Dictionary of Organic Compounds

SIXTH EDITION

VOLUME ONE

A–Bromebric acid

Dictionary of Organic Compounds

SIXTH EDITION

VOLUME ONE

A–Bromebric acid

A-0-00001–B-0-03701

CHAPMAN & HALL
Electronic Publishing Division
London · Glasgow · Weinheim · New York · Tokyo · Melbourne · Madras

Published by Chapman & Hall, 2–6 Boundary Row, London SE1 8HN, UK

Chapman & Hall, 2–6 Boundary Row, London SE1 8HN, UK

Blackie Academic & Professional, Wester Cleddens Road, Bishopbriggs, Glasgow G64 2NZ, UK

Chapman & Hall, Pappelallee 3, 69469 Weinheim, Germany

Chapman & Hall USA, 115 Fifth Avenue, New York, NY 10003, USA

Chapman & Hall Japan, ITP-Japan, Kyowa Building, 3F, 2-2-1 Hirakawacho, Chiyoda-ku, Tokyo 102, Japan

Chapman & Hall Australia, 102 Dodds Street, South Melbourne, Victoria 3205, Australia

Chapman & Hall India, R. Seshadri, 32 Second Main Road, CIT East, Madras 600 035, India

First edition in three volumes published by
Eyre & Spottiswoode (Publishers) Ltd:
Volume I published 1934
Volume II published 1936
Volume III published 1937
Second edition of Volume I published 1943
Volumes II and III reprinted with supplement 1944
Volumes I, II and III reprinted 1947
Third edition in four volumes published 1953
Fourth edition in five volumes published 1965
Fifth edition in seven volumes published 1982 by Chapman and Hall

This Sixth edition in nine volumes published 1996

Typeset and printed in Great Britain at the University Press, Cambridge

ISBN 0 412 54090 8 (Nine volume set)

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

Library of Congress Catalog Card Number: 94-69994

♾ Printed on acid-free text paper, manufactured in accordance with ANSI/NISO Z39.48-1992 (Permanence of Paper).

Dictionary of Organic Compounds

SIXTH EDITION

VOLUME ONE

A–Bromebric acid

A-0-00001–B-0-03701

Note to Readers

Annual supplements keep *Dictionary of Organic Compounds* up-to-date

Supplements to *Dictionary of Organic Compounds* will be published each year subsequent to publication of the Main Work. The first annual supplement will be published in Autumn 1996.

Each volume will contain about 2,000 new and revised entries derived from the primary literature of the preceding year, and will provide a complete updating service for all organic chemistry. Cumulative indexes will appear in supplements as space permits.

***Dictionary of Organic Compounds* on CD-ROM**

A CD-ROM version is also available, which is searchable by text or by substructure. For more details of either the CD-ROM or printed version, please write to:

Product Manager
Electronic Publishing Division
Chapman & Hall
2–6 Boundary Row
London SE1 8HN

Product Manager
Chapman & Hall USA
115 5th Avenue, 4th Floor
New York
NY 10003

New Compounds for *Dictionary of Organic Compounds*

The Editors are always pleased to receive comments on the Dictionary, and in particular to receive specific suggestions for compounds, or groups of compounds, for inclusion.

Write to:
The Editors, *Dictionary of Organic Compounds*
Chapman & Hall
2–6 Boundary Row
London SE1 8HN

Caution

Treat all chemical substances as if they have dangerous properties.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this Dictionary, and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

The specific information in this publication on the hazardous and toxic properties of certain substances is included to alert the reader to possible dangers associated with the use of those compounds. The absence of such information should not however be taken as an indication of safety in use or misuse.

Preface

We are delighted to present this new, updated and considerably enlarged Edition of the *Dictionary of Organic Compounds*, the latest in its 60-year history.

The previous Edition, DOC5, published in 1982, was the first to benefit from computerised methods of compilation and printing which laid the foundation for an authoritative databank that could be continually updated and revised. During the succeeding ten years, annual supplements appeared reporting all that was best from the preceding year's literature together with the results of retrospective searches on many more useful starting materials and intermediates selected for inclusion by the Editors. In the three-year period since the tenth supplement to DOC5, the whole dataset has been very considerably further expanded and thoroughly checked and revised. The result of this intensive effort is that DOC6 represents and extremely well-organised and easy-to-use reference source containing far more data than any previous edition.

The aims of DOC remain as they have always been, to allow the user (whether chemist or non-chemist) to access the key data on the most important organic compounds as rapidly and easily as possible. The majority of queries are thus answered accurately and quickly but for those needing more detailed information about a particular compound, or group of compounds, DOC provides the key information necessary to delve deeper into the literature, for example by reading a review cited in DOC6 or by carrying out an online search.

The coverage of DOC has been sharpened by the decision in the late 1980s to publish the separate and complementary *Dictionary of Natural Products* (DNP), which provides comprehensive coverage of the very active research field of natural product chemistry. DOC now contains entries for the most significant and typical natural products, sometimes abbreviated compared with DNP so as to allow space for greatly expanded coverage of core DOC compounds. This sharpening of focus has been combined with a space-saving three-column layout which has enabled the overall increase in the number of pages to be kept to approximately 15% over DOC5.

The worlds of chemistry and of information science have both changed dramatically in the period since the last Edition. Many comments received from users of DOC during this period have assured us that its role in rationalising and organising the vast chemical literature is welcomed now more than ever.

J.I.G.C., S.V.L., G.P., R.A.R., C.W.R.

