Chemical Nomenclature

Edited by K. J. Thurlow

KLUWER ACADEMIC PUBLISHERS

	DATE DUE		

Chemical Nomenclature

Chemical Nomenclature

Edited by

K. J. Thurlow
Laboratory of the Government Chemist
Teddington, Middlesex
UK

KLUWER ACADEMIC PUBLISHERS
DORDRECHT/BOSTON/LONDON

A C.I.P. Catalogue record for this book is available from the Library of Congress

ISBN 0 751 404756

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Sold and distributed in North, Central and South America
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands

Printed in Great Britain

For my parents

Contributors

J. Buckingham	Chapman & Hall, 2–6 Boundary Row, London SEI 8HN, UK
P. E. Childs	Dept. of Chemical and Environmental Sciences, University of Limerick, Limerick, Ireland
E. W. Godly	10 Lawrence Avenue, New Malden, Surrey KT3 5LY, UK
A. D. Jenkins	The School of Chemistry, Physics and Environmental Science, University of Sussex, Falmer, Brighton BN1 9QJ, UK
K. J. Thurlow	Chemical Nomenclature Advisory Service, LGC, Queen's Road, Teddington, Middlesex TW11 0LY, UK
R. B. Trigg	British Pharmacopoeia Commission, Market Towers, 1 Nine Elms Lane, London SW8 5NQ, UK
S. B. Walker	Mr. S.B. Walker, 45 Windmill Avenue, Wokingham, Berks RG41 3XA, UK
D. Weisgerber	Chemical Abstracts Service, 2540 Olentangy River Road, PO Box 3012, Columbus, Ohio 43210–0012, USA

Contents

Preface

My interest in nomenclature started when I was child. I was listening to a recording of Tom Lehrer performing *The Elements*. The original was almost worn out as I tried to write down the words and make sense of them. Praseodymium, proctactinium, dysprosium... What could they mean? The more I learned, the more fascinating it became. A twin interest in chemistry and words fuelled the desire for more knowledge. I joined the Laboratory of the Government Chemist (LGC) and worked in a variety of areas. When a member of the Chemical Nomenclature Advisory Service (CNAS) was approaching retirement, I was asked to replace him. Part of the duties included writing articles on nomenclature for the *VAM Bulletin*, a publication supporting the Department of Trade and Industry's Initiative on 'Valid Analytical Measurement' Nomenclature underpins all analysis, because if you do not know what you are analysing, you will not get very far. These articles were written for non-specialists and proved to be the genesis for this book.

The idea was to produce a book which would give a general introduction to forms of nomenclature, without reading too much like a text book. Nomenclature can be interesting, not just a 'necessary evil'. Some people even think it is an 'unnecessary evil'! Chemical names do not have to be deadly serious. There have been many books on specific aspects of nomenclature, but this volume deals with both 'systematic' and 'trivial' names. Apart from a general discussion on why you need nomenclature, readers can compare CAS and IUPAC styles, which have many similarities, but important differences. Specialised naming systems are needed for polymers and natural products, and of course no such work would be complete without a chapter on the elements. Computers are playing an increasing role in nomenclature, and work continues in the quest to find a program which will name a given structure correctly and quickly.

It is impossible to deal with these subjects without some complicated material, indeed it is necessary to demonstrate how difficult some problems are. However, I hope the non-specialist will appreciate the material herein, and that the specialist will appreciate information on unfamiliar areas. Rivalry between nomenclators is frequently intense, but we all have the same aim, and perhaps this book will engender better understanding and cooperation.

I would like to thank everyone involved in this book. The authors and publishers have put a great deal of effort into its production.

Dr Richard Worswick, Chief Executive of LGC, has kindly given permission for CNAS material to be used in the preparation of this book. I must also thank my CNAS colleagues, without whose assistance I would not have been in a position to participate in this project. Norman Soutar (my first trainer), Gary Sayers, Ivor Cohen and Ted Godly have all offered helpful advice over the last few years, and pointed out errors! Thanks also to IUPAC.

I hope this book helps explain why nomenclature is necessary and how it works.

Kevin Thurlow
February 1998

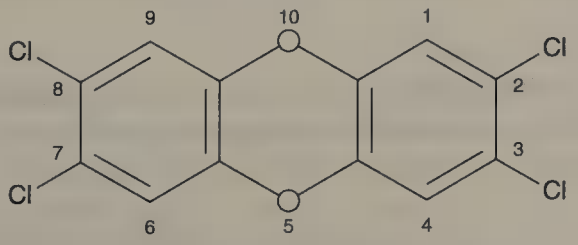
1 The need for good nomenclature

E. W. GODLY

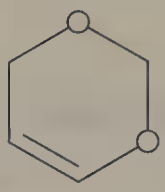
. . . killing's the matter. – but he can tell you the perpendiculars.

R. B. Sheridan "The Rivals", Act V, Scene 1

Mrs. Malaprop is there to be enjoyed. Her wrong notes are mostly easily recognized and amended or else the joke falls flat. However, the primary purpose of Chemistry is not comic entertainment and the fact that its every aspect is (or should be) underpinned by good nomenclature is seldom the subject of mirth. After all, the main object of giving names to chemicals is to distinguish them from other chemicals. Although this is a truism, it can be startlingly disregarded. For example, the systematic name for the substance causing the evacuation of the Italian town of Seveso in 1976 is 2,3,7,8-tetrachlorodibenzo[*b,e*][1,4]dioxin (**1.1**). It has since featured in many publications and the editor of a well-known Chemistry journal, finding it too much of a mouthful, published an anarchic statement to the effect that it was thereafter to be referred to in his journal as 'dioxin', so dropping everything before the last three syllables. This Procrustean corrective recourse (which mercifully has not started a fashion) ignores the fact that the name 'dioxin' applies only to the central ring of structure (**1.1**). It has also been used for the relatively harmless preservative methyl (2,6-dimethyl-1,3-dioxan-4-yl)acetate, so it could be considered thoroughly 'used up' already. The situation has deteriorated: 'dioxin' has become a generic name for any derivative of the three-ringed skeleton of structure (**1.1**) having one or more of its H atoms substituted by chlorine. As toxicity is of prime interest within this group, it is interesting to note that it covers a range with some of its members being a little more toxic than common salt right up to substance (**1.1**) which has powerful effects even at tiny concentrations.



(1.1)



1,3-Dioxin
(1.2)

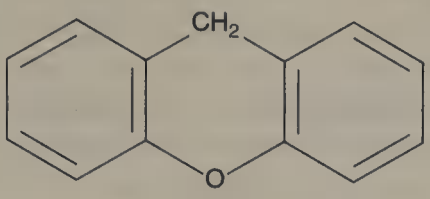
In view of the emotive concerns which the name 'dioxin' generates, it is unfortunate that it has been applied so blithely to include such relatively innocuous substances along with the scourge of Seveso.

This is an extreme case of a chemical name signifying more than one substance, so inviting confusion of a most serious kind, although it was the subject of a hilarious episode of the BBC TV-programme 'Yes, Minister' in which an MP heard of a plan to manufacture '*meta*-dioxin' (1,3-dioxin; structure (**1.2**)) in her constituency.

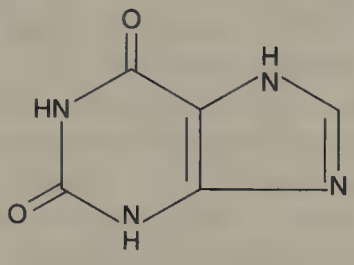
For obvious reasons, such conflicts are normally avoided but a few are in common usage. It has long embarrassed teachers of chemical nomenclature that 'ethylene' means both the hydrocarbon $H_2C{=}CH_2$ and the divalent group $-CH_2-CH_2-$. Although it is usually clear from the context which is meant, it still rankles and the IUPAC [1] Commission on Nomenclature of Organic Chemistry (CNOC) is trying to encourage 'ethene' for the unsaturated hydrocarbon and 'ethane-1,2-diyl' for the divalent group[1a]. ('Polyethylene' could be said to mean both things at once depending on whether this name is intended to be structurally descriptive or to be based on the monomer (cf. polystyrene and PVC); both approaches are accepted by IUPAC's Commission on Macromolecular Nomenclature [2].)

A slightly less obvious source of potential confusion is the occurrence of homophonic names, those that are spelled differently but sound the same. Examples are fluorine the element and fluorene the aromatic hydrocarbon, xanthene (**1.3**) and xanthine (**1.4**). Neither of these seem to cause as much trouble as might be expected. Such conflict is most serious when both names are used in the same field of study or application; it is particularly dangerous when they are both drugs prescribable to the same patient, most especially when one has much more powerful effects than the other. I offer no example of this because great trouble is taken to ensure that it never happens.

A current case of homophonic confusion is that anticipated as a result of the IUPAC recommendation (following an animated exchange of views within the Chemical Community) for naming element 107 Bohrium in honour of the distinguished Danish chemist Niels Bohr. When it was realized that salts and esters of the oxoacids of this element would be termed 'bohrates', it was feared that confusion with similar compounds of element



Xanthene
(1.3)



Xanthine
(1.4)

number 5 might result. The suggested remedy was to name the heavier element 'Nielsbohrium', a solution which seemed unappealing. Unlike xanthine and xanthene, which are both stable compounds, element 107 has a very short half-life and its compounds inhabit a very different world from the borates. True, this doesn't prevent references to them in the appropriate literature, but there the 'h' in the centre of 'bohrate' precludes confusion with 'borate'. In speech, it is surely no great labour to add 'with an h' when necessary. To avoid such rare and comparatively far-fetched instances of potential confusion by generating the 'nielsbohrate' moiety seemed simultaneously over-zealous and lacking in proper respect for one of the founders of modern Atomic Theory. Niels is not an uncommon name in Denmark and Nielsbohrium called to mind possible analogous treatment for some other elements and the atomic symbol so generated: Mariecurium (Mc), Enricofermium (Ef), Dimitrimendelevium (Dm). Even if IUPAC had embraced the fondness prevailing in some quarters for such easy-going first-name familiarity, 'Nielsbohrium' could not have been assigned the symbol Nb without considerable upheaval in niobium nomenclature.

It is perhaps unreasonable to expect those responsible for coining new names to take account of dialectical variations in pronunciation. There is a story of an Australian who made a telephonic order for a cylinder of acetylene under the more modern name 'ethyne' and was not pleased to receive ethane instead.

Chemical names are often mispronounced. One has only to listen to the efforts of radio announcers and newsreaders to get their tongues round

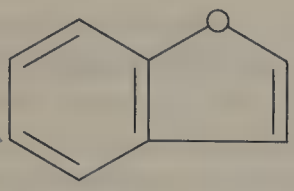
them to realize that correct pronunciation of chemical entities is not a widely shared skill. The confusion between silicon and silicone is constantly being manifested by reference to the uncomfortable-sounding silicon implant. The pesticide Dieldrin is more often than not rendered as though it were the second member of some supposed 'eldrin' series rather than deriving from the first of the two chemists who described the Diels–Alder reaction. A colleague of mine was once quite bemused to receive a query concerning 'twelve pro-pan-needy-ol' until he managed to decipher this as 1,2-propanediol. (Placement of the locants immediately before the functions they refer to means that propane-1,2-diol is now more likely to be used and that may ease some of this oral difficulty.) I was similarly mystified to hear a name quoted over the telephone which appeared to end with '1' and I began to assume it was a French attempt as their locants tended to come at the end. This line of reasoning collapsed when I realized that the speaker was describing a ketone (not a 'ket-l').

A number of different chemical pronunciations prevail within the so-called English-speaking world. 'Ethyl' is spoken in Brit-speak to rhyme with 'free-style' whereas in the USA it sounds more like the girl's name Ethel. Even the word 'nomenclature' carries the stress on 'men' in the UK whereas the first and third syllables are stressed in American English. The moral of all this is that Chemistry is best communicated in writing and discussion by telephone should be confirmed by post.

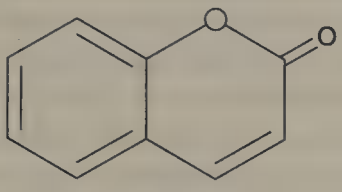
This invokes the alternative possibility of visual confusion and pharmaceutical names in particular come in for criticism for starting or finishing with the same group of letters as other names, particularly within series of similar compounds. This happens a lot with such beginnings as Brom . . ., Cycl . . ., Dex . . ., Iso . . ., Meth . . ., Phen . . . and Tetra . . ., and such endings as . . . amide, . . . cillin, . . . olol and . . . zepam. The tradition for doctors' prescriptions to tail off into indeterminate squiggles for the rest of the name is invoked to sustain the case for risk of conflict.

The number of generic names (whose very purpose is defeated if they have too many syllables) is such that devising new and also distinctive names for new drugs whilst conveying some indication of clinical class, would be impossible if allowance for poor handwriting had also to be made. Surely the time has come when the GP might reasonably be expected to ensure that the pharmacist can read his prescription without doubt or difficulty. In Switzerland it is, I am told, obligatory for prescriptions to be typewritten. (This leads to a new area of conflict – that generated by the finger landing on an adjacent key. It came as a surprise to me to find that French and English typewriters do not have the same keyboard.)

I have preferred writing as being less prone to error than speech, but even written chemical terminology should be accurate as regards both spelling and the avoidance of mis-description. If you miss the 'n' out of 'menthyl', don't be surprised if you get the wrong result. Some spelling



Coumarone
(1.5)



Coumarin
(1.6)

errors are commonplace but seem not to matter too much: one or other of the 'h's is often omitted from 'naphthalene'; 'methly' and 'flouro' recall Mrs. Malaprop again – but the reader usually makes the right correction and this is an argument against replacing chemical names by numbers. It doesn't seem to matter that glycerol trinitrate has long been named 'nitroglycerine' even though it is not a nitro compound; the usage is so well established that it is amended or ignored. More baffling are such apparent paradoxes as coumarone/coumarin (structures (**1.5**) and (**1.6**)): why is the oxo compound not the one with the '-one' ending? Why is 'copperas' an iron salt? How did zinc sulfide, which is white, come to be called 'Black Jack'?

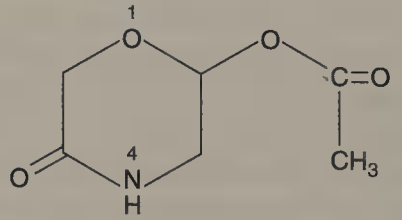
Some variations in spelling remain unresolved: alumin(i)um, (o)estrane, sulfur/sulphur, as proponents of each alternative continue to cling to their established practice. Such variations afflict only those concerned with constructing alphabetical indexes.

The case for avoiding the application of the same name to different compounds has, I hope, been made and also that for avoiding conflict, whether through error or ambiguity.

What of the assignment of more than one name to the same compound? This is both inevitable and desirable although it can be carried too far, as I shall show. Chemical names are analogous in some ways to those given to people. Let us consider the progress of a new substance whether isolated from a natural source or synthesized. In the case of the natural product, it would once have been named as an active principle in a plant extract, say, even before being isolated, e.g. quinine, obtained from cinchona bark. In the case of a synthetic with no such botanical ancestry it might be referred to in the research laboratory by a convenient partly

descriptive but cryptic title as 'diketone-4' or perhaps DK-4. This sort of description may survive into publication texts where the structure is supplied and numbered, e.g. 'the diketone (999)'. When the publication is abstracted by the Chemical Abstracts Service at Columbus, Ohio, USA, the substance will be assigned a CAS Registry number and entered in the names index under a systematic name devised by the CAS according to their procedures (more fully described in a later chapter). Like a child dying at birth, it may never acquire any other names. However, it may be necessary to patent it and/or to report it for legal purposes, e.g. to the European Commission. Now, the laboratory nickname will not suffice. Instead, an internationally acceptable systematic name will be needed to convey the exact structure (so far as is known; stereo-descriptors for example may have to be added later) – this is analogous to the official christening or registration of a person's birth. Such a name should convey unambiguously to those capable of deciphering it (and that includes the Patent Examiner and/or the relevant EC officials or their technical advisors) the exact chemical structure concerned. The CAS Index name does all that but registration may well have to precede publication. Moreover, the CAS name is likely to take the form of a parent name, followed by periphrases describing its modifications. Such index names can readily be de-inverted so as to start at the beginning and finish at the end. Alongside structure (**1.7**) name (i) illustrates the CAS style and (ii) is an acceptable IUPAC name for it. Comparison of these two names shows that the de-inversion process is not invariably straightforward: name (i) uses the '-one' suffix whilst name (ii) expresses the =O group by a prefix. The locant-numbering in each case is appropriate to the name-style, i.e. different. Why this is so will be explained later.

During testing and development, the new compound may acquire one or more tentative synonyms before being presented as a marketable chemical substance. Less than 1% of substances reported in the Chemical literature reach this stage; many are reported and seldom if ever heard of again. For a new chemical of possible commercial interest, the manufacturer may suggest a convenient short name to replace the systematic one, which may well be long and cumbersome. In the fields of pesticides



(i) morpholin–3–one, 6–hydroxy, acetate ester

(ii) 5–oxomorpholin–2–yl acetate

(1.7)

and pharmaceuticals, there are both national and international committees who examine them with a view to avoiding possible conflicts whilst conveying some clue as to their class, e.g. fungicide or herbicide in the case of a pesticide, perhaps analgesic or anti-inflammatory in the case of a drug. In the UK the status of British Approved Name (BAN) is given after such vetting and possibly discussion with the manufacturer; after similar appraisal by the US Adopted Name Council, the status of USAN is conferred. The World Health Organization in Geneva considers all such applications and, after consultation in an international committee, awards the title of International Non-proprietary Name (INN) to the name it finally decides on. This is often the same as, or very similar to, the national submission and may even be what the manufacturer proposed, though sometimes modifications or even complete replacements are made.

Similar procedures apply to pesticides, the International Organization for Standardization assigning to its approved names the title (ISO). The ISO and the INN names are legally acceptable in any of the member states, the orthography and spelling of the official WHO languages being imposed where appropriate, e.g. languages having no 'y' can use 'i' instead; those not using 'ph' can substitute 'f'. For INNs Latin is also given and the nearest Russian equivalent conveyed by the Cyrillic alphabet.

Now the substance is made up into formulations, possibly by various manufacturers and each gives a trade-name to his formulation. This is the name by which the active ingredient is most widely known, just as an actor or actress often wins fame under a concocted stage name. Their official name stays on their birth certificate and the systematic name for the chemical will often be supplied in tiny print on the package. It need cause no confusion that one substance may have several names and designations; the important thing is that the name used should be appropriate to the context. Thus, when a truck-load is spilled on the highway and the emergency services ask its identity, they do not expect or need a five-line systematic name kicking off with a set of numbers and stereo-designators, decorated with Greek letters and superscripts. They are more concerned to see an appropriate HAZCHEM code symbol rather than a name or to know that none applies to the liquid oozing towards them across the tarmac.

On the other hand, the paramedic treating a poisoning needs the active ingredient to be identified and the trade-name may not be sufficient. Here is where the ISO or the INN is appropriate. Neither of these conveys much about the structure to a research chemist; he needs to see a systematic name and to be able to translate it into a correct structural diagram, whether by himself or using a suitable computer program.

Every 15 years or so, someone goes on record to hail the end of chemical nomenclature, usually in reaction against such discrepancies as that

shown beside structure (**1.7**) or due to changes in the rules, rare though these are. The mid-century devising of line notations was seen as a liberation from all the complexity of systematic naming. Unfortunately, the Wiswesser Line Notation [3] – the one which most widely caught on – requires mastery of a set of encoding and decoding rules filling well over 200 pages of an instruction manual. The system is easy to apply in the sense that playing the violin is easy; in other words the required facility is achieved only by constant practice. WLN is supplied for the entries in the Pesticide Manual [4] (along with much other useful information) but it tends to be practised by a hard core of devotees rather than adopted, as once hoped, by chemists world-wide as a convenient means of communication. It is not likely ever to be adopted in texts for legislative purposes in the way that names are currently used.

Later, the widespread use of computers was thought likely to remove the need for systematic chemical nomenclature as increasingly sophisticated and powerful programs for structure-handling were developed. These new techniques undoubtedly have a valuable place in the world of communication in Chemistry but there still comes a point at which a structure needs to have a name assigned to it, and what is obtainable from a computer depends, as we are constantly reminded, upon what has been fed in.

The Beilstein Institute has developed and marketed a naming program called AUTONOM after about 10 years of work. This addresses an input chemical structure and issues a systematic name for it in the majority of cases. Where its program fails to cope, it acknowledges the fact rather than trying to generate something rather than nothing. This is a useful work and capable of improvement but it should be noted that the 10 or 15% of cases for which it fails may well be considered inadequate for certain purposes.

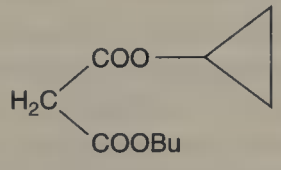
Moreover its output, although based on an IUPAC philosophy of name formation, consists of names couched in what must be called the Beilstein 'dialect'. To clarify this remark, I should explain that the IUPAC Commission on the Nomenclature of Organic Chemistry has, through a series of publications on which they have worked during the latter half of the 20th century, sought to codify what is acceptable and to recommend against deplorable usages. Thus, except for certain special areas where the IUPAC rules are specific, such as procedures for fusion names, use of stereodescriptors, and special rules for naming certain classes of natural product (e.g. carbohydrates and steroids), their rules offer a choice of acceptable naming style.

For example, butyl(ethyl)amine and *N*-ethylbutan-l-amine are both acceptable IUPAC names. Such choice is accommodated in neither the CAS Names Index nor in the Beilstein rationales. Each of these agencies has been obliged to construct in-house rationales based on the IUPAC

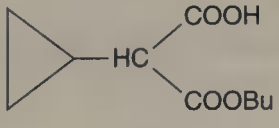
philosophy but limiting the choice to one name per structure. Their names often agree in all but presentation but sometimes they have followed divergent paths in making their choice-rationale.

Another point to be made about Beilstein is the somewhat special use of punctuation in their names. This does not sound like a major point but the effects can be powerful when a comma or space is or is not used. For example, in English names esters are named in two parts separated by a space. Structures (**1.8**), (**1.9**) and (**1.10**) have names differing only in the position of the spacing. (The names for (**1.9**) and (**1.10**) should have 'hydrogen' inserted as a middle word but such refinements have been known to be omitted and the resulting names would be likely to be interpreted as shown.) This example illustrates the potential significance of punctuation in terms of chemical structure; the effect of overlooking the comma in '1,2-' having been already noted (the absence of a 12-position in propane helped in locating the error).

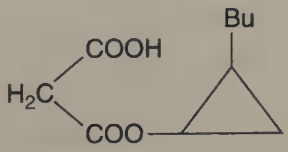
A suitable name for all occasions is advantageous but a multiplicity of synonyms can, as I said, be taken too far. The European Commission set out in the early 1980s to construct a register of all chemicals of practical importance to the European Communities. The intention was to list all the synonyms actually used for each chemical. Unfortunately it tried to be too all-embracing and the trawl included some fish which should have been thrown back, including some very parochial names and some downright errors. This proved self-defeating and the position was exemplified



butyl cyclopropyl malonate
(1.8)



butyl cyclopropylmalonate
(1.9)

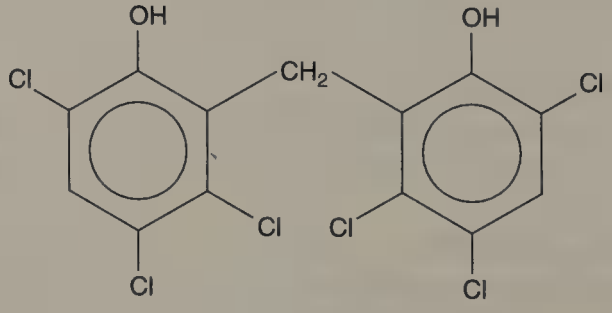


butylcyclopropyl malonate
(1.10)

by the simple trichloroethane, for which over 30 different names were collected.

Structure (**1.11**) is an anthelmintic (fatal to flukes) also used in the manufacture of germicidal soaps. The following assortment of names has been used for it. In addition to the registered trademarks G-11, pHisoHex, Gamophen and Surgi-cen, it is also known as AT-7, Dial, Surofene, Hexosan, Bilevon, Dermadex, Exofene and Hexachlorophane. Hexachlorophene (INN) is the name given to it by the WHO and that is internationally agreed and accepted. In addition to these 13 names the following have also been applied: (14) Bis(3,5,6-trichloro-2-hydroxyphenyl)methane, (15) Dihydroxyhexachlorodiphenylmethane, (16) 2,2′-dihydroxy-3,5,6,3′,5′,6′-hexachlorodiphenylmethane, (17) 3,3′,4,4′,6,6′-hexachloro-2,2′-methylenediphenol and (18) 2,2′-methylenebis(3,4,6-trichlorophenol). These last five differ in kind from the first 13. Obviously, they are longer, but the more important distinction between these two groups is that the short names convey no information about chemical structure. They are referred to as 'trivial' names. Names (14)–(18) are made up of fragments (such as di, tri, hexa, chloro, hydroxy, meth, ane, ol) which have a structural significance which survives unchanged when they are used in any other such name. They are like building blocks which go to make up the edifice of systematic chemical nomenclature and names (14)–(18) are referred to as **systematic** names. The idea is that rudiments of the system are taught in Chemistry classes at school and learning relatively few of these name-fragments enables the student to construct complete names for a given structure. Correct application of the rules of the nomenclature system permits whole classes of compound to be named in a consistent and meaningful way. This is contrasted with the alternative of learning individual trivial names for a vast and growing number of compounds, such names being unlikely to indicate kinship within classes.

