

INTRODUCTION TO THE DATABASE

The Chapman & Hall/CRC Chemical Database is a structured database holding information on chemical substances. It includes descriptive and numerical data on chemical, physical and biological properties of compounds; systematic and common names of compounds; literature references; structure diagrams and their associated connection tables. The *Dictionary of Organic Compounds on CD-ROM* is a subset of this and includes all compounds contained in the *Dictionary of Organic Compounds* (Sixth Edition and Supplements), *Dictionary of Organophosphorus Compounds* and some other publications (with the omission of the majority of natural products, which are available in the companion *Dictionary of Natural Products on CD-ROM*).

COMPOUND SELECTION

In general, DOC includes the following compounds:

- The basic fundamental organic compounds of simple structure.
- Compounds of industrial or commercial value including currently important pharmaceuticals, pesticides, monomers, etc.
- Compounds frequently encountered in the laboratory as solvents, reagents or starting materials.
- 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 on CD-ROM*.
- Important biochemicals.
- Other compounds of particular interest because of their chemical, structural or biological properties, including many newly synthesised compounds of active research interest.

DOC also 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 Inorganic and Organometallic Compounds on CD-ROM*

DOC does not in general document isotopically labelled variants except for a few isotopically labelled drug variants used in chemotherapy or tracer work.

DATA PRESENTATION AND ORGANISATION

Derivatives and variants

In the database, closely related compounds are grouped together to form an entry. Stereoisomers and derivatives of a parent compound are all listed under one entry. The compounds in the *Dictionary of Organic Compounds* are grouped together into approximately 71,000 entries.

The structure of an entry is shown below.

<p>Entry (parent compound) Derivatives</p> <p>Variants (stereoisomers or other closely-related compounds) Derivatives of the variant</p>
--

In a simple entry, there is just one compound, with no derivatives or variants.

Variants are commonly stereoisomers, e.g., (*R*)-form, *endo*-form. *Derivatives* may include hydrates, complexes, salts, classical organic derivatives, substitution products and oxidation products, 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.

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 entries under the parent compound.

Carbonyl compounds, RR'C=O

Oxime, RR'C=NOH

Hydrazone, RR'C=NNH₂

Phenylhydrazone, RR'C=NNHPh

2,4-Dinitrophenylhydrazone, RR'C=NNH[C₆H₃(NO₂)₂]

Semicarbazone, RR'C=NNHCONH₂

Di-Me acetal, RR'C(OMe)₂

Di-Et acetal, RR'C(OEt)₂

Ethylene acetal

Carboxylic acids, RCOOH

Salts, e.g., Na salt, NH₄ salt

Esters, RCOOR'

Chloride (=acid chloride), RCOCl

Amide, RCONH₂

Alkylamides, RCONHR'

Anilide, RCONHPh

Dialkylamides, RCONR'R''

Nitrile, RCN

Anhydride, RCO-O-COR

Imide, RCO-NH-COR

Amines, RNH₂, RR'NH, RR'R''N

Salts, e.g., hydrochloride, methiodide

Picrates and other complexes

Ac (*N*-Acetyl) deriv., RNHAc (Ac=COCH₃)

N-Benzoyl, RNHCOPh

N-Alkyl, RNHR'

N,N-Dialkyl, RNHR'R'

N,N,N-Trialkyl, RR',R'',R'''N[⊕]

N-Oxide, RN(O)H

Alcohols, ROH

Ac (acetate), ROAc (Ac=COCH₃)

Benzoyl, ROCOPh

Benzenesulfonyl and 4-methylbenzenesulfonyl (tosyl) ROSO₂Ar

Alkyl ethers ROR'

Thiols, RSH

S-Ac, RSAc

S-Benzoyl, RSCOPh

Benzenesulfonyl and 4-methylbenzenesulfonyl (tosyl) RSSO₂Ar

S-Alkyl, RSR'

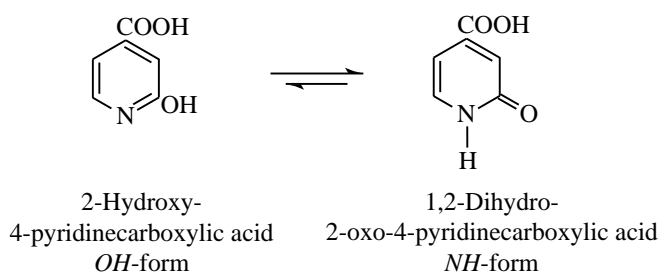
Disulfide, R-S-S-R (sometimes included; usually a readily obtainable oxidative dimer of the thiol)

Molecular formulae are included in DOC for nearly all of these derivatives and so are readily searchable, 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).