Introduction

For a fuller treatment of many of the topics covered by this Introduction, see the companion volume *The Organic Chemist's Desk Reference.*

1. Using DOC 6

As with previous Editions the arrangement of Entries is alphabetical by DOC Name. Thus for many users the *Dictionary* acts as its own index. The criteria governing the choice of DOC Name are described below. Because most organic compounds have at least two alternative names, DOC 6 also contains a Name Index (Volume 7) which includes all DOC Names, synonyms and alternative names given throughout the *Dictionary*, including those applicable to derivatives. Volumes 8 and 9 contain the Molecular Formula and Chemical Abstracts Service (CAS) Registry Number indexes respectively.

Every Entry is numbered to assist ready location. The DOC Number consists of a letter of the alphabet followed by a six-digit number. The first digit (separated from the rest by a hyphen) identifies the supplement number and in these Main Work volumes is invariably 0. Entries in the First Supplement will carry the first digit 1, and so on. All index entries refer to the DOC Number. Each Index is described in detail at the beginning of the appropriate volume.

2. Compound selection and presentation

2.1 *General*

In general, the following compounds are included:

(*a*) The basic fundamental organic compounds of simple structure.

(*b*) Compounds of industrial or commercial value including currently important pharmaceuticals, pesticides, monomers, etc.

(*c*) Compounds frequently encountered in the laboratory as solvents, reagents or starting materials.

(*d*) A limited selection of the most important and well-documented natural products. For a comprehensive coverage of *all* natural products, see the companion *Dictionary of Natural Products* (published 1994). Some of the natural product entries in DOC 6 have been abbreviated by the omission of data referring to minor congeners and their literature. If this is the case the entry in DOC 6 is headed by the message 'For a fuller version of this Entry see *Dictionary of Natural Products*'.

(*e*) Important biochemicals.

(*f*) Other compounds of particular interest because of their chemical, structural or biological properties, including many newly synthesised compounds of active research interest.

The Editors believe that the great majority of users will find that this coverage corresponds to their needs. We are always pleased to receive comments on the selection policy, and in particular to receive specific suggestions for compounds for inclusion.

A few entries cover groups of homologous compounds having similar properties and uses, e.g. there is a single entry for 4-alkyl-4′-cyanobiphenyls (A-0-00763). All of the compounds covered by such a composite entry have their names and molecular formulae given and can be readily traced using the indexes.

2.2 *Organometalloids*

DOC 6 covers the principal organic compounds of Se, Te, As, Sb and Bi. There is also selective coverage of organo-B and organo-Si compounds, especially those of importance in synthetic organic chemistry. For a fuller coverage of the organometallic compounds of B, Si, As, Sb and Bi, see the companion *Dictionary of Organometallic Compounds*, which also gives a full coverage of organometallic compounds of the true metals, no longer included in DOC.

2.3 *Derivatives*

Derivatives of compounds containing the common functional groups are normally given under the parent compound, except when the derivative has a relatively extensive literature of its own or where the derivative requires special stereochemical description (e.g. some simple chiral sulfoxides have individual entries and therefore are not shown under the parent sulfide). In such cases the derivative entry is cross-referenced from the parent compound.

In response to user requests, derivative descriptors have been made clearer in this Edition, e.g. B, 2HCl (DOC 5) → Hydrochloride (1:2) (DOC 6).

The most common functional groups and their commonest derivatives are listed below. The order in which the derivatives are listed here is the order in which they appear in DOC 6 entries under the parent compound:

Carbonyl compounds, RR′C=O

Oxime, RR′C=NOH
Hydrazone, RR′C=NNH_2
Phenylhydrazone, RR′C=NNHPh
2,4-Dinitrophenylhydrazone, RR′C=NNH[$C_6H_3(NO_2)_2$]
Semicarbazone, RR′C=$NNHCONH_2$
Di-Me acetal, RR′$C(OMe)_2$
Di-Et acetal, RR′$C(OEt)_2$
Ethylene acetal

Carboxylic acids, RCOOH

Salts, e.g. Na salt, NH_4 salt
Esters, RCOOR′
Chloride (= acid chloride), RCOCl
Amide, $RCONH_2$
Alkylamides, RCONHR′
Anilide, RCONHPh
Dialkylamides, RCONR′R″
Nitrile, RCN
Anhydride, RCO–O–COR
Imide, RCO–NH–COR

Amines, RNH_2, RR′NH,RR′R″N

Salts, e.g. hydrochloride, methiodide
Picrates and other complexes
Ac (*N*-Acetyl) deriv., RNHAc (Ac = $-COCH_3$)
N-Benzoyl, RNHCOPh
N-Alkyl, RNHR′
N,N-Dialkyl, RNHR′R′
N,N,N-Trialkyl, RR′,R″,R‴$N^{\oplus}$
N-Oxide, RN(O)H

Alcohols, ROH

Ac (acetate), ROAc (Ac = $-COCH_3$)
Benzoyl, ROCOPh
Benzenesulfonyl and 4-methylbenzenesulfonyl (tosyl) $ROSO_2Ar$
Alkyl ethers ROR′

Thiols, RSH

S-Ac, RSAc
S-Benzoyl, RSCOPh
Benzenesulfonyl and 4-methylbenzenesulfonyl (tosyl) $RSSO_2Ar$
S-Alkyl, RSR′
Disulfide, R-S-S-R (sometimes included; usually a readily obtainable oxidative dimer of the thiol)

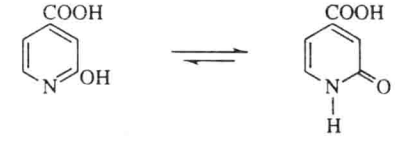
Molecular formulae are included in DOC for nearly all of these derivatives and they can therefore be readily traced using the molecular formula index, whether they are documented as derivatives or have their own individual entry. Molecular formulae are not in general given for salts, hydrates or complexes (e.g. picrates) nor for most "characterisation" derivatives of carbonyl compounds such as 2,4-dinitrophenylhydrazones and semicarbazones.