That, in a very bald and over-simplified form, is the case for systematic nomenclature and against most trivial approaches to naming. In example (**1.11**) none of the first 13 names conveys any hint of the structure, though



(1.11)

each will doubtless have its devotees among those who know it already; coterie-names are handy familiars within a parochial group but not very communicative outside it. Those having 'hex' in the name suggest six of something; 'Dial' misleadingly suggests a dialdehyde. The two beginning 'Hexachloro' would qualify for the systematic group but for their final syllable: 'phane' is an ending conveying a structural system of chains linked to ring systems, for which (**1.11**) is too simple to qualify, while 'phene' by itself has no meaning.

Names (14)–(16) exemplify an attempt launched over 50 years ago to base substitutive names on parent hydrocarbons. Name (14) uses the central methane as parent while (15) and (16) are based on 'diphenylmethane' (so setting up a double-tiered substitution), the difference between them being that (15) cannot be bothered to indicate where the Cl atoms are. Names (17) and (18) attempt to apply the IUPAC published Recommendations which have been extant during the latter half of this century. There is present in the structure a functional group expressible as a suffix in the name (-OH: '-ol'). When attached to a benzene ring this hydroxy compound is known as 'phenol' (not benzeneol or even benzenol). As it occurs twice and the linking group $-CH_2-$ is symmetrical, the two can be collected in the ending '. . .-diphenol'. As the central divalent group is named 'methylene', the formation 'methylenediphenol' is appropriate. The -OH site is given the number 1 on one ring and 1′ on the other and the linkage point on each ring is given as low a number as possible after that assignment, i.e. 2 and 2′. Numbering for the Cl sites is obtained by proceeding round each ring sequentially, hence 3, 4, 6 and 3′, 4′, 6′. The only difference is whether the chloro substituents should be collected inside the parentheses along with their parent 'phenol' as in name (18), or put at the beginning as in name (17). Opinions differ on this but the IUPAC 'Blue-Book' Rule C-73.1 clearly prefers name (17). This rule is more general since it enables collection of functional parent structures in the suffix even when they are differently substituted, so making this convenient style available in more cases. Some still contend that, when the pattern of substitution is identical, the whole parcel should be taken inside the parentheses.

If *o*-cresol were available as a substitutive parent name without restriction (as it was until the publication of the 1993 'Guide to IUPAC Nomenclature of Organic Compounds', which restricts use of this parent name to derivatives having ring-substitution only [1b], we could have added a 19th synonym, i.e. (19) 3,4,6-trichloro-α-(2,3,5-trichloro-6-hydroxyphenyl)-*o*-cresol as an allowed IUPAC name. Names (14)–(19) all provide the right structure but only name (17) is currently recommended under the IUPAC Rules. Incidentally, structure (**1.11**) would perplex the AUTONOM program, certainly in its initial form. Whether phenol or *o*-cresol were identified as the senior parent available in its reference library

it would, due to the symmetry, be unable to choose which one to go for. As they are equally valid according to the hierarchical principles imposed, it would choose one and express the other moiety as a prefix, as in name (19), so failing to generate the agreed IUPAC preference. At present IUPAC and CAS use differing criteria in deciding when the style of name (18) is applicable, the IUPAC rule requiring symmetry in the central divalent (or polyvalent) radical, while CAS allows certain relaxation in this respect but restricts its use to identically substituted parent units. Doubtless, the AUTONOM programmers are waiting for a resolution of this divergence. It may not be long in coming as the often heard plea for '**the** IUPAC name' for a new substance may be met in the near future. As pointed out above, IUPAC rules often provide a choice of names. A name or procedure may be recommended or even preferred; on the other hand, it may be allowed as non-preferred or deprecated. Occasionally something is condemned as erroneous or misleading and it is declared as abandoned. In between, alternatives may be offered without any indication of preference. The IUPAC Rules are the only internationally promulgated recommendations for chemical nomenclature but no power exists for their enforcement beyond the fact that some journal editors (but by no means all) make efforts to ensure IUPAC usage in the papers they publish, and use of non-IUPAC-approved names in submissions to such bodies as the European Commission may well cause expensive inconvenience and delay to be imposed on the perpetrator. The case for certain non-preferred IUPAC names, which parallels that for many of the special index-name procedures of the CAS, is that they can offer advantages for the particular purpose in hand. For example, a study of the variation of a certain property over a range of halogen-substituted diphenylmethanes might cause name (16) to be preferred for structure (**1.11**) so that each compound studied will be named in a consistent way and differ only in the appropriate prefixes. If a cyano group were to be included in the study, it would not help matters to change the form of the name to end in '. . . benzonitrile', as both the CAS and IUPAC Rules require. So long as the name gives the correct structure unambiguously, it has some validity. Formal errors and omissions such as incorrect use of a hyphen can be categorized as ungrammatical but they need not matter too much. Name (15) omits all the locants and is thus too unspecific to be acceptable. All possible positional isomers consistent with this name are covered, as are all conceivable intermixtures of them. Such names are of possible use only when such vagueness is intentional and even then they are unlikely to serve adequately without explanatory periphrases. Thus, name (15) has very little utility although it is in a systematic form. Databases world-wide tend to be cluttered with such rather useless synonyms because they have obtained publication somewhere, somehow. For example, in the hypothetical comparative study of halogenated diphenylmethanes postulated above, the non-preferred

systematic name used for convenience there would probably be distinctly inconvenient outside it and to copy it uncritically on to a general-use database helps hardly anyone.

Perhaps even more irritating is the kind of variation illustrated by the different citations for the Cl-atom locants used in names (16) and (17). As it happens, these two names are different but the order variation in these six locants might well be the only difference in otherwise identical names. It is in issuing a recommendation for world-wide acceptance to settle just such an issue that IUPAC would hope to avoid such tedious waste of space and effort. In such cases, let us all adopt one and drop the other.

In asking for 'the IUPAC name', chemical manufacturers and legislators citing chemicals want all such issues to be settled. They need to be supplied with or instructed how to derive for themselves a name that will be not only internationally acceptable but also unchallengeably authoritative. As already explained, this is something of an abstraction, but the IUPAC Commission on Organic Nomenclature (CNOC) is currently hard at work on the production of just such an instruction-set to complement the 1993 'Guide', which sought to improve the earlier 'Blue Book' [5] presentation while taking on board certain perceived trends in usage since 1979. (The separate generation of the various Blue Book sections, e.g. Section A (Hydrocarbons), Section B (Fundamental Heterocyclic Systems), Section C (Characteristic Groups Containing C, H, O, N, X, S, Se and/or Te), had inevitably led to certain aspects being dispersed due to historical accident. For example, it is a little startling to find the instructions for citing locants for substituents on the CH_3 groups of xylene in a footnote to Section C-511 entitled 'Thiols' rather than under Section A-12.1 'Substituted Aromatic Compounds', as might have been expected. Doubtless, the xylene example occurred in the writing of the Thiol section and the name needed explanation.)

The current situation on the General Rules for Nomenclature of Organic Chemistry is that the 1993 'Guide' supersedes some of the rules of Sections A, B and C of the Blue Book (1979) but refers to others as supplying a more detailed treatment. For those that want 'the IUPAC name' the work in preparation is designed to effect selection of a unique result from among the various naming possibilities currently offered. To quote from the Preamble to the 'Guide': ". . . formulating rules not having general support is a futile exercise; such rules will be widely ignored". This sentiment reflects an accumulation of experience. IUPAC Nomenclature Commissions usually 'test the water' by submitting drafts of their proposed recommendations to a set of referees including specialist researchers in the field concerned and editors of appropriate journals. Apart from matters of detail which are thereby aired and re-examined, the general reaction usually encourages the Commission to proceed

towards publication of the mature document. On a few occasions, however, the strength of unfavourable reaction has brought about a change of heart. This happened with the first set of CNOC proposals for systematic names for hetero-monocycles; the revised document [6] represents the results of the subsequent reconsideration.

The recent furore over the set of names initially proposed for elements 101-109 by IUPAC's Commission on Nomenclature of Inorganic Chemistry (CNIC) and the consequent reconsideration resulting in a revised name-set affords another example. In this connection it is only fair to point out that CNIC never pretended to any authority or indeed to any standing as arbitrating between the rival claims deriving from experimental work by the research schools concerned. In trying to confine their attention to nomenclatural considerations on the basis of all the correspondence sent in, the Commission was faced with something like the Judgement of Paris, as it proved impossible to assign a name for an element without implying (however unintentionally) some value judgement about the manner of its discovery. The situation was further complicated by the fact that the only bodies competent to make a critical evaluation of the various submissions were the rival schools themselves. Whatever was to be decided, it was unlikely to meet much difficulty in offending all three; so it proved.

Another inorganic topic that stimulated lively reactions from the user-public was the Periodic Table. When the Red Book (Recommendations 1990) [7] was in preparation, there was a clear need for a Periodic Table to be placed conveniently for reference inside the front cover so that citation could be made by group-number within the text of the book. Discrepancies in the designation of the A and B sub-groups across the world had already been noted prior to the 1970 Definitive Inorganic Rules (Red Book 2nd edition) [8], in section 1.21 of which these were assigned for 1A–7A and 1B–7B, as exemplified by 1A: K, Rb, Cs and 1B: Cu, Ag, Au. The reverse usage was predominant in the Americas and those who had used 1A for Cu, Ag and Au continued to do so, papers continuing to be published with such references as 'the elements of Group 3A' with nowhere any indication as to whether Sc, Y and La or Ga, In and Tl were intended. This situation was anathema to IUPAC as it impeded and frustrated international communication. Accordingly CNIC proposed that a 'long' form for the Reference Table of the Elements should be used and all further reference in its rules to A and B sub-groups abandoned. This drew indignant protests from both sides of the Atlantic, mainly from outraged teachers who claimed to have used the 8-column Table with unvarying success and freedom from confusion throughout a long and distinguished teaching career serving many generations of Chemistry students. The texts of letters from both sets of complainants were virtually identical, save only that the letters 'A' and 'B' for any given element

were interchanged. The recommended form of Periodic Table afforded another source of divergence, the number of Tables being almost as great as that of the correspondents. A letter from the USSR deplored the fact that CNIC/IUPAC had not preferred the original form of the Periodic Table due to Dimitri Ivanovitch Mendeleev. It was interesting to learn that his first published Table took a 'long' form. The Appendix to the 1990 Red Book contains as representatives of the various possibilities, a 'short form', an 18-column form and a 32-column form.

These episodes of lively interaction between CNIC and sections of its user-public are not typical; more normal is the general absence of feedback on its proposals prepared for publication. On the criterion of sustained usage, which is the converse of the quotation from the 'Guide' preamble, the record of achievement of CNIC is best assessed on the strength of such widely applied recommendations as those of the 'Red Book' Part I (1990) [7], covering General Inorganic Chemistry, including Coordination Compounds, Stereochemistry and Boron Hydride clusters. More specialized topics are covered in Part II, soon to go to press.

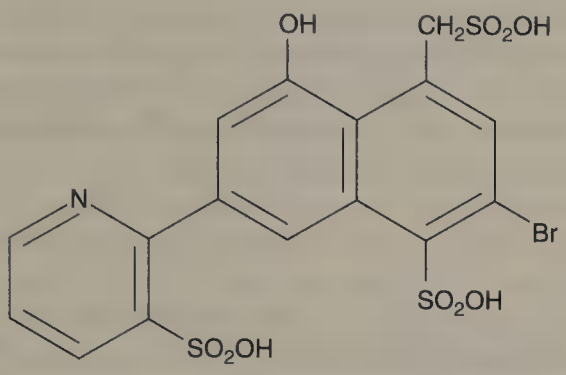
A glance at the Chemistry journals now and for the past twenty or so years will readily show that most modern Chemistry research activity takes place in organo-inorganic chemistry. ('Organometallic' used to mean what it says but organo-boron compounds have invaded that domain and its one-time significance has begun to widen into imprecision and eventual uselessness.) It is a little ironic that this vast and growing area of activity has to be served by a sort of shotgun wedding between the two philosophies of organic and inorganic nomenclature. These developed separately for the usual historical reasons. Chemical nomenclature has always been devised to cope with the practical knowledge and the theoretical ideas on the subject as they were acquired or developed. This is why names still survive from the days when Chemistry was in a fairly primitive state. In the absence of any coherent theoretical structure, chemicals were named according to source, e.g. Spirits of Hartshorn (ammonia), or from a property, e.g. Green Vitriol (iron(II) sulfate – a green glassy solid), or personalized by its discoverer, e.g. Glauber's Salt (sodium sulfate decahydrate – Johann Rudolph Glauber (1604–1670) obtained this as a residue from HCl formation and he called it '*Sal Mirabilis*' because he thought it cured just about everything).

'Glauber's Salt' is still used for the decahydrate – also still used as a purgative, but I doubt whether 'Spirits of Hartshorn' enjoys much currency any more. Some survivors are surprising; the British Standard BS 2474 *Recommendations for Names for Chemicals used in Industry* [9] is (or at any rate was) renewed about every decade and a comparison of successive editions provides a convenient insight into how usage develops. The preferred name is printed in bold-face, the tolerated synonyms in ordinary, obsolescent names are in parentheses and deprecated names in

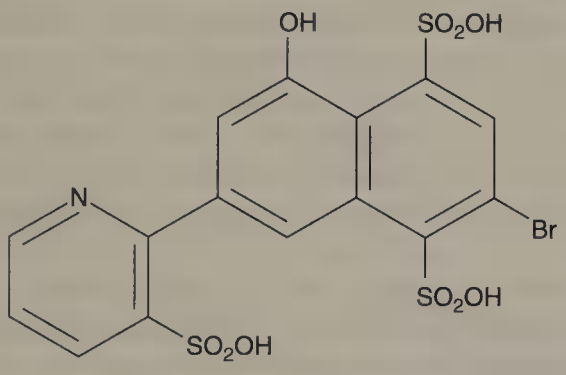
square brackets. Thus, it is possible to follow the ageing of a name from bold-face to square brackets before it is dropped altogether. As I remember, a post-war edition still had as a synonym for $CC1_4$ the name 'Tetramuriate of Coke', which meant that it was still in use somewhere.

August Wilhelm von Hofmann remarked on the similarity between various compounds formed when one or more H atoms of ammonia was replaced by organic groups. In 1850, he put forward the idea of chemical classes although his friends begged him not to risk his reputation with such an outlandish idea. It still underpins accepted approaches to general systematic organic chemical nomenclature. This approach based on chemical class uses the name endings acid, -oate, -amide, -nitrile, -one, -ol, among others, all set forth in an order of preference for naming. It has to be reconciled with structural factors and with such operational concepts as hydrogenation and dehydrogenation. Thus, the IUPAC 'Blue Book' rules are something of a patchwork quilt. It is intended to cover the entire bed of Organic Chemistry although some parts are covered more than once. The most generally useful of its methods is the so-called Substitutive Nomenclature, which expresses a molecule in terms of a parent hydride, be it an unbranched chain of one or more atoms or a ring-system of one or more rings. The unit bearing the group expressible as a name-ending from the aforementioned seniority-list in the greatest number is chosen as stem, the senior group(s) considered to replace H atoms, and each other group present is expressed in alphabetical order as a prefix.

Thus, in structure (**1.12**) the contest between methanesulfonic acid, naphthalenesulfonic acid and pyridinesulfonic acid is won by the last of the three because $-SO_2OH$ is the senior group present and each structure unit bears only one. In structure (**1.13**) the loss of the $-CH_2$ -group means that the two on the naphthalene win over the one on the pyridine and the two names accordingly differ. This may seem deplorably inconsistent but the simpler the system the more numerous its inconsistencies; the more exceptions it accommodates the greater the complexity of its general application. This substitutive system is basic to the CAS Names Index entries and the Beilstein AUTONOM Program. It deals with monomeric carbon-based structures without mentioning the name for the skeletal element carbon; it is assumed. Thus 'hexane' means an unbranched chain of six C-atoms, each joined to its neighbour by a single covalent C–C bond and the valency of carbon is assumed to be 4 throughout. Thus, the population of H atoms can be deduced and this never finds mention in the name either. Hexa-2,4-diene is $H_3C\text{-}CH{=}CH\text{-}CH{=}CH\text{-}CH_3$ whichever end we number from. In structures (**1.12**) and (**1.13**) the names naphthalene and pyridine convey the ring systems shown with the maximum number of non-cumulative double bonds (this means no C atom has two double bonds attached to it; in allenes the double bonds are cumulative). Thus the '-ane' ending conveys C–C bonding analogous to that in



2-[7-bromo-4-hydroxy-8-sulfo-5-(sulfomethyl)-2-naphthyl]pyridine-3-sulfonic acid
(1.12)



2-bromo-5-hydroxy-7-(3-sulfo-2-pyridyl)naphthalene-1,4-disulfonic acid
(1.13)

diamond, but modified by unsaturation when the endings are '-ene', '-diene', '-yne' and the like. The carbocyclic and heterocyclic systems are all distinctively named and convey known states of bonding whether saturated (e.g. tricyclo[6.4.2]tetradecane), aromatic (e.g. naphthalene, pyrene, calling to mind fragments chipped from a graphite sheet), or partly hydrogenated in an organized way (e.g. indane).

Inorganic nomenclature developed along different lines. It began in Alchemy with names identifying substances from mineral sources and remained preoccupied with such considerations as winning of metals from their ores. There were no notions of bonding or molecular structure at first, but as the acid–base–salt relationships came to be better understood, names were formed accordingly. After the work of Lavoisier and Priestley in identifying oxygen and its acceptance as an element, the Phlogiston Theory came to be supplanted by more rational ideas of bonding; possibly helped by a tendency to choke on the nomenclature generated by Phlogistonists. Thereafter, it usually sufficed to identify an inorganic substance in terms of its constituent elements with an indication of their proportions. Names such as trilead tetraoxide and chromium sesquioxide described the proportions

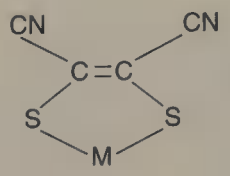
of the named elements in a 3D-crystal array as well as in the molten state, but they did not provide information on the spatial arrangement in the solid. When a structural name is required for an inorganic compound, the built-in assumptions for carbon and hydrogen in organic names are not available. Instead the entire range of chemical elements is possible and every atom present has to be accounted for in the name. Thus, for a case such as Me_3SnH, the quasi-organic name trimethylstannane is based on the parent hydride SnH_4 but some inorganic chemists blench at the '-ane' ending for cases where there is no H atom attached to the metal, e.g. tetraphenylplumbane. The inorganic approach cites all the ligands on the central metal and would name Me_3SnH as hydridotrimethanidotin, although this is commonly relaxed to hydridotrimethyltin as the alkyl group names are retained for simple anionic organic ligands. Faced with the wider range of atomic association, the assignment of integral bond orders (single, double or even triple) can no longer be automatically assigned for nomenclature purposes. Bond character may also range from covalent to ionic (although this is not unknown in certain organic compounds, their nomenclature recognizes only the extreme states; the only kind of organic name applicable to intermediate bonding states would be a trivial name, such as sydnone, for which various canonical forms can be written).

In inorganic structural names, the concern is not so much with classical valency concepts (which were in any case clearly inadequate to cover the coordination complexes identified by Werner) as with connectivity between near neighbours, and charge counting. During the first half of the 20th century the common compounds of elements exhibiting mainly two valency states were generally described by means of the '-ous' and '-ic' suffixes to signify the lower and higher states, respectively. Cuprous/cupric, ferrous/ferric, mercurous/mercuric, etc. had their day but they were inadequate to cover all the various oxoacids of non-metals, which had to be supplemented by such prefixes as meta, ortho, pyro and hypo. Elements such as vanadium, rhenium and osmium exhibited so many oxidation states that '-ous' and '-ic' were clearly inadequate and so the modern system was generally adopted of placing a charge symbol or the so-called Stock number, which is an upper-case Roman numeral inside parentheses immediately after the name of the central element to signify the oxidation state, e.g. tin(IV). (This resembles the fate of the London 3-letter telephone exchanges, which had to go when increased demand could no longer be met by pronounceable new ones.) Names for organo-inorganic structures tend as far as possible to use organic naming for the organic parts and inorganic (i.e. Coordination) methods for the inorganic parts. This has worked well enough by and large, but not without the devising of some extra symbols conveying structural information. Structure (**1.14**) is named *trans*-[μ-(1,2,3,4-η-:5,6,7,8-η-cycloocta-1,3,5,7-tetraene)]bis(tricarbonyliron). The μ-symbol indicates a bridging group between the two

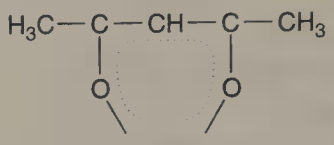
(1.14)

Fe atoms. This is a carbocyclic ring named as such, but its conjugated double bonds are given locants corresponding to their position before delocalization for the ligation to the metal atom in two separate zones, as indicated by the preceding sets of hapto (η) locants, their distinctness being conveyed by the colon. The '*trans*' puts the Fe atoms on opposite sides of the ring. The CO groups are neutral ligands as usual.

In the case of structure (**1.15**) the bidentate ligand has been named 2,3-dimercaptomaleonitrilato-κ^2S,S'. This exemplifies the derivation of ligand names from their parent reagent, in this case 2,3-dimercaptomaleonitrile (IUPAC considers mercapto somewhat archaic and now recommends sulfanyl instead, but that is by the way). Names for simple anionic organic ligands can readily be derived from that of the ligand precursor by changing the anion name-ending from -ide to -ido, -ite to -ito or -ate to -ato. This works when the ligating group is also that referred to in the name-ending, e.g. oxalate → oxalato. In the case of structure (**1.15**) the nitrile groups are remote from the ligation sites and the name misleads by referring to -SH groups which are no longer present. This, in my opinion, is not good nomenclature – even though we are rescued by the use of the kappa symbol indicating specific connectivity, so cancelling the false impression given by the 'dimercapto' (kappa is used to remove possible ambiguity; here it corrects a falsity). It is an important principle of systematic nomenclature that it should indicate the actual structure and not something else. Application of that principle here would name the ligand from the ligating anion rather than its neutral precursor; after all, the hierarchy of organic suffix groups places anion at the head of the list, well above nitrile. Here, the anion is 1,2-dicyanoethene-1,2-dithiolate and a change of its final 'e' to 'o' would result in a ligand name free from false implications and dispensing with the kappa symbol. This procedure might sometimes give longer names and in some such cases the existing method has become firmly established, most users making the required molecular rearrangement mentally (shades of 'nitroglycerine'). For example 'pentane-2,4-dionato' is the established name for the ligand (**1.16**). A structural name for this or even for one canonical enol form would now be regarded as unacceptably cumbersome. However, an anion name might well come into play when the routine -ato conversion gives misleading results.



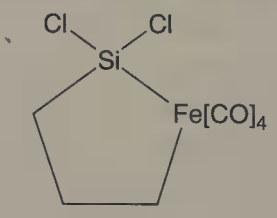
(1.15)



(1.16)

It is in such areas as metallacycles that the combined organic/inorganic approach runs into difficulties. Structure (**1.17**) for example is not readily nameable by substitutive organic methods because the concept of a parent hydride (in this case the ring) would require a 'normal bonding number' for the Fe atom in order to establish how many H atoms it would carry in the unsubstituted parent. As the carbonyls are neutral ligands, they cannot be treated as substituents and so thought has to be given as to how best to reconcile the substitutive process for the tetracovalent Si position with a possible Coordination approach to the situation on the Fe atom.

The foregoing ramble around the rather broad landscape of chemical nomenclature has, I hope, served to indicate in a general way how the naming of chemicals is intended as an aid to communication as well as to teaching and learning about Chemistry. Essentially, it is the language of Chemistry, its vocabulary, grammar, syntax and idiom. If systematic nomenclature has been emphasized at the expense of trivial names, it is because these tend to have a limited acceptance outside their original parish, as the shorter synonyms for structure (**1.11**) show. Trivial names having the status of INN or ISO are carefully tailor-made for their field of use and are internationally accepted. That puts them on a very different footing from other trivial names, only a few of which have achieved similar status under their own steam (e.g. aspirin and saccharin), and those which have been accepted into the systematic area because they were too well established to be replaced, e.g. aniline, naphthalene, butane, quinoline.



(1.17)

They go to make up so-called semi-systematic names such as 2,4-dichloroaniline, 2-naphthol, but-2-ene and quinolin-4(1*H*)-one. Contrasting these with, say, bicyclo[6.2.0]decan-2-one brings home the fact that not all names are equally systematic in the sense discussed below structure (**1.11**). In fact there was some controversy about 15 years ago concerning use of the terms 'trivial' and 'systematic' to describe types of chemical name. It was objected that 'trivial' bore the connotation of insignificant or trifling and was therefore not suitable for certain non-structural names of widespread importance. Some dictionaries do give 'unimportant' as the primary meaning, 'commonplace' as the secondary and then they consider specialized meanings in mathematics, biotaxonomy and chemistry. Older dictionaries first give the term 'trivial' its original meaning of the three subsidiary subjects of the mediaeval university, then cover the special meanings, including that given below structure (**1.11**), and lastly mention 'commonplace' and 'unimportant'. At the same time, the word 'systematic' was criticized on the grounds that it did not convey accurately the principle stated earlier that such names were made up of transferable fragments, each preserving its structural significance in any context. The counter-argument advanced was the, to my mind, unhelpful contention that any name forming part of a nomenclature system was systematic. On that ground benzene is a 'systematic' name and the distinction becomes blurred.