Tautomerism

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. 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.



(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 compound 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.

(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 may only 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 ester (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 *1H*-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.

DATA TYPES

The format of a typical entry is given in Fig.1, and shows the individual types of data that may be present in an entry.

Chemical names and synonyms

All of the names discussed below can be searched using the Chemical Name field.

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 conventions 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.

The American spelling sulf- for organosulfur compounds, is used throughout DOC in preference to the British sulph-. For a fuller treatment of nomenclature principles and details, see *The Organic Chemist's Desk Reference* (Chapman & Hall/CRC, 1995).

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.

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 '*obso*' or '*incorrect*' respectively. Several obsolescent systems such as the carbinol and hydroxyalkane alcohol nomenclatures have

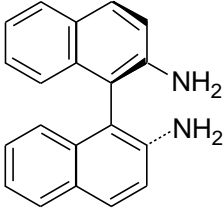
DOC name	→	2,2'-Diamino-1,1'-binaphthyl
Structural formula and stereochemical description	→	 (<i>R</i>)-form
Alternative names	→	[1,1'-Binaphthalene]-2,2'-diamine, 9CI. 2,2'-Diamino-1,1'-dinaphthyl. 1,1'-Bi[2-naphthylamine]
CAS Registry Number	→	FNC76-Y [4488-22-6]
Molecular Formula	→	C ₂₀ H ₁₆ N ₂
Use	→	Intermediate for chiral auxiliaries.
Hazard alert symbol and description of hazards	→	▶ Exp. tumourigen by skin contact. Dec. with emission of toxic fumes. DU3090000
Supplier information	→	<p>(<i>R</i>)-form: FNC77-Z [18741-85-0] Mp 242.5-243°. [α]_D^{21.4} + 155.5° (c, 1 in Py). [α]_D¹⁸ + 46.8° (2<i>M</i> HCl). Supplier: Aldrich 38242-6; Fluka 32787. <i>N,N</i>-Di-Me: MMX23-Z [93713-30-5] Cryst. (EtOH). Mp 143-144° [α]_D²³ + 182° (c, 1.09 in C₆H₆). <i>N,N,N',N'</i>-Tetra-Me: MMX24-A [135029-77-5] Cryst. (EtOH/C₆H₆). Mp 216-218°.</p>
Stereoisomer heading	→	<p>(<i>S</i>)-form: FNC78-A [18531-95-8] Cryst. Mp 243° (235-239°). [α]_D²⁰ - 149° (Py). [α]_D¹⁹ - 46° (2<i>M</i> HCl). Supplier: Aldrich 38243-4; Fluka 32788.</p>
Derivative Subheading	→	<p><i>N,N</i>-Di-Ac: FNC80-V C₂₄H₂₀N₂O₂ M 368.434. Prisms (C₆H₆). Mp 226-227°. [α]_D²⁵ + 10.8° (c, 1 in THF). (±)-form: FNC81-W [79082-81-8] Silvery plates (EtOH). Mp 193.2-194.5° (191°). Picrate: FNC84-Z Brownish-yellow plates (C₆H₆). Mp 185° (dec.). <i>N,N</i>-Di-Ac: FNC82-X Cubes (EtOH). Mp 235-236°. <i>N,N</i>-Dibenzoyl: FNC83-Y C₃₄H₂₄N₂O₂ M 492.576. Prisms (PhNO₂). Mp 235°.</p>
Additional CAS Registry Numbers	→	[93621-61-5] [97644-73-0]
Bibliographic references	→	<p>Kuhn, R <i>et al.</i>, <i>Annalen</i>, 1929, 470, 183 (<i>synth, resoln</i>) Cumming, WM <i>et al.</i>, <i>J.C.S.</i>, 1932, 528 (<i>synth</i>) Clemo, GR <i>et al.</i>, <i>J.C.S.</i>, 1939, 1114 (<i>synth</i>) Mislow, K <i>et al.</i>, <i>J.A.C.S.</i>, 1962, 84, 1455 (<i>uv, ord</i>) Akimoto, H <i>et al.</i>, <i>Tetrahedron</i>, 1971, 27, 5999 (<i>resoln, abs config</i>) Miyano, S <i>et al.</i>, <i>Bull. Chem. Soc. Jpn.</i>, 1984, 57, 2171 (<i>pmr, ir, deriv</i>) Brown, KJ <i>et al.</i>, <i>J.O.C.</i>, 1985, 50, 4345 (<i>synth, resoln</i>) Benson, SC <i>et al.</i>, <i>J.O.C.</i>, 1988, 53, 5335 (<i>synth, N-tetra-Me</i>) <i>Fieser and Fieser's Reagents for Organic Synthesis</i>, Wiley, 1989, 14, 32 (<i>use</i>) Franzini, L <i>et al.</i>, <i>Acta Cryst. C</i>, 1991, 47, 1259 (<i>cryst struct, N-tetra-Me</i>) Smrcina, M <i>et al.</i>, <i>J.O.C.</i>, 1992, 57, 1917 (<i>synth, resoln, bibl</i>) Lewis, RJ <i>et al.</i>, <i>Sax's Dangerous Properties of Industrial Materials</i>, 8th edn., Van Nostrand Reinhold, 1992, BGB750</p>

