourea	-Nur [e]
-nitrosamine	-Nam [f]
-fluorene	-Fln
Aminofluorene	NH_2Fln
Acetylaminofluorene	AcNHFln [g]
Acetoxyacetylaminofluorene	Ac(AcO)NFln
N-Acetylneuraminic acid	AcNeu [h]

[a] Replaces DMSO.
[b] Replaces TNM.
[c] Replaces Gu, Gd, and G.
[d] Replaces CTAB (similarly for other ammonium compounds).
[e] Replaces NU.
[f] Replaces NA.
[g] Replaces AAF.
[h] Not NANA.

I. Phosphorylated Compounds (Reference A1)

$-PO_3H_2$ (or its ions)	*-P*(or *P-*) ("p" in Nucleic Acids; see IV)
$-PO_2H-$(or its ion)	*-P-*(hyphen in Nucleic Acids; see IV)
$-PO_2H-PO_3H_2$ (or ions)	*-P-P* or *-PP* or *PP-* (cf. PP_i in Abbreviations)

Examples:[a]

Glucose 6-phosphate	Glucose-6-*P* (or Glc-6-*p*; see II below).
Phosphenolpyruvate (pyruvenol phosphate)	*P-enol*Pyruvate or *enol*Pyruvate-*P* or *e* Prv-*P*[b]
Fructose 1,6-bisphosphate (not di)	Fructose-1,6-P_2 (or Fru-1,6-P_2; see II below).
Creatine phosphate	Creatine-*P*
Phosphocreatine	*P*-Creatine

[a]Note that symbols are hyphenated even where names are not.
[b]Recommended by OBN.

II. Peptides and Proteins (References A1–A5)

A. Symbols (Reference A2–A5)

1. Common amino acids

Name	Symbol		Name	Symbol	
	Three-letter[a]	One-letter[b]		Three-letter[a]	One-letter[b]
Alanine	Ala	A	Lysine[e]	Lys	K
Arginine	Arg	R	Methionine	Met	M
Asparagine	Asn [c,d,e]	N	Phenylalanine	Phe	F
Aspartic acid	Asp [d,e]	D	Proline	Pro	P
Cysteine	Cys [e]	C	Serine[e]	Ser	S
Glutamic acid	Glu [f]	E	Threonine[e]	Thr	T
Glutamine	Gln [e,f,g]	Q	Tryptophan[e]	Trp	W
Glycine	Gly	G	Tyrosine[e]	Tyr	Y
Histidine	His [e]	H	Valine	Val	V
Isoleucine	Ile	I	Unknown or "other"	AA [h]	X
Leucine	Leu	L			

[a]One capital, two small letters, at all times.
[b]For special uses and with special conventions; see III, *I* following.
[c]Or Asp (NH_2); see Footnotes e and g.
[d]Uncertainty as between Asp and Asn may be designated by Asx (or B).
[e]Substitution on a functional group may be indicated, as shown in Footnotes c and g, by parenthesis following the symbol, e.g., Cys (Cme), Ser (*P*); see C 2 below.
[f]Uncertainty, as between Glu and Gln, may be designated by Glx (or Z); pyroglutamate is pGlu or <Glu, not PCA.
[g]Or Glu (NH_2); see Footnotes c and e.
[h]See Abbreviations, Part 4 (AA-tRNA).

2. Less common amino acids

Name	Symbol
β-Alanine	βAla
Alloisoleucine	aIle
2-Aminoadipic acid	Aad
3-Aminoadipic acid	βAad
2-Aminobutyric acid	Abu
6-Aminocaproic acid[a]	ϵAhx [a]
2-Aminopimelic acid	Apm
2,4-Diaminobutyric acid	A_2bu [b]
2,2′-Diaminopimelic	A_2pm [b]
2,3-Diaminopropionic acid	A_2pr [b]
N-Ethylglycine, etc.	EtGly, etc.
Hydroxylysine	Hyl
allo-Hydroxylysine	*a*Hyl
3-Hydroxyproline	3Hyp [c]
4-Hydroxyproline	4Hyp [c]
N-Methylglycine (sarcosine)	MeGly or Sar
N-Methylisoleucine	MeIle
N-Methylvaline, etc.	MeVal, etc.
Norleucine	Nle
Norvaline	Nva
Ornithine	Orn

[a] 6-Aminohexanoic acid is preferred; see Reference C2.
[b] The use of D (for di), T (for tri or tetra), etc., is undesirable. Hence, in this context, A_2 for diamino is recommended (cf. II F below).
[c] Or Pro(PH) for hydroxyproline.

B. Sequence, Direction, and Bonding (Reference A2)

→ peptide bond, originating in peptide -CO-
– peptide bond, originating in peptide -CO- of residue at left.
, separates symbols in unknown sequence (the entire unknown sequence is enclosed in parentheses).
| bond originating in first letter of symbol of a residue having a substituted functional group (-SH, 3- or 4-COOH, -OH, 6-NH_2, etc., or the remaining H of a peptide bond; see C2 below).

C. Substitution (Reference A2)

Groups substituted for hydrogen or for hydroxyl may be indicated either by their structural formulae, or by symbols, or by combinations of both, e.g.,

Benzoylglycine (hippuric acid)	PhCo-Gly or C_6H_5CO-Gly or Bz-Gly
Glycine methyl ester	Gly-OCH_3 or Gly-OMe
Trifluoroacetylglycine	CF_3CO-Gly

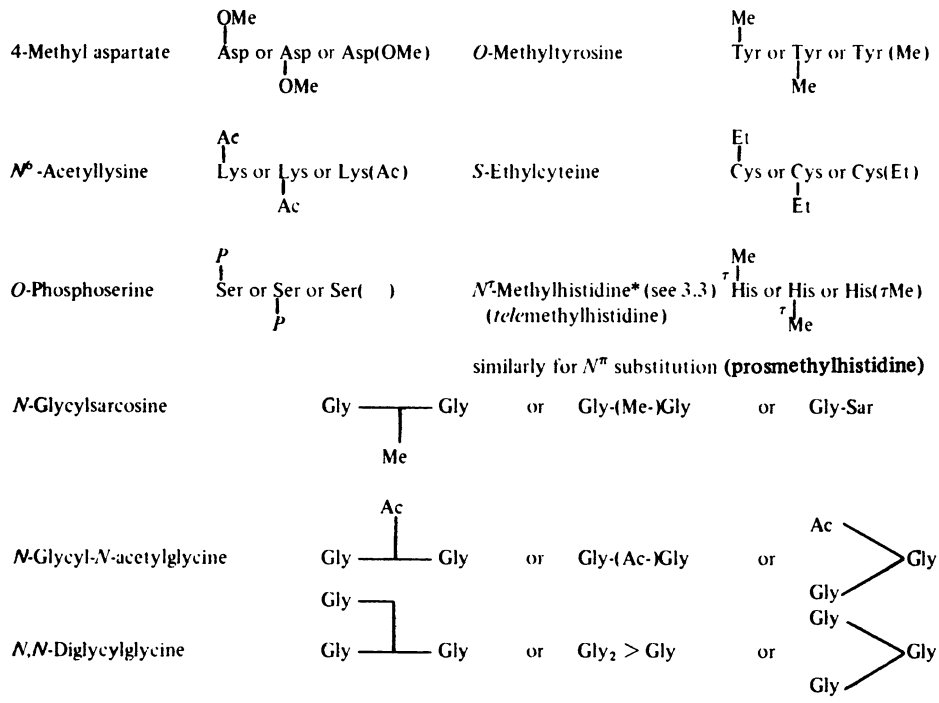
1. In α-NH_2 or α-COOH groups: horizontal dash (hyphen) to left or right, respectively.