IUPAC were sufficiently moved by these representations to issue an edict recommending that, as both terms were tainted, they should not be used in published IUPAC documents on nomenclature. Unfortunately, they omitted to suggest any replacement terms; there matters have rested. At that time I suggested the word 'acribic' instead of systematic and 'anacribic' instead of trivial because it is difficult to discuss name construction formally without some appropriate terminology. The Greeks, as usual, had a word for it: ακριβειν means 'to give exact details of'. However, these terms have never caught on; probably nomenclaturists are not sufficiently numerous to constitute a significant user group. However, this is not an IUPAC document and I have explained what I mean by 'systematic name' in terms of a meaningful distinction. Nevertheless, it must be admitted that there are degrees of systematicity. The tendency for valeric acid to be replaced by pentanoic acid exemplifies a steady process of discarding trivial names. The stems 'meth', 'eth', 'prop' and 'but' survive but will the next step be to replace butane by tetrane (or quadrane)? As more generalized approaches come to be used in this computer age, will exceptional treatment for carbon compounds be abandoned? This would result in such names as tetracarbane for butane and cyclopentacarbane-1,3-diene instead of cyclopentadiene. These names are more systematic but, at present anyway, it does not seem that they are more useful than those currently accepted.

Name assignment has always tried to meet the changing needs of chemistry and continues to do so. Even if a beautifully simple and appealing new system were devised and promulgated now, it would not result in immediate abandonment of existing procedures nor even a gradual supplanting of them. An attempt was made to re-think the whole approach to systematic nomenclature following the discernment by the author of a lack of logicality in the patchwork of revision and extension to organic nomenclature over the past 200 years. This is the so-called HIRN system [10], the principles for which were published in 1984. Its author claims it to be simple, logically consistent and to use far fewer rules than those of IUPAC. So far, it has not won many converts and this characterizes the general attitude to chemical nomenclature, even among chemists, i.e. almost total lack of interest. Apart from a few fanatics such as myself who, somewhat like crossword addicts (I plead guilty again), cannot drop the habit of dabbling in systematic name construction, the great majority of chemists shun involvement in nomenclature unless it is unavoidable. Only then are they driven to consult the IUPAC Rules or perhaps the CAS Index Guide [11] in the hope of finding relevant instructions for naming their new compound (or if they can fund it, consult the LGC Chemical Nomenclature Advisory Service (CNAS)). It probably comes as a shock to discover that the rules do not constitute a beginner's instruction manual but are couched in terse language and require months if not years of study and practice before a fairly complicated molecule can be tackled with confidence. A commonplace reaction is to riffle through the Blue Book looking for an example that looks fairly close to the molecule in question and try to name it by analogy. That is the raison d'être for such undertakings as the AUTONOM program. I tried to meet certain aspects of this demand by writing a stand-alone instruction manual [12] on naming new organic compounds. It took the form of a dichotomous tree, at whose outermost branches, reached by a pathway determined by YES-or-NO answers to questions on characteristics of the input structure, were directions to an appropriate section where specific naming instructions were to be found. Initial trials on user panels gave promising results, the only failures being due to certain subjects inserting bits of their own knowledge instead of blindly following the instructions.

Perhaps one day it will be possible to machine-read a more comprehensive version of such an approach and generate 'the IUPAC name' from an input structure without human intervention. Full retirement might then become a realistic prospect for me, but some hurdles remain to be crossed first.

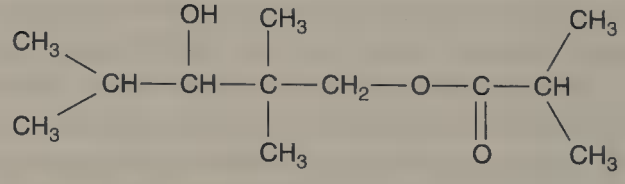
What next? Without doubt, the most challenging event to confront the intellectual apparatus of systematic nomenclature is the identification of [60]fullerene and its analogues. In the initial phase following this event there was a feeling that this was chiefly significant for being another

allotropic form of carbon. Its implications were thought to lie mainly in Materials Science and Astrophysics; it was part of the inorganic realm and the chemistry seemed severely limited. This is no longer the case; the proliferation of derivative compounds in the intervening few years has been explosive. Every few months a new structural class is reported and systematic or semi-systematic nomenclature will be needed for a growing range of bowls, cylinders – open or closed at one or both ends, hemi-cylinders, tauroidal shapes and doubtless many others, together with combinations of any or all of these and their associative combinations with existing systems, together with an unlimited range of structural, atomic replacement and substitutive derivatives. Organic chemistry is in the process of inflating rapidly and nomenclature will have to face some novel challenges. Even at this comparatively early stage, devising systematic names for fullerene derivatives posed unprecedented problems, not the least of which was numbering their atomic sites in a logically based continuous sequence. A whole range of problems looms in the designation of various stereochemical isomers, not only those already confronted in designating enantiomers of some of the larger fullerenes, but in the proliferation of forms such as dumb-bells and other possible fullerene-with-ring-and/or-chain combinations. Fullerenes, as they stand, have an inside and an outside but no edges. It seemed logical to use existing procedures for name formation where they could be applied, rather than devise unnecessary new ones. However, all fullerene bonds resemble the central bond of indene rather than its peripheral bonds and the normal methods of generating fusion names could not be applied using existing rules. For such compounds the alternative of devising bridging names also posed problems for more complicated cases and so a special fusion procedure was devised by the CAS. This, in some respects, resembles the special procedure for fusion to steroidal components previously published in the Biochemical Nomenclature Compendium [13]. Several reviews of bowl-shaped fullerene sub-units have been published and names such as circulene and semi-fullerene have been coined, while corannulene has been used as a fusion component (applying the conventional method) in naming similar bowl-shaped polycyclic aromatic hydrocarbons (PAHs). Coming from the other direction, a fullerene ring has been enlarged to nine atoms and it is possible to envisage a goldfish-bowl PAH receiving a trivial name that becomes accepted for use as a fusion component. If it were to accommodate fusion components at its rim as well as elsewhere and one of these was a steroid structure, we would be faced with the possibility of three different modes of fusion, each governed by a separate set of rules, all in the same molecular structure. Watch this space.

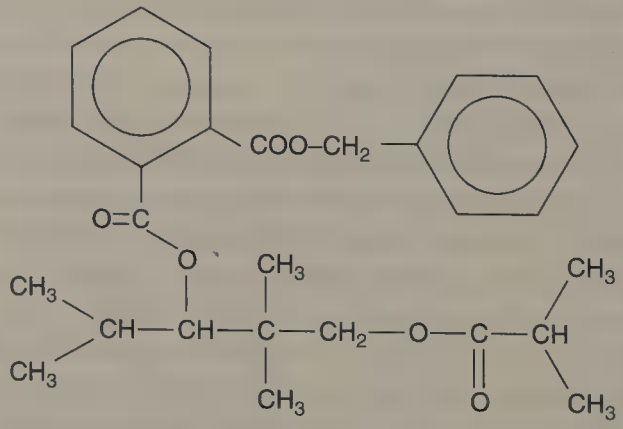
In the course of this somewhat discursive survey of a large and complex subject, I hope I have succeeded in explaining how chemical nomenclature fulfils a range of important functions in a way that numbers, codes and

pictures do not, though each can usefully complement the others. Examples of poor nomenclature should have served to show that communication failure should be avoided. This may be due to vagueness, ambiguity, omission of key information, confusion or downright error. Before handing over to the more structured presentations and the greater detail of treatment of more specialized aspects in the ensuing chapters, I will end, not by trying to define 'good nomenclature', which would be like trying to define virtue, but take the easier option of describing sin, by giving an example of bad nomenclature. The ester (**1.18**) may be given the systematic name 3-hydroxy-2,2,4-trimethylpentyl isobutyrate, but has been assigned the registered trademark (RTM) 'TEXANOL'. Structure (**1.19**) was produced by another manufacturer by esterifying the -OH group of TEXANOL to form the triester (**1.19**). Doubtless finding it troublesome to name their product from scratch, they hijacked the RTM and came up with 'Texanol benzyl phthalate'. This applies a systematic procedure to a newly coined trivial name (and an RTM at that), so interbreeding between two species to give an unholy hybrid. More legitimate and altogether more 'acribic' would have been: benzyl 3-isobutyryloxy-1-isopropyl-2,2-dimethylpropyl phthalate. Admittedly this is longer but complicated structures tend to have lengthy names if they are to be structurally descriptive.

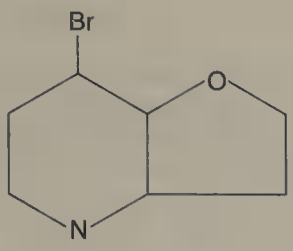
Such blatant mixing of systems is not good nomenclature. More common is using inappropriate numbering. Structure (**1.20**) might be named 7-bromoperhydrofuro[3,2-*b*]pyridine. This is a fusion name, so its



(1.18)



(1.19)



(1.20)

final numbering starts at the O atom. Alternatively, a replacement name based on the hydrocarbon parent would give 5-bromo-7-oxa-2-azabicyclo[4.3.0]nonane, for which the numbering starts at the lower bridgehead. Each name style commands its own numbering and to use the wrong one will either cause bafflement or generate an incorrect structure. That is why the two names beside structure (**1.7**) use different numberings. Morpholine assigns locants 1 and 4 to its O and N atoms, respectively, but the path from 1 to 4 can be clockwise or anti-clockwise. Name (i) gives the lower number 3 to the =O group rather than 5 as it lists the compound under morpholin-3-one to spare users the labour of searching under the densely crowded 'acetic acid' entry. Name (ii), uninfluenced by such indexing considerations, prefers to give lowest numbering to the ester-site, following IUPAC rule C-15.21 governing the numbering of cyclic radicals. The principle of 'horses for courses' is central to 'good nomenclature'. For further perpendiculars read on.

References

1. The International Union of Pure and Applied Chemistry (1993) *A Guide to IUPAC Nomenclature of Organic Compounds* Recommendations 1993, Blackwell Scientific Publications, Oxford.
1a. The International Union of Pure and Applied Chemistry (1993) *A Guide to IUPAC Nomenclature of Organic Compounds* Recommendations 1993, Blackwell Scientific Publications, Oxford, pp. 57 and 163.
1b. The International Union of Pure and Applied Chemistry (1993) *A Guide to IUPAC Nomenclature of Organic Compounds* Recommendations 1993, Blackwell Scientific Publications, Oxford, R-9.1 Appendix, Table 26 (a), p. 173.
2. IUPAC (1991) *Compendium of Macromolecular Nomenclature*, Blackwell Scientific Publications, Oxford, Ch. 5, pp. 94 and 108.
3. *Wiswesser Line Formula Chemical Notation* (1975) 3rd edn (1st edn 1954), Chemical Information Management Inc., New Jersey, USA.
4. *The Pesticide Manual* (1994) 10th edn, British Crop Protection Council, The Bath Press, Bath, UK.
5. IUPAC (1979) *IUPAC Nomenclature of Organic Chemistry Sections A,B,C,D,E,F and H*, Pergamon Press, Oxford.
6. IUPAC (1983) *Revision of the Extended Hantzsch–Widman System of Nomenclature for Heteromonocycles* Recommendations 1982. *Pure Appl. Chem.*, **55 (2)**, 409–416.
7. IUPAC (1990) *IUPAC Nomenclature of Inorganic Chemistry* Recommendations 1990, Blackwell Scientific Publications, Oxford.

8. IUPAC (1971) *IUPAC Nomenclature of Inorganic Chemistry* 2nd edn, Pergamon Press, Oxford.
9. BS 2474 (1993) *Recommendations for Names for Chemicals used in Industry*, HMSO, London.
10. Kenzo Hirayama (1984) *The HIRN System*, Nomenclature of Organic Chemistry: Principles, Maruzen, Springer-Verlag, Tokyo.
11. American Chemical Society (1973) *Naming and Indexing of Chemical Substances for CHEMICAL ABSTRACTS during the 9th Collective Period (1972–1976)* American Chemical Society, Washington, DC.
12. Godly E.W. (1989, 1995) *Naming Organic Compounds*, a Systematic Instruction Manual. Ellis Horwood, Chichester, UK, 2nd edn 1995.
13. International Union of Biochemistry and Molecular Biology (1992) The Nomenclature of Steroids, Recommendations 1989, 3S-10.2, in *Biochemical Nomenclature and Related Documents*, A Compendium, 2nd edn, International Union of Biochemistry and Molecular Biology, Portland Press, London. pp. 219–220.

2 From hydrogen to meitnerium: naming the chemical elements

P. E. CHILDS

2.1 Introduction

> The world of chemical reactions is like a stage. On which scene after scene is ceaselessly played. The actors on it are the elements.
>
> *C. Winkler*[1]

It might be thought that naming the chemical elements presents no problems and has a relatively simple history compared to other areas of chemical nomenclature. This chapter will look at the way the concept of a chemical element developed and how naming these elements moved from an idiosyncratic process, strongly influenced by history and personalities, to one regulated internationally by IUPAC. However, the naming of the elements still arouses strong nationalistic and personal feelings, as the recent controversy over naming the transuranium elements shows. This chapter will show that this state of affairs is nothing new in the history of the chemical elements.

The purpose of systematic nomenclature is to remove uncertainty and confusion, to simplify and remove unnecessary complexity and to dispel ambiguity, so that everyone (whatever their nationality) can communicate chemistry clearly and meaningfully. The naming of the elements is probably the simplest and least controversial aspect of chemical nomenclature. Naming the 112 elements presently known is also the foundation for the naming of the 15 million plus compounds known today. Until we can name the elements correctly, we cannot name any other chemical substance.

Thanks to J. J. Berzelius, chemists moved away from the confusing and irrational symbolism of alchemy, and other more complex systems, to agreement on a set of rational and universally agreed symbols for the chemical elements. This built on the earlier work of A. Lavoisier and colleagues who systematized the way compounds were named. This means that now any chemist, whatever their nationality and language, can 'converse' in chemical formulae and equations without any ambiguity or misunderstanding. Like mathematical equations, chemical equations cross the language barrier and enable the chemist to 'read' the essence of papers in languages he/she does not know.

Each language, however, still has its own names for most of the elements and even in English there are minor differences between American and

English names. Thus, it is not possible to write the names of the elements and be understood everywhere, as it is with the symbols. However, English is rapidly becoming the universal *lingua franca* of science, written and oral. This has improved international communication in chemistry and, of course, IUPAC's recommendations are made in English.

After symbols and names, the remaining item of elemental nomenclature is the numbering of groups in the Periodic Table. Even today if you mention 'group four' to an older chemist you may well get different responses depending on their background. Do you mean 4 or IV, IVA or IVB? Since 1989 IUPAC has introduced a standard numbering system from 1 to 18 in Arabic numerals to replace the previous confusion of Roman numerals (I to VIII) and A and B subgroups (whose definition was opposite in America and Europe).

The announcement in September 1997 that the IUPAC Council had accepted the recommendations of the Committee for Nomenclature in Inorganic Chemistry (CNIC) for the names of elements 101–109 ended over a decade of indecision and 4 years of wrangling over the list of names proposed in 1993. This has restored one's faith that it **is** possible to achieve a rational compromise in international disputes. Now they only have to agree the names for elements 110–112 and subsequent elements yet to be made.

In this chapter, I will survey the evolution of elemental nomenclature, starting with the idea of an element itself and ending with the naming of the newest transuranium elements. This short account will show that even in the realm of the elements, the simplest substances, the work of reaching international agreement on their names was hardly an elementary exercise.

2.2 The search for simplicity: the concept of an element

There are an enormous number of different materials in the world: solids, gases and liquids, rocks and minerals, plants and animals. They differ widely in appearance and properties, but do they have anything in common? What is matter made of? Is everything we see unique and different, or is everything made from a finite number of simple building blocks or from variations on one universal substance? Until these questions were properly settled chemistry couldn't develop as a science, nor could the many chemically based technologies develop beyond a certain point. It seems obvious to us today that the world is made by combining a limited number of different types of atoms together to produce a vast variety of substances. This wasn't at all obvious and it took millennia for scientists to come to this understanding of the natural world. The atomic theory of matter wasn't fully accepted in science until this century and in 1887 Henry Roscoe could still write disparagingly: "Atoms are round bits of wood invented by Mr. Dalton."

The Greek philosophers were the first to address these questions about the nature and constitution of matter. There were two rival schools of thought: the atomists (who believed everything was made from a finite number of different particles or atoms, associated with Leucippus *ca.* 40 BC, Democritus 460–*ca.*370 BC, Epicurus 341–270 BC and Lucretius 95–55 BC) and those who believed everything was made from the four elements – air, earth, fire and water (due to Empedoceles 490–430 BC, Plato 428–348 BC and Aristotle 384–322 BC). Aristotle's name is most associated with the four elements theory (Figure 2.1) and this triumphed over the atomic theory and became the dominant theory of matter for nearly 2000 years. It was adopted by the early Christian church and by the Arabs and became the dominant scientific philosophy in Western Europe and the Middle East until the middle ages. It stifled scientific development, though not technological advance, until the Renaissance and Reformation in Europe broke the traces of authority and asserted the role of experiment in determining the truth or falsity of scientific theory.

All substances were thought to be composed of the four elements, air, earth, fire and water, in different proportions, as a result of the combination of four properties or qualities: hot, dry, wet and cold. Thus, hot and dry produced fire and cold and wet produced water. This was a very simple theory and led to the idea of transmutation, and the search for the

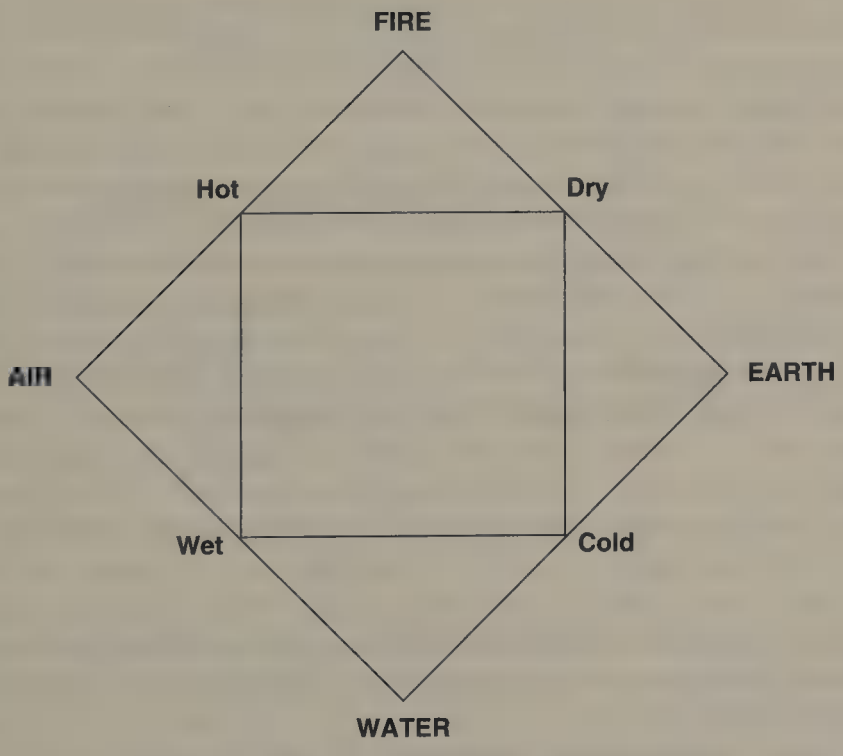


Figure 2.1 The four elements and four qualities.

philosopher's stone that would turn everything into gold, since if one could alter the proportions of the four elements one could (in principle) turn anything into anything. The Islamic alchemists applied the idea of the four elements to the human body and Jabir, an Arab alchemist, also proposed that mercury and sulphur were the basic principles that made up metals. All metals were made of mercury and sulphur in different proportions. Other people added salt as a basic principle, so that we had four elements and three principles as the basic theory of matter by the late Middle Ages.

The most important person in the development of chemistry in making the break with the past, and rejecting the old ideas of the four elements (from Aristotle) or the three principles – mercury, sulphur and salt (from the alchemists) was Robert Boyle (1626–1691). He was an Irishman, born in Lismore, Co. Waterford the youngest son of the Earl of Cork. He was wealthy and was able to devote his life to scientific pursuits, in Oxford and then in London. He is often referred to as the Father of Chemistry because of his important book, *The Sceptical Chymist* published in 1661 (Figure 2.2).

In this book Boyle rejected the old theories and showed that they were not tenable in the light of new experimental evidence. He gave a new definition of an element, which is similar to what we hold today (Appendix 2.A). However, he didn't try to identify which substances were elements and so his ideas were really speculations. He stressed the importance of experiment over authority in determining the worth of a theory, and initiated the process of separating a scientific study of chemistry from alchemy. He made many contributions to science and to chemistry, but the seeds of doubt that he sowed about the validity of the Four Elements and the Three Principles was probably his major contribution. By the time Boyle was researching and writing only 13 substances were known that we would consider to be elements (iron, gold, silver, copper, tin, sulphur, mercury, lead, zinc, carbon, bismuth, arsenic, antimony) but the next 125 years or so were to see the discovery of phosphorus, platinum, cobalt, nickel, hydrogen, nitrogen, oxygen, chlorine, manganese, molybdenum, tellurium and tungsten (Appendix 2.B). Antoine Lavoisier, a French chemist who was assisted by his wife Marie, firmly established the concept of an element and was able in 1789, in his book *The Elements of Chemistry*, to list 33 elements (Figure 2.3). Note that he included heat and light, and the oxides (calxes) of the alkali and alkaline earth metals as elements. Within 20 years, the introduction of the voltaic pile and the work of Humphry Davy were to reveal the elements lurking in these oxides. Lavoisier's definition (Appendix 2.A) was a pragmatic definition and he was open to the idea that some simple substances would in future be shown to be compounds.

The idea of a chemical element as the simplest substance from which compounds are made, was consolidated through the 19th Century, despite

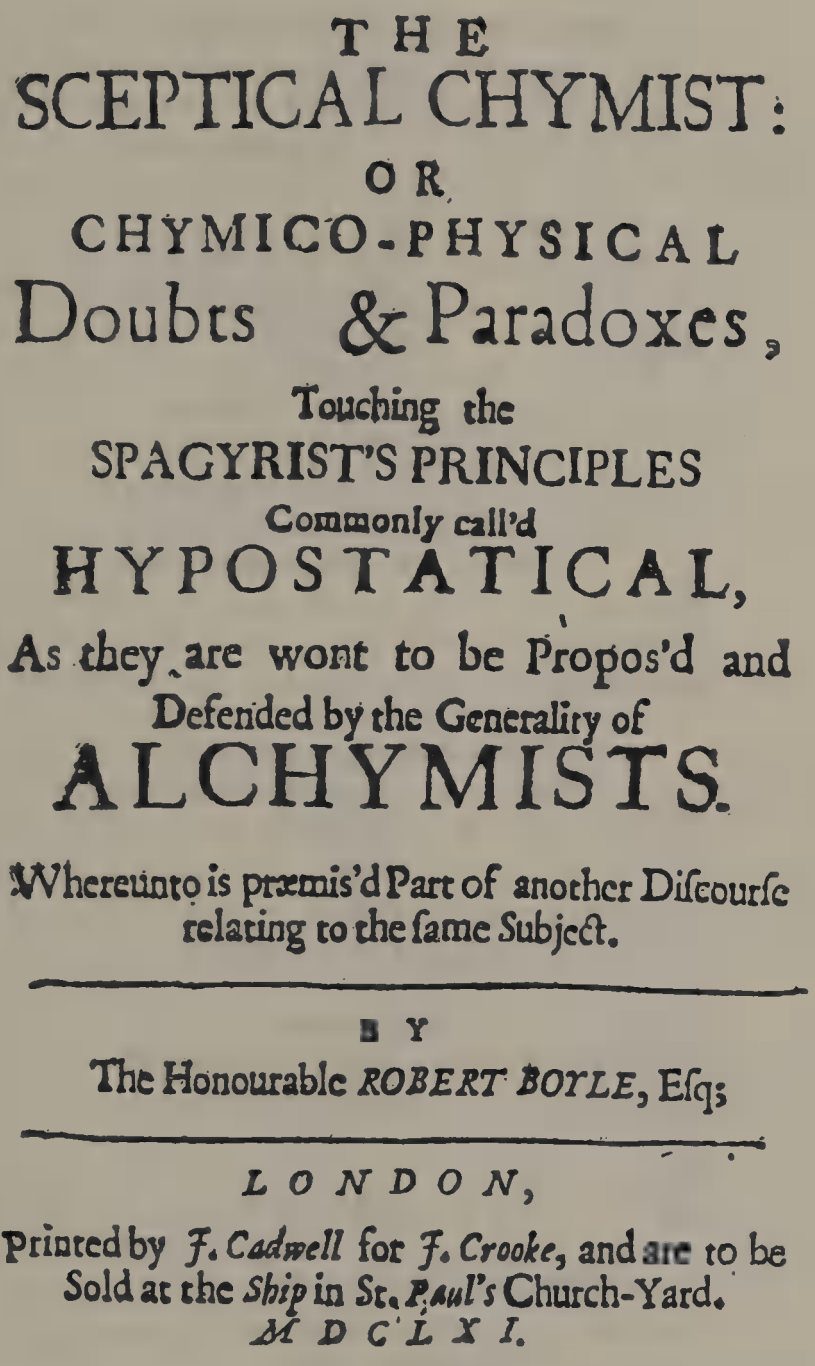

THE
SCEPTICAL CHYMIST:
OR
CHYMICO-PHYSICAL
Doubts & Paradoxes,
Touching the
SPAGYRIST'S PRINCIPLES
Commonly call'd
HYPOSTATICAL,
As they are wont to be Propos'd and
Defended by the Generality of
ALCHYMISTS.
Whereunto is præmis'd Part of another Diſcourſe
relating to the ſame Subject.
BY
The Honourable *ROBERT BOYLE*, Eſq;
LONDON,
Printed by *J. Cadwell* for *J. Crooke*, and are to be
Sold at the *Ship* in St. *Paul's* Church-Yard.
M DC LXI.