Fig. 1. Sample entry from database

been almost completely discarded, since although they are still occasionally met with, users should have no difficulty in converting these to the normal nomenclature.

Names which are known to be duplicated within the chemical literature (not necessarily within DOC), 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.

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, 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.

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 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 *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 represents 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 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 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 entry.

(f) Numbers may refer to derivatives similar to those described in the DOC 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 (\pm)-*erythro*- β -Hydroxyphenylalanine.

This particular compound is not documented in the DOC 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 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.

Diagrams

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 current best 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 formulae the pseudoatom abbreviations Me, Et and Ac for methyl, ethyl and acetyl respectively, are used only when attached to a heteroatom. Ph is used throughout whether attached to carbon or to a heteroatom. Other pseudoatom abbreviations such as Prⁱ for isopropyl and Bz for benzoyl are not used in DOC.

It should be noted that in each entry display there is a single diagram which applies to the parent entry. Separate diagrams are not given for variants or derivatives.

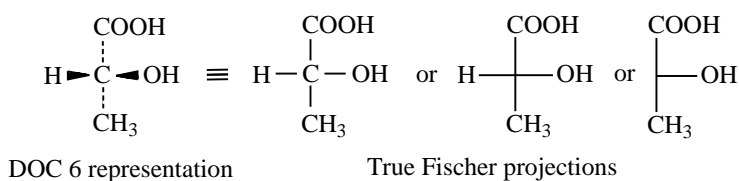
Structures for derivatives can be viewed in **Structure Search**, but remember that these structures are generated from connection tables and may not always be oriented consistently.

Stereochemical conventions

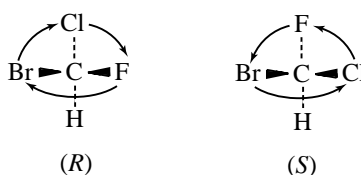
Where the absolute configuration of a compound is known or can be inferred from the published literature without undue difficulty, this is indicated.

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.

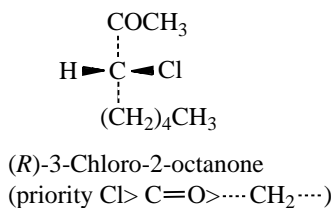
The structure diagrams for compounds containing one or two chiral centres are given in DOC as Fischer-*type* diagrams showing the stereochemistry unequivocally. True Fischer diagrams in which the configuration is implied by the North-South-East-West positions of the substituents are widespread in the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



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.

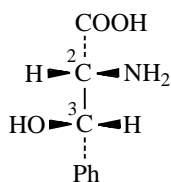


Extensions of the (*R,S*)-system refer to situations such as axial and planar chirality (biaryls, cyclophanes, etc.) and to molecules with central atoms other than carbon (e.g. chiral sulfoxides).

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 (\pm).

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

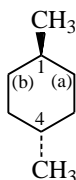
based on the sequence rule priority and not on any numbering scheme. For example in DOC, the following compound:



is (2*R*,3*S*)-2-Amino-3-hydroxyphenylpropanoic 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 > NH₂). 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.

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*R*,4*R*) 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) > CH₃ > H at C(1) leading to (*R*)-configuration. The chirality at C(4) is then (a) > (b) > CH₃ > H or (*R*-). The configuration of the compound can therefore be described as (1*R*,4*R*)(DOC) (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.