N-Acetylglycine	Ac-Gly
Glycine ethyl ester	Gly-OEt
N^2-Acetyllysine	Ac-Lys
Serine methyl ester	Ser-OMe
O^1-Ethyl *N*-acetylglutamate	Ac-Glu-OEt
Isoglutamine	Glu-NH_2

N-Tosylphenylalanyl chloromethyl ketone (TPCK) Tos-Phe CH_2Cl

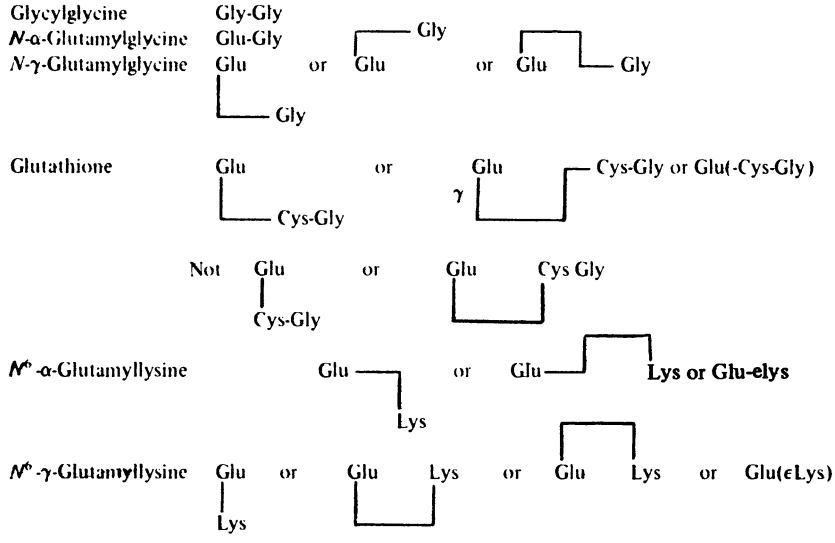
N-Ethyl-*N*-methylglycine Et-(Me-)Gly or 

2. On functional group: Vertical bond (see B above) or parentheses (see II A1 above, Footnote e).



*The prolonged and well-entrenched ambiguity in the nomenclature of the *N*-1 being the biochemist's *N*-3 and *vice versa*) led to a new trivial system for designating these substances: The imidazole N *nearer* the alanine residue is designated *pros* (symbol π) and the one *farther* *tele* (symbol τ), to give the following names and symbols: *pros*methylhistidine or π-methylhistidine, His(πMe); *tele*methylhistidine or τ-methylhistidine, His(τMe).

D. Polypeptides: Follow Rules for Substitution (C above) (Reference Az)



Glycyllysylglycine dihydrochloride $+H_2$— Gly-Lys-Gly—OH · 2HCl

Its N^{ε}-formylderivative Gly-Lys-Gly - or Gly-Lys(CHO)-Gly

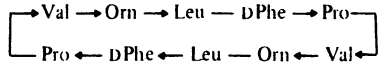
CHO

E. Cyclic Polypeptides (Reference A2)

1. Homodetic: Gramicidin *S*

cyclo-Val-Orn-Leu-DPhe-Pro-Val-Orn-Leu-DPhe-Pro

Val-Orn-Leu-DPhe-Pro-Val-Orn-Leu-DPhe-Pro



2. Heterodetic:

Oxytocin Cys-Tyr-Ile-Asn-Gln-Cys-Pro-Leu-Gly-NH_2

Cyclic ester of threonylglycylglycylglycine

Thr-Gly-Gly-Gly or H-Thr-Gly-Gly-Gly

F. Substituents (Reference A2)

1. NH_2 protecting groups of the urethan type (partial list)

Benzyloxycarbonyl-	Z- or Cbz-[a]	*p*-Methoxyphenylazobenzyloxycarbonyl-	Mz-
p-Nitrobenzyloxycarbonyl-	$Z(NO_2)$-	*p*-Phenylazobenzyloxycarbonyl-	Pz-
p-Bromobenzyloxycarbonyl-	Z(Br)-	*t*-Butoxycarbonyl-	Boc-[b] or Bu^tOCO-
p-Methoxybenzyloxycarbonyl-	Z(OMe)-	Cyclopentyloxycarbonyl-	Poc- or cPeOCO-

2. Other N- protecting groups (partial list)

Acetyl-	Ac-	Maleoyl- (–OC–CH=CH–CO–)	Mal-[c] or Mal<
Benzoyl-(C_6H_5CO-)	PhCO- or Bz-	Maleyl- (HOOC–CH=CH–CO–)	Mal-
Benzyl- ($C_6H_5CH_2$-)	$PhCH_2$- or Bzl	Methylthiocarbamoyl-	MeNHCS-[a] or Mtc-[f]
Benzylthiomethyl-	$PhSCH_2$- or Btm-	*o*-Nitrophenylthio-	Nps-
Carbamoyl-	NH_2CO- (preferred to Cbm)	Phenylthiocarbamoyl-	PhNHCS-[a] or Ptc-[f]
		Phthaloyl-	–Pht- or Pht<
1-Carboxy-2-nitrophenyl-5-thio-	Nbs-	Phthalyl-	Pht-
3-Carboxypropionyl- (HOOC–CH_2–CH_2–CO–)	Suc-	Succinyl- (–OC–CH_2–CH_2–CO–)	–Suc- or Suc<
		Tetrahydropyranyl-	H_4pyran-[c,g]

Dansyl-(5-dimethylamino-naphthalene-inonaphthalene-1-sulfonyl)	Dns-[c] or dansyl[a,3]	Tosyl- (*p*-tolylsulfonyl)	Tos- or tosyl
Dinitrophenyl-	N_2 ph-[a,c] or Dnp	Trifluoroacetyl-	CF_3CO- or F_3Ac-[a,h]
Formyl-	HCO-[d] or CHO-	Trityl- (triphenylmethyl)	Ph_3C-[a,i] or Trt-
p-Iodophenylsulfonyl (pipsyl)	Ips or pipsyl		

3. Substituents at carboxyl group

Benzyloxy- (benzyl ester)	$-OCH_2Ph$ or -OBzl	*p*-Nitrophenoxy- (*p*-nitrophenyl ester)	-ONph
Cyanomethoxy- (cyanomethyl ester)	$-OCH_2CN$ or -OMeCN	*p*-Nitrophenylthio-	-SNph
Diphenylmethoxy- (benzhydryl ester)	$-OCHPh_2$ or -OBzh	Phenylthio- (phenylthiolester)	-SPh
Ethoxy- (ethyl ester)	-OEt	1-Piperidino-oxy-	-OPip
Methoxy- (methyl ester)	-OMe	8-Quinolyloxy-	-OQu
		Succinimido-oxy-	-ONSuc
		Tertiary butoxy- (*t*-butyl ester)	$-OBu^t$

[a] Preferred.
[b] Not BOC or *t*BOC.
[c] The use of D for di and T for tri (or tetra) is discouraged. Recognized symbols with numerical subscripts are recommended.
[d] fMet is approved for formylmethionine.
[e] MalNEt is recommended for *N*-ethylmaleimide (not NEM).
[f] Mtc and Ptc have been used to denote methyl- and phenylthiohydantoins (e.g., Ptc-Leu). Since this incorrectly implies the substitution of an amino acid by a "phenyl (or methyl) thiohydantoyl" group, the correct representation, CS-Leu-NPh or PhNCS-Leu, or, in text, Leu>PhNCS, is recommended.
[g] Not THP or Thp (see Footnote c).
[h] Not TFA (see Footnote c).
[i] Or trityl.