Where a derivative appears to have been characterised only as a salt, the properties of the salt may be given under the heading for the derivative. In such cases the data is clearly labelled, e.g. Mp 179° (as hydrochloride).

2.4 *Tautomerism in DOC 6*

Entries for tautomeric substances have been rearranged and improved for this Edition.

A completely consistent scheme for covering all such entries is not possible or desirable. Some variation is necessary in the way the appropriate DOC entries are organised in order to cover the various possibilities, but it is hoped and believed that the maximum possible clarity has been achieved by the changes made for DOC 6. The general principles which have been followed are described here using as an illustration one of the commonest types of tautomerism exhibited by simple organic compounds, which is heterocyclic NH $\rightleftharpoons$ OH prototropy as exemplified by 2- and 4-hydroxypyridines.



2-Hydroxy-4-pyridinecarboxylic acid *OH*-form

1,2-Dihydro-2-oxo-4-pyridinecarboxylic acid *NH*-form

(*a*) Although in most simple cases the *NH*-form is the predominant tautomer in solution, the equilibrium is influenced by electronic and steric factors as the structure of the heterocycle varies. DOC entries often give a statement about tautomerism for a particular compounds with supporting reference(s), but in many cases the individual compound will not have been studied closely and the probable tautomerism will have to be inferred.

(*b*) In the great majority of cases, the two (or more) tautomers and their derivatives are included in the same entry (this is a distinction from previous Editions).

(*c*) All synonyms applicable to the tautomeric forms are given at the head of the entry.

(*d*) The entry name may refer to an unfavoured tautomer, for ease of presentation of a series of entries, e.g. in the above case the entry name is 2-Hydroxy-4-pyridinecarboxylic acid. Note that in such a series of isomers some (i.e. those with a 3-OH substituent) will not be capable of NH $\rightleftharpoons$ OH tautomerism of the type shown (although they may tautomerise to zwitterionic tautomers).

(e) For important compounds such as that shown above, structures are shown (or implied) for both tautomers. For less important compounds, the probable predominant tautomer only may be illustrated but synonyms are still given for all possible reasonable tautomers.

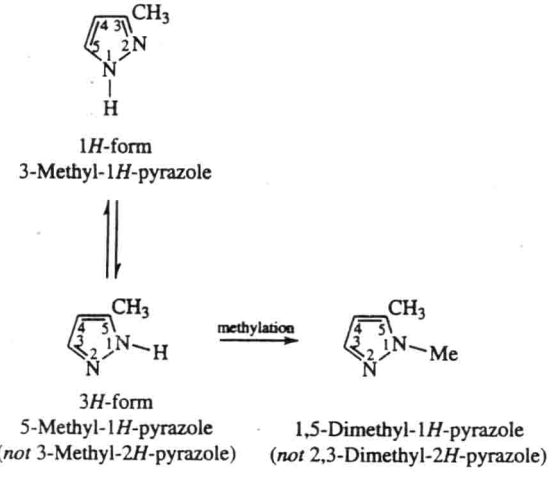
(*f*) Derivative data is given where appropriate under subheadings for the various tautomers. Derivatives may themselves be capable of tautomerism (e.g. Me ester, Methyl 2-hydroxy-4-pyridinecarboxylate), or may be blocked by substitution so that they clearly belong to one or other tautomer, e.g. Me ether (2-Methoxy-4-pyridinecarboxylic acid, derivative of the *OH*-form; *N*-Me, derivative of the *NH*-form). They appear in the appropriate place in the DOC entry.

(*g*) More complex examples where there are several possible tautomers not greatly differing in energy (e.g. purines, pteridines) are treated pragmatically to give the clearest possible presentation within the DOC entry structure. The situation is complicated by the fact that some derivatives may be partially blocked and capable of tautomerism to fewer tautomeric structures than the parent. Such situations are usually covered by notes within the entry.

(*h*) CAS frequently indexes compounds where the tautomerism is unclear under a default structure, frequently the unfavoured 1*H*-form. If this is the case, a note is given in the DOC entry.

Other very common types of tautomerism encountered include the P(O)SH $\rightleftharpoons$ P(S)OH interconversion shown by many organophosphorus compounds and the degenerate NH $\rightleftharpoons$ NH tautomerism of pyrazoles and imidazoles. The same general principles have been followed and should be clear from inspection of the individual DOC entries. For some types of organophosphorus tautomerism, e.g. phosphinic acids RPH(O)OH $\rightleftharpoons$ phosphonous acids $RP(OH)_2$, separate entries for the two substances have been retained.

Special care is often necessary in numbering and naming tautomeric compounds. Indicated hydrogen normally takes priority. For example;



1*H*-form
3-Methyl-1*H*-pyrazole

3*H*-form
5-Methyl-1*H*-pyrazole
(*not* 3-Methyl-2*H*-pyrazole)

1,5-Dimethyl-1*H*-pyrazole
(*not* 2,3-Dimethyl-2*H*-pyrazole)

2.5 *Isotopically labelled compounds*

DOC does not in general document isotopically labelled variants (except for a few isotopically-labelled drug variants used in chemotherapy or tracer work). Information on isotope variants can be obtained from Chemical Abstracts or from specialised sources.