Figure 2.2 Title page of Robert Boyle's *The Sceptical Chymist* (1661).

the realization that elements could have different forms (allotropes). These allotropes often differed widely in their physical properties, and sometimes in their chemistry, but contained the same sort of atoms. Then came the discovery of radioactivity and the realization that elements could contain unstable atoms. Not only that, but one element could change into

TABLE OF SIMPLE SUBSTANCES.

Simple ſubſtances belonging to all the kingdoms of nature, which may be conſidered as the elements of bodies.

New Names.	*Correſpondent old Names.*
Light	Light.
Caloric	Heat. / Principle or element of heat. / Fire. Igneous fluid. / Matter of fire and of heat.
Oxygen	Dephlogiſticated air. / Empyreal air. / Vital air, or / Baſe of vital air.
Azote	Phlogiſticated air or gas. / Mephitis, or its baſe.
Hydrogen	Inflammable air or gas, / or the baſe of inflammable air.

Oxydable and Acidifiable ſimple Subſtances not Metallic.

New Names.	*Correſpondent old names.*
Sulphur	The ſame names.
Phoſphorus	
Charcoal	
Muriatic radical	Still unknown.
Fluoric radical	
Boracic radical	

Oxydable and Acidifiable ſimple Metallic Bodies.

New Names.		*Correſpondent Old Names.*
Antimony	Regulus of	Antimony.
Arſenic		Arſenic.
Biſmuth		Biſmuth.
Cobalt		Cobalt.
Copper		Copper.
Gold		Gold.
Iron		Iron.
Lead		Lead.
Manganeſe		Manganeſe.
Mercury		Mercury.
Molybdena		Molybdena.
Nickel		Nickel.
Platina		Platina.
Silver		Silver.
Tin		Tin.
Tungſtein		Tungſtein.
Zinc		Zinc.

Salifiable ſimple Earthy Subſtances.

New Names.	*Correſpondent old Names.*
Lime	Chalk, calcareous earth. / Quicklime.
Magneſia	Magneſia, baſe of Epſom ſalt. / Calcined or cauſtic magneſia.
Barytes	Barytes, or heavy earth.
Argill	Clay, earth of alum.
Silex	Siliceous or vitrifiable earth.

Figure 2.3 The list of elements from Lavoisier's *Elements of Chemistry* (1789).

another by nuclear transmutation. The dream of the alchemists had become reality. Frederick Soddy then discovered that all the atoms of a given element were not the same, but could differ in mass, and he coined the name isotopes to describe them. The concept of what truly defines an element was finally established in 1913–1914 in a series of brilliant experiments by the young physicist Henry Moseley [2]. His studies of X-ray spectra showed that an element was uniquely defined by its atomic number, which was the charge on its nucleus or the number of protons in the nucleus. This was true, irrespective of its mass or stability or its allotropic forms. Later the development of particle accelerators was to allow scientists to turn one element into another, even lead into gold, or to make new elements at will.

2.3 The naming of the elements before 1789

The names of the earliest elements were not systematic but grew up by common usage. In some cases there is still dispute about the origin of the names, e.g. zinc, sulphur, as their origins are uncertain, but until 1789 Latin names for elements were widely used and formed the basis of Berzelius' symbols. Thus, the elements known from antiquity such as gold (Au), silver (Ag), lead (Pb), mercury (Hg), iron (Fe), copper (Cu), tin (Sn) have English names but the symbols are derived from their Latin names. The same is true for the later element tungsten (W), which was originally given the symbol Tn. These early elements have their own names in other languages. Later elements were discovered or isolated by particular individuals who then had the right to name the new elements. Thus, from the 1700s onwards the number of new elements and individually coined names increases dramatically as chemists developed the tools to uncover the elements lurking in nature [3]. The rate of discovery of the elements is shown in Figure 2.4 and the chronology of the discovery of the elements is listed in Appendix 2.C.

2.4 Lavoisier's contribution

Although Lavoisier's name is most commonly associated with the 'new chemistry' and the overthrow of the phlogiston theory, he also played a major role in the reform of chemical nomenclature in the late 18th Century. It was in fact a team effort and it was initiated by Guyton de Morveau who published a memoir on chemical terminology in 1781–1782. He laid down the principles of a revised nomenclature and enlisted the help of Antoine Lavoisier, Claude-Louis Berthollet and Antoine Fourcroy in his endeavour. Lavoisier's main contribution was to help popularize

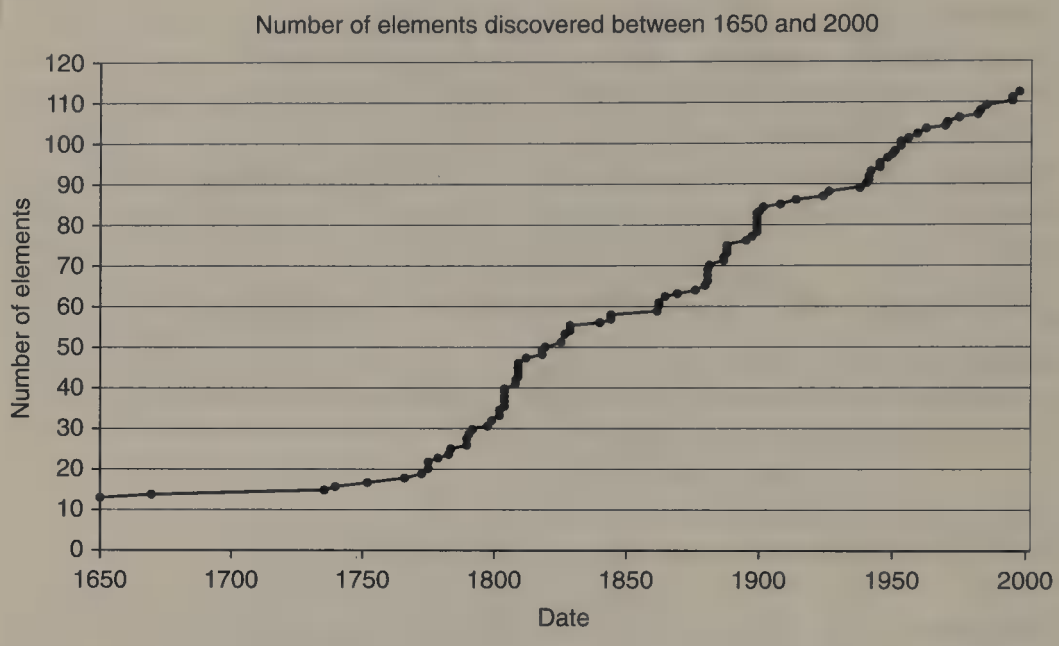


Figure 2.4 The chronology of the discovery of the elements.

the new system of naming compounds, where a name replaced a phrase or several phrases, and also gave an indication of the composition. It provided a systematic way of using the existing names of the elements in naming compounds. In the *Methode de Nomenclature Chimique* (1787), Lavoisier (together with du Morveau, Bertollet and Fourcroy) made an eloquent case for a new language for chemistry:

> . . . we shall have three things to distinguish every physical science: the series of facts that constitute the science, the ideas that call these facts to mind and the words that express them. The word should give birth to the idea; the idea should depict the fact.

The basic ideas were these: first name the elements (the basic constituents of matter), then replace the phlogiston theory with an experimentally based theory of combustion, and then coin new names showing just the elements present and their relative quantities. Substances that could not be decomposed were considered to be 'simple' (i.e. elements) and their names were to be used as the basis for naming compounds. Thus, chlorine replaced dephlogisticated muriatic acid (a confusing, chemically inaccurate name) and nitrogen dioxide and nitrogen pentoxide, for example, clearly represented two different compounds made from the same elements.

The similarity of English and French, with the system of suffixes inherited from the Normans, meant that the new names were easily translated

into English, which did much to popularize the new ideas in the world outside France. Laurence Hogben comments:

> The hybrid character of the English vocabulary thus played a decisive role in promoting the acceptance of the French reforms in countries other than France. Prompt acceptance of the new nomenclature by Britain, at a time when Britain and France were in the vanguard of chemical discovery and chemical industry, confronted the international scene with a *fait accompli*. It did so by the unforeseen accident that two national languages had the same battery of suffixes.
>
> *[4]*

William Brock comments on Lavoisier's nomenclature: "Perhaps the most significant assumption in the nomenclature was that substances that could not be decomposed were simple (i.e. elements) and that their names should form the basis for the entire nomenclature." [5]. Thus, it is clear that the proper naming of the elements is fundamental to chemical nomenclature, and that is still true today.

2.5 Berzelius and the agreement on symbols

Before the early 1800s there was no agreed system for representing the elements in symbolic form. There were the symbols used by alchemists to denote different elements and compounds but no single system prevailed. The symbols were meant as much to conceal and mystify as to simplify and illuminate readers. Lavoisier's system also included a set of symbols devised by Hassenfratz and Adet to represent the elements and various compounds. John Dalton from 1807 onwards had devised his own system of representing elements and compounds. These were all complex and idiosyncratic systems which were difficult to write and even more difficult to typeset. Table 2.1 compares these different systems with Berzelius' symbols.

Something better was needed that was easier to understand, was universally accepted and most importantly, was easy to write and print for the increasing number of chemical papers and books. In 1813, Berzelius submitted a paper to the *Annals of Philosophy* on a new system of chemical nomenclature. He proposed that letters should be used for chemical symbols, as they were easy to write and print, and he proposed that the initial letter of the Latin name for an element should be used, and if two elements started with the same letter, he proposed that non-metals (which he called metalloids) should have only one letter, and for the metals the first two letters should be used if they had a common initial with another metal or non-metal. If the first two letters are in common, then he proposed adding the first consonant they did not have in common. Thus,

Table 2.1 Representing the elements as symbols

Element	Alchemical symbols	Dalton's symbols	Hassenfratz & Adet's symbols	Berzelius' symbols
Silver	☽	Ⓢ	Ⓐ	Ag
Gold	☉	Ⓖ	⊙	Au
Iron	♂	Ⓘ	Ⓕ	Fe
Copper	♀	Ⓒ	Ⓒ	Cu
Lead	♄	Ⓛ	Ⓟ	Pb
Tin	♃	Ⓣ	Ⓢ	Sn
Mercury	☿	◌	Ⓗ	Hg

he proposed: S = sulphur; Si = silicium (silicon); St = stibium (antimony); Sn = stannum (tin); C = carbonicum (carbon); Co = cobaltum (cobalt); Cu = cuprum (copper); O = oxygen; Os = osmium, etc. He made minor changes over the years to his symbols, e.g. palladium changed from Pl to Pa to Pd. He called chlorine 'the muriatic radical', symbol M, but by 1827 had converted to Cl. The 'fluoric radical' was F and iodine was I (often J in German), so that iridium was changed from I to Ir, although this had not been in agreement with the principle of giving the non-metals single symbols as far as possible.

One can see the effect of Berzelius' reforms in our present Periodic Table (*cf.* Table 2.2), where many of the non-metals have single letters as symbols. A number of Berzelius' symbols have changed, e.g. Cr for chromium instead of Ch, W for wolfram (tungsten) instead of Tn, Sb for antimony instead of St, Nb for niobium instead of Cl (columbium), Pt for platinum instead of Pl, Pd for palladium instead of Pa (now protactinium), Mn for manganese instead of Ma, Be for beryllium instead of Gl for glucinium, Mg for magnesium instead of Ms, Na for sodium (natrium) instead of So, K for potassium (kalium) instead of Po (now polonium), Cl for chlorine instead of M (for muriatic radical). In addition, our English names (in parentheses above) are not the same as the Latin names and, in some cases, the English names have changed since Berzelius' time, e.g. for beryllium.

Crosland [6] points out that the lack of agreement on the names of the elements made it difficult to assign symbols to certain elements:

> – Berzelius favoured certain names current in Germany, but different from those in France or England. Thus he insisted on using the Latin names Wolframium (tungsten), Beryllium (the glucinum of Vauquelin), Kalium (potas-

sium) and Natrium (sodium), since he considered these names superior on etymological grounds to the alternative names. Berzelius also disagreed with the British chemists who, loyal to their compatriot Hatchett, preferred to call columbium what Berzelius referred to as tantalum, a name bestowed on a similar metal by the Swede Ekeberg who had first examined it in the metallic state.

New reforms are rarely popular and Berzelius' ideas were not well received by many chemists and took several decades to win acceptance, although today they look so familiar and unobjectionable. John Dalton complained about them in 1837 in a letter to Thomas Graham: "Berzelius' symbols are horrifying; a young student of chemistry might as well learn Hebrew as make himself acquainted with them. They appear like a chaos of atoms ... to equally perplex the adept of Science, to discourage the learner as well as to cloud the beauty of the Atomic Theory." Perhaps Dalton was disgruntled that his own even more esoteric system had failed to catch on. The French persisted in using Gl for glucinium (beryllium), Tu for tungsten (wolframium) and Az for azote (nitrogen). Crosland [7] commented in 1962 that the use of Az "persisted in France to this day and ... had only recently been replaced by N for the sake of international conformity." Azote is still the name for nitrogen in France.

2.6 Organizing the elements

The discovery of the elements until 1869 was very much like a treasure hunt, where the seekers believed but did not know for sure that there was any buried treasure to be found. Chemists worked away all round the world looking for new substances and new elements. A new mineral was always a hot favourite for a new element. There was no rationale in the pursuit, since no one knew how many elements there were, or where to find them. Any unidentified mineral or compound was game for the element hunters and some chemists were remarkably successful. Dimitri Mendeleev's statement of the Periodic Law and his organization of the 63 known elements into the first proper Periodic Table in 1869 enabled chemists for the first time to predict the existence and properties of new elements. Mendeleev's greatest triumph was to predict with uncanny accuracy the properties of three elements which were missing in his table eka-boron, eka-silicon and eka-aluminium – when they were discovered, scandium, germanium and gallium had chemical and physical properties almost identical to those predicted. (This terminology of denoting a new element as eka-, from the Sanskrit for one, by relating it to the neighbouring element in its group continues until today. Mendeleev also referred to the second and third undiscovered elements, adjacent to a

known one, as dwi- and tri-, respectively.) In 1889, in his Faraday Lecture to the Royal Society of Chemistry, Mendeleev said this:

> Before the promulgation of this [Periodic] law the chemical elements were mere fragmentary, incidental facts in Nature; there was no special reason to expect the discovery of new elements, and the new ones which were discovered from time to time appeared to be possessed of quite novel properties. The law of periodicity first enabled us to perceive undiscovered elements at a distance which formerly was incredible to chemical vision, and long ere they were discovered new elements appeared before our eyes, possessed of a number of well defined properties.
>
> *[8]*

Even the reason for the structure of the Periodic Table was not known as it had been based solely on the periodicity of physical and chemical properties, i.e. it was purely empirical. Thus although the Periodic Table could organize existing elements and identify gaps, it could not predict how many more elements there were. Chemists had to wait until Henry Moseley's studies on X-ray spectra in 1913–1914 established the correct sequence of elements, as defined by their atomic number, and the identification of missing elements:

> The prevalence of lines due to impurities suggests that this may prove a powerful method of chemical analysis. Its advantage over ordinary spectroscopic methods lies in the simplicity of the spectra and the impossibility of one substance masking the radiation from another. It may even lead to the discovery of missing elements, as it will be possible to predict the position of their characteristic lines.
>
> *(Henry Moseley, 1913) [9]*

One problem Moseley was able to solve was that of the rare earths – in 1913 chemists were not sure how many there were, and between 14 and 23 elements had been accepted, proposed or anticipated! But no one knew how many to expect or even how many there were. William Crookes said:

> The rare earths perplex us in our very dreams. They stretch like an unknown sea before us, mocking, mystifying and murmuring strange revelations and possibilities.
>
> *[10]*

By early 1914 Moseley had them sorted out, something which decades of activity by able chemists had failed to do. In a letter to Ernest Rutherford on 4 March 1914, Moseley wrote this:

> The rare earths I am now in the middle of. They give the usual spectra, but the commercial salts which I have been using are fearful mixtures, so that I am not yet sure of the results. I have tried neodymium, praseodymium, samarium, gadolinium, erbium. . . . I find that there are not enough places to accommodate all the elements to which names have been given in this region. It will be a great clearance to put each element in its right place, and weed out the superfluous, as the subject is still in terrible confusion. There are some who

> would split every one of these rare earths into 3 or 4. It is surprising that the only [atomic] numbers not yet occupied are probably 39 (celtium)*, 44 (canadium) and 76. And also possibly one or two between 60 and 73. It says much for the industry of chemists.
>
> *[11]*

(*This appears to be a mistake by Moseley as element 39 (yttrium) was well known and the element named celtium by Urbain in 1911 was thought to be the missing no. 72. So celtium was another name destined to fade into obscurity. Hafnium was the name given to element 72 when it was unambiguously identified by George Hevesy in 1923.)

The distinguished French rare earth chemist Georges Urbain came over to see Moseley in Oxford to try and sort out the rare earths, and after Moseley's tragic early death in the Gallipoli campaign he wrote to Rutherford:

> I had been very much surprised when I visited him in Oxford to find such a young man capable of accomplishing such a remarkable piece of work. The Law of Moseley confirmed in a few days the conclusions of my efforts of twenty years of patient work ... His Law substituted for Mendeleev's somewhat romantic classification a completely scientific accuracy.
>
> *[12]*

Frederick Soddy said this about Moseley's work:

> For the first time Moseley had called the roll of the elements and we could now say definitely the number of possible elements between the beginning and the end, and the number that still remained to be found.
>
> *[13]*

It is hard to over-estimate the importance of Moseley's work, done over a period of a few months in 1913–1914, which provided the necessary framework for later theories of electronic structure. Quantum theory was to provide the theoretical rationale for the arrangement of elements and their periodicity of properties based on their electronic structure. This led in turn to the ability to predict the positions and properties of as-then undiscovered transuranium elements, which from the 1940s onward were to be made in nuclear laboratories in the USA and then later around the world. Theoreticians would even predict the existence of superheavy elements and an 'island of stability' around element 114, predictions still to be realized. The number of elements was no longer a mystery to be discovered by chance, but was now open to systematic investigation and synthesis.

2.7 Isotopes and radioactivity

Radioactivity was discovered in 1896 by Henry Becquerel and X-rays were discovered by Wilhelm Roentgen in 1897. These provided fundamental

insights into the nature of atoms. In 1913, Frederick Soddy established the concept of an isotope: all atoms of a given element need not be identical – they could differ in mass and this explained the non-integer atomic masses. In 1920, Francis Aston, using the new tool the mass spectrograph, was able to identify and measure the mass of isotopes. This explained the atomic weights of the elements and the reason why some elements (e.g. tellurium) appear out of sequence in the Periodic Table when ranked by atomic weight rather than atomic number. It also explained why some elements, obtained from different sources, had different atomic weights.

The study of radioactivity was to lead to the discovery of new elements, first by the Curies who isolated radium and polonium (1898), followed by M. A. Debierne who isolated actinium (1899), and F. Dorn who in 1900 identified radon in the radioactive emanations from radium and other radioactive substances. (Radon was at various times called niton, thoron, acton and even emanation!) Element 91 was first identified by Fajans and Gohring in 1913 and was first called brevium, because of its brief existence. Its existence was later confirmed in 1917–1918 and renamed protoactinium, shortened in 1949 to protactinium. Several names were related to the idea of radiation, e.g. radium, radon, actinium (from the Greek for ray) and protactinium.

Between 1913 and 1946 elements 43, 61, 85 and 87 were discovered using nuclear transmutation and identified as new elements but not named. They had not been discovered in nature despite much searching. In 1946, Professor F. A. Paneth pointed out this omission, which was partly due to the reluctance of chemists to recognize unstable isotopes as true elements. He said:

> The denial of full citizenship to artificial elements seemed justified in those days. They had only been produced in invisible amount only, and they were unstable and usually not present on earth; whereas in the case of all the natural elements, we could be sure that, even if they belonged to the radioactive families and were only represented by short-lived isotopes, very considerable quantities also existed. . . . The limited importance attributed until a few years ago to the artificially produced elements was reflected also by the absence of names suggested for them . . .
>
> *[14]*

Following this intervention, early in 1947, the respective discoverers announced that element 43 would be called technetium, element 85 astatine, element 87 francium, and element 61 promethium. Earlier names proposed for the elements occupying the missing slots in the Periodic Table (43, masurium; 61, illinium; 85, alabamine; 87, virginium) were quietly dropped.

The technique of making new elements by controlled nuclear bombardment was now well established and during World War 2 the nuclear alchemists in the USA, led by Glenn Seaborg at Berkeley, quietly 'synthesized' and named several new elements (93–96), whose discovery was only

revealed after the war. They later went on to make elements 97–103. These were later named by their makers: elements 93 and 94 were named neptunium and plutonium, respectively, because like the eponymous planets (Neptune and Pluto) they come after uranium (Uranus). Element 95 was named americium after its country of discovery because of its chemical similarity to europium, named after Europe. Element 96 resembled the rare earth gadolinium, named after a rare earth mineral named after the Swedish chemist Johan Gadolin. It was thus named curium after Pierre and Marie Curie who had opened up the field of radioactive elements. Element 97 (1949) was like terbium (named after the place of discovery Ytterby), so element 97 was named berkelium after its place of discovery. Similarly element 98 (1950) was named californium after the state where it was discovered. Elements 99 and 100 were both discovered around 1952 and named einsteinium and fermium, after two pioneering physicists. Element 101 was discovered in 1955 and named mendelevium in honour of the founder of the Periodic Table.

From element 102 onwards controversy crept in as the Americans no longer had the monopoly on synthetic elements, as other nuclear laboratories in Russia, Sweden and Germany started to play the 'make an element' game. Different laboratories made competing claims to have discovered a new element, often by different methods, and the syntheses could not always be repeated by others. This is a recipe for controversy over priority of discovery. Although element 102 was synthesized in 1957/1958 and named nobelium, it was not until **1997**, fully 40 years later, that this and subsequent names up to element 109 were agreed internationally by IUPAC (see below). The current tally of elements is 112 as three new elements were made in 1994–96 by the Darmstadt laboratory in Germany. Initially the discoverers wanted the symbols Lw, E and Mv for lawrencium, einsteinium and mendelevium. IUPAC decided on Lr, Es and Md. This is the reason for the discrepancies in older chemical textbooks and articles.

2.8 Choosing names for new elements [15]

How does one go about choosing names for new elements? Several elements were known from antiquity and so had traditional names, often from Latin or they had names local to where they were first used. The origin of several of these names is still disputed (e.g. sulphur) but they passed into common usage and stuck. When new elements started to be discovered from the 1700s onwards, the invariable rule was that the discoverer named the element, although there was no agreed system of how to name an element, and there was often argument as to who got there first. Is was quite common too for a new element to be identified and named long before it could be isolated in a pure form, for example aluminium.

Thus, several elements have different dates for discovery and isolation. Fluorine is a good example as its existence was known for nearly 100 years before Moissan was able to isolate it.

Several naming systems have been used as the honour traditionally goes to the discoverer who could do as he/she pleased, depending on the fashions of the day, and the scientific community then ratified their choice, or sometimes modified the name and/or symbol. Many elements were discovered simultaneously or were named independently by several people, and sometimes the identification of elements could get very messy as scientists argued about priority. There was no international body like IUPAC to arbitrate disputes and recommend accepted names. Thus, we had the dual names columbium/niobium (see below), glucinum/beryllium and more recently rutherfordium/kurchatovium.

Elements have been named after people, after places, after mythological figures, after planets, or after their distinctive properties, often as the whim took the discoverer. The opportunity to name one of the fundamental building blocks of matter must surely be one of the greatest prizes in science. However, there can be no greater claim to fame than having an element named after you while you are still alive and Glenn Seaborg has just had the honour to be the first living scientist to give his name to an element. The earliest element named in this way after a person was gadolinium in 1886 but since element 96 (1944) it has been the most common method. In the sections below, we will summarize the various naming systems that have been used.

2.8.1 Elements named after people

Name	**Z**	**Person**
Gadolinium	64	Johan Gadolin
Curium	96	Pierre and Marie Curie
Einsteinium	99	Albert Einstein
Fermium	100	Enrico Fermi
Mendelevium	101	Dimitri Mendeleev
Nobelium	102	Alfred Nobel
Lawrencium	103	Ernest Lawrence
Rutherfordium	104	Ernest Rutherford
Seaborgium	106	Glenn Seaborg
Bohrium	107	Niels Bohr
Meitnerium	109	Lise Meitner

Following this precedent, one might expect that elements 110–112 might be named after famous nuclear scientists. One could speculate who should be honoured in this way: Frederic Joliot has been suggested already and rejected, as has Kurchatov. Henry Moseley surely deserves commemoration

as it was his work that finally sorted out the identity of the elements and solved the problem of the rare earths (see above). Albert Ghiorso who succeeded Seaborg and whose team discovered several elements, is also a strong contender. We will have to wait and see.

2.8.2 Elements named after countries or places

Naming the elements after the place (town or country) where they were discovered or after the country or town of their discoverer is one of the most common naming methods. One small town in Sweden, Ytterby, has no less than four elements named after it as they all occur together in a rare earth mineral found there. The elements are ytterbium, erbium, terbium and yttrium. Not bad for an obscure Swedish town. Often the Latin name for a place is used following Berzelius' preference for classical, particularly Latin, names, e.g. Hafnia (Copenhagen). Others were named after planetary or other bodies in the solar system.