4. Other substituents (and reagents)

2-Aminoethyl-[a]	-$(CH_2)_2NH_2$ or Aet-[a,b]	Chloroethylamine	$Cl(CH_2)_2NH_2$ or AetCl
Carbamoylmethyl-	$-CH_2CONH_2$ or Ncm-	Chloroacetamide	$ClCH_2CONH_2$ or NcmCl
Carboxymethyl-	$-CH_2CO_2H$ or Cxm-	Chloroacetic acid	$ClCH_2CO_2H$ or CxmCl
p-Carboxyphenylmercuri-	-HgBzOH	*p*-Chloromercuribenzoate	$ClHgBzO^-$[c]
1-Carboxy-2-mitrophenyl-5-thio- nitrophenyl-5-thio-	Nbs-	5,5′-Dithiobis(2-nitrobenzoic acid) (2-nitrobenzoic acid)	Nbs_2[d]
Diazoacetyl-	N_2CHCO- or N_2<Ac-		
-Diisopropylphosphor	iPr_2P-[e]	Diisopropylfluorophosphate	iPr_2P-F[f]
Dinitrophenyl-	N_2ph[g]	Fluorodinitrobenzene	N_2ph-F[b]
Hydroxyethyl-	-$(CH_2)_2OH$ or HOEt-	Ethylene oxide	$(CH_2)_2O$ or Et>O
		N-Ethylmaleimide	MalNEt[i]
		Tetrahydrofuran	H_4furan[j]
		Tosyllysyl chloromethyl ketone	Tos-LysCH_2Cl[k]

		Tosylarginine methyl ester	Tos-ArgOMe[l]
Trifluoroacetyl-	F_3Ac[m]	Trifluoroacetic acid	F_3AcOH
Trimethylsilyl-	Me_3Si-[n]	Tetramethylsilane	Me_4Si[n]

[a]For -ethylamine, -Etn; for -ethanolamine (see Lipids), -OEtn.
[b]Not AET.
[c]Replaces PCMB, *p*CMB, and CMB.
[d]Replaces DTNB.
[e]Replaces, DIP and Dip.
[f]Replaces DPF, DFP, DIPF, etc.
[g]Replaces DNP and Dnp.
[h]Replaces FDNB.
[i]Replaces NEM.
[j]Replaces THF. Similarly, H_4 folate.
[k]Replaces TLCK (similarly for TPCK, etc.).
[l]Replaces TAME (similarly for other N-substituted amino-acid esters. See C1 above).
[m]Replaces TFA.
[n]Not TMS- or TMS. Similarly, Me_2SO, not DMSO; NAc_3, not NTA.

G. Polymerized Amino Acids (Synthetic Polypeptides) (Reference A4)

1. Linear polymers (only normal peptide links are involved).

a. Homopolymer: polylysine; poly(Lys) or $(Lys)_n$ (n may be replaced by a number).

b. Copolymer, alternating sequence: poly(alanine-lysine); poly(Ala-Lys) or $(Ala\text{-}Lys)_n$.

c. Copolymer, random sequence, composition unspecified: poly(alanine, lysine); poly(Ala,Lys) or $(Ala,Lys)_n$.

d. Copolymer, random sequence, molar percentages (Σ = 100%) known: poly(DLGlu^{56}Lys^{38}DTyr^{6}) or $(DLGlu^{56}Lys^{38}DTyr^{6})_n$ (only lysine is L).

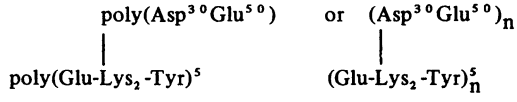
e. Block polymer of poly(Glu) linked via α-COOH to α-NH_2 of poly(Lys): poly(Glu^{56})-poly(Lys^{44}) or $(Glu^{56})_n$-$(Lys^{44})_n$.

f. Block polymer, a repeating series of the known sequence Glu-Lys-Lys-Tyr: poly(Glu-Lys_2-Tyr) or $(Glu\text{-}Lys_2\text{-}Tyr)_n$.

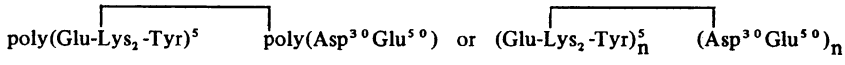
g. Block polymer of two repeating series: poly $(Glu\text{-}Lys)^{25}$-poly$(Ala\text{-}Tyr_2\text{-}Glu)^{12\text{-}5}$ or $(Glu\text{-}Lys)_n^{25}$-$(Ala\text{-}Tyr_2\text{-}Glu)_n^{12\text{-}5}$ (molar percentages = 100).

2. Branched graft polymers (functional groups are involved).

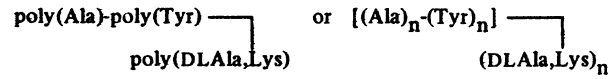
a. Main chain is a repeating sequence (see 1f above), sidechain is of random sequence, connection is from ϵ-NH_2 of a lysine to an unknown group in the sidechain



or



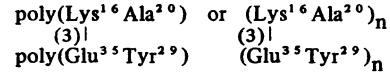
b. Main chain of unknown sequence, linked via ϵ-NH_2 of a lysine to the α-COOH of an L-tyrosine in the sidechain, which is a block polymer (no analytical data):



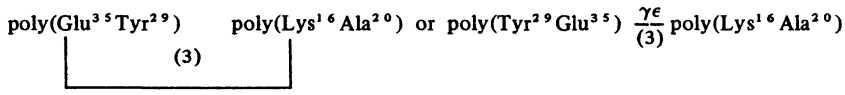
or

Note: The points of attachment of Lys and Tyr cannot be specified in the last example. This system, depending on double hyphens to express functional group involvement, is not recommended.

c. Two linear copolymers of unknown sequence, triply linked between ϵ-NH_2 groups of lysines and γ-COOH residues of glutamates:

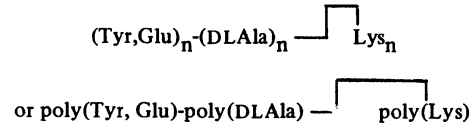


or



(See comment under b with respect to last example).

d. Linear, random-sequence chain attached via terminal α-COOH group (of either Tyr or Glu) to NH_2 terminal of poly(DLalanine) in turn connected, via terminal COOH, to ϵ-NH_2 groups(s) of poly(L-lysine) (no analytical data):



H. Synthetic Modifications of Natural Peptides (Reference A3)

Modification and Name	Abbreviation
1. Replacement:	
a. of 8th residue in vasopressin by citrulline:	
[8-Citrulline] vasopressin	[Cit^8] vasopressin
b. at 5 and 7 positions in hypertensin II:	
[5-Isoleucine, 7-alanine] hypertensin II	[Ile^5, Ala^7] hypertensin II
2. Extension of X by valyl residue, at N and C terminals:	
valyl-X	Val-X
X(yl)-valine	X(yl)-Val
3. Insertion of tyrosine residue between 4th and 5th residue	
4a-endo-tyrosine-hypertensin II	endo-Tyr^{4a}-hypertensin II
4. Removal of proline from position 7 in oxytocin:	
des-7-proline-oxytocin	des-Pro^7-oxytocin

5. Substitution of valine on ϵ-nitrogen of a lysine at position 2 in peptide X:
   $N^{\epsilon 2}$-valyl-X — $N^{\epsilon 2}$-Val-X
6. Substitution of valine on γ-carboxyl of glutamate at position 3 in peptide X:
   $C^{\gamma 3}$-X(yl)-valine — $C^{\gamma 3}$-X(yl)-Val
7. Fragments, or partial sequences: fragments from α-MSH.

Ac-Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val-NH_2 1 2 3 4 5 6 7 8 9 10 11 12 13	α-MSH
Met-Glu-His-Phe-Arg-Trp-Gly 6 10	α-MSH-(4–10)-heptapeptide
His-Phe-Arg-Lys-Pro-Val-NH_2 6 8 11 13	α-MSH-(6–8)-(11–13)-hexapeptide amide

I. One-letter Notation[a] (Reference A5)

1. Symbols: see II.A.1 above (NH_2 terminal at left, COOH terminal at right).
2. Known sequence: space[b] between symbols.
3. Unknown sequence: comma[c] between symbols, parentheses[d] enclosing.
4. Adjacent unknown sequences: = replaces)(.
5. Uncertainty as to sequence or terminus: / (see examples b and c).