2.6 *Anions and cations*

Ionic substances such as quaternary ammonium salts are presented under the name of the naked anion or cation and the molecular formula and molecular weight given are those of the ion. (See for example, the entry for 1-Acenaphthenyltriphenylphosphonium(1+), A-0-00060, $C_{30}H_{24}P^{\oplus}$). This agrees with CAS practice. The various salts of the cation are treated as derivatives, and their molecular formulae are given. Where a substance such as a dye is normally prepared and handled as, for example, a sodium salt, the entry usually refers to the parent acid.

3. Literature coverage

In compiling this Edition the primary literature has been surveyed to mid-1994. The first annual Supplement, which will appear in 1996, will survey the literature up to mid-1995.

Extensive further searching of the older literature has been carried out in preparing this Edition, for the purposes of increasing the coverage and for checking and expanding the data given in existing entries. Special attention was given to the entries for the commonest substances where users familiar with the Fifth Edition will find the addition of many additional carefully checked and labelled references.

4. Organisation of entries

In general the format of individual entries remains similar to that of previous Editions. Fig. 1 illustrates the format of a typical entry within which the individual types of data have been labelled. During the later editorial stages of DOC 6, every entry has been inspected to ensure the greatest possible uniformity of organisation of the information and nomenclature within and between entries. The range of information included within entries is described in detail below.

4.1 *Chemical names and synonyms*

4.1.1 *DOC Names*

The DOC Name is that chosen to head each entry and is that which, in the opinion of the Editors, is most likely to be known by, and of use to, most readers. Systematic DOC Names following IUPAC convention are used wherever convenient, but trivial names may be used for more complex structures such as pharmaceuticals and natural products. In cases where no one name stands out as being clearly more familiar or convenient than others, the Chemical Abstracts name is normally used as the entry name.

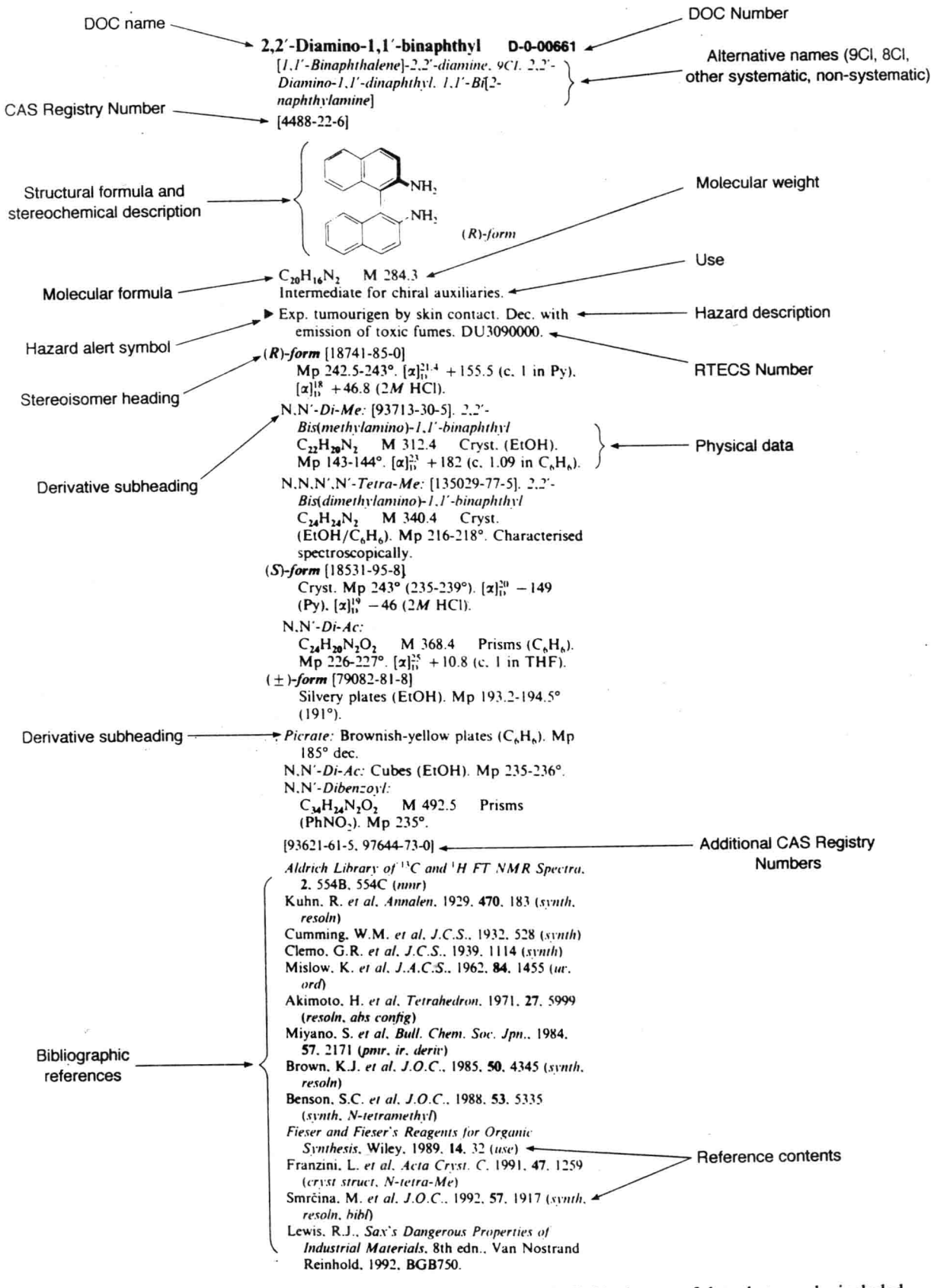


Figure 1 The format of a typical entry in DOC 6, showing the individual types of data that may be included.