Element	**Z**	**Town/Region**	**Country/Continent**	**Planet**
Helium	2			Sun (helios)
Magnesium	12	Magnesia		
Scandium	21		Scandinavia	
Manganese	25	Magnesia		
Copper	29		Cyprus (Cuprum)	
Gallium	31		France (Gallia)	
Germanium	32		Germany (Germania)	
Selenium	34			Moon (selene)
Strontium	38	Strontian		
Yttrium	39	Ytterby		
Ruthenium	44		Russia (Ruthenia)	
Palladium	46			Pallas (asteroid)
Cadmium	48	Cadmia		
Tellurium	52			Earth (tellus)
Cerium	58			Ceres (asteroid)
Europium	63		Europe	
Terbium	65	Ytterby		
Holmium	67	Stockholm (Holmia)		
Erbium	68	Ytterby		
Thulium	69		Scandinavia (Thule)	
Ytterbium	70	Ytterby		
Lutetium	71	Paris (Lutecia)		
Hafnium	72	Copenhagen (Hafnia)		
Rhenium	75	Rhine (Rhenus)		
Polonium	84		Poland	
Francium	87		France	

Element	Z	Town/Region	Country/Continent	Planet
Uranium	92			Uranus
Neptunium	93			Neptune
Plutonium	94			Pluto
Americium	95		America	
Berkelium	97	Berkeley		
Californium	98	California		
Dubnium	105	Dubna		
Hassium	108	Hesse		

The clear winner is Scandinavia with seven elements named after it, mainly because this was where the rare earth elements were discovered, many of them in Swedish minerals.

2.8.3 Named after a distinctive property

Many elements were named after a distinctive property, of which colour is one of the most common – either directly, or after 1860, after the colour of its spectral lines. Other properties such as smell or weight or the difficulty of obtaining the element have been used.

Name	Z	Property
Hydrogen	1	Water former
Lithium	3	Stone (Gk., *lithios*)
Nitrogen	7	Soda former
Oxygen	8	Acid former
Fluorine	9	To flow (L., *fluere*) (also fluorspar)
Neon	10	New (Gk., *neos*)
Sodium	11	Headache remedy (L., *Sodanum*)
Phosphorus	15	Bringer of light (Gk., *phosphorus*)
Chlorine	17	Greenish–yellow (Gk., *chloros*)
Argon	18	Inactive (Gk., *argos*)
Chromium	24	Colour (Gk., *chroma*)
Manganese	25	Magnetism
Bromine	35	Stench (Gk., *bromos*)
Krypton	36	Hidden (Gk., *kryptos*)
Rubidium	37	Deep red (L., *rubidus*)
Technetium	43	Artificial (Gk., *technetos*)
Rhodium	45	Rose-red (Gk. *Rhodon*)
Indium	49	Indigo (L. *indicum*)
Iodine	53	Violet (Gk. *Iodes*)
Caesium	55	Blue (L. *caesius*)
Lanthanum	57	To lie hidden
Praseodymium	59	Green twin

Name	Z	Property
Neodymium	60	New twin
Dysprosium	78	Hard to obtain
Osmium	76	Smell (Gk. *Osme*)
Iridium	77	Rainbow (Gk. *Iris*)
Thallium	81	Green twig (L. *thallus*)
Astatine	85	Unstable
Radon	86	Gas from radium
Radium	88	L. for ray
Actinium	89	Gk. for ray
Protactinium	91	First + actinium

2.8.4 Named after a mineral

A number of elements were named after the mineral from which they were isolated or some other mineral connection.

Name	Z	Mineral
Lithium	3	Stone (*lithios*)
Beryllium	4	Beryl
Boron	5	Borax
Carbon	6	Charcoal (*Carbo*)
Fluorine	9	Fluorspar
Sodium	11	Soda
Aluminum	13	Alum
Silicon	14	Flint (L., *silicus*)
Potassium	19	Potash
Calcium	20	Lime (L., *calx*)
Arsenic	33	L., *Arsenicum* (pigment)
Zirconium	40	Zircon
Molybdenum	42	Lead (Gk., *molybdos*)
Barium	52	Barytes (heavy spar)
Samarium	62	Samarskite
Gadolinium	64	Gadolinite (after J. Gadolin)
Tungsten	74	Heavy stone (Swedish)

2.8.5 Named after a mythological figure

The interest in classical mythology in the 19th Century led to a number of elements being named after a suitable figure from mythology or folklore, although some go back even earlier like cobalt and nickel.

Name	Z	Figure
Titanium	22	Titans

Name	Z	Figure
Vanadium	23	Vanadis
Cobalt	27	Goblin (Ger., *kobold*)
Nickel	28	Old Nick
Niobium	41	Niobe
Tantalum	73	Tantalus
Mercury	80	Mercury (also a planet)
Palladium	46	Pallas Athene (also an asteroid)
Cerium	58	Ceres (also an asteroid)
Promethium	61	Prometheus
Thorium	90	Thor

2.9 Some cases of elemental confusion

In a number of cases, particularly elements known from antiquity, the path to an agreed name was long and tortuous. Some examples are given below.

2.9.1 Manganese and magnesium

Many students of chemistry confuse the elements magnesium and manganese as it is easy to misplace the 'g' and 'n', e.g. to get mangesium. The interesting thing is that this confusion goes right back to the discovery and naming of these elements, whose names both derive from the same source.

Iron occurs naturally as a magnetic oxide (Fe_3O_4, magnetite) which was known as lodestone or loadstone = leading stone, because a small rod of it seeks the North Pole and also attracts iron. This was also known as the '*ho Magnes lithos*', Greek for 'the Magnesian stone', since it was found near Magnesia in Thessaly. Pliny, the Roman naturalist, incorrectly said there were two forms of lodestone: one magnetic (magnetite) and one non-magnetic. The non-magnetic mineral was regarded as the feminine (passive) lodestone and was in fact manganese dioxide (MnO_2), which is a heavy, dark mineral similar in outward appearance to magnetite. In the Middle Ages the masculine or active lodestone was called 'magnes' (from which comes magnet) and the feminine lodestone became known as magnesia.

Around 1700, an Italian discovered a valuable mineral salt with medicinal properties. He thought it was related to magnesia (MnO_2) and so he called it *magnesia alba* (white magnesia) and manganese dioxide became *magnesia nigra* (black magnesia). *Magnesia alba* was in fact hydrated magnesium carbonate and it was initially confused with lime, since they are both basic substances. In 1740, Friedrich Hoffmann showed that magnesia (MgO) and lime (CaO) were different. In 1808, Sir

Humphry Davy showed that they contained two different metals, which he isolated by electrolysis of their fused oxides. The metal from lime he called calcium and that from magnesia he initially called magnium.

> I shall venture to denominate the metals from the alkaline earths barium, strontium, calcium, and magnium: the last of these words is undoubtedly objectionable, but magnesium has already been applied to metallic manganese. *[16]*

Thus was the confusion complete. Manganese was known as magnesium and magnesium as magnium. Two elements with no chemical similarity and no magnetic compounds had become almost totally confused. In 1812, Davy renamed his metal magnesium. Magnesia, which had originally been MnO_2 and then also $MgCO_3$, now became applied to MgO and $MgCO_3$ became magnesite, hence magnesian limestone.

But what about manganese? Where did that name come from?

> Mangania is the Greek word for magic or in modern parlance voodooism. If this is the root of the metal's name, it reflects some reality in the biology of manganese which is rich in phenomena and lacking in guiding principles. *[17]*

The word 'magnesia' became corrupted in the Middle Ages into *manganese* in both Italian and French and became *mangan* in German. Thus, manganese and magnesia were used interchangeably for the same substance and were also applied to different substances. The glassmaker's stone, which was used to clarify glass, became known as black manganese or glassmaker's manganese. We met this already as *magnesia nigra*, also known as the mineral pyrolusite = fire + to wash, because of its use to wash out or remove the yellow and green colours from glass (which were due to iron impurities). It was also later called 'glassmakers soap'. It is still used for this purpose in glassmaking. In 1686, Sir T. Browne said "In the making of glass it have been an ancient practice to cast in pieces of magnet, or perhaps manganes."

The metal in this compound (MnO_2) was known as magnesium and also variously as manganese, manganesum or manganesium. In 1774 C. W. Scheele, a Swedish chemist, established that manganese was an element and Johan Gahn, a compatriot, first isolated the metal later that year by reducing manganese dioxide with charcoal. It wasn't until 1812 that Davy resolved the confusion by fixing the name magnesium to element 12 and manganesum (later manganese) to element 25, where they have remained ever since, except when students or typewriters get their letters mixed up and the confusion returns. But, given their history, this is hardly surprising. So, both magnesium and manganese derive their names from Magnesia and also from magnetism, though neither metal nor their compounds are magnetic.

2.9.2 Plumb crazy over lead

Dioscorides, the Greek writer (40–49 AD), used the word *molybdaina* (from the root *molybdos* = the metal lead), to describe a mineral which was in fact lead monoxide (PbO), also known as litharge. (The word litharge comes from the Greek *litharguros* (*lithos* = stone + *arguros* = silver.) It was called this because it was the stony residue formed in the purification of silver-bearing lead ores. Pliny the Elder (50–59 AD) latinized the name to 'molybdaena' and also gave the name 'plumbago' to lead monoxide, from the Latin for lead, *plumbum*.

In 1572, a French translation of a commentary on Dioscorides gave *molybdaina* as *plombagine*, and identified it as litharge; 16th century German chemists, for example Agricola, gave the name *plumbago* to the mineral galena (lead sulphide, PbS) and to other substances which could make marks on paper. Thus, stibnite (Sb_2S_3), molybdenite (MoS_2) and graphite were also all called *plumbago*. They are all soft black substances that mark paper. *Plumbago* was also used to refer to 'red lead' (Pb_3O_4), the oxide formed by oxidation of litharge, just to compound the confusion.

In 1567, Enkel in Germany distinguished between fertile species of *plumbago* which could be reduced to metallic lead (i.e. lead compounds) and infertile species, such as graphite, which didn't give lead. In 1599, the Italian Ferrante Imperato described the use of graphite in making the *grafio plombino* or 'leaden pencil'. In 1779 Scheele, the Swedish chemist, showed that barren *plumbago* (graphite) gave carbon dioxide when burnt in air and distinguished 'black lead' (graphite) from *molybdaena* (lead monoxide).

> The black lead or plumbago which is generally known in commerce, is very different from molybdaena .... Hence, I am convinced, that plumbago is a kind of mineral sulphur or charcoal ...
>
> *[18]*

In 1789 'black lead' was called graphite (from the Greek word for writing, *graphos*) by Werner and Karsten, although it was only finally proved to be a form of carbon in 1829. A graphite crucible has been referred to as a plumbago crucible. To be 'plumbagoed' meant to be covered with black lead, and some of you may still remember black-leaded ranges in old houses.

The Romans called the element lead, *plumbum*, from which comes our symbol for lead, Pb. The confusion between lead (which makes black marks on paper) and graphite continues with our use of 'pencil leads' to refer to the mixtures of graphite and clay used in making pencils. The Latin word also survives in plumber and plumbing, going back to the use of lead in making pipes and joints, which goes right back to the Romans. Lead is also a neurotoxin and can cause mental disturbance, hence the phrase 'plumb crazy'. Interestingly, certain diseases caused by lead are still referred to using the prefix molybdo-, as for example, molybdo-colic.

But what about the word 'molybdaena'? How did this come to be attached to another element entirely, molybdenum? We have already seen that a mineral of molybdenum, molybdenite (MoS_2) was called plumbago because it made black marks like graphite and lead compounds. The root *moly* was the name given to a herb with a white flower and a black root, thought to be endowed with magic properties. This may be the root of the word used for lead and other materials making a black mark. In 1778 Scheele recognized that molybdenite was the ore of a new element and it was first prepared in an impure state by P. J. Hjelm in 1782, who gets the credit for its discovery. From about 1790 the word 'molybdaena' became applied just to molybdenum disulphide and the element was called molybdenum.

> The ore containing molybdenum has almost the appearance of plumbago. *[19]*

Before we leave this involved saga we must ask why do English-speaking nations refer to element 82 as lead, rather than say *plumbum*? The word 'lead' is Anglo-Saxon and is found in old German as *lod* and Dutch as *lood*, also in Irish as *luaidhe*. The alchemists, of course, wrote in Latin and used the Latin name *plumbum*, which became the basis of the symbol for lead and was also used in the archaic terms 'plumbic' and 'plumbous' for the higher and lower oxidation states of lead.

2.9.3 A Hatchett job on niobium [20]

One still finds advertisements in the N. American metallurgical press for columbium and columbium alloys. What are they talking about? Some older American periodic tables give Cb (or Cl) instead of Nb for element 41. This confusion goes back to the discovery of the element in 1801 by Charles Hatchett, an English chemist. Very few chemists today have heard of Charles Hatchett (born 2/1/1765 and died 10/3/1847 in London). In his day, Charles Hatchett had a reputation as a skilled chemical analyst. He was the son of a wealthy coach builder in London, who built coaches for royalty and he was brought up in luxury.

Charles Hatchett's important scientific work was done in the period 1796 to 1806. He was elected a Fellow of the Royal Society in 1797, which is a measure of the esteem in which he was held by his fellow scientists. In 1801 he described in a paper his analysis of a mineral called columbite, named after the location where it had been found in N. America. This mineral sample from Massachusetts had lain in the British Museum since 1753. He described the mineral as "a heavy black stone with golden streaks . . . from Mr. Winthrop": John Winthrop was the first Governor of Connecticut, the source of the mineral. Hatchett showed that it contained a new element and he called it columbium and the mineral columbite, after its place of origin.

Niobium (columbium) always occurs with tantalum because of the similarity in their atomic size and chemistry. The minerals columbite and tantalite contain both elements, differing only in their proportion: columbite contains more niobium and tantalite more tantalum. In 1802, Ekeberg discovered a new metal in a rare earth mineral called yttrotantalite, and called it tantalum because like Tantalus in the Greek myths who could not drink, this new element would not react with acids. He wrote: "This metal I call tantalum . . . partly in allusion to its incapacity, when immersed in acid, to absorb any and be saturated."

Both columbite and tantalite were analysed by William Wollaston in 1809 (after Hatchett had effectively given up science and taken over his father's business following the latter's death). Wollaston was confused by the similarity in the physical and chemical properties of the two elements and he thought they were the same, i.e. he stated that Hatchett's columbium and Ekeberg's tantalum were in fact the same element and consequently the two elements were confused until 1844. Hatchett's claims and name were disregarded, although Ekeberg's name survived.

The matter was not resolved until 1844, not long before Hatchett died, when Heinrich Rose 'rediscovered' columbium, but he called it niobium after the Greek nymph Niobe, who was the daughter of Tantalus – thus recognizing the close relationship between the elements. Hatchett had already shown that the oxide of niobium, Nb_2O_5, and the corresponding oxide of tantalum, Ta_2O_5, had different properties. Wollaston's reputation had ensured that his erroneous views prevailed, especially as Hatchett had given up science by then. Even Rose's researches failed to clarify the chemistry of niobium and tantalum, because of their complexity and similarity to each other. The Swedish chemist Blomstrand (1866) and the Swiss chemist Marignac (1866) finally sorted out the chemistry of the two elements. It was Blomstrand who first isolated metallic niobium by reducing niobium(V) chloride with hydrogen. Roscoe obtained a better yield by reducing the trichloride. The first pure samples of both niobium and tantalum were not made until 1907 when W. von Bolton reduced K_2NbOF_5 and K_2TaOF_5 with sodium metal.

Both names are still in use. The official IUPAC name is niobium (Nb), which was adopted as late as 1950, but the North Americans always called it columbium and in the metals industry they still do. Why this name should have persisted in this way across the Atlantic is a mystery, as is the insistence on using it still despite international agreement to call it niobium. The fact that it was named after America and its American origin is probably sufficient explanation.

2.10 Hydrogen: a special case

Hydrogen is the first and lightest of the elements; indeed, it is the primordial element from which all other elements are synthesized in the stars. It was not isolated and identified as an element until 1766 by Henry Cavendish, although it must have been made many times over the centuries whenever an acid met a metal and Robert Boyle, for example, describes its production. Its lightness gives it the unusual property that the three isotopes, atoms with the same nuclear charge but different masses, are quite different in their physical and some chemical properties. Their masses are roughly in the ratio 1:2:3 and this makes them easily separable and identifiable. The names protium, deuterium and tritium were suggested by Harold Urey in 1933 for these three isotopes, with the symbols H, D and T. No other element has different names and symbols for its isotopes, although these are only used when isotopic properties are important. Deuterium is almost twice the mass of protium, hence the names 'heavy hydrogen' and 'heavy water' for D_2O.

For other elements their isotopes only differ slightly in mass, although their nuclear properties may differ considerably, e.g. stability (radioactivity) and nuclear magnetic moment. Thus, instead of giving isotopes different names, as with hydrogen, they are identified by their mass number and in their symbols by showing atomic number and mass numbers where necessary, e.g. chlorine-35 and chlorine-37; uranium-235 and uranium-237.

2.11 Numbering the columns and groups

There are many ways in which the Periodic Table is drawn and there is no single accepted form. We have the long and short forms, the eight-column short form *á la* Mendeleev, spiral, 3-dimensional forms, etc. IUPAC does not discourage the use of any alternate forms of the Periodic Table although it recommends the 18-column form. Numbering the columns and groups is another matter, since there has been considerable confusion between Arabic and Roman numerals and between A and B sub-groups. The A and B sub-group designation is opposite in Europe and the USA, although IUPAC ruled in 1970 in favour of using A for groups 1 to 10 and B for groups 11 to 18. The USA went their own way with the opposite usage. In addition, the main groups alone were often referred to as groups I to VIII in Roman numerals, to avoid the A and B confusion. In 1989, IUPAC proposed a uniform 18-column numbering system using Arabic numerals. After 10 years of debate and much resistance from traditionalists, this has now largely been accepted, and should reduce the confusion of what one means by group four, or group four A

Table 2.2 The Periodic Table and different numbering systems

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18[1]
I	II											III	IV	V	VI	VII	VIII[2]
IA	IIA	IIIA	IVA	V	VIA	VIIA		VIIIA		IB	IIB	IIIB	IVB	V	VIB	VIIB	VIIIB[3]
IA	IIA	IIIB	IVB	V	VIB	VIIB		VIIIB		IB	IIB	IIIA	IVA	V	VIA	VIIA	VIIIA[4]
H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	La-Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Ac-Lr	Rf	Db	Sg	Bh	Hs	Mt	Uun	Uub	Uuu						

[1]IUPAC, 1989.
[2]Main groups only.
[3]IUPAC, 1970.
[4]USA usage.

(Table 2.2). Old habits die hard, however, and group five will mean either V, Nb, Ta or N, P, As, Sb, Bi (new group 15), depending on the age of the chemist.

2.12 Naming groups, columns and rows

A number of groups have their own names which are in common usage. Thus, group 1 are the alkali metals (because they react with water to give alkalis); group 2 the alkaline earth metals (because they form alkaline oxides or earths); group 17 are the halogens (or salt-formers) and group 18 the noble or rare gases (the term inert gases was dropped in 1962 when their first compounds were made). Group 16 are sometimes called the chalcogens, which comes from the Greek *chalkos*, meaning copper or brass. This name is used because of the tendency of copper to occur in nature bound to the group 16 elements, particularly sulphur and oxygen. Chalcogens thus means copper-formers because copper was extracted from sulphide or oxide ores. Much less familiar is the name pnicogens for group 15. This is not common and many chemists have never heard or used it. The name comes from the Greek *pniktos*, meaning strangled or stifled. The reference being to nitrogen's inability to support life or combustion, also reflected in the German *Stickstoff* and the French *Azote*. The name pnicogen seems to be quite recent in coinage but was rejected by IUPAC in 1971. Groups 14 (the carbon group) and 13 (the boron group) don't appear to have attracted their own names, but I suppose there's still time.

The various blocks of the periodic table are often referred to as the s, p, d and f blocks, indicating the type of orbitals being filled. The d block

elements are also referred to variously as the transition elements, transition metals or transition series, as they come **between** the s and p blocks. The s and p blocks together are commonly referred to as the Main Group Elements or sometimes as the Typical Elements. With the f block elements, we come back to familiar names. The 4f series is referred to as the Lanthanides, but the variants Lanthanons and Lanthanoids have also been used. The 5f series are now usually known as the Actinides, although again the terms Actinons and Actinoids have been used. Elements beyond uranium are known as the transuranium, artificial or synthetic elements.

The elements are divided into two main classes, metals and non-metals. This would seem to have no place for confusion, but originally the non-metals were referred to as metalloids, a term now reserved for the elements lying on the borderline between the metals and non-metals. These halfway elements are also known as the semi-metals.

2.13 Spurious elements

Many elements have been discovered and named that either didn't exist or were previously known and named elements and were 'rediscovered'. The number of 'discovered' elements which later turned out to be spurious is probably greater than the existing list of elements! Some of these lost elements are: erythronium (Del Rio, 1801) now named vanadium (Sefstrom, 1830); virginium (Papish, 1931) and moldavium (Hulubei, 1936) were suggested as names for element 87 (eka-caesium) before it was finally called francium (Percy, 1939); element 61 was 'discovered' several times and given the names illinium (Hopkins, 1926), cyclonium and eventually promethium; didymium turned out to be a mixture of two rare earth elements, neodymium and praseodymium (Welsbach, 1885). Barium was called after barytes by H. Davy, a very dense mineral, but the metal has quite a low density and for a time a rival name plutonium was used (see Thomson, 1817), later to be applied to a very different element. Bromine (Gk. *bromos* = stench) was first called muride by Balard, bringing it in line with the other halogens. Osmium (from Gk. *osme* = smell) was first called ptene (Gk. *ptenos* = winged) because of its volatile, and smelly oxide, OsO_4 – the smell won. In 1828, Osann thought he had discovered three new metals in a metal sample from a Russian ore: he named them pluranium, polinium and ruthenium. The existence of a new element was confirmed by Klaus (1842) and the name ruthenium from *Ruthenia* = Russia, was accepted. As mentioned above G. Urbain thought he had discovered element 72 in 1911 and called it celtium. He was wrong and when element 72 was finally found it was called hafnium.

Thorium is a good example of an element that has had several names over the years [21]. In 1817, Berzelius thought he had found the oxide of

a new metal in gadolinite and he called it thorium, after the Scandinavian god Thor. He then decided he was wrong but in 1828 he isolated an oxide of thorium from another mineral and then isolated the metal itself. In 1851, Bergemann announced a new element in a thorium mineral, which he called donarium, but it was in fact, thorium. Bahr 'discovered' wasium in 1862 but he showed it to be thorium himself. Much later Baskerville thought he had found two new elements in a sample of thorium chloride and called them berzelium and carolinium, both spurious. The radioactivity of thorium was not discovered until 1898

There have been so many wrong 'discoveries' that we can't list them all, but as the means of identifying elements developed, particularly using the visible spectroscope and then the X-ray spectroscope, then the number of false positives drops, although disputes of priority still arise. Thus element 104 was called rutherfordium by the Americans and kurchatovium by the Russians – the name has just been decided in favour of the American claim to discovery. Unfortunately, most people outside Russia or the nuclear physics community had never heard of the Russian physicist Kurchatov.

2.14 No universal names

Unfortunately, although the symbols for the chemical are now universal, this is not true of the names which are different in different languages. Thus, iron is *fer* in French, *Eisen* in German. Sulphur is *soufre* in French, *Schwefel* in German, *azufre* in Spanish, *solfer* in Dutch, etc. The move towards standardization of names and their authorization by IUPAC means that the new names will be universally used whatever the language. For the older elements, there is no universal name and to read foreign chemical papers one must learn the language. There are also a few variant English spellings.

2.15 Variant English spellings and symbols

Although English is becoming the universal language of scientific communication there are still some differences between English and American spellings. The most obvious is sulphur and sulfur. The official IUPAC recommendation is now that sulfur should be used, despite the long history of English usage of sulphur. Americans persist in calling an element aluminum, although most of the world calls it aluminium. This confusion seems to go back to the identification of the element. Humphry Davy initially called it aluminum but changed this to aluminium, to agree with the developing convention that metals end in -ium. This is almost universally

true, the exceptions being molybdenum, tantalum and lanthanum (apart from metals named before 1800) and all new names follow the convention that metals end in -ium. The other variant spelling, again an American foible, is to simplify the original spelling of caesium to cesium thus losing its derivation from the Greek *caesius* for blue sky. None of these variants can be mistaken for anything else, unlike the persistence of the use of columbium (Cb) instead of niobium (Nb) in American textbooks until the 1950s and metallurgical publications still use it in America (see above). The French call nitrogen 'azote' and continued to use the symbol Az until quite recently. The Russians called element 104 'kurchatovium' (Ku) and the Americans called it 'rutherfordium' (Rf). Element 105 had been named 'hahnium' (Ha) and is now named 'dubnium' (Db). Lawrencium hasn't changed its number (103) but, in many older books, its symbol was Lw and it is now accepted as Lr.

Beryllium was originally called 'glucinum' (G or Gl) after the sweet taste of its salts. Both beryllium and glucinum were in use until 1924 when the Chemical Society in England decided on beryllium (Be), derived from the mineral beryl, which in turn came from the Greek *beryllos*. This is now accepted internationally.

Why do these variant names matter? In the case of cesium and sulfur or aluminum it is a matter of annoyance, as the symbols are common and they cannot be mistaken for anything else. Where names or symbols have changed it makes it very difficult to use older literature without making mistakes. There will undoubtedly be much confusion for a few years as the new names for the transuranium elements become familiar and the older books fall out of use.