Examples: a. (Ala, Cys, Asp) (Arg, Ser) (Gly, His, Ile) Lys-Leu-Met-Asn-Pro-Gln
becomes (A, C, D = R, S = G, H, I) K L M N P Q

b. A C D E F G H I K L M N P Q
c. (A. C. D = R, S = G. H. I) K L/ M N/ P Q/

In c., the tripeptides A . C . D and G . H . I are not of known sequence, but are inferred by analogy with the known peptide b.; the inference is expressed by periods instead of commas. The comma between R and S indicates that no inference as to sequence can be drawn for this dipeptide. The internal slashes indicate that no connection between L and M, and N and P, has been proven, although KL, MN and PQ are each of known internal sequence. The final slash indicates that Q has not been proven to be the COOH terminal residue of the entire peptide, although it is the terminus of the PQ dipeptide.

[a] For display of very long sequences or computer use only.
[b] In place of hyphen in three-letter system. Spaces must be equal to characters, as in typing. So must commas, dots, and all other symbols.
[c] As in three-letter system; becomes a dot (period) when sequence is inferred but not demonstrated (see example c).
[d] The double symbol,)(, is replaced by = (see 4) to preserve equal spacing.

NOMENCLATURE OF LABELED COMPOUNDS

The statement below was adopted by the IUB Commission of Editors of Biochemical Journals* (CEBJ) and appears, in the same or in similar form, in the Instructions to Authors of their journals. This system originated with the Chemical Society (London) and was subsequently adopted by the American Chemical Society (Handbook for Authors, 1967). It was adopted by CEBJ in 1971 and is the only system currently permitted in the pages of their journals.

ISOTOPICALLY LABELED COMPOUNDS

The symbol for the isotope introduced is placed in *square* brackets directly attached to the front of the name (word), as in [^{14}C]urea. When more than one position in a substance is labeled by means of the same isotope and the positions are not indicated (as below), the number of labeled positions is added as a right-hand subscript, as in [$^{14}C_2$]glycollic acid. The symbol "U" indicates uniform and "G" general labeling, e.g., [U-^{14}C]glucose (where the ^{14}C is uniformly distributed among all six positions) and [G-^{14}C]glucose (where the ^{14}C is distributed among all six positions, but not necessarily uniformly); in the latter case it is often sufficient to write simply "[^{14}C]glucose."

The isotopic prefix precedes that part of the name to which it refers, as in sodium [^{14}C]formate, iodo[$^{14}C_2$]acetic acid, 1-amino[^{14}C]methylcyclopentanol ($H_2N-{}^{14}CH_2-C_5H_8-OH$), α-naphth[^{14}C]oic acid ($C_{10}H-{}^{14}CO_2H$), 2-acetamido-7-[^{131}I]iodofluorene, fructose 1,6-[1-^{32}P]diphosphate, D-[^{14}C]glucose, 2*H*-[2-^{2}H]pyran, *S*-[8-^{14}C]adenosyl[^{35}S]methionine. Terms such as "^{131}I-labeled albumin" should not be contracted to "[^{131}I]albumin" (since native albumin does not contain iodine), and "^{14}C-labeled amino acids" should similarly not be written as "[^{14}C]amino acids" (since there is no carbon in the amino group).

When isotopes of more than one element are introduced, their symbols are arranged in alphabetical order, including ^{2}H and ^{3}H for deuterium and tritium, respectively.

When not sufficiently distinguished by the foregoing means, the positions of isotopic labeling are indicated by Arabic numerals, Greek letters, or prefixes (as appropriate), placed within the square brackets and before the symbol of the element concerned, to which they are attached by a hyphen; examples are [1-^{2}H]ethanol ($CH_3-C^2H_2-OH$), [1-^{14}C]aniline, L-[2-^{14}C]leucine (or L-[α-^{14}C]-leucine), [*carboxy*-^{14}C]leucine, [*Me*-^{14}C]isoleucine, [2,3-^{14}C]maleic anhydride, [6,7-^{14}C]xanthopterin, [3,4-^{13}C,^{35}S]-methionine, [2-^{13}C; 1-^{14}C]acetaldehyde, [3-^{14}C; 2,3-^{2}H; ^{15}N]serine.

The same rules apply when the labeled compound is designated by a standard abbreviation or symbol, other than the atomic symbol, e.g. [γ-^{32}P]ATP.

For simple molecules, however, it is often sufficient to indicate the labeling by writing the chemical formulae, e.g. $^{14}CO_2$, $H_2{}^{18}O$, $^{2}H_2O$ (not D_2O), $H_2{}^{35}SO_4$, with the prefix superscripts attached to the proper atomic symbols in the formulae. The square brackets are not to be used in these circumstances, nor when the isotopic symbol is attached to a word that is not a chemical name, abbreviation or symbol (e.g. ^{131}I-labeled).

*CEBJ consists of the Editors-in-Chief of the following journals: *Archives of Biochemistry and Biophysics, Biochemical Journal, Biochemistry, Biochimica et Biophysica Acta, Biochimie, European Journal of Biochemistry, Hoppe-Seyler's Zeitschrift für Physiologische Chemie, Journal of Biochemistry, Journal of Biological Chemistry, Journal of Molecular Biology, and Molekulyarnaya Biologiya;* corresponding members include *Proceedings of the National Academy of Sciences* (U.S.A.) and approximately 40 others.

THE CITATION OF BIBLIOGRAPHIC REFERENCES IN BIOCHEMICAL JOURNALS RECOMMENDATIONS (1971)*

IUB Commission of Editors of Biochemical Journals (CEBJ)

These Recommendations were reviewed by the Commission in August 1972, when it was decided to publish them.

PREAMBLE

Two basic systems for the citation of references are used at present. The so-called Harvard System (where names of authors and the date are cited in the text, and the reference list is in alphabetical order) and the Numbering System (where numbers, but not necessarily names of authors, are cited in the text, and the reference list is in order of citation in the text). Several ways of quoting references in the list are in current use.

The Commission is of the opinion, arrived at as a result of much consultation between many senior editors, that it is unlikely that all journals would accept a recommendation to use either the Harvard or the Numbering System to the exclusion of the other. It believes, however, that most biochemists will accept the need for, and indeed welcome, a substantial degree of unification of practices, there being no strong case for the individuality of each journal on this issue. Accordingly, the Commission makes the following Recommendations to all biochemical journals; the reasons for some of them are given. The Recommendations deal first with the way in which references should be cited in the list; the proposal is suitable for journals adopting either the Harvard or the Numbering System. Secondly, there are Recommendations about the way in which each of these systems is used. Thirdly, abbreviations for titles of journals and a few other points are considered. Implementation of the Recommendations would mean that any very small differences between journals in their practices would be of the type that can be attended to at the redactory stage of preparation for press. The Commission recognizes that it cannot deal with a number of smaller problems concerning citations that arise from time to time.

RECOMMENDATIONS

1. Citations of References in the List of References Should Be as Follows

Braun, A., Brown, B. & LeBrun, C. (1971) *Journal*, 11, 111–113.

Notes: (a) This form can be used by both systems.

(b) Journals using the Numbering System should arrange the references in numerical order beside the number (which can be italicized or in brackets according to the house custom of the journal).

(c) Journals using the Harvard System should arrange the references in alphabetical order, whatever the language, except in certain situations (see Recommendation 4a below).

(d) This recommendation incorporates the following points:

i. Initials after surnames (full first names are not given in the list).

ii. The use of the symbol "&" is recommended if at all possible because of its widespread usage and the fact that it is independent of the language. No comma before "&."