Where appropriate, alternative stereochemical descriptors may be given using the D, L or α,β -systems. For a fuller description of these systems, consult *The Organic Chemist's Desk Reference*.

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.

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. Data in this field may be searched under **Use/Importance** or **All Text**.

Physical Data

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 phenomenon and there is no reason to believe that it is not widespread among organic compounds generally.

Melting points and boiling points

The policy followed in the case of conflicting data is as follows:

- (a) Where the literature melting points are closely similar, only one figure (the highest or most probable) is quoted.
- (b) Where two or more melting points are recorded and differ by several degrees (the most likely explanation being that one sample was impure), the lower figure is given in parentheses, thus: Mp 139°(134–135°).
- (c) Where quoted figures differ widely and some other explanation such as polymorphism or incorrect identity seems to be the most likely explanation, both figures are quoted without parentheses, thus Mp 142°, Mp 205–206°.
- (d) Known cases of polymorphism or double melting point are noted.

Boiling point determination is less precise than that of melting points and conflicting boiling point data is not usually reported except when there appears to be a serious discrepancy between the different authors.

Optical rotations

These are given whenever possible, and normally refer to what the DOC contributor believes to be the best-characterised sample of highest chemical and optical purity. Where available an indication of the optical purity (op) or enantiomeric excess (ee) of the sample measured now follows the specific rotation value.

Specific rotations are dimensionless numbers and the degree sign which was formerly universal in the literature has been discontinued.

Densities and refractive indexes

Densities and refractive indexes are now of less importance for the identification of liquids than has been the case in the past, but are quoted for common or industrially important substances such as solvents, or where no boiling point can be found in the literature.

Densities and refractive indexes are not quoted where the determination appears to refer to an undefined mixture of stereoisomers.

Solubilities

Solubilities are given only where the solubility is unusual for an organic compound. Typical organic compounds are soluble in the usual organic solvents such as ether and chloroform, and virtually insoluble in water. The presence of polar groups (OH, NH₂ and especially COOH, SO₃H, NR₃⁺) increases water solubility.

pK_a values

pK_a values are given for both acids and bases. The pK_b of a base can be obtained by subtracting its pK_a from 14.17 (at 20°) or from 14.00 (at 25°).

Spectroscopic data

Spectroscopic data such as uv wavelengths and extinction coefficients, are given only where the spectrum is a main point of interest, or where the compound is unstable and has been identified only by spectroscopic data.

In many other cases, spectroscopic data can be rapidly located through the references quoted.

Hazard and toxicity information

General

Toxicity and hazard information is displayed in red type and additionally is highlighted by the sign ►. It has been selected to assist in risk assessments for experimental, manufacturing and manipulative procedures with chemicals.

Physical, reactive and **toxic** properties all contribute to the hazard associated with a particular chemical. As part of the **physical data**, flash points, explosive limits and autoignition temperatures have been included (where appropriate). Flammability classifications, which are based on flash point measurements and boiling points, are also mentioned, and the opportunity has been taken to include UK occupational exposure limits, or for some compounds threshold limit values published by the American Conference of Governmental Industrial Hygienists (ACGIH).

For the **reactive** hazards, a brief comment is made on any explosive (or violent polymerisation) properties and aspects of the chemical reactivity of a substance which are of concern. These include the potential for peroxidation, oxidizing/reducing properties and incompatibility with commonly available chemicals.

Toxicity information has been chosen to show hazardous effects from short-term or long-term exposure. Observations from human exposure are summarised if available (including possible adverse effects of drugs), otherwise experimental (exp.) tests are quoted. Included in the toxicity data are the results of irritancy tests, acute lethality data, target organ toxicity, and carcinogenic and reproductive properties where appropriate. Those chemicals which have been classified by the International Agency for Research on Cancer (IARC) as *human carcinogens*, *probable human carcinogens* or *possible human carcinogens* have been identified in DOC accordingly.

The Publishers cannot be held responsible for any inaccuracies in the reported information, neither does the omission of hazard data in the **Dictionary** imply an absence of this data from the literature. Widely recognised hazards are included however, and where possible key toxicity reviews are identified in the references. Further advice on the storage, handling and disposal of chemicals is given in *The Organic Chemist's Desk Reference*.