2.16 IUPAC tries to bring order

The right of naming an element belongs by long tradition to the discoverer. The transuranic elements were made by three groups: in the USA at Berkeley, in Russia at Dubna, and in Germany at Darmstadt. In some cases more than one group claimed and named the same elements, thus creating confusion. The Cold War didn't help as it was matter of national pride that an element be claimed as an American or a Russian element. Thus, element 104 was known as both kurchatovium and rutherfordium. IUPAC stepped in to bring order and imposed interim systematic names for elements 104 onwards in 1970. Thus, these became known for several years as Unp, unnilpentium (105), etc. This is what happens when names are chosen by committees! The Committee for Nomenclature in Inorganic Chemistry (CNIC) of IUPAC was given the job of adjudicating the claims to first discovery, and dealing with the problem that some elements had names in common usage. However, the principle that the undisputed

discoverer should name an element has not been over-turned, it's just that deciding who made what could be quite difficult to decide. Apart from prior discovery, they also introduced an arbitrary rule in 1994 that no element could be named after a living scientist! Thus, the Americans were put out when their name for element 106, seaborgium (Sg), in honour of Glenn Seaborg, was rejected by CNIC in 1994 for this reason, in favour of rutherfordium, previously attached to element 104 (Table 2.3). This caused immense uproar, heated correspondence in the chemical press and the threat of the USA unilaterally using their own names. After many submissions and further deliberation on the matter, the committee came up with a compromise set of names in May 1997 and these were ratified by IUPAC's General Assembly in August 1997. The agreed list and symbols is given below in Table 2.3, balancing national pride with scientific claims. The table also shows the earlier suggestions.

The only one that has stayed the same throughout is element 109, meitnerium, named in honour of Lise Meitner. The story isn't finished as elements 110, 111 and 112 have all been reported and will presumably carry the 'systematic' names ununnilium, Uun; unununium, Uuu and ununbisium, Uub until IUPAC gets around to agreeing on new names. They were all made at Darmstadt in Germany and so there should be no international disputes in their naming, providing the claims are verified. Following an earlier convention they could be called eka-platinum, eka-gold and eka-mercury, as they belong to the fourth transition series (6d block).

Unlike the older elements, whose names are different in different languages by virtue of historical usage, the new names should be the same in every language.

2.17 Conclusion

In this chapter, I have attempted to look at some of the issues involved in the naming of the elements. Although we are only dealing with a relatively

Table 2.3 The evolution of the names and symbols for elements 104–109

	IUPAC (1970)		IUPAC (1994)		ACS (1994)		IUPAC (1997)	
104	Unnilquadium	Unq	Dubnium	Db	Rutherfordium	Rf	Rutherfordium	Rf
105	Unnilpentium	Unp	Joliotium	Jl	Hahnium	Ha	Dubnium	Db
106	Unnilhexium	Unh	Rutherfordium	Rf	Seaborgium	Sg	Seaborgium	Sg
107	Unnilseptium	Uns	Bohrium	Bh	Nielsbohrium	Ns	Bohrium	Bh
108	Ununoctium	Uno	Hahnium	Hn	Hassium	Hs	Hassium	Hs
109	Unnilennium	Une	Meitnerium	Mt	Meitnerium	Mt	Meitnerium	Mt
110	Unnilnilium	Unn						
111	Unununium	Uuu						
112	Ununbium	Uub						

small number of elements (112 at the last count), over the years there has been some degree of uncertainty and ambiguity in the naming of the elements, and in labelling the Periodic Table. The recent controversy between IUPAC and the creators of new elements shows that there is still life left in the old debates over priority and prestige and national pride. The symbols and labelling are now agreed, but each language has its own set of elemental names which introduces an unhelpful 'element' of confusion into the chemical literature. The swelling tide towards the universal use of English as a medium for scientific communication, both orally and in writing, will probably remove this last obstacle as well. Although we are dealing with simple substances, an idea that is the foundation of modern chemistry, the history of the development of the concept of an element, the identification of the elements and their ordering and naming, has hardly been a simple process.

Appendix 2.A Definitions of 'element' down the ages

Aristotle (384–322 BC)

He identified four 'simple bodies', fire, air, earth and water, of which all other substances are composed. There were also four elementary 'qualities', the moist, the dry, the hot, and the cold, which are combined in pairs in the simple bodies. Fire is hot and dry; air is hot and moist; water is cold and moist; earth is cold and dry. All compound bodies are composed of all the 'simple' bodies in different proportions. Behind the four 'elements' or 'simple bodies' lay a primitive matter – *prima materia* – which was common to all so that they can pass one into the other. This matter had to be combined with one of four different forms to become one of the four elements. The element of the stars was ether.

Paracelsus (1493–1541)

He stated that there were three elementary 'principles' in chemistry: salt (an earthy kind of substance), sulphur (the principle of combustion) and mercury (the liquid principle). A correct balance of the three brought health; an imbalance resulted in sickness.

Robert Boyle (1627–1691)

He rejected the four elements and the three principles. In *The Sceptical Chymist* (1661) Boyle defines elements as: "I mean by elements, as those

chymists that speak plainest do by their Principles, certain primitive and simple, or perfectly unmingled bodies; which not being made of any other bodies, or of one another, are the ingredients of which all those perfectly mixt bodies are immediately compounded, and into which they are ultimately resolved .... I must not look upon any body as a true principle or element, which is not perfectly homogeneous, but is further resolvable into any number of distinct substances."

Antoine Lavoisier (1743–1794)

He defined an element as follows: "A substance that cannot be split up by any known means into something simpler" or as chemicals that could neither be produced from other chemicals nor broken down into other chemicals. "The principal object of chemical experiment is to decompose natural bodies, so as separately to examine the different substances which enter into their composition ... Thus as chemistry advances towards perfection, by dividing and subdividing, it is impossible to say where it is to end; and these things we at present suppose simple may soon be found quite otherwise. All we dare venture to affirm of any substance is, that it must be considered as simple in the present state of our knowledge, and so far as chemical analysis has hitherto been able to show."

The modern definition

A substance which cannot be decomposed into simpler substances by chemical means or a substance where all the atoms have the same atomic number.

Appendix 2.B Discovery of the elements

Z	Name	Discoverer(s) and date discovered or isolated
1	Hydrogen	Isolated and identified by H. Cavendish 1766
2	Helium	Discovered independently by P. Janssen in 1868
3	Lithium	Discovered by J. A. Arfwedson in 1817 (isolated independently by W. T. Brande and H. Davy)
4	Beryllium	Oxide discovered by L-N. Vauquelin in 1798; isolated independently by F. Wohler and A. Bussy in 1828
5	Boron	Compounds known from antiquity; element isolated in 1808 by J-J. Gay-Lussac and L-J. Thenard in 1808, and independently by H. Davy
6	Carbon	Known as charcoal and diamond since antiquity

Z	Name	Discoverer(s) and date discovered or isolated
7	Nitrogen	Discovered by D. Rutherford in 1772; also discovered independently by J. Priestley, H. Cavendish and C. W. Scheele
8	Oxygen	Discovered around 1772 by C. W. Scheele and in 1774 by J. Priestley. A. Lavoisier identified it as an element and gave it its name 1775–1777.
9	Fluorine	Existence predicted early 17th century, first isolated by H. Moissan in 1886
10	Neon	Discovered by W. Ramsay and M. W. Travers in 1898
11	Sodium	Isolated and identified as an element by H. Davy in 1807
12	Magnesium	Isolated by H. Davy in 1808
13	Aluminium	Proposed by A. Lavoisier 1787, named by H. Davy 1807 and isolated by H. C. Oersted in 1825
14	Silicon	Isolated and identified by J. J. Berzelius in 1824
15	Phosphorus	Isolated by H. Brand in 1669
16	Sulphur	Known from antiquity; classified as an element by A. Lavoisier 1777
17	Chlorine	Made by C. W. Scheele 1774; shown to be an element by H. Davy 1810
18	Argon	Isolated and identified by Lord Rayleigh and W. Ramsay in 1894
19	Potassium	Discovered and isolated by H. Davy 1807
20	Calcium	Isolated and identified by H. Davy 1808
21	Scandium	Predicted by Mendeleev 1871; confirmed by L. F. Nilson 1879 (as oxide)
22	Titanium	Discovered by W. Gregor in 1791; isolated and purified by M. A. Hunter in 1910
23	Vanadium	Discovered by A. M. del Rio in 1801; rediscovered by N. G. Sefstrom in 1830; isolated by H. E. Roscoe 1867
24	Chromium	Discovered by L-N. Vauquelin 1797, isolated 1798
25	Manganese	Recognized as an element by C. W. Scheele 1774, isolated later that year by J. G. Gahn
26	Iron	Identified and used since antiquity
27	Cobalt	Isolated by G. Brandt 1739; compounds known since antiquity
28	Nickel	Isolated by A. F. Cronstedt 1751
29	Copper	Known from antiquity
30	Zinc	Compounds known since antiquity; metal used since Middle Ages; first production in the West 1738 by W. Champion
31	Gallium	Existence predicted by Mendeleev; discovered and isolated by P-E. Lecoq de Boisbaudran in 1875

Z	Name	Discoverer(s) and date discovered or isolated
32	Germanium	Existence predicted by Mendeleev; discovered by C. Winkler 1886
33	Arsenic	Possibly identified by Albertus Magnus in 1250
34	Selenium	Identified by J. J. Berzelius 1818
35	Bromine	Isolated by A-J. Balard in 1826
36	Krypton	Discovered by W. Ramsay and M. W. Travers in 1898
37	Rubidium	Discovered by R. Bunsen and G. Kirchoff in 1861
38	Strontium	Existence noted by A. Crawford 1790; isolated by H. Davy 1808
39	Yttrium	Discovered by J. Gadolin 1789; isolated by F. Wohler 1828
40	Zirconium	Identified by M. H. Klaproth 1789; isolated by J. J. Berzelius 1824
41	Niobium	Discovered by C. Hatchett 1801 and named columbium; rediscovered by H. Rose in 1844 and named niobium
42	Molybdenum	Identified by C. W. Scheele 1778, isolated by P. J. Hjelm 1781
43	Technetium	Discovered by E. G. Segre and C. Perrier in 1937
44	Ruthenium	Existence predicted by G. W. Osann in 1828; isolated by K. K. Klaus in 1844
45	Rhodium	Discovered by W. H. Wollaston in 1803
46	Palladium	Discovered by W. H. Wollaston in 1803
47	Silver	Known from antiquity
48	Cadmium	Discovered by F. Strohmeyer in 1817, and independently in 1817 by K. S. L. Hermann and J. C. H. Roloff
49	Indium	Discovered by F. Reich and T. Richter in 1863
50	Tin	Known from antiquity
51	Antimony	Compounds known from antiquity; element known by early 17th century
52	Tellurium	Discovered by F. J. Muller in 1782
53	Iodine	Discovered by B. Courtois in 1811
54	Xenon	Discovered by W. Ramsay and M. W. Travers in 1898
55	Caesium	Discovered by R. Bunsen and G. Kirchoff 1860
56	Barium	Isolated by H. Davy 1808
57	Lanthanum	Identified as a rare earth by C. G. Mosander in 1839
58	Cerium	Oxide discovered by J. J. Berzelius and W. Hisinger in 1803, and independently by M. Klaproth.
59	Praseodymium	Isolated and identified by C. A. von Welsbach in 1885
60	Neodymium	Discovered by C. A. von Welsbach in 1885
61	Promethium	Predicted 1912, existence confirmed by J. A. Marinsky, L. E. Glendenin and C. D. Coryell in 1947
62	Samarium	Isolated and identified by Lecoq de Boisbaudran in 1879

Z	Name	Discoverer(s) and date discovered or isolated
63	Europium	Discovered by E-A. Demarcay in 1896 and isolated in 1901
64	Gadolinium	Discovered by J. de Marignac (1880) and isolated by Lecoq de Boisbaudran (1886)
65	Terbium	Discovered by G. Mosander in 1843
66	Dysprosium	Discovered by Lecoq de Boisbaudran in 1886, isolated by G. Urbain in 1906
67	Holmium	Discovered by P. T. Cleve in 1879
68	Erbium	Discovered by C. G. Mosander 1843
69	Thulium	Discovered by P. T. Cleve in 1879
70	Ytterbium	Discovered by J-C-G. de Marignac in 1878
71	Lutetium	Discovered independently by C. A. von Welsbach and G. Urbain in 1907–1908
72	Hafnium	Discovered by D. Coster and G. C. de Hevesy in 1923
73	Tantalum	Discovered by A. G. Ekenberg in 1802
74	Tungsten	First isolated by J. J. and F. Elhuyar in 1783
75	Rhenium	Discovered by I. and W. Noddack and O. C. Berg in 1925
76	Osmium	Discovered by S. Tennant in 1803
77	Iridium	Discovered by S. Tennant in 1803
78	Platinum	Discovered independently by A. de Ulloa in 1735 and C. Wood around 1741
79	Gold	Known since antiquity
80	Mercury	Known since antiquity
81	Thallium	Discovered by W. Crookes in 1861
82	Lead	Known since antiquity
83	Bismuth	Isolated by B. Valentine in 1450
84	Polonium	Discovered by M. Curie in 1898
85	Astatine	Made in 1940 by D. R. Corson, K. R. MacKenzie and E. Segre (USA)
86	Radon	Discovered by F. E. Dorn in 1900
87	Francium	Discovered by M. Percy in 1939
88	Radium	Discovered by P. and M. Curie 1898
89	Actinium	Discovered by A-L. Debierne in 1899, independently by O. Giesel in 1902
90	Thorium	Discovered by J. J. Berzelius 1828
91	Protactinium	Discovered by K. Fajans and O. H. Gohring in 1913; isolated by A. V. Grosse in 1934
92	Uranium	Discovered by M. Klaproth in 1789 and isolated by E-M. Peligot in 1841
93	Neptunium	Discovered by E. M. McMillan and P. H. Abelson in 1940
94	Plutonium	Discovered by G. T. Seaborg in 1941

Z	Name	Discoverer(s) and date discovered or isolated
95	Americium	Discovered by G. T. Seaborg, R. A. James, L. O. Morgan and A. Ghiorso in 1944
96	Curium	Discovered by G. T. Seaborg, R. A. James and A. Ghiorso in 1944
97	Berkelium	Discovered by S. G. Thompson, A. Ghiorso and G. T. Seaborg in 1949
98	Californium	Discovered by S. G. Thompson, K. Street, A. Ghiorso and G. T. Seaborg in 1950
99	Einsteinium	Discovered by A. Ghiorso *et al.* in 1952
100	Fermium	Discovered by A. Ghiorso *et al.* in 1952
101	Mendelevium	Discovered by A. Ghiorso, S. G. Thompson, G. T. Seaborg and others 1955
102	Nobelium	Discovered by A. Ghiorso, G. T. Seaborg *et al.*, 1958
103	Lawrencium	Discovered by A. Ghiorso, T. Sikkland, A. E. Larsh and R. M. Latimer in 1961
104	Rutherfordium	Synthesis reported by Dubna 1964 and Berkeley 1969; US claim upheld and name confirmed 1997
105	Dubnium	Synthesis reported by Dubna 1967 and Berkelely 1970; name confirmed 1997
106	Seaborgium	Synthesis reported by Dubna and Berkeley in 1974; US claim upheld and element name confirmed in 1997
107	Bohrium	Synthesis reported by Russians in 1976 and Germans in 1981
108	Hassium	Synthesis reported by Germans in 1984
109	Meitnerium	Synthesis reported by Germans in 1982
110	Ununnilium	Synthesis reported by Germans in 1994
111	Unununium	Synthesis reported by Germans in 1994
112	Ununbisium	Synthesis reported by Germans in 1996

Appendix 2.C Chronology of the discovery of the elements

Order	Date	Name	Symbol	Origin of name
	Antiquity:			
1		Carbon	C	L., *carbo* = coal
2		Sulphur	S	
3		Gold	Au	L., *aurum*
4		Copper	Cu	L., *cuprum* (Cyprus)
5		Silver	Ag	L., *argentum*
6		Lead	Pb	L., *plumbum*
7		Iron	Fe	L., *ferrum*
8		Tin	Sn	L., *stannum*

Order	Date	Name	Symbol	Origin of name
	Antiquity:			
9		Mercury	Hg	L., *hydrargyrum* = liq. Silver; Mercury
10	Middle Ages	Zinc	Zn	Ger., *Zink*
11	1250	Arsenic	As	Gk., *arsenikon* =yellow orpiment
12	1450	Bismuth	Bi	Ger., *bisemetum*
13	C17th?	Antimony	Sb	L., *stibium*; Gk., *Anti+monos* = not alone
14	1669	Phosphorus	P	Gk., *Phosphros* = bringer of light
15	1735	Platinum	Pt	Sp., *platina* = little silver
16	1739	Cobalt	Co	Ger., *Kobold* = goblin
17	1751	Nickel	Ni	Ger., *Nickel* = Satan
18	1766	Hydrogen	H	Gk., *Hydro genes* = water forming
19	1772	Nitrogen	N	Gk., *Nitron genes* = nitre forming
20	1772–1774	Oxygen	O	Gk., *oxy genes* = acid forming
21	1774	Chlorine	Cl	Gk., *Chloros* = pale green
22	1774	Manganese	Mn	L., *Magnes* = magnet
23	1778	Molybdenum (isolated 1781)	Mo	Gk., *molybdos* = lead
24	1782	Tellurium	Te	L., *Tellus* = earth
25	1783	Tungsten	W	L., *Wolframium*
26	1789	Yttrium	Y	Ytterby
27	1789	Zirconium (isolated 1824)	Zr	Ar., *zargun* = gold
28	1789	**Uranium** (isolated 1841)	U	Uranus
29	1790	Strontium (isolated 1808)	Sr	Strontian, Scotland
30	1791	Titanium	Ti	Titans
31	1797	Chromium	Cr	Gk., *Chroma* = colour
32	1798	Beryllium (isolated 1828)	Be	Gk., *Beryllos* = Beryl
33	1801	Vanadium	V	Vanadis
34	1801	Niobium	Nb	Niobe
35	1802	Tantalum	Ta	Tantalus
36	1803	Palladium	Pd	Pallas (asteroid)
37	1803	Rhodium	Rh	Gk., *Rhodon* = rose
38	1803	Cerium	Ce	Ceres (asteroid)
39	1803	Osmium	Os	Gk., *Osme* = smell
40	1803	Iridium	Ir	L., *Iris* = rainbow
41	1807	Sodium	Na	L., *Natrium*

Order	Date	Name	Symbol	Origin of name
	Antiquity:			
42	1807	Potassium	K	L., *Kalium*
43	1808	Boron isolated	B	Ar., *buraq* = borax
44	1808	Magnesium	Mg	Gk., *Magnesia* (place)
45	1808	Calcium	Ca	L., *Calx* =lime
46	1808	Barium	Ba	Barytes
47	1811	Iodine	I	Gk., *Iodes* = violet
48	1817	Lithium	Li	Gk., *Lithos* = stone
49	1817	Cadmium	Cd	L., *Cadmia* = calamine
50	1818	Selenium	Se	Gk., *Selene* = moon
51	1824	Silicon	Si	L., *Silicis* = flint
52	1825	Aluminium (isolated 1825)	Al	L., *Alumen* = Alum
53	1826	Bromine	Br	Gk., *Bromos* = stench
54	1828	Ruthenium (isolated 1844)	Ru	L., *Ruthenia* = Russia
55	1828	**Thorium**	Th	Thor
56	1839	Lanthanum	La	Gk., *Lanthanein* = to lie hidden
57	1843	Terbium	Tb	Ytterby
58	1843	Erbium	Er	Ytterby
59	1860	Caesium	Cs	L., *Caesius* = sky-blue
60	1861	Rubidium	Rb	L., *Rubidius* = deep red
61	1861	Thallium	Tl	Gk., *Thallos* = green twig
62	1863	Indium	In	Indigo
63	1868	Helium	He	Gk., *Helios* = sun
64	1875	Gallium	Ga	L., *Gallia* = France
65	1878	Ytterbium	Yb	Ytterby
66	1879	Scandium	Sc	L., *Scandia* = Scandinavia
67	1879	Samarium	Sm	Samarskite (mineral)
68	1879	Holmium	Ho	L., *Holmia* = Stockholm
69	1879	Thulium	Tm	Thule = Scandinavia
70	1880	Gadolinium (isolated 1886)	Gd	J. Gadolin/gadolinite
71	1885	Praseodymium	Pr	Gk., *prasios didymos* = green twin
72	1885	Neodymium	Nd	Gk., *neos didymos* = new twin
73	1886	Fluorine	F	L., *fluere* = to flow
74	1886	Germanium	Ge	Germany
75	1886	Dysprosium	Dy	Gk., *dysprositos* =hard to obtain
76	1894	Argon	Ar	Gk., *argos* = inactive
77	1896	Europium	Eu	Europe
78	1898	Krypton	Kr	Gk., *kyrptos* = hidden
79	1898	Neon	Ne	Gk., *neos* = new

Order	Date Antiquity:	Name	Symbol	Origin of name
80	1898	Xenon	Xe	Gk., *xenos* = stranger
81	1898	**Radium**	Ra	L., *radius* = ray
82	1898	**Polonium**	Po	Poland
83	1899	**Actinium**	Ac	Gk., *aktinos* = ray
84	1900	**Radon**	Rn	Gas from radium
85	1907	Lutetium	Lu	L., *Lutetia* = Paris
86	1913	**Protactinium**	Pa	Gk., *protos* = first + actinium
87	1923	Hafnium	Hf	L., *Hafnia* = Copenhagen
88	1925	Rhenium	Re	L., *Rhenus* = Rhine
89	1937	**Technetium**	Tc	Gk., *technikos* = artificial
90	1939	**Francium**	Fr	France
91	1940	**Astatine**	At	Gk., *astatos* = unstable
92	1940	**Neptunium**	Np	Neptune
93	1941	**Plutonium**	Pu	Pluto
94	1944	**Americium**	Am	America
95	1944	**Curium**	Cm	Pierre and Marie Curie
96	1947	**Promethium**	Pm	Prometheus
97	1949	**Berkelium**	Bk	Berkeley
98	1950	**Californium**	Cf	California
99	1952	**Einsteinium**	Es	Albert Einstein
100	1952	**Fermium**	Fm	Enrico Fermi
101	1955	**Mendelevium**	Md	Dimitri Mendeleev
102	1958	**Nobelium**	No	Alfred Nobel
103	1961	**Lawrencium**	Lr	Ernest Lawrence
104	1969	**Rutherfordium**	Rf	Ernest Rutherford
105	1967/70	**Dubnium**	Db	Dubna
106	1974	**Seaborgium**	Sg	Glenn Seaborg
107	1976/81	**Bohrium**	Bh	Niels Bohr
108	1982	**Meitnerium**	Mt	Lise Meitner
109	1984	**Hassium**	Hs	Hesse
110	1994	**unnamed**		
111	1994	**unnamed**		
112	1996	**unnamed**		

Names in bold type are radioactive elements.

References

1. Quoted in Weeks, M. (1968) *The Discovery of the Elements*, 7th edn, Chemical Education Publishing Co., New York.
2. Jaffe, B. (1972) *Moseley and the Numbering of the Elements*, Science Studies Series, Heinemann, London.

3. Ringnes, V.V. (1989) Origin of the names of the chemical elements. *J. Chem. Educ.*, **66**, 731–738.
4. Hogben, L. (1969) *The Vocabulary of Science*, Heinemann, London, p. 36.
5. Brock, W. H. (1992) *The Fontana History of Chemistry*, Fontana, HarperCollins, London, p. 116.
6. Crosland, M. P. (1962) *Historical Studies in the Language of Chemistry*, Heinemann, London, p. 275.
7. Crosland, M. P. (1962) *Historical Studies in the Language of Chemistry*, Heinemann, London, p. 276.
8. Davis, H. M. (1961) *The Chemical Elements*, Ballantine Books, New York, p. 11.
9. Jaffe, B. (1972) *Moseley and the Numbering of the Elements*, Science Studies Series, Heinemann, London, p.93.
10. Jaffe, B. (1972) *Moseley and the Numbering of the Elements*, Science Studies Series, Heinemann, London, p. 112.
11. Jaffe, B. (1972) *Moseley and the Numbering of the Elements*, Science Studies Series, Heinemann, London, p. 117–118.
12. Jaffe, B. (1972) *Moseley and the Numbering of the Elements*, Science Studies Series, Heinemann, London, p. 119.
13. Jaffe, B. (1972) *Moseley and the Numbering of the Elements*, Science Studies Series, Heinemann, London, p. 141.
14. Davis, H. M. (1961) *The Chemical Elements*, Ballantine Books, New York, p. 28.
15. Ringnes, V. V. *op. cit.*
16. Davy, H. (1808) *Phil.Trans.* **XCVIII**, 346.
17. Cotzias, G. C. quoted in Weiss, S. A. (1977) *Manganese: the other uses*, Metal Bulletin Books.
18. Scheele, C. W. (1786) *Chemical Essays*, J. Murray, London.
19. Smith, J. (1816) *Panorama of Arts and Science*.
20. Friend, J. N. (1961) *Man and the Chemical Elements*, Charles Griffin, London, pp. 239–241.
21. Friend, J. N. (1961) *Man and the Chemical Elements*, Charles Griffin, London, p. 234.

3 Chemical Abstracts Service Chemical Substance Index nomenclature

D. W. WEISGERBER

3.1 Introduction

Suppose I am a chemist interested in learning more about the use of ethanolamine in detergent compositions and I have decided to look in *Chemical Abstracts (CA)* for citations to journal articles or patents that discuss such applications. Where should I begin to look?