*From IUB Commission of Editors of Biochemical Journals (CEBJ), *J. Biol. Chem.*, 248(21), 7279–7280 (1973). With permission.

iii. Year in parentheses (this follows immediately after the authors' names because it is essential to the Harvard System).

iv. Journal title (abbreviated). This can be in italics according to house practice (see Recommendation 7 below concerning journal title abbreviations).

v. Volume number. This can be in heavy type or italics according to house practice.

vi. A few journals do not have volume numbers in which case the page numbers should follow immediately after the abbreviated journal title.

If it is necessary to quote both a volume and a part number, the reference should read: Brown, B. (1971) *Journal,* 11, pt 1, 121–123.

vii. First and last pages should be given. The Commission decided to make this Recommendation mainly on the basis of evidence that the additional information provided by quoting the last page was being required increasingly in many types of library and information retrieval services. Citation of the last page (as well as the first) has been requested for some time by the secondary and abstracting journals. Citation of both first and last pages is also an aid in the prevention of errors.

viii. The number of stops and commas is kept as small as possible.

(e) Authors' names and the abbreviated name of the journal when repeated in the next reference should be spelled out in full; ibid. and similar terms should not be used.

(f) Recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature (CBN) and similar documents should be referred to as: Commission on Biochemical Nomenclature (1970) followed by a journal reference.

(g) Junior should be abbreviated to "Jr," not "jun."

2. Numbering System in the Text

The use of authors' names is permissible as authors wish; only the initial letter of the name should be in capital type. Numbers can be inserted in parentheses or as superscripts according to house custom. The printing of references at the foot of the page on which they are first quoted is considered to be helpful with the Numbering System but is not part of the Recommendation because the extra cost is generally considered to be prohibitive.

3.Harvard System in the Text

For multi-author papers, it is recommended that:

a. Not more than two authors to be named either on the first or any subsequent occasion;

b. et al. should be used for three or more authors on every occasion;

c. Each name to have the initial letter in capital type only.

Examples (Harvard System style):

Braun et al. (1969) did some work that was confirmed by LeBrun (1970).

These results (Braun et al., 1969; LeBrun, 1970) have been discussed by Brown & Braun (1971).

The same Recommendation (without the year) applies when authors are quoted in the text in the Numbering System.

4. Harvard System in the List of References

a. A special problem arises in the list when there are several papers by, e.g., Green et al. in the same or over several years. While the list could be in strict alphabetical order of the full reference, the reader will find no clue in the text to the alphabetical status of the names of the second and subsequent authors (see Recommendations 3a and 3b). It is therefore recommended that all the papers by Green et al. (that is by Green and more than one co-author) should be arranged, irrespective of the names of the other

authors, in chronological order (over many years if necessary) and designate tham a, b, c, etc.

Examples:

Green, G. (1970) etc.
Green, G. & Brown, B. (1971) etc.
Green, G. & White, W. (1969) etc.
Green, G., White, W. & Black, B. (1968a) etc.
Green, G., Brown, B. & Black, B. (1968b) etc.
Green, G., White, W., Black, B. & Brown, B. (1969) etc.
Green, G., Black, B. & Brown, B. (1970) etc.

sequence governed by order or date of publication, as far as can be ascertained.

b. Names beginning with "Mc" should be listed under "Mc" and not under "Mac," to decide alphabetical order.

c. Names beginning with "De," "Van," or "von," etc. should be arranged under D or V/v, etc.

5. Reference to Books

These should appear in text like any reference to a journal paper. The reference in the list should read: Brown, B. & Braun, A. (1971) in *Book Title* (LeBrun, C., ed.), pp. 1–20, Publisher, Town.

Notes:

a. If a volume number has to be quoted, this would appear before the pp. as, e.g., "vol. 2," with the number in Arabic numerals (even when Roman numerals are printed on the cover of the book).

b. Where an author wishes to refer to a specific page within a book reference, this should be given in the text.

Example (in text): ". . . discussed on p. 21 of Braun et al.(1971)."

6. Other Forms of References

a. *In the press.* It is recommended that (i) this should mean that the paper has been finally accepted by a journal, (ii) it is quoted in the text (both systems) just as any other paper, (iii) the year quoted should be the best estimate revised if necessary at proof stage, and (iv) the full citation in the list to read: Braun, A. & Brown, B. (1971) *Journal,* in the press.

b. *Submitted for publication* should be used in a typescript only when it is reasonable to expect that it will be possible to alter the quotation to a final form at a stage before publication; if such alteration cannot be made then the name of the journal involved should be stated.

c. The use of *in preparation* and *private communication* should not be allowed because they have no real value.

d. *Personal communication* and *unpublished work* should be permitted in the text only, i.e., not in the list of references. Editors may require to see written evidence of the former.

7. Abbreviations for Journal Titles

Most biochemical journals use the *Chemical Abstract** system but a few use the World List, 4th Edition. The Commission noted that the latest information available (International List of Periodical Title Word Abbreviations prepared for the UNISIST/ ICSU-AB Working Group on Bibliographical Descriptions) suggests that the abbreviations that will be recommended finally by ICSU will be very similar to those now used by *Chemical Abstracts.*

Believing that complete uniformity on this issue is highly desirable now and estimating that it may be a few more years before ICSU finally reports, the Commission recommends that all biochemical journals should now use the *Chemical Abstracts* (American Chemical Society) system. The Commission believes that any changes that will be required when ICSU eventually issues recommendations on this point will be comparatively minor ones.

8. Implementation of these Recommendations

The Commission at its meeting in Menton, May 7 to 8, 1971, has taken the view that the degree of uniformity envisaged in the Recommendations is highly desirable and therefore further recommends to all biochemical journals that the changes required should be made as soon as possible. The Commission recognizes that all journals will have to make some changes (in most cases these are minor) from their present established practices to implement these Recommendations in full. It considers that the possible objections of difficulties even for a commercial publisher with an established "house style" are outweighed by the advantage that conformity of style in the citation of references will prove to the authors, editors, and readers upon whom all journals depend for their existence.

*The journal-title abbreviations in *Biological Abstracts* are essentially the same in *Chemical Abstracts.* A *List of Serials with Title Abbreviations* is available from BioSciences Information Service of Biological Abstracts, 2100 Arch Street, Philadelphia, PA 19103.

IUPAC TENTATIVE RULES FOR THE NOMENCLATURE OF ORGANIC CHEMISTRY SECTION E. FUNDAMENTAL STEREOCHEMISTRY*

International Union of Pure and Applied Chemistry

INTRODUCTION

This Section of the IUPAC Rules for Nomenclature of Organic Chemistry differs from previous Sections in that it is here necessary to legislate for words that describe concepts as well as for names of compounds.

At the present time, concepts in stereochemistry (that is, chemistry in three-dimensional space) are in the process of rapid expansion, not merely in organic chemistry, but also in biochemistry, inorganic chemistry, and macromolecular chemistry. The aspects of interest for one area of chemistry often differ from those for another, even in respect to the same phenomenon. This rapid evolution and the variety of interests have led to development of specialized vocabularies and definitions that sometimes differ from one group of specialists to another, sometimes even within one area of chemistry.

The Commission on the Nomenclature of Organic Chemistry does not, however, consider it practical to cover all aspects of stereochemistry in this Section E. Instead, it has two objects in view: To prescribe, for basic concepts, terms that may provide a common language in all areas of stereochemistry; and to define the ways in which these terms may, so far as necessary, be incorporated into the names of individual compounds. The Commission recognizes that specialized nomenclatures are required for local fields; in some cases, such as carbohydrates, amino acids, peptides and proteins, and steroids, international rules already exist; for other fields, study is in progress by specialists in Commissions or Subcommittees; and further problems doubtless await identification. The Commission believes that consultations will be needed in many cases between different groups within IUPAC and IUB if the needs of the specialists are to be met without confusion and contradiction between the various groups.

The Rules in this Section deal only with Fundamental Stereochemistry, that is, the main principles. Many of these Rules do little more than codify existing practice, often of long standing; however, others extend old principles to wider fields, and yet others deal with nomenclature that is still subject to controversy.