A survey of DOC users, editors, and editorial board members was carried out in 1992. The results indicated almost no support for a complete change to current CAS nomenclature for DOC entry names, and a fair degree of variation between respondents as to which minor changes they would like to see. On the whole, respondents were happy with the current DOC nomenclature policy and it was decided therefore to make only a few minor changes in DOC 6 compared with DOC 5 in order to bring names closer to CAS in relatively noncontroversial areas. The following table includes some of the commonest variations between DOC 5, DOC 6 and current CAS (9CI) nomenclature

DOC 5 *(Entry names)*	***DOC 6*** *(Entry names)*	*CAS, 9–13CI* *(1973–)*
Aniline	**Aniline**	Benzenamine
1,4-Benzoquinone	**1,4-Benzoquinone**	2,5-Cyclohexadiene-1,4-dione
Naphthyl	**Naphthalenyl**	Naphthalenyl
Isopropyl	**Isopropyl**	1-Methylethyl
tert-Butyl	***tert*-Butyl**	1,1-Dimethylethyl
1,4-Naphthoquinone	**1,4-Naphthoquinone**	1,4-Naphthalenedione
Pyridyl	**Pyridinyl**	Pyridinyl
Quinuclidine	**Quinuclidine**	1-Azabicyclo[2.2.2]-octane
Vinyl	**Vinyl**	Ethenyl
Benzyl	**Benzyl**	Phenylmethyl
Naphthol	**Naphthol**	Naphthalenol
Naphthylamine	**Naphthylamine**	Naphthalenamine

In the above cases where a change has been made on going from DOC 5 to DOC 6 (e.g. Naphthyl to Naphthalenyl), the old forms have usually not been retained as synonyms because users should have little difficulty in adjusting to the change and locating compounds in the indexes under their new names.

The American spelling sulf- for organosulfur compounds, is used throughout DOC in preference to the British sulph-.

DOC does not seek to establish standards of nomenclature, and the adoption of a particular name as the DOC Name does not imply endorsement, other than on the purely pragmatic grounds outlined above. However, the DOC 6 dataset was carefully reviewed before publication in order to give a high degree of internal consistency in the choice of entry names.

Remember that the DOC nomenclature policy outlined above describes only how we select the DOC *entry name* and that all reasonable alternatives appear as synonyms and can rapidly be found using the Name Index.

For a fuller treatment of nomenclature principles and details, see *The Organic Chemist's Desk Reference*.

4.1.2 *Synonyms*

An important function of DOC is to present a wide range of synonyms. In general the selection is made as useful as possible, but no attempt is made to provide exhaustive lists of proprietary names for pharmaceuticals and other commercial substances. Extensive work over the last 12 years has greatly increased the coverage and consistency of synonyms in the DOC database.

Archaic systematic names are in general not given, but obsolete synonyms have often been retained where there has been a change in numbering of the parent ring system and these synonyms could assist readers who have to consult the older literature. In a few cases incorrect synonyms from the literature have also been reported. Synonyms in these classes are distinguished as '*obsol*' or '*incorrect*' respectively. Several obsolescent systems such as the carbinol and hydroxyalkane alcohol nomenclatures have been almost completely discarded, since although they are still occasionally met with, users should have no difficulty in converting these to the usual nomenclature.

Names which are known to be duplicated within the chemical literature (not necessarily within DOC 6), are marked with the sign †. These are usually duplicate trivial names for natural products or pharmaceuticals, but there are a few cases (of organophosphorus compounds) where two or more compounds of different structure have been allocated the same CAS name.

Synonyms are presented in the following order:

(*a*) ***CAS Names***: names corresponding to those used by Chemical Abstracts Service during the 8th and 9th–13th Collective Index periods (1967–1971 and 1972 onwards respectively) are labelled with the suffixes 8CI and 9CI. Note that 9CI nomenclature is defined as that brought into use by CAS at the beginning of the 9th Collective Index period (1972), and that for organic compounds it has been carried over essentially unchanged into subsequent Collective Index periods (there are some variations for organophosphorus and organoboron compounds). Therefore the suffix '9CI' does *not* mean that the compound can necessarily be found in the 9th Collective Index, as it may have been indexed only since 1976. In addition, 9CI names have been allocated for consistency to some simple compounds which have not been indexed by CAS since 1972, where this could be done without any possibility of error.

Where the 9CI name is not given as a synonym in the DOC 6 entry (usually because it is too cumbersome), it can usually be found by looking it up under the CAS registry number in the CAS Registry Number Handbook section.

(*b*) ***Other systematic names***

(*c*) ***Recommended names***: those recommended by the British Pharmaceutical Commission as British Approved Names (BAN), the United States Adopted Name Review Board (USAN), the British Standards Institute (BSI) or the World Health Organisation as an International Nonproprietary Name (INN). No distinction is made between International Nonproprietary Names which are

currently proposed (pINN) and those which are recommended (rINN). The BAN listing includes compounds approved only for veterinary use.