Even as simple a compound as ethanolamine, $H_2NCH_2CH_2OH$, can be named in many different ways – 2-aminoethanol, β-aminoethanol, 2-aminoethyl alcohol, β-aminoethyl alcohol, 2-hydroxyethylamine, β-hydroxyethylamine, 2-hydroxyethanamine, and 1-amino-2-hydroxyethane – all of which describe its structure unambiguously. Trivial names such as ethanolamine, monoethanolamine, and colamine are names that might also be applied to this substance. Under which name should I look in the *Chemical Abstracts* index, or must I look under all of the possibilities?

Suppose now that I am a *Chemical Abstracts* indexer faced with a similar question. How should this same compound, $H_2NCH_2CH_2OH$, be named for indexing purposes? What would be the best name to use in an index, and why? Whereas the author of a paper might use any of the names mentioned above since any one would most likely be understood and be generally unambiguous (with the possible exception of colamine), I, as an indexer, want to assign a single preferred name to each substance, one that will consistently and reliably guide the reader to the appropriate references.

Chemical Abstracts Service (CAS) indexers face this kind of question repeatedly, often hundreds of times each day. In 1997, CAS indexers had to name more than one million new chemical substances that they had not encountered before. They also indexed more than two million substances that they had encountered previously, often hundreds of times. How can all those substances be named reliably, consistently, and efficiently?

This chapter will review the history of CAS, the special challenges, nature, and evolution of its chemical substance indexing and the CAS Chemical Registry that began as a computer support tool for indexing and became a world-wide authority for chemical substance identification.

3.2 Mission of CAS

CAS, a division of the American Chemical Society (ACS), located in Columbus, Ohio, began essentially as an American information service in 1907. Today, it is international in the primary literature sources that it covers, the databases it creates, the audience it serves, and the reach of its printed, microform, and electronic services.

The mission of CAS today, as it has been for over 90 years, is to promote the advancement of science and technology through a conscientious effort to meet the information needs of the world-wide scientific community. CAS seeks to accomplish this by selecting, reviewing, abstracting, indexing, and providing comprehensive access to the world's chemical literature – journals, conference and symposia proceedings, technical reports, patents, dissertations, and books.

Indexing the subject matter of chemical literature by CAS has been the key to providing efficient access to this literature, and the major focus of this indexing is the chemical substances described in the original reports. Chemistry is a unique science in that it benefited early from a relatively systematic and international language useful for the exchange of information – the chemical molecular structure. CAS chemical substances indexing has been built around those structures and their corresponding systematic nomenclature descriptions.

3.3 History of CAS

The history of CAS goes back to 1895 when Arthur A. Noyes, a patriarch of a distinguished family of American chemists, founded the *Review of American Chemical Research* at the Massachusetts Institute of Technology. Noyes felt that American chemists were not being recognized adequately for their accomplishments – German chemists seemed to be getting all the attention – and he decided to publish summaries or abstracts of American chemical research papers.

In 1897 the *Review of American Chemical Research* became a part of another publication, the *Journal of the American Chemical Society*. William A. Noyes, Sr, a distant cousin of Arthur Noyes, was editor of the *Journal of the American Chemical Society*. He argued strongly that simply publishing these abstracts as a supplement to the journal was not enough and that the American Chemical Society (ACS) should publish a more comprehensive abstracting journal in the field of chemistry. In 1906, the Society authorized publication of *Chemical Abstracts (CA)* and charged it with the mission of abstracting the world's literature of chemistry. CA began publication in 1907 with William Noyes as the first Editor and a budget of $15 500. Just under 12 000 abstracts were published in the first year.

Noyes edited *CA* through its first 2 years, working first from a corner of a laboratory at the National Bureau of Standards in Washington and then from the University of Illinois after he became chairman of the chemistry department in late 1907. He was succeeded as Editor of *CA* by Austin M. Patterson in 1909. Dr Patterson had joined Noyes as Associate Editor in 1908 and brought with him a strong literary focus having been editor for chemistry of *Webster's New International Dictionary* since 1903.

At the invitation of the head of the chemistry department at The Ohio State University (OSU), the *CA* editorial office was moved to Columbus and the OSU campus so that Patterson could be nearer the family home in Xenia, Ohio. Patterson remained Editor until 1914 when poor health forced him to retire, but returned in 1916 to assist with the compilation of the first *CA Decennial Index*.

The next four decades in the history of *CA* are those of Evan J. Crane. Crane was Editor from 1915 until 1958. Crane had the additional title of Director of CAS for the last 2 years. He developed and nurtured *CA* through some difficult financial times, and built *CA* into a model for other scientific abstracting and indexing services. He established the *CA* indexes as an equally if not even more important part of the service than the abstracts.

The *CA* editorial office occupied various buildings on the OSU campus for almost 60 years. In 1965, CAS moved into its own building on 50 acres just north of the OSU campus. A second adjoining building was added in 1973.

Until 1956, *CA* production expenses were funded in part from dues paid by the individual members of the ACS. In 1956, Chemical Abstracts Service became established as a self-supporting division of the Society. All CAS operations are now supported through fees charged for its publications and services.

In the beginning, *CA* was basically an abstracting service, with the abstracts provided by volunteers world-wide and the editing and indexing done by a limited full-time staff. Today all abstracts are prepared by CAS editorial staff in Columbus, or by special input center groups in Berlin and Tokyo. This shift away from volunteer abstractors to in-house staff occurred in the 1960s. It resulted from the need for greater timeliness in abstracting and indexing and the introduction of computer processing of the data.

Central to the development and growth of CAS has been the evolution from the production of a printed service to the generation of a computer-readable database from which a variety of printed, microform, and electronic services can be readily derived. The computer-based production system developed by CAS in the late 1960s and early 1970s that supported this growth and the development of new services was primarily intended to provide CAS with the ability to cope with the increasing

volume of the chemical literature. The system consisted of essentially three components: integrated, computer-assisted input of all information comprising *Chemical Abstracts*, computer databases to store and organize the information and a system to package the information in ways most useful to information users.

Perhaps the most far-reaching development to come out of CAS's work on mechanized information handling, the CAS Chemical Registry System, was installed in 1965. The CAS Chemical Registry System is a computer-based system that uniquely identifies chemical substances on the basis of their molecular structures. This system will be discussed in more detail later in this chapter.

In 1968 the first computer-readable service to cover the full range of documents abstracted by *CA*, *CA Condensates*, was introduced. *CA Condensates* contained the full range of bibliographic information, plus the natural-language keyword indexing. This was followed in the early 1970s by the *CA Subject Index Alert*, a biweekly computer-readable file that included segments of the computer file of in-depth controlled-vocabulary index entries. *CA Search*, introduced in 1978, combined all of the information available in *CA Condensates* and the *Subject Index Alert* into a single file. *CA Search* has become the most widely used file of chemical information in the world and is the basis for the offerings of several online search and retrieval services. The introductions of these latter services followed directly from the development and extensions of the CAS computer-based publication system.

Beginning in the late 1960s, CAS started to license some of its computer-readable files to a number of organizations for the purpose of their providing information services based on local batch searching of the files. In the early 1970s, remote online access was extended to some of these files.

In 1980, CAS introduced its own online service, CAS Online. The initial offering provided an online substructure search of the CAS Chemical Registry database. This online service continued to grow and expand to include the full *CA* bibliographic, abstract, and index information.

A significant accomplishment of CAS research and development efforts during the 1980s was the development of Messenger. This versatile software was designed to permit online searching in a wide variety of technical databases and became the foundation of STN International, an online network today offering access to more than 200 international scientific and technical databases. STN International is an online search service offered jointly by the American Chemical Society, FIZ Karlsruhe, and the Japan Science and Technology Corporation (JST). It links together the CAS computers in Columbus, the FIZ computers in Karlsruhe, Germany, and the JST computers in Tokyo.

CAS provides a wide range of printed, microform, and computer-readable chemical information services. *Chemical Abstracts* is still the principal

printed service. But there are others such as *Chemical Industry Notes (CIN),* containing chemical business information, *CA Selects* and *CAS BioTech Updates*, containing selected abstracts in special limited subject areas such as organofluorine chemistry and biosensors, and *Chemical Titles* listing just the titles of articles from the core chemical journals.

Just as important as the printed products are the computer-based services. Today, CAS offers access to many databases, including the CA File, CAplus, Registry File, CAOLD File, CASREACT, MARPAT, CIN, CHEMLIST, and CHEMCATS, which are available for searching on STN International. *Chemical Abstracts*, the *CAS Source Index (CASSI),* and the *CASurveyor* current awareness services are also now available on CD-ROM.

In October 1994, CAS announced SciFinder, a new generation research tool to assist scientists and researchers world-wide with access to the CAS databases. SciFinder, a client-server application which works on Macintosh or Windows desktop computer systems, places information ranging from chemical structures to chemical-related literature at the fingertips of scientists who have no or little online search expertise. SciFinder Scholar, introduced in 1997, extended this application to the academic community by enabling campus-wide access to the CAS databases.

CAS recently introduced several Web-based products that expand access to chemical information. STN Easy is an interface to STN, featuring point-and-click access to selected STN databases. Chemical Patents Plus provides easy, cost-effective access to text and full-page images for all classes of US patents issued since 1975, enhanced with *CA* indexing. FirstSearch: CA Student Edition, a joint effort of CAS and OCLC, is an information product customized to serve the chemical information needs of undergraduate students.

3.4 Chemical Substance Indexes

Chemical substances, their structures, their syntheses, their properties, and their applications, are the core of chemistry and the main occupation of chemists. Chemists build on the experiences of their many predecessors as recorded in the primary and secondary literature, and in compendia, textbooks, and handbooks, sometimes referred to as tertiary literature. From the very beginning, recording, storing, and retrieving information on chemical substances have been paramount to progress in chemistry. CAS indexing has sought to ensure the efficiency of this information retrieval and utilization.

The selection of entries for the *CA* Subject Indexes – the *Chemical Substance Index* and the *General Subject Index* – is the most important function performed by the CAS information analyst. (Prior to 1972, chemical

substances and general subjects were combined in a single index, the *Subject Index*.) Indexes serve to guide users to the information that they seek. At the same time, depending on the depth of indexing, they provide the searcher with a basis for judging relevance before seeking the particular abstracts or original documents. In large measure, access to the chemical literature by users of CAS services depends critically on the index entries chosen by the CAS analysts.

'Indexing' is the intellectual process of selecting the ideas that reflect the important and novel aspects of the original reports, and then translating those ideas (both chemical substances and concepts) into the appropriate controlled-vocabulary index terms, and the formulation of the accompanying text that describes the contexts of studies as they relate to the headings. The purpose of these index entries is to provide access to literature citations and abstracts, and thereby to the technical contents of the original documents themselves. CAS has long had the philosophy of indexing subjects from the complete original documents and not solely from the abstracts. In some cases, hundreds of substances are indexed, whereas the abstract may only cite a few or identify them generically.

The driving force in CAS indexing is always author intent, i.e. the main purpose of the author for carrying out the work reported. Statements of intent are usually expressed by authors in the titles, introductions, and conclusions of their papers. The index entries reflect the author intent. They may be either very specific or quite general, or both, depending on the author's purpose. While it has long been CAS indexing practice to select the most specific substance(s) or concept(s) being studied, this does not override an author's stated intent.

To be selected for CAS indexing, a particular substance should itself be new or have new information reported about it. Novelty is usually recognized on the basis of author statements or emphasis. In the absence of references to the previous literature, the work is presumed to be new. New information can take such forms as:

- structure elucidation,
- sources or preparative methods,
- reactions that are newly described or are carried out under conditions that differ substantially from those that had been used before,
- reaction kinetics or mechanisms,
- properties or processes (physicochemical, biochemical, mechanical, etc.),
- biological effects, or
- uses or applications.

Specificity of indexing refers to the extent to which a substance in a document is identified by a precise index name. The general rule is to index chemical substances to the maximum degree of specificity that can be derived from the original document, including:

- specific locants,
- complete stereochemical information,
- definite molecular formulas, and
- specific substituents rather than class designations.

Mere mention of a substance or a potential use or application does not in itself constitute sufficient justification for indexing in *CA*. To be indexed, new and useful data about that compound or use should appear in the original document. The novelty claimed must be substantiated by some documented data. Such data can include elemental analyses, physical data such as melting points, boiling points, and yields, spectral and other types of optical, electric, magnetic, and thermodynamic information, reaction rates, biological activities and documented uses.

Products of chemical syntheses may be characterized by a variety of data, any one of which suffices to justify indexing them:

- yield;
- a single physical constant for a new compound – typical constants include melting point, boiling point, optical rotation and refractive index;
- R_f: chromatographic data (e.g. separation factors, retention times); such data should be accompanied by some separation conditions (e.g. identification of mobile or stationary phases);
- IR, NMR and mass spectroscopic results;
- chemical analysis – results of combustion analysis either for carbon and hydrogen or for nitrogen (or any other heteroatom).

E. J. Crane was the first editor of *CA* to decide that indexes were a highly important part of an abstracting service. Somewhat meager subject indexes were published from the beginning of *CA* and serious attention was not given to indexing until 1916 when it was decided to publish the first decennial index to *CA*, cumulating the indexing of the first 10 years. Crane personally did much of the work on this first collective index, including soliciting enough advance subscriptions to convince the ACS Board that the publication would be a financial success.

3.5 Chemical Substance Index nomenclature

In the first nine volumes of *CA* (1907–1915), chemical substances were usually indexed using the names that the authors had used for them in the original papers. This sometimes resulted in the same substance being indexed under two or more different names, and under names that a user might never have anticipated. An example of the indexing of chemical substances in these early volumes is shown in Figure 3.1.

Methyl-*p*-phenetidine (Wedekind, Fröhlich), 1553.
***p*-Methylphenylacetaldehyde** and semicarbazone (Auwers), 317.
Methylphenylacetone (Tiffeneau), 328.
Methylphenylamino - *d* - *p* - bromphenylpyrrodiazole (Stolle), 2479.
Methylphenylaminodiphenylpyrrodiazole (Stolle), 2479.
***p*-Methylphenylaminopropylmethylcarbnol** (Markwalder), 2477.
***p*-Methylphenylaminopropylmethylketone** and picrate, oxime hydrochloride, and oxime (Markwalder), 2477.
1 - Methyl - 3 - phenyl - 4 - amino - 5 - pyrazolone benzylidene deriv. hydrochloride, furfuraldehyde, and cinnamic aldehyde condensation product, *o*-hydroxybenzylidene, *p*-methoxybenzylidene, *p*-nitrobenzylidene deriv. (Michaelis, Wrede), 1856.
1 - Methyl - 3 - phenyl - 4 - azobenzene - 5 - pyrazolone (Michaelis, Dorn), 1853.
1 - Methyl - 3 - phenyl - 4 - brom - 5 - chlorpyrazole and perbromide (Michaelis, Dorn), 1854.
3 - Methyl - 5 - phenyl - 1 - *p* - bromphenyl - 4 - [α - cyan - 4 - nitrobenzalamino] - pyrazole (Sachs, Alsleben), 1289.
1 - Methyl - 3 - phenyl - 5 - chlorpyrazole hydrochloride, chlorplatinate, methyliodide and chloride addition product, periodide (Michaelis, Dorn), 1853.
1 - Methyl - 5 - phenyl - 3 - chlorpyrazole (Michaelis, Dorn), 1854.
3 - Methyl - 5 - phenyl - 4 - [α - cyano - 5 - nitro-benzalamino] - pyrazole, and alkali salts (Sachs, Alsleben), 1289.
1 - Methyl - 3 - phenyl - 4,5 - dichlorpyrazole (Michaelis, Dorn), 1854.
1 - Methyl - 3 - phenyl - 4 - [2,4 - dinitrobenzalamino]-pyrazole, and alkali salts (Sachs, Alsleben), 1289.
Methylphenyldithiocarbamate, phenyl (Rivier), 3005.

Figure 3.1 Examples of substance index entries from *CA Volume 1 Subject Index* (1907). (Copyright American Chemical Society.)

Once it was decided to publish a 10-year cumulative index, it soon became evident that some systematic means of naming and indexing substances should be used to avoid scattering of references to the same substance at various names in the index. Grouping of related substances was also identified as an important objective.

In the preparation of an original scientific paper, an author usually assigns names to the substances that serve to highlight those features that are of major interest to the author. For example, the simple compound $HOCH_2CH_2NH_2$ mentioned earlier may be named: (1) as an alcohol derivative – 2-aminoethanol, 2-aminoethyl alcohol, (2) an amine derivative – 2-hydroxyethylamine, β-hydroxyethylamine, 2-hydroxyethanamine, (3) an ethane derivative – 1-amino-2-hydroxyethane, or (4) by a 'trivial' name – ethanolamine or colamine. The author's choice of name is generally based on consideration of the entire series of substances in the paper and not on individual members.

This approach to naming works fine in a scientific paper or presentation as long as the names are unambiguous and precisely identify the substances. However, in an alphabetically arranged index, names of substances must be based on a systematic nomenclature scheme that results in the assignment of one name, and only one, for each chemical structure, and the name must be predictable to the user. Only this will assure maximum usefulness of the index by bringing all studies on the same substance together at one name in the index; otherwise, the information regarding the substances will be scattered in the index.

A systematic method of naming and indexing substances in *CA* was devised in 1916, primarily by Carleton E. Curran and former *CA* Editor Austin Patterson, whose health had improved and who returned to serve as a highly valued consultant for *CA*. This system was used for the first time with the Volume 10 (1916) subject index. It has stood the test of time and had a profound effect on chemical nomenclature in general. The *CA* offices subsequently became an international center for nomenclature development.

Patterson became recognized as a world authority on chemical nomenclature and is considered by many as the father of organic nomenclature as we know it today. Crane later described Patterson's contributions as made possible "by a high order of scholarship, a comprehensive knowledge of chemistry, a fair and judicious attitude, extraordinary ability to see the whole picture and thus to keep decisions consistent, and a most generous willingness to work for the common good".

Patterson's philosophy of nomenclature was clearly expressed in a paper that he and Curran wrote in 1917:

> We have not tried to invent a new system. Our aim has been to follow existing usage as far as it could be made consistent, choosing what appeared to us good practices and rejecting bad, and introducing new features only when some very positive advantage was to be gained. In striving for names which would be suitable for general use, and not only for the Index, our tendency has been frankly toward compact naming and we have freely used short names for complex compounds as fresh starting points in naming derivatives of them. However, mere briefness is not sufficient to recommend a name if it is not consistent with general principles, and so it sometimes happens that the preferred name is not the shortest which might be chosen.

The introduction to the First *CA Decennial Subject Index* (1907–1916) noted that the "system [for naming] organic compounds is based on existing usage and follows this as far as is practicable, so that a great many names are unaffected". The principal characterizing features of this first system included expression of the chief function of a compound in the main part of the name wherever possible, and not as a substituent (ethanol not hydroxyethane), expression of multiple chief functions where feasible as 'diol', 'dicarboxylic acid', etc., placement of the 'main part' of the name with its functional ending first followed by the names of the substituents in alphabetical order, use of enclosing marks where necessary to clearly identify complex (i.e. multipart) radical names and use of numbering schemes for substituents that give the smallest number or numbers when two or more numberings were possible. Inorganic compounds were indexed under the 'usual names'.

Patterson's philosophy became the basis for both CAS index nomenclature and the nomenclature of the International Union of Pure and Applied Chemistry (IUPAC).

In order to develop the systematic scheme for naming substances, Patterson first conducted a survey of the organic chemical literature to determine how most chemists named a selected group of chemical molecules. He used this survey to establish an order of precedence of functions. A function (or functional group) is any atom or group of atoms that causes a substance to behave in a characteristic way, for example, a carboxylic acid, amine, or alcohol group. The order of precedence of functions determines the chief function of a compound and is the key to the assignment of a unique name.

The order of precedence of functions is basic to today's CAS name selection principles, and has been revised and extended over the years for new molecular systems and the evolution of the chemical sciences. The first order of precedence of functions consisted of: 'onium' compounds, acid (carboxylic first), acid halide, amide, imide, aldehyde, nitrile, ketone, alcohol, phenol, mercaptan, amine, imine, ether, sulfide (and sulfoxide and sulfone). While this list was extended and revised through the years, the basic order has met the test of time and has remained largely unchanged. Table 3.1 shows the current order of precedence of compound classes used by CAS.

The principle of inversion is another significant feature of the indexing system. Simply placing substance names in an index in alphabetical order results in many names beginning with such common radicals as 'methyl', 'ethyl', and 'propyl'. This was done in the early volumes of *CA* and is shown in Figure 3.1. While this might not be much of a problem in a small index, it creates an overwhelming searching problem as the index becomes ever bigger. The major problem with this type of arrangement of names is that it does not usually organize them on the basis of any distinctive feature.

Patterson instituted the use of inverted index names for organic substances to group related substances together in the printed index at what was termed 'index heading parents'. This is similar to the practice of indexing author names at the surname (e.g. Priestley, Joseph).

The index heading parent usually consists of the principal function, as determined by the order of precedence of functions, expressed as a suffix along with the basic acyclic or cyclic skeleton to which it is attached (e.g. 1-butanesulfonic acid). The full substance index names appear in an inverted format that successively identifies the parent structure, chemical substituents, chemical modifications (usually of the principal functional groups), and stereochemistry, as appropriate (e.g. Propanoic acid, 2-hydroxy-, methyl ester, (*R*)- for (+)-methyl lactate or Ethanol, 2-amino- for ethanolamine). Figure 3.2 shows examples of inverted index name entries from the first decennial subject index that show the grouping of related substances.

While the inverted names serve an especially useful purpose in a printed index, the inverted names can also be uninverted for use in normal text

Table 3.1 *CA* order of precedence of compound classes in descending order

1. Free radicals and other compound classes for which there is no accepted method of designation as substituent prefixes and certain functional parent compounds, e.g. Sulfur diimide.
2. Cationic compounds: coordination cations, substitutive cations, e.g. carbonium ions.
3. Neutral coordination compounds, including metallocenes, e.g. Ferrocene.
4. Anionic compounds, e.g. Borate (1-).
5. Acids: peroxy acids expressed as principal groups, in the order of the corresponding normal acids; acids expressed as principal groups in the general order: carbon (e.g. carboxylic) acids, sulfur acids, selenium acids, tellurium acids; acids expressed as functional parent compounds in the general order: carbon acids (e.g. Carbonic acid, Carbamic acid, Formic acid).
6. Acid halides and halogenides, first in the order of the corresponding acids, then in the order of the class term: fluoride, chloride, bromide, iodide, azide, isocyanate, isothiocyanate, isocyanide, cyanide.
7. Amides (in the order of the corresponding acids).
8. Nitriles (in the order of the corresponding acids).
9. Aldehydes, followed by chalcogen analogs in the order S, Se, Te.
10. Ketones, Thiones, Selones, Tellones.
11. Alcohols (Phenols), Thiols, Selenols, Tellurols.
12. Hydroperoxides
13. Amines
14. Imines (including Sulfilimine, Sulfoximine and related compounds)

Compound classes 1–14 above are considered to be functional and, except for coordination compounds and some cations, are expressible as functional suffixes. The remaining classes below are regarded as non-functional:

15. Nitrogen compounds: heterocyclic; acyclic, e.g. Triazane, Diazene, Hydrazine, Hydroxylamine, Thiohydroxylamine.
16. Phosphorus compounds: heterocyclic; acyclic, e.g. Diphosphine, Phosphine oxide, Phosphine sulfide, Phosphine imide, Phosphorane, Phosphine.
17. Arsenic compounds (in similar order).
18. Antimony compounds (in similar order).
19. Bismuth compounds (in similar order).
20. Boron compounds: carbapolyboranes; heteropolyboranes; polyborane cages; heterocyclic, e.g. 1,3,2-Benzoxaborole; acyclic, e.g. Diborane(6), Diborane(4), Borane.
21. Silicon compounds: cyclic; acyclic, e.g. Disiloxane, Disilathiane, Trisilane, Disilane, Silane.
22. Germanium compounds (in similar order).
23. Tin compounds (in similar order).
24. Lead compounds (in similar order).
25. Oxygen compounds: heterocyclic; acyclic, e.g. Tetraoxide, Trioxide, Peroxide.
26. Sulfur compounds: heterocyclic; acyclic, e.g. Trisulfone, Trisulfide, Disulfoxide, Disulfide.
27. Selenium compounds (in similar order).
28. Tellurium compounds (in similar order).
29. Carbon compounds: cyclic, acyclic.

(e.g. (*R*)-methyl 2-hydroxypropanoate). These uninverted names can often be found in *CA* index names involving addition compounds, polymers, and mixtures (e.g. 1,4-Benzenedicarboxylic acid, polymer with 2,5-diamino-1,4-benzenediol).

Pyrazole *(1,2-diazole)*,

NH.N:CH.CH:CH
1 2 3 4 5

book: Die Pyrazolfarbstoffe, **5**:1680[2].
prepn. of, **5**:1091[7].
from pyrazolones, **1**:2128[1].
reaction with SO_2Cl_2, **1**:857[4].