Rule E-0

The stereochemistry of a compound is denoted by an affix or affixes to the name that does not prescribe the stereochemistry; such affixes, being additional, do not change the name or the numbering of the compound. Thus, enantiomers, diastereoisomers, and *cis–trans* isomers receive names that are distinguished only by means of different stereochemical affixes. The only exceptions are those trivial names that have stereochemical implications (for example, fumaric acid, cholesterol).

Note: In some cases (see Rules E-2.23 and E-3.1) stereochemical relations may be used to decide between alternative numberings that are otherwise permissible.

E-1. Types of Isomerism

E-1.1. The following nonstereochemical terms are relevant to the stereochemical nomenclature given in the Rules that follow.

*From *IUPAC Inf. Bull. Append. Tentative Nomencl. Sym. Units Stand.*, No. 35, August 1974, pp. 36–80. With permission.

(a) The term structure may be used in connection with any aspect of the organization of matter.

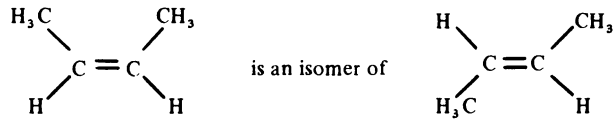
Hence: structural (adjectival)

(b) Compounds that have identical molecular formulas but differ in the nature or sequence of bonding of their atoms or in arrangement of their atoms in space are termed isomers.

Hence: isomeric (adjectival)
isomerism (phenomenological)

Examples:

$H_3C-O-CH_3$ is an isomer of H_3C-CH_2-OH



(In this and other Rules a broken line denotes a bond projecting behind the plane of the paper, and a thickened line denotes a bond projecting in front of the plane of the paper. In such cases a line of normal thickness denotes a bond lying in the plane of the paper.)

(c) The constitution of a compound of given molecular formula defines the nature and sequence of bonding of the atoms. Isomers differing in constitution are termed constitutional isomers.

Hence: constitutionally isomeric (adjectival)
constitutional isomerism (phenomenological)

Example:

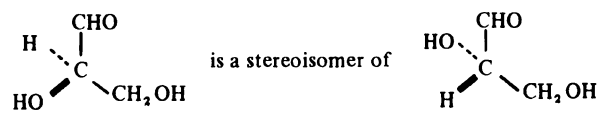
$H_3C-O-CH_3$ is a constitutional isomer of H_3C-CH_2-OH.

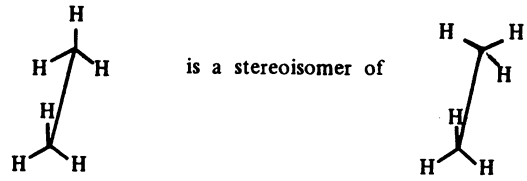
Note: Use of the term "structural" with the above connotation is abandoned as insufficiently specific.

E-1.2. Isomers are termed stereoisomers when they differ only in the arrangement of their atoms in space.

Hence: stereoisomeric (adjectival)
stereoisomerism (phenomenological)

Examples:

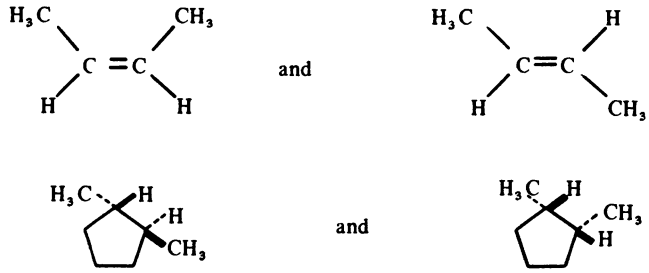




E-1.3. Stereoisomers are termed *cis–trans* isomers when they differ only in the positions of atoms relative to a specified plane in cases where these atoms are, or are considered as if they were, parts of a rigid structure.

Hence: *cis–trans* isomeric (adjectival)
cis–trans isomerism (phenomenological)

Examples:



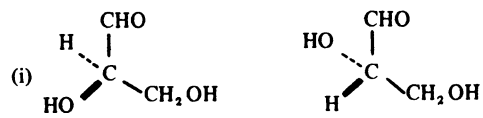
E-1.4. Various views are current regarding the precise definition of the term "configuration." (a) Classical interpretation: The configuration of a molecule of defined constitution is the arrangement of its atoms in space without regard to arrangements that differ only as after rotation about one or more single bonds. (b) This definition is now usually limited so that no regard is paid also to rotation about π bonds or bonds of partial order between one and two. (c) A third view limits the definition further so that no regard is paid to rotation about bonds of any order, including double bonds.

Molecules differing in configuration are termed configurational isomers.

Hence: configurational isomerism

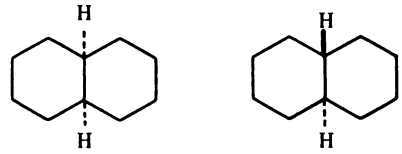
Notes: (1) Contrast conformation (Rule E-1.5). (2) The phrase "differ only as after rotation" is intended to make the definition independent of any difficulty of rotation, in particular independent of steric hindrance to rotation. (3) For a brief discussion of views (a) to (c), see Appendix 1. It is hoped that a definite consensus of opinion will be established before these Rules are made "Definitive."

Examples: The following pairs of compounds differ in configuration:

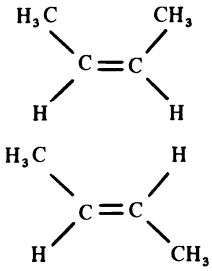


(ii)

(iii)



(iv)



These isomers (iv) are configurational in view (a) or (b) but are conformational (see Rule E-1.5) in view (c)

E-1.5. Various views are current regarding the precise definition of the term "conformation." (a) Classical interpretation: The conformations of a molecule of defined configuration are the various arrangements of its atoms in space that differ only as after rotation about single bonds. (b) This is usually now extended to include rotation about π bonds or bonds of partial order between one and two. (c) A third view extends the definition further to include also, rotation about bonds of any order, including double bonds.

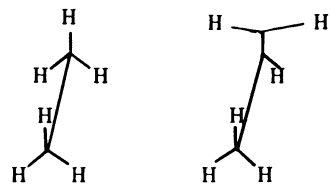
Molecules differing in conformation are termed conformational isomers.

Hence: conformational isomerism

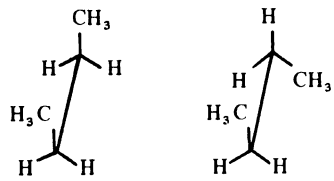
Notes: All the Notes to Rule E-1.4 apply also to E-1.5.

Examples: Each of the following pairs of formulas represents a compound in the same configuration but in different conformations.

(a, b, c)

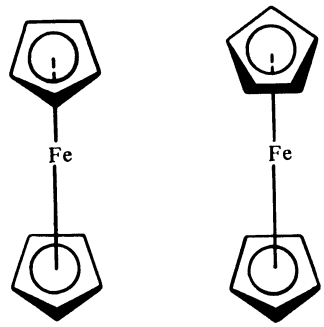


(a, b, c)



(a, b, c)

(b, c)



(c) See Example (iv) to Rule E-1.4.

E-1.6. The terms relative stereochemistry and relative configuration are used with reference to the positions of various atoms in a compound relative to one another, especially, but not only, when the actual positions in space (absolute configuration) are unknown.