(*d*) ***Other trivial names***

4.2 *CAS Registry Numbers*

CAS Registry Numbers are identifying numbers allocated to each distinctly definable chemical substance indexed by the Chemical Abstracts Service since 1965 (plus retrospective allocation of numbers by CAS to compounds from the sixth and seventh collective index periods). The numbers have no chemical significance but they provide a label for each substance independent of any system of nomenclature.

In DOC 6, much effort has been expended to ensure that accurate CAS numbers are given for as many substances as possible. If a CAS number is not given for a particular compound, it may be (*a*) because CAS have not allocated one, (*b*) very occasionally because an editorial decision cannot be made as to the correct number to cite, or (*c*) because the substance was added to the DOC database at a late stage in the compilation process, in which case the number will probably be added to the database soon.

4.3 *Additional registry numbers*

At the foot of the DOC entry, immediately before the references, may be shown additional registry numbers. These are numbers which have been recognised by the DOC editors or contributors as belonging to the entry concerned but which cannot be unequivocally assigned to any of the compounds covered by the entry. Their main use will be in helping those who need to carry out additional searches, especially online searches in the CAS or other databases and who will be able to obtain additional hits using these numbers. Clearly, discretion is needed in their use for this purpose.

Additional registry numbers may arise for a variety of reasons:

(*a*) A number may refer to stereoisomers or other variants of the main entry compound which may or may not be mentioned in the entry but for which no physical properties or other useful information is available. For example, the DOC entry for 4,4,5,5-Tetramethyl-2-(1-propenyl)-1,3,2-dioxaborolane [72824-05-6] states that it has so far been obtained only as an inseparable mixture of (*E*) and (*Z*)-forms. The additional registry numbers given are those of the (*E*)-[83947-58-4] and (*Z*)-[83947- 59-5] isomers.

(*b*) A CAS number may refer to a mixture, in which case it is added to the DOC entry referring to the most significant component. It may refer to a hydrate, salt, complex, etc. which is not described in detail in the DOC entry.

(*c*) The number may refer to an undefined isomer, e.g. there is a CAS number for "Chlorobenzoic acid" for use where the original document does not make it clear which isomer is referred to, in addition to the separate numbers for 2-, 3- and 4-chlorobenzoic acids. Such a number may appear as an 'additional registry number' in all three appropriate entries.

(*d*) Replaced numbers, duplicate numbers and other numbers arising from CAS indexing procedure or, occasionally, from errors or inconsistencies by CAS are also reported. For example the DOC entry for *scyllo*-Inositol [488-59-5] contains an additional registry number for D-*scyllo*-Inositol [41546-32-1]. Since *scyllo*-inositol is a meso-compound, the number is erroneous. More generally, CAS frequently replace a given number with one that more accurately represent what they now know about a substance, and the replaced number remains on their files for searching.

(*e*) In the case of compounds with more than one stereogenic centre, additional registry numbers frequently refer to levels of stereochemical description which cannot be assigned to a particular stereoisomer described in the DOC entry.

For example the DOC 6 entry for 2-Amino-3-hydroxy-3-phenylpropanoic acid (β-Hydroxyphenylalanine, 9CI) has a general CAS number [1078-17-7] and CAS numbers for all four optically active diastereoisomers [7352-06-9, 32946-42-2, 109120-55-0, 6524-48-4] as well as the two possible racemates [2584-74-9] [2584-75-0]. However, among the additional registry numbers quoted in DOC 6 are the following:

[7687-36-7] – number for *erythro*-β-Hydroxyphenylalanine
[50897-27-3] – number for β-Hydroxy-L-phenylalanine
[68296-26-4] – number for β-Hydroxy-D-phenylalanine
[39687-93-9] – general number for the methyl ester, hydrochloride which cannot be placed under any of the individual stereoisomers of this compound described in the DOC 6 entry.

(*f*) Numbers may refer to derivatives similar to those described in the DOC 6 entry for which no data is available, or has not yet been added to the entry. Thus in the above example, another additional number [64792-93-4] refers to the methyl ester hydrochloride of (±)-*erythro*-β-Hydroxyphenylalanine. This particular compound is not documented in the DOC 6 entry owing to lack of readily available data.

(*g*) Some DOC entries refer to families of compounds such as natural products. An example is the entry for Calcitonin C-0-00023 where only the porcine and human variants are described in detail. The additional registry numbers given in this entry are those of a number of other species variants which appear to have been identified according to CAS but for which no attempt has been made to collate full data for DOC.

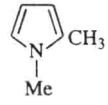
4.4 *Structural formulae*

An extensive guide to the conventions used in representing and describing organic structures is given in the companion volume *The Organic Chemist's Desk Reference*.

Every attempt has been made to present the structures of chemical substances as accurately as possible according to best current practice and IUPAC recommendations. In drawing the formulae, as much consistency as possible between closely related structures has been aimed at. Thus, for example, sugars have been standardised as Haworth formulae and wherever possible in complex structures the rings are oriented in the standard Haworth manner so that structural comparisons can quickly be made.

In a series of closely-related compounds the structural formula is given only for the first member provided that the structures of all isomers following can be unambiguously inferred.

In formulae the pseudoatom abbreviations Me, Et and Ac for methyl, ethyl and acetyl respectively are used only when attached to a heteroatom. Thus 1,2-dimethylpyrrole is

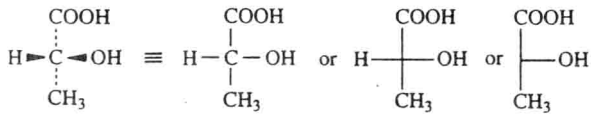


Ph is used throughout whether attached to carbon or to a heteroatom. Other pseudoatom abbreviations such as Pr^i for isopropyl and Bz for benzoyl are not used in DOC.