Finally, it should be emphasised that any chemical has the potential for harm if it is carelessly used. For many newly synthesised materials (e.g. new synthetic reagents), hazardous properties may not be apparent or may have been cited in the literature. In addition, the toxicity of some very reactive chemicals may not have been evaluated for ethical reasons, and these substances in particular should be handled with caution.

RTECS® Accession Numbers

RTECS® Accession Numbers are compiled and distributed by the National Institute for Occupational Safety and Health Service of the U.S. Department of Health and Human Services of The United States of America. All rights reserved. (1996)

Many entries in DOC contain one or more RTECS® Accession Numbers. Possession of these numbers allows users to locate toxicity information on relevant substances from the NIOSH Registry of Toxic Effects of Chemical Substances, which is a compendium of toxicity data extracted from the scientific literature.

For each Accession Number, the RTECS® database provides the following data when available: substance prime name and synonyms; date when the substance record was last updated; CAS Registry Number; molecular weight and formula; reproductive, tumorigenic, and toxic dose data; and citations to aquatic toxicity ratings, IARC reviews, ACGIH Threshold Limit Values, toxicological reviews, existing Federal standards, the NIOSH criteria document program for recommended standards, the NIOSH current intelligence program, the NCI Carcinogenesis Testing Program, and the EPA Toxic Substances Control Act inventory. Each data line and citation is referenced to the source from which the information was extracted.

Bibliographic References

The selection of references is made with the aim of facilitating entry into the literature for the user who wishes to locate more detailed information about a particular compound. Thus, in general, recent references are preferred to older ones, particularly for chiral compounds where optical purity and absolute configuration may have been determined relatively recently. The number of references quoted cannot therefore be taken as an indication of the relative importance of a compound, and the references quoted for important substances may not be the most significant historically. For very common compounds which are nowadays readily available from bulk suppliers, long lists of syntheses are not presented, but the emphasis is on references to spectra, chromatography, etc.

References are given in date order except for references to spectroscopic library collections, which sort at the top of the list, and those to hazard/toxicity sources which sort at the bottom.

The content of many references are indicated by means of suffixes. A list of the most common ones is given in Table 1.

Some reference suffixes are now given in **boldface** type, where the editors consider the reference to be particularly important, for example the best synthesis giving full experimental details and often claiming a higher yield than previously reported methods.

In some entries, minor items of information, particularly the physical properties of derivatives, may arise from references not cited in the entry.

Journal abbreviations

In general these are uniform with the *Chemical Abstracts Service Source Index* (CASSI) listing except for a short list of very common journals:

DOC ABBREVIATION

Acta Cryst.
(and sections thereof)
Annalen
Chem. Comm.
J.A.C.S.
J.C.S. (and various
subsections thereof)
J. Het. Chem.
J.O.C.
Tet. Lett.

CASSI

Acta Crystallogr.
(and sections thereof)
Justus Liebigs Ann. Chem.
J. Chem. Soc., Chem. Commun.
J. Am. Chem. Soc.
J. Chem. Soc. (and various
subsections thereof)
J. Heterocycl. Chem.
J. Org. Chem.
Tetrahedron Lett.

Table 1. Reference tags

The following is a selection of the most common reference tags that are used

Abbreviation	Meaning
abs config	absolute configuration
anal	analysis
bibl	bibliography
biodistribn	biodistribution
biosynth	biosynthesis
cd	circular dichroism
chromatog	chromatography
cmr	¹³ C nuclear magnetic resonance spectrum
config	configuration
conformn	conformation
cryst struct	X-ray crystal structure determination
deriv(s)	derivative(s)
detn	determination, detection
epr	electron paramagnetic (spin) resonance spectrum
glc	gas-liquid chromatography
haz	hazard
hplc	high performance liquid chromatography
ir	infrared spectrum
isol	isolation
isom	isomerism
manuf	manufacture
metab	metabolism
ms	mass spectrum
nmr	nuclear magnetic resonance spectrum
occur	occurrence
ord	optical rotatory dispersion
pharmacol	pharmacology
pmr	proton (¹ H) nuclear magnetic resonance spectrum
props	properties (chemical or physical)
resoln	resolution

Table 1. Reference tags *continued*

<i>Abbreviation</i>	<i>Meaning</i>
rev	review
sepn	separation
spectra	
struct	structure
synonyms	
synth	synthesis
tautom	tautomerism
tlc	thin layer chromatography
tox	toxicity
use(s)	
uv	ultra-violet visible spectrum