E-1.7. The terms absolute stereochemistry and absolute configuration are used with reference to the known actual positions of the atoms of a molecule in space.*

E-2. *cis–trans* Isomerism†

Preamble. The prefixes *cis* and *trans* have long been used for describing the relative positions of atoms or groups attached to nonterminal doubly bonded atoms of a chain or attached to a ring that is considered as planar. This practice has been codified for hydrocarbons by IUPAC.** There has, however, not been agreement on how to assign *cis* or *trans* at terminal double bonds of chains or at double bonds joining a chain to a ring. An obvious solution was to use *cis* and *trans* where doubly bonded atoms formed the backbone and were nonterminal and to enlist the sequence-rule preferences to decide other cases; however, since the two methods, when generally applied, do not always produce analogous results, it would then be necessary to use different symbols for the two procedures. A study of this combination showed that both types of symbols would often be required in one name and, moreover, it seemed wrong in principle to use two symbolisms for essentially the same phenomenon. Thus it seemed to the Commission wise to use only the sequence-rule system, since this alone was applicable to all cases. The same decision was taken independently by Chemical Abstracts Service who introduced *Z* and *E* to correspond more conveniently to *seqcis* and *seqtrans* of the sequence rule.

It is recommended in the Rules below that these designations *Z* and *E* based on the sequences rule shall be used in names of compounds, but *Z* and *E* do not always correspond to the classical *cis* and *trans* which show the steric relations of like or similar

*Determination of absolute configuration became possible through work by Bijvoet, J. M., Peerdeman, A. F., and van Bommel, A. J., *Nature*, 168, 271 (1951); cf. Bijvoet, J. M., *Proc. Kon. Ned. Akad. Wetensch.*, 52, 313 (1949).

†These Rules supersede the Tentative Rules for olefinic hydrocarbons published in the Comptes rendus of the 16th IUPAC Conference, New York, N.Y., 1951, pp. 102–103.

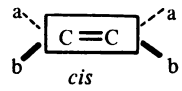
**Blackwood, J. E., Gladys, C. L., Loening, K. L., Petrarca, A. E., and Rush, J. E., *J. Amer. Chem. Soc.*, 90, 509 (1968); Blackwood, J. E., Gladys, C. L., Petrarca, A. E., Powell, W. H., and Rush, J. E., *J. Chem. Doc.*, 8, 30 (1968).

groups that are often the main point of interest. So the use of *Z* and *E* in names is not intended to hamper the use of *cis* and *trans* in discussions of steric relations of a generic type or of groups of particular interest in a specified case (see Rule E-2.1 and its Examples and Notes, also Rule E-5.11).

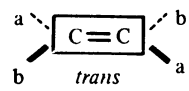
It is also not necessary to replace *cis* and *trans* for describing the stereochemistry of substituted monocycles (see Subsection E-3). For cyclic compounds the main problems are usually different from those around double bonds; for instance, steric relations of substitutents on rings can often be described either in terms of chirality (see Subsection E-5) or in terms of *cis–trans* relationships, and, further, there is usually no single relevant plane of reference in a hydrogenated polycycle. These matters are discussed in the Preambles to Subsections E-3 and E-4.

E-2.1. *Definition of cis–trans.* Atoms or groups are termed *cis* or *trans* to one another when they lie respectively on the same or on opposite sides of a reference plane identifiable as common among stereoisomers. The compounds in which such relations occur are termed *cis–trans* isomers. For compounds containing only doubly bonded atoms, the reference plane contains the doubly bonded atoms and is perpendicular to the plane containing these atoms and those directly attached to them. For cyclic compounds, the reference plane is that in which the ring skeleton lies or to which it approximates. When qualifying another word or a locant, *cis* or *trans* is followed by a hyphen. When added to a structural formula, *cis* may be abbreviated to *c*, and *trans* to *t* (see also Rule E-3.3).

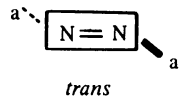
Examples: (Rectangles here denote the reference planes and are considered to lie in the plane of the paper.)



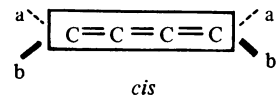
cis



trans

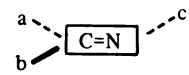


trans

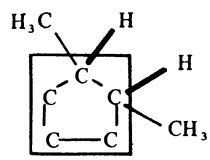


cis

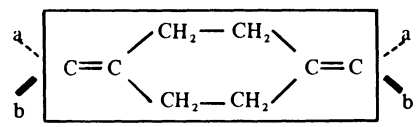
The groups or atoms a,a are the pair selected for designation but are not necessarily identical; b,b are also not necessarily identical but must be different from a,a.



cis or *trans* according as a or b is taken as basis of comparison

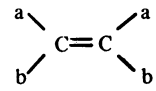


cis



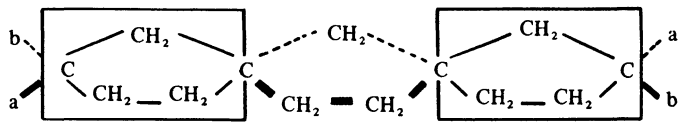
trans

Notes: The formulas above are drawn with the reference plane in the plane of the paper, but for doubly bonded compounds it is customary to draw the formulas so that this plane is perpendicular to that of the paper; atoms attached directly to the doubly bonded atoms then lie in the plane of the paper and the formulas appear as, for instance

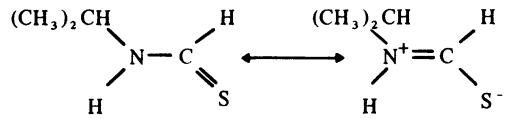


cis

Cyclic structures, however, are customarily drawn with the ring atoms in the plane of the paper, as above. However, care is needed for complex cases, such as



The central five-membered ring lies (approximately) in a plane perpendicular to the plane of the paper. The two a groups are *trans* to one another; so are the b groups; the outer cyclopentane rings are *cis* to one another with respect to the plane of the central ring. *cis* or *trans* (or *Z* or *E*; see Rule E-2.21) may also be used in cases involving a partial bond order when a limiting structure is of sufficient importance to impose rigidity around the bond of partial order. An example is

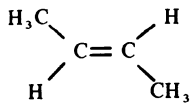


trans (or E)

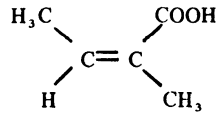
E-2.2. *cis-trans Isomerism around Double Bonds.*

E-2.21. In names of compounds steric relations around one or more double bonds are designated by affixes *Z* and/or *E*, assigned as follows. The sequence-rule-preferred* atom or group attached to one of a doubly bonded pair of atoms is compared with the sequence-rule-preferred atom or group attached to the other of that doubly bonded pair of atoms; if the selected pair are on the same side of the reference plane (see Rule 2.1) an italic capital letter *Z* prefix is used; if the selected pair are on opposite sides an italic capital letter *E* prefix is used.† These prefixes, placed in parentheses and followed by a hyphen, normally precede the whole name; if the molecule contains several double bonds, then each prefix is immediately preceded by the lower or less primed locant of the relevant double bond.

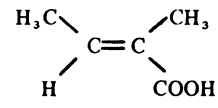
Examples:



(*E*)-2-Butene



(*Z*)-2-Methyl-2-butenoic acid** *or* (*Z*)-2-methylisocrotonic acid (see Exceptions below)



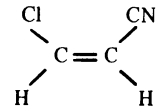
(*E*)-2-Methyl-2-butenoic acid†† *or* (*E*)-2-Methylcrotonic acid (see Exceptions below)

*For sequence-rule preferences see Appendix 2.

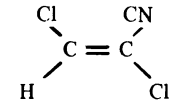
†These prefixes may be rationalized as from the German *zusammen* (together) and *entgegen* (opposite).

**The name angelic acid is abandoned because it has been associated with the designation *trans* with reference to the methyl groups.

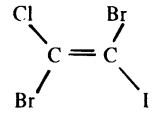
††The name tiglic acid is abandoned because it has been associated with the designation *cis* with reference to the methyl groups.