4.5 *Stereochemical conventions*

Where the absolute configuration of a compound is known or can be inferred from the published literature without undue difficulty, this is indicated. Where only one stereoisomer is referred to in the text, the structural diagram indicates that stereoisomer. Wherever possible, stereostructures are described using the Cahn-Ingold-Prelog sequence-rule (*R,S*) and (*E,Z*) conventions but in cases where these are cumbersome or inapplicable, alternatives such as the α,β-system are used instead. Alternative designations are frequently presented in such cases.

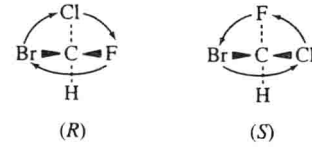
The structure diagrams for compounds containing one or two chiral centres are given in DOC 6 as Fischer-*type* diagrams showing the stereochemistry unequivocally. True Fischer diagrams in which the configuration is implied by the North-South-East-West positions of the substituents are widespread in the literature: they are quite unambiguous but need to be used with caution by the inexperienced. They can not be reoriented without the risk of introducing errors.



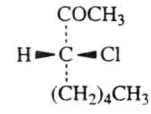
.DOC 6 representation True Fischer projections

4.5.1 *The (R,S)-system*

In the simplest case, the four substituent atoms about a tetrahedral carbon atom are placed in order of decreasing atomic number and the molecule is then viewed from the side remote from the substituent of lowest priority. The configuration is (*R*) (*rectus*) if the order of the three other groups from highest to lowest is clockwise, and (*S*) (*sinister*) if it is anticlockwise.



If two or more of the four atoms attached to the central atom are identical, the molecule is explored outwards by a process of comparing atom with atom.

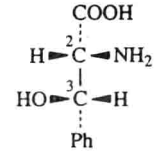


(*R*)-3-Chloro-2-octanone
(priority Cl> C=O>···· CH_2····

Extensions of the (*R,S*)-system refer to situations such as axial and planar chirality (biaryls, cyclophanes, etc.) and to molecules with central atoms other than carbon (e.g. chiral sulfoxides).

Where only the relative configuration of a compound containing more than one chiral centre is known, the symbols (*R**) and (*S**) are used, the lowest-numbered chiral centre being arbitrarily assigned the symbol (*R**). For racemic modifications of compounds containing more than one chiral centre the symbols (*RS*) and (*SR*) are used, with the lowest-numbered chiral centre being arbitrarily assigned the symbol (*RS*). The racemate of a compound containing one chiral centre only is described in DOC as (±)-.

In comparing CAS descriptors with those given in DOC, it is important to remember that the order of presentation of the chirality labels in CAS is itself based on the sequence rule priority and not on any numbering scheme. For example in DOC 6 the following compound:

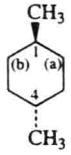


is (2*R*,3*S*)-2-Amino-3-hydroxy-3-phenylpropanoic acid. In CAS it is [*S*-(*R**,*S**)]-β-Hydroxyphenylalanine. The relative stereochemical label (*R**,*S**) is first applied with the *R* applying to chiral centre 3* because it has higher priority than centre 2 (OH > NH_2). The absolute

stereochemical descriptor (*S*)- is then applied changing *R** to *S* for chiral centre 3 and *S** to *R* for chiral centre 2. For further details, see the current CAS Index Guide.

In this Edition of DOC, for simplicity, the enantiomers of bridged-ring compounds such as camphor are described simply as (+)- and (–)-. Although camphor has two chiral centres, steric restraints mean that only one pair of enantiomers can be prepared.

The (*R*,*S*) descriptor system can be extended to describe the configurations of many types of symmetrical compound, e.g. the 1,4-Dimethylcyclohexanes.



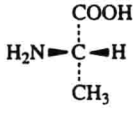
(1*RS*,4*RS*) 1,4-Dimethylcyclohexane

At chiral centre 1, an arbitrary choice is made between the two equivalent sequence chains (a) and (b). Choosing (a) arbitrarily gives (a) > (b) > CH_3 > H at C(1) leading to (*R*)-configuration. The chirality at C(4) is then (a) > (b) > CH_3 > H or (*R*-). The configuration of the compound can therefore be described as (1*RS*,4*RS*) (DOC 6) (or 1*R**,4*R**). This is independent of the arbitrary choice made.

For further information on the (*R*,*S*)-system see Cahn, R.S. *et al*, *J. Chem. Soc.*, 1951, 612; *Experientia*, 1956, **12**, 81; *Angew. Chem. Int. Ed. Engl.*, 1966, **5**, 383.

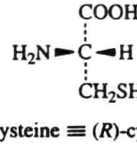
4.5.2 *The* D,L-*system*

This is an older system which is now redundant except for sugars, amino acids and closely related substances. In the simplest case, the Fischer-type projection of the molecule is oriented such that, for an amino acid, the alkyl group is at the bottom and the carboxyl group at the top. If the amino-group is at the left the compound belongs to the L-series, and if to the right, to the D-series.



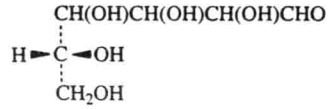
L-alanine ≡ (*S*)-alanine

L corresponds to (*S*) except when the alkyl group is a group of higher priority than COOH, according to the (*R*,*S*)-system.