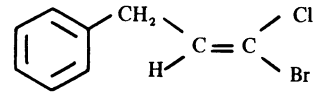
(*Z*)-3-Chloroacrylonitrile



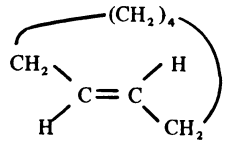
(*E*)-2,3-Dichloroacrylonitrile



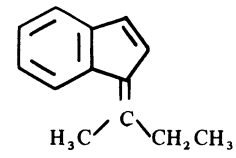
(*Z*)-1,2-Dibromo-1-chloro-2-iodoethylene
(By the sequence rule, Br is preferred to Cl, but I to Br)



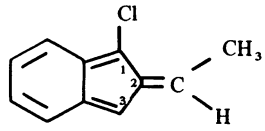
(*E*)-(3-Bromo-3-chloroallyl)benzene



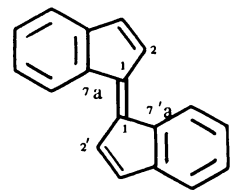
(*E*)-Cyclooctene



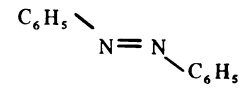
(*E*)-1-*sec*-Butylideneindene



(*Z*)-1-Chloro-2-ethylidene-2*H*-indene

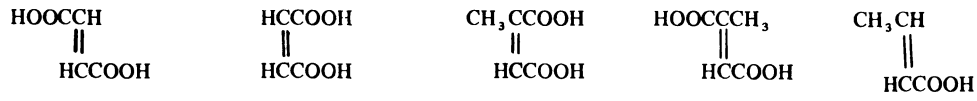


(*E*)-1,1′-Biindenylidene

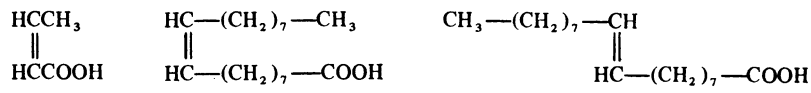


(*E*)-Azobenzene

Exceptions to Rule E-2.21. The following are examples of accepted trivial names in which the stereochemistry is prescribed by the name and is not cited by a prefix.



Fumaric acid | Maleic acid | Citraconic acid* | Mesaconic acid* | Crotonic acid



Isocrotonic acid | Oleic acid | Elaidic acid

E-2.22 (*Alternative to Part of E-2.21*). (a) When more than one series of locants starting from unity is required to designate the double bonds in a molecule, or when the name consists of two words, the *Z* and *E* prefixes together with their appropriate locants may be placed before that part of the name where ambiguity is most effectively removed.

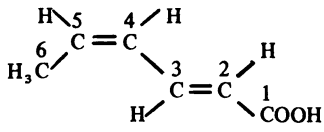
(b) [Alternative to (a)] When several *Z* or *E* prefixes are required they are arranged in

*Systematic names are recommended for derivatives of these compounds formed by substitution on carbon.

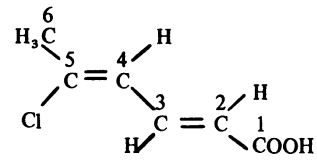
order as follows: Of the four atoms or groups attached to each doubly bonded pair of atoms, that one preferred by the sequence rule is selected; the single atoms or groups thus selected are then arranged in their sequence rule order (determined in respect of their position in the whole molecule), and the prefixes *Z* and/or *E* for the respective double bonds are placed in that order, but *without* their locants.

Note: In method (a) the final choice is left to an author or editor because of the variety of cases met and because the problems are not always the same in different languages. The presence of the locants usually eases translation from the name to a formula, but this method (a) may involve the logical difficulty explained for the third example below. Method (b) always gives a single unambiguous order and is not subject to the logical difficulty just mentioned, but translation from the name to the formula is harder than for method (a). Method (a) may be more suitable for cursive text, and method (b) for compendia. If method (b) is used it should be used whenever more than one double bond is involved, but method (a) is to be used only under the special conditions detailed in the rule.

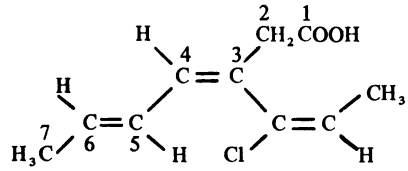
Examples:



(a) (2*E*,4*Z*)-2,4-Hexadienoic acid
(b) (*E*,*Z*)-2,4-Hexadienoic acid

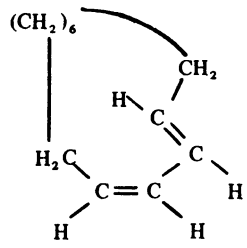


(a) (2*E*,4*Z*)-5-Chloro-2,4-hexadienoic acid
(b) (*Z*,*E*)5-Chloro-2,4-hexadienoic acid



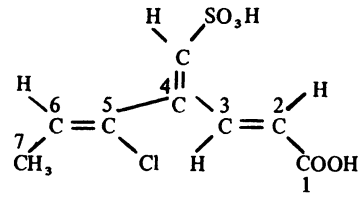
(a) 3-[(*E*)-1-Chloropropenyl]-(3*Z*,5*E*)-3,5-heptadienoic acid
(b) (*E*,*Z*,*E*)-3-(1-Chloropropenyl)-3,5-heptadienoic acid

[The last example shows the disadvantages of both methods. In method (a) there is a fault of logic, namely, the 3*Z*,5*E* are not the property of the unsubstituted heptadienoic acid chain, but the 3*Z* arises only because of the side chain that is cited before the 3*Z*,5*E*. In method (b) it is some trouble to assign the *E*,*Z*,*E* to the correct double bonds.]



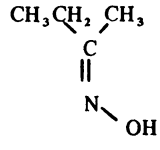
(a) (1*Z*,3*E*)-1,3-Cyclododecadiene
(b) (*Z*,*E*)-1,3-Cyclododecadiene

[The lower locant is assigned to the *Z* double bond.]

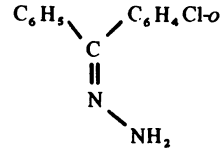


(a) 5-Chloro-4-(*E*-sulfomethylene)-(2*E*,5*Z*)-2,5-heptadienoic acid
(b) (*Z*,*E*,*E*)-5-Chloro-4-(sulfomethylene)-2,5-heptadienoic acid

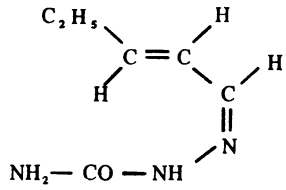
[In application of the sequence rule, the relation of the SO_3H to CCl (rather than to C-3), and of the CH_3 to Cl, are decisive.]



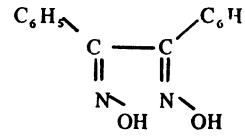
(a) Butanone (*E*)-oxime*
(b) (*E*)-Butanone oxime



(a) 2-Chlorobenzophenone (*Z*)-hydrazone
(b) (*Z*)-2-Chlorobenzophenone hydrazone



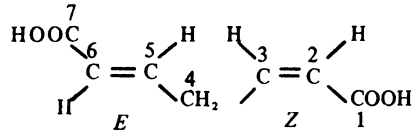
(a) (*E*)-2-Péntenal (*Z*)-semicarbazone
(b) (*Z*,*E*)-2-Pentenal semicarbazone



(a) Benzil (*Z*,*E*)-dioxime
(b) (*Z*,*E*)-Benzil dioxime

E-2.23. When Rule C-13.1 or E-2.22(b) permits alternatives, preference for lower locants and for inclusion in the principal chain is allotted as follows, in the order stated, so far as necessary: *Z* over *E* groups; *cis* over *trans* cyclic groups; *R* over *S* groups (also *r* over *s*, etc., as in the sequence rule); if the nature of these groups is not decisive, then the lower locant for such a preferred group at the first point of difference.

Examples:



(a) (2*Z*,5*E*)-2,5-Heptadienedioic acid
(b) (*E*,*Z*)-2,5-Heptadienedioic acid

[The lower numbers are assigned to the *Z* double bond.]

*The terms *syn*, *anti*, and *amphi* are abandoned for such compounds.