L-cysteine ≡ (*R*)-cysteine

The D,L assignment for carbohydrates is based on the configuration at the penultimate carbon atom of the sugar chain. Thus for example all hexoses having the configuration



are D-.

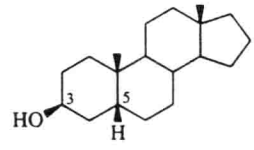
In this Edition, the (*R*,*S*)-system is given precedence over the D,L-system even for amino-acids but the alternative D,L-descriptors are also given.

For a fuller description of the D,L-system and its relationship to (*R*,*S*), see *The Organic Chemist's Desk Reference* and references quoted therein.

4.5.3 *The α,β-system*

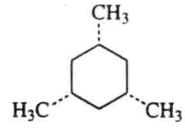
Substituents labelled α- are considered to be below the plane of the molecule and those labelled β- above. This convention is particularly used in DOC for the following types of compound:

(*a*) Steroids and related compounds



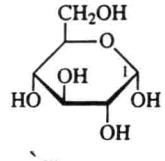
5β-Androstan-3β-ol

(*b*) Cyclic molecules of high symmetry where the (*R*,*S*)-system is inapplicable or cumbersome;

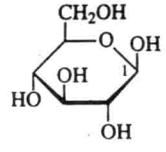


(1α,3α,5α)-1,3,5-Trimethylcyclohexane

(*c*) For denoting the anomeric configuration of sugars



α-D-Glucopyranose



β-D-Glucopyranose

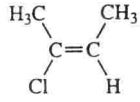
4.5.4 *The (E,Z)-system*

This is an extension of the (*R*,*S*)-system for specifying configurations at alkene double bonds. The substituents are

ordered as in the (*R*,*S*)-system and if the two of higher priority are on the same side of the double bond the configuration is (*Z*) (*zusammen*), while if they are on the opposite side it is (*E*) (*entgegen*).

(*E*)- (*Z*)-

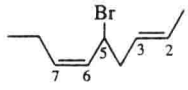
Note that (*E*)- does not always correspond to the *trans*- of the earlier literature.



cis-(w.r.t. CH_3 groups)
(*E*)-(priority Cl>CH_3;CH_3>H)

DOC gives a linear typeset formula, e.g. $H_3CCH_2CH{=}CHCH_2OH$, for alkenes carrying two vicinal hydrogen substituents. The configurations of these alkenes are easy to interpret, with (*E*) = *trans*- and (*Z*) = *cis*-.

Note that in the case of polyalkenes CAS uses sequence-rule descriptors based on sequence-rule priorities (cf. *R** and *S** conventions above). Thus:



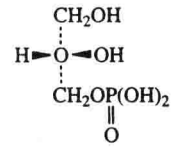
would be (2*E*,5*Z*)-5-Bromo-2,6-nonadiene in DOC but (*Z*,*E*)-5-Bromo-2,6-nonadiene in CAS (the 1-bromo-3-pentenyl substituent on C-6 takes sequence priority over the 2-bromo-3-hexenyl substituent on C-3, so the 6,7-double bond stereodescriptor precedes the 2,3-).

4.5.5 *The* ent-*convention*

Where there is configurational inversion of all of the chiral centres whose configurations are implied in a name, the prefix *ent*- is placed in front of the complete name of the compound. Thus the enantiomer of 11α-Hydroxy-16*S*-kauran-15-one is *ent*-11α-Hydroxy-16*S*-kauran-15-one. This convention is used for natural products (terpenoids) only and is therefore rare in DOC 6: for many more examples, see the *Dictionary of Natural Products*.

4.5.6 *The* sn-*convention*

This convention is sometimes used for glycerides and related substances and takes recognition of the fact that $C_{(2)}$ of the parent substance, glycerol, is prochiral and mono-substitution at *O*-1 or *O*-3 produces distinguishable compounds.



sn-Glycerol 3-(dihydrogen phosphate)

The system has the advantage over the (*R*,*S*)-system that closely related compounds always have the same *sn*-configuration, whereas they may belong to opposite series according to the (*R*,*S*)-convention. However, only limited reference to this system is made in DOC.

4.6 *Molecular formula and molecular weight*

The elements in the molecular formula are given according to the Hill convention (C, H, then other elements in alphabetical order. The molecular weights given are formula weights (or more strictly, molar masses in daltons) and are rounded to one place of decimals. In the case of some high molecular mass substances such as proteins the value quoted may be that taken from an original literature source and may be an aggregate molar mass.

4.7 *Importance/use*

Care has been taken to make the information given on the importance and uses of chemical substances as accurate as possible. Many substances have now been patented for a wide variety of uses but this does not imply that the patented uses are of widespread applicability or even of established utility. In general, information on a particular use is given prominence only when it is documented in a critical source, such as *Kirk-Othmer* or *Ullmann*, when it is protected by numerous patents, or when a reference is quoted which will assist the reader to assess the value of the claimed application.

4.8 *Physical data*

4.8.1 *Appearance*

Organic compounds are considered to be colourless unless otherwise stated. Where the compound contains a chromophore which would be expected to lead to a visible colour, but no colour is mentioned in the literature, the DOC entry will mention this fact if it has been noticed by the contributor. An indication of crystal form and of recrystallisation solvent is often given but these are imprecise items of data: most organic compounds can be crystallised from several solvent systems and the crystal form often varies. In the case of the small number of compounds where crystal behaviour has been intensively studied (e.g. pharmaceuticals), it is found that polymorphism is a very common