—— **4 - acetamido - 5 - anilino - 3 - methyl - 1 - phenyl -**, **6**:601[3].
—— **5 - acetamido - 4 - antipyryl - 3 - methyl - 1 - phenyl -**, **1**:260[9]
—— **2,5[*m*(and *p*) - acetamidophenylimino] - 2,3 - dimethyl - 1 - phenyl -**, and derivs., **6**:603[2].
—— **5 - (*N* - acetylanilino) - 4 - antipyryl - 3 -methyl - 1 - phenyl -**, **6**:601[7].
—— **5 - (*N* - acetylanilino) - 4 - antipyryl - 3 - methyl - 1 - *p* - tolyl -**, **6**:602[2].
—— **acetyldiphenyl -**, **10**:1526[3].
—— **4 - alkyl - 5 - amino -**, derivs., **9**:60[5].
—— **4-amino-**, and dihydrochloride,**8**:1749[8].
—— **5 - amino -**, derivs., **6**:600[4]; **7**:2202[6].
—— **4 - (*p* - aminoanilino) - 3 - methyl - 1 - phenyl -**, and derivs. **9**:1030[9].
—— **4 - amino - 4 - antipyryl - 1 - methyl - 1- phenyl -**, **1**:2602[8].
—— **5 - amino - 4 - benzoyl - 3 - methyl - 1 -phenyl -**, derivs., **2**:3352[1].
—— **4 - amino - 5 - chloro - 3 - methyl - 1 - phenyl -**, derivs., **8**:691[3].
and derivs., **9**:1031[1].
—— **5 - amino - 3, 4 - dimethyl - 1 - phenyl -**, and derivs., **9**:60[8].
—— **4 - amino - 5 - (ethylanilino) - 3 - methyl - 1 -phenyl-**, and derivs., **6**:601[7].
—— **5 - amino - 4 - ethyl - 3 - methyl - 1 - phenyl -**, hydrochloride, **9**:61[2].

Figure 3.2 Examples of inverted substance index names from *CA First Decennial Index* (1907–1916). (Copyright American Chemical Society.)

The principle of inversion and the desire to group related substances are also reflected in the practice of citing functional derivatives of the principal reactive groups in the last part of the index name, the chemical modification. These derivatives, as defined for *CA* indexing purposes, are restricted to salts and addition compounds, counter ions, acyclic anhydrides, esters, hydrazides, hydrazones, oximes, and polymers. Examples of inverted names for some of these derivatives are shown in Table 3.2.

A special challenge faced by an indexer that is not usually encountered by a researcher writing a technical paper is the need to assign a name to a substance whose molecular structure is not fully defined. Authors of papers are sometimes lax in reporting the complete structures of chemical

Table 3.2 Examples of *CA* Index Names of Functional Derivatives

Salts and addition compounds:
Acetic acid, trifluoro-, silver(1+) salt
Benzenamine, 2-methoxy-, hydrochloride
2,5-Cyclohexadiene-1,4-dione, compd. with 1,4-benzenediol (1:1)
Counter ions:
Methanaminium, *N,N,N*-trimethyl-, bromide
Acyclic anhydrides:
Benzoic acid, anhydride
Esters:
Acetic acid, chloro-, 1,1-dimethylethyl ester
Hydrazides and hydrazones:
Acetaldehyde, (2,4-dinitrophenyl)hydrazone
Oximes:
Benzaldehyde, 4-nitro-, oxime
Polymers:
2-Propenoic acid, polymer with ethene

substances, perhaps because the structures are so well known to them that they do not realize that the names they have chosen to use are incomplete in some respect. Since it is impractical for the *CA* indexer to consult with each author for which there is some uncertainty about the nature of the substance reported, the indexer must quickly make a decision on the identity and index name of the substance based on the information reported.

For example, when an indexer encounters a paper that describes some use of 'xylene', what isomer should the indexer assume? *ortho-*, *meta-*, or *para-*, or is it possible that the researcher had a mixture of all three? In practice, the *CA* indexer does not usually assume any specific isomer unless there is other information in the document that serves to suggest clearly a specific form, perhaps via a reaction scheme or physical property data. In most cases, the indefinite substance is assigned a name that discloses all of the known features while retaining the ambiguity. This is often achieved using a name in which locants for the indefinite portions of the structure are omitted. For example, 'xylene' is given the index name 'Benzene, dimethyl-'. Sometimes the uncertainty extends to only certain positions and the index name identifies alternative structures, for example 'Naphthalene, 1(or 2)-ethyl-2(or 1)-methyl-'.

In a few cases, the *CA* indexer makes certain structural assumptions for some frequently indexed substances whose common names are imprecise. For example, the indexer assumes that most naturally occurring amino acids are the L-isomers if this has not been specified in the original document and there is nothing to suggest that they are the uncommon D-forms or racemates. Thus, 'alanine' is indexed as 'L-alanine'. Similarly, the naturally occurring (+)-isomer of tartaric acid is assumed unless there is evidence to the contrary.

In some cases, an author's name for a substance is used as the index name when there is insufficient information to determine the identity of the substance. This occurs most frequently with natural products whose structures have yet to be elucidated or reported, or for substances identified only by their commercial trade-names.

One of the more challenging aspects of chemical substance nomenclature is the systematic assignment of a name and numbering to a ring system. In 1921, Dr Patterson reported to the Board of Editors of the *Journal of the American Chemical Society* that data about organic ring systems and their systematic numbering had been gathered in connection with the indexing of *CA*, and he asked whether it would be advisable to publish a 'catalog of such systems' in the Journal or elsewhere. The Board decided that it would be useful, but suggested that it would be desirable to secure beforehand a wide agreement among chemists as to the numbering of the positions of the ring systems.

Accordingly, a joint committee of the ACS and the National Research Council was established in 1922. It was given the title 'Committee on the

Preparation and Publication of a List of Ring Systems Used in Organic Chemistry'. Patterson became chairman of this group. He was assisted in this effort by Leonard Capell who later became a worthy successor to Patterson in the development of chemical nomenclature.

Rules for numbering ring systems were devised and submitted to this Committee and later to the Commission on the Reform of Organic Nomenclature of IUPAC. They were also shown unofficially to some of the German chemists. The group spent considerable time devising a set of simple and clear rules that would be applicable to all systems of single or fused rings. They also tried to preserve as many of the existing numbering schemes as possible. The rules went through several drafts before being approved by the International Commission. In the meantime *CA* began using the rules for numbering newly discovered ring systems and adopted the rules for all rings systems for the *Third Decennial Index* (1927–1936).

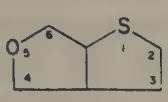
Patterson's innovations in ring nomenclature included:

- Adaptation of proposals of von Baeyer and of Bruckner and Wiegand for naming bi- and polycyclic bridged ring systems.
- Consolidation of proposals of Hantzsch, Widman, and Hale for naming heteromonocycles with his own syllables for describing rings with three and seven through ten ring members.
- Development of 'fusion nomenclature', which he called 'skeletal nomenclature'.
- Development of proposals from von Baeyer for simple spiro ring systems and devising a method for other spiro ring systems, that is, those having other unions in addition to the spiro union.

Patterson's nomenclature scheme for rings was incorporated into the *CA* nomenclature system after a reasonable time for comments to be received had elapsed. However, it was not immediately accepted by the IUPAC Organic Commission. Nevertheless, Patterson, with the help of Dr Capell and encouragement from the ACS, compiled and published in 1940 the first edition of *The Ring Index*. In it, he again presented his ideas, now polished by their use in *CA* index nomenclature during the intervening years. Nearly 4000 rings were included. It should be realized that *The Ring Index* is a compilation of skeletal ring structures and not necessarily of compounds.

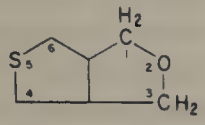
Patterson's work was finally rewarded by its appearance in the 1957 edition of the *IUPAC Organic Rules*. There were a few changes and modifications, but none of a fundamental nature. The second edition of *The Ring Index* was published in 1960. This edition included the changes and modifications adopted for the 1957 *IUPAC Organic Rules*. Three supplements were published and Patterson's basic ideas were extended to more than 14 000 rings. Figure 3.3 shows an example of a page from *The Ring Index*, Second Edition.

II 5,5 C_4O-C_5

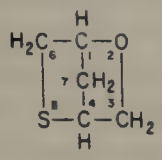


C_4O-C_4S

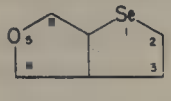
1006.[657]. C_6H_4OS. **Thieno[2,3-*c*]furan.** J. Chem. Soc. 1937, 916("anhydride").



1007.[658]. C_6H_6OS. **1*H*,3*H*-Thieno[3,4-*c*]furan.** Roczniki Chem. 16, 475(1936); CA 31, 2186[9].

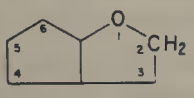


1008. C_5H_8OS. **2-Oxa-5-thiabicyclo[2.2.1]heptane.** Helv. Chim. Acta 27, 142(1944).



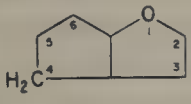
C_4O-C_4Se

1009. C_6H_4OSe. **Selenolo[2,3-*c*]furan.** Ann. 562, 23(1949).

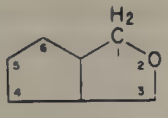


C_4O-C_5

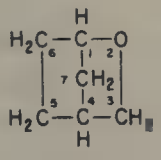
1010.[659]. C_7H_6O. **2*H*-Cyclopenta[*b*]furan.** Bull. soc. chim. [3] 27, 404(1902); B_4 17, 301[8].



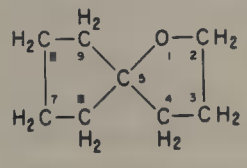
1011. C_7H_6O. **4*H*-Cyclopenta[*b*]furan.** Rec. trav. chim. 70, 40(1951).



1012.[660]. C_7H_6O. **1*H*-Cyclopenta[*c*]furan.** J. Chem. Soc. 65, 587(1894); B_4 17, 449[8].



1013.[661]. $C_6H_{10}O$. **2-Oxabicyclo[2.2.1]heptane.** J. Chem. Soc. 79, 1289(1901); B_4 17, 259[9], 18, 401[6], 403.



1014.[662]. $C_8H_{14}O$. **1-Oxaspiro[4.4]nonane.** J. prakt. Chem. [2] 89, 337(1914); B_4 17/19, (484[8]).

Figure 3.3 Sample page from *The Ring Index* (2nd edn) (1960). (Reproduced with permission from the American Chemical Society.)

The system for assigning *CA* index names to chemical substances evolved slowly, but remained essentially the same as that in use starting with the 1916 volume subject index. *CA* took into account the 'Definitive Report of the Commission on the Reform of the Nomenclature of Organic Chemistry' adopted by the Commission and by the Council of the International Union of Chemistry in 1930 in Liege and the 'Rules for Naming Inorganic Compounds' issued by the Committee for the Reform of Inorganic Chemical Nomenclature in 1940. The *CA* indexing system was generally in conformance with these recommendations although it did not conform in all points. It should be pointed out that whereas the 1892 Geneva Conference sought to establish a single, systematic, official name for every compound, this concept was abandoned at Liege and replaced by the idea of unifying existing practice as far as possible.

In 1945, when chemical publication was at a low ebb because of World War II, the *CA* staff took advantage of the lull to prepare an extensive treatise on the nomenclature system used in generating index names. This appeared as part of the Introduction to Volume 39 of *CA* under the title 'The Naming and Indexing of Chemical Compounds by Chemical Abstracts'. Many thousands of reprints of this treatise were sold demonstrating a high level of interest in systematic chemical nomenclature.

At the time, this publication was the only comprehensive manual of chemical nomenclature in existence in English. The only IUPAC nomenclature instructions were the 1930 Liege Rules for organic compounds and the 1940 *Inorganic Rules*, both of which were far from complete.

In subsequent years, *CA* indexing practice moved toward more systematic nomenclature and general consistency with IUPAC recommendations. *CA* discontinued the use of trivial names that were not countenanced by IUPAC. Carbohydrate nomenclature was systematized according to the 1953 *Rules of Carbohydrate Nomenclature* resulting from a joint study of British and American chemists. Terpene and steroid names were revised in accordance with the 1955 and 1957 IUPAC rules, respectively, and ring names were revised in accordance with the 1957 IUPAC recommendations.

It is generally the practice of CAS to revise its indexing policies and index terms, including the index names for chemical substances, only at the beginning of each new collective period so that the index entries in the Collective Indexes would be self-consistent. Over the years since the first decennial index, most improvements in *CA* index names had taken one of two forms, (a) conversion of 'trivial' names into more 'systematic' names, and (b) unification and simplification of naming principles for all chemical substances.

With the start of the Ninth Collective Index Period, covering 1972–1976, CAS accelerated these improvements and introduced the most extensive changes in index names in its history since CAS had first begun to use systematic names during the second Decennial Index Period. The rather

large number of trivial names that were approved by IUPAC and used in CAS index nomenclature often complicated the index nomenclature requirement that the same name be generated for each compound every time it was encountered. While remaining within the framework of IUPAC and other nomenclature rules, CAS generally chose to use the most systematic recommended names and to simplify the rules for selection of the index names.

The Ninth Collective Index Period nomenclature revision was prompted by the need to generate index names quickly, consistently, and with confidence by different indexers working independently, and to more reliably group names of structurally related substances together in the index. The revision was also intended to ensure that users of CAS services worldwide would be able to more quickly and easily interpret the index names.

A classic example of the change in trivial names effective with the Ninth Collective Index Period is the group of names used for the isomeric dihydroxybenzoic acids shown in Table 3.3. Entries for these related substances were scattered in the previous indexes, but pulled together at 'Benzoic acid' with the new systematic practice. In addition, the new names were recognizable by many more users than the previous trivial names.

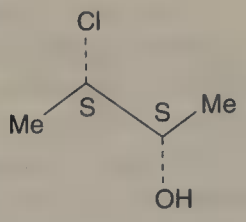
Another important change occurred in the area of stereochemical nomenclature. For compounds with two or more chiral centers, the revised CAS system was based on the principle of describing the relative stereochemistry of all centers and then indicating the absolute configuration, where known, by the appropriate absolute descriptor, at a single reference center. Figure 3.4 shows a few examples of such descriptors.

Other changes in CAS index nomenclature made in 1972 include:

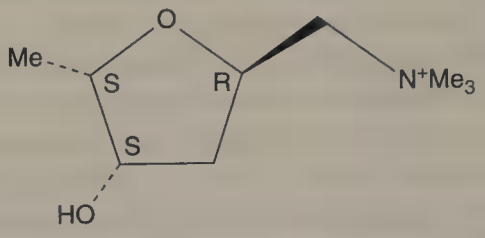
- Emphasis on the largest parent structure in preference to 'like treatment' and double bond functionality. IUPAC continues to prefer functionality of double bonds over largest parent.
- Names for amines are formed by adding the suffix 'amine' to the name of a parent hydride; thus, amines are named just like hydroxy compounds, for example, 2-butanol and 2-butanamine.
- Cation names for quaternary acyclic nitrogen compounds are derived

Table 3.3 Examples of CAS Ninth Collective Index Period (1972–1976) switch from trivial to systematic nomenclature

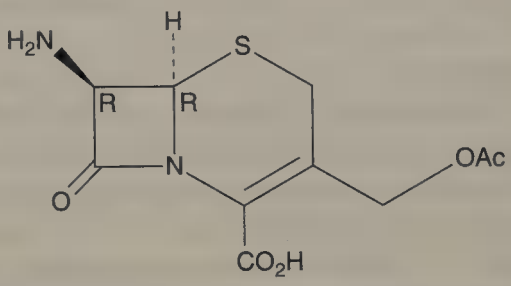
Pre-1972 *CA* Index Name	*CA* Index Name introduced in 1972
o-Pyrocatechuic acid	Benzoic acid, 2,3-dihydroxy-
β-Resorcylic acid	Benzoic acid, 2,4-dihydroxy-
Gentisic acid	Benzoic acid, 2,5-dihydroxy-
γ-Resorcylic acid	Benzoic acid, 2,6-dihydroxy-
Protocatechuic acid	Benzoic acid, 3,4-dihydroxy-
α-Resorcylic acid	Benzoic acid, 3,5-dihydroxy-



2-Butanol, 3-chloro-, [S-(R^*,R^*)]-



2-Furanmethanaminium, tetrahydro-4-hydroxy-
N,*N*,*N*,5-tetramethyl-, [2*R*-(2α,4β,5β)]-



5-Thia-1-azabicyclo[4.2.0]oct-2-ene-2-carboxylic acid,
3-[(acetyloxy)methyl]-7-amino-8-oxo-, (6*R*-*trans*)-

Figure 3.4 (Copyright American Chemical Society.)

from the names of the preferred amine, rather than as derivatives of the parent cation ammonium, for example, methanaminium.

- Conjunctive nomenclature is used for monosubstituted benzenes, for example, benzenemethanol.
- Detailed rules for naming tautomeric compounds are used to avoid multiple index names based on the various tautomeric forms.

Initial reactions from users to the more systematic names were mixed. Some welcomed the simplifications, while others found the changes to be too sweeping and felt that CAS was trying to change the names of substances by fiat. CAS responded by noting that an index requires a formalized, rigidly controlled, alphabetic listing of names that must be quite specific and unique for each substance. The only way to ensure that

both the indexer and user can consistently arrive at the same names for the same substances is through the use of fully systematic names. CAS emphasized that it was not trying to introduce its index names into common usage nor to suggest that authors or lecturers use *CA* index names if they did not serve their purposes.

To assist its users in making the transition to the new names, CAS introduced cross-references in the *Chemical Substance Index* from the previous index names to the new names for the former trivially named index heading parents. It also provided indexing notes in the *CA Index Guide* at many substance class and common substance headings summarizing briefly the appropriate nomenclature policies. A 120-page summary of the complete set of name selection policies was added to the *Index Guide* for those users who were interested in a learning more about *CA* nomenclature policies, and a comprehensive two-volume name selection manual was also made available.

The preferred *CA* index names for most chemical substances have been continued unchanged since 1972. Changes in name selection policies for the Twelfth (1987–1991), Thirteenth (1992–1996), and Fourteenth (1997–2001) Collective Index Periods affected alloys, carbohydrates (lactams), coordination compounds, formazans, fullerenes, inorganic compounds (line formulas for clusters, intermetallic compounds), onium compounds (free radicals), peptides, phosphonium ylides, phosphoryl halides and halogenoids, polymers (block, graft, hydrolytic), and stereochemical practices (optical rotation, racemates, stereoparents). These changes are detailed in Appendix IV of the most current *CA Index Guide.*

Successor publications to *The Ring Index* have continued to assist users of CAS services to identify the names and numberings of ring systems, including cage systems. The *Parent Compound Handbook* was introduced in 1977 and replaced by the *Ring Systems Handbook* in 1984. The *Ring Systems Handbook* offers several routes of access to all reported ring systems indexed in *CA*, including those in the former *Ring Index* and *Parent Compound Handbook*, and presents structural diagrams illustrating the ring system numbering schemes, the current *CA* index names, molecular formulas, and CAS Registry Numbers. The most recent inventory of ring systems is the 1993 edition of the Handbook with the November 1997 supplement, together listing over 100 000 rings.

3.6 Relationship of CAS nomenclature to national and international authorities

The historical development of systematic chemical nomenclature at CAS has involved a number of organizations. Internationally, nomenclature commissions of the International Union of Pure and Applied Chemistry

(IUPAC) and the International Union of Biochemistry (IUB) approve and publish detailed recommendations. CAS nomenclature specialists have been especially active participants in the IUPAC commissions, with early CAS participants being Drs Patterson and Crane.

In the United States, the American Chemical Society established a Committee on Nomenclature and Notation as early as 1886 that was later discontinued. This was followed in 1911 by the ACS Committee on Nomenclature, Spelling, and Pronunciation. A motion to establish this committee was appropriately made by Austin Patterson, who also was appointed to serve as its first chair. This committee is now known simply as the ACS Committee on Nomenclature. In addition, special subject nomenclature subcommittees exist under seven ACS divisions.

Obviously, because of its special relationship to the ACS, CAS has worked closely with the ACS nomenclature bodies over the years in the development and refinement of nomenclature generally and that of CAS specifically. For much of its history, CAS had a Director of Nomenclature who also chaired the ACS Committee on Nomenclature. Dr Crane chaired the committee for 44 years, followed by Drs Capell and Loening.

Dr Kurt Loening joined the staff at CAS in 1951, and in 1964 became the Director of Nomenclature and Chairman of the ACS Nomenclature Committee, replacing Dr Capell. He also continued the long-standing participation of CAS in IUPAC activities. During his tenure with IUPAC, which ended officially in 1987, he participated in a number of IUPAC Commissions and Committees and carried on the tradition of ACS leadership in the development of chemical nomenclature both nationally and internationally.

CAS has actively participated in national and international chemical nomenclature bodies for the purposes of collectively addressing the diverse needs for systematically and unambiguously naming substances. Additionally, CAS has looked towards and drawn on the nomenclature standardization efforts of other national and international organizations for guidance in naming substances in specialized areas. These bodies have included the Enzyme Commission of the IUB and the Commission on New Minerals and Mineral Names of the International Mineralogical Association.

The rules for generating the *CA* index names for substances for the most part parallel the nomenclature rules of the IUPAC where the latter are available, but CAS has extended the IUPAC rules so that a single, invariant name results for each chemical structure and substances with closely related structures appear together in an alphabetically arranged index. Generally speaking, IUPAC names are not necessarily unique and a given structure may have two or more IUPAC names. When alternatives are permitted by IUPAC rules, the *CA* index name is usually the more systematic one.

Perhaps the main difference between *CA* and IUPAC names is one of format. As noted earlier, a conventional substance name is written consecutively from left to right. It sometimes begins with a stereochemical descriptor, which is followed by substitutive prefixes, the parent itself, suffixes describing further modifications to the parent and the suffix describing the principal characteristic group.

On the other hand, *CA* index names are inverted; that is, the name of the parent compound and its suffixes is written first, followed by a comma, and then prefixes describing substituents of the parent structure, followed by functional derivatives, such as esters, salts, anhydrides, hydrazones and oximes. Finally, following any functional derivatives, stereochemical information and other descriptive phrases that may be needed are cited. This procedure allows collection of many compounds related to the parent structure at one point in an alphabetical arrangement.

Some differences between IUPAC nomenclature and CAS index nomenclature resulted from the changes that CAS made beginning with the Ninth Collective Index Period in 1972. Most of these are now under study by IUPAC's Commission on Nomenclature of Organic Chemistry (CNOC). Although not yet published, some differences have already been removed and it is hopeful that other differences will be eliminated in the coming years. A few of the differences are described below.

As long as both are substituted by the same number of the principal characteristic group, CAS prefers a hydrocarbon ring or ring system over a hydrocarbon chain no matter what the length of the chain or size of the ring system, for example, hexadecylbenzene (CAS) vs. 1-phenylhexadecane (IUPAC). CAS prefers the longest chain whereas IUPAC prefers a chain with unsaturation, for example, 3-methylenehexane (CAS) vs. 2-ethylpent-1-ene (IUPAC).

In numbering of chains named by replacement nomenclature, CAS assigns low numbers to the heteroatoms, but IUPAC gives low numbers to the principal characteristic groups, as in 2,5,8-trioxa-11-azatetradecan-14-ol (CAS) vs. 4-aza-7,10,13-trioxatetradecan-1-ol (IUPAC).

CAS uses the suffix name carboxaldehyde whereas IUPAC uses carbaldehyde; both use -carbothialdehyde. In 1972, CAS began to name amidines as amides of imidic acids, for example benzenecarboximidamide. CNOC made this change in 1993. IUPAC names most peroxyacids using the prefix peroxy whereas CAS names peroxyacids by means of the infix peroxo, for example, propaneperoxoic acid (CAS) vs. peroxypropionic acid (IUPAC).

CAS and IUPAC both use infixes in functional replacement nomenclature for phosphorus and arsenic acids, for example, phosphorochloridic acid. CAS has extended this technique to carbonic acids, for example, carbonothioic acid. IUPAC continues to use acyl and amido substituent prefix names whereas CAS names these, except for acetyl and benzoyl,

as compound prefixes, for example, the name propionyl and propionamido are used by IUPAC but these prefixes are named 1-oxopropyl and (1-oxopropyl)amino, respectively, by CAS. IUPAC continues to use trivial names for branched chain prefixes, such as *tert*-butyl whereas CAS uses substituted alkyl prefix names, for example, 1,1-dimethylethyl.

There are recent recommendations by CNOC that have not been incorporated into CAS index nomenclature. The λ-convention, developed in order to extend substitutive nomenclature systematically to compounds with heteroatoms in valence states other that the normal or standard one, has been applied by CAS only to rings and ring systems. The δ-convention, developed to name structures having formal cumulative double bonds, has not been adopted for use in CAS index nomenclature. The Hantzsch–Widman system for naming heteromonocycles was refined and extended by CNOC in 1983 but CAS still uses the system as described in the 1979 edition of the *IUPAC Nomenclature of Organic Chemistry*. The main reason that CAS has not adopted the CNOC revisions and extension is that rings named by the Hantzsch–Widman system are used in naming many of the more than 100 000 rings systems in the CAS files.

There are a number of natural product classes for which IUPAC/IUB has published recommendations, some of which are closely followed in CAS index nomenclature, for example, carbohydrates, steroids, and amino acids, but other recommendations, such as those for prenols and certain vitamin classes, are not.

Finally, there are many instances where there are no IUPAC recommendations but since CAS must provide an index name for each substance encountered it has developed its own practices where needed. Some of these areas include delocalized radicals and ions, tautomers, incompletely defined compounds, interferons, and many other biochemically significant compound classes.

3.7 Growth of the chemical literature

From the beginning the focus of *CA* was on the preparation and publication of abstracts of documents that were published world-wide that related to chemistry or chemical engineering. In its first year, 1907, there were 11 847 abstracts published in *CA*. In comparison to the 502 abstracts published in the first issue of *CA*, each *CA* weekly issue today averages about 13 500 abstracts.

Since 1907, CAS has published more than 17 million abstracts, keeping pace with the rapid growth of the scientific and technical literature (Figure 3.5). The number of abstracts published annually did not reach 100 000 until 1957. It reached 200 000 in 1966, 400 000 in 1977, 600 000 in 1994, and passed 700 000 in 1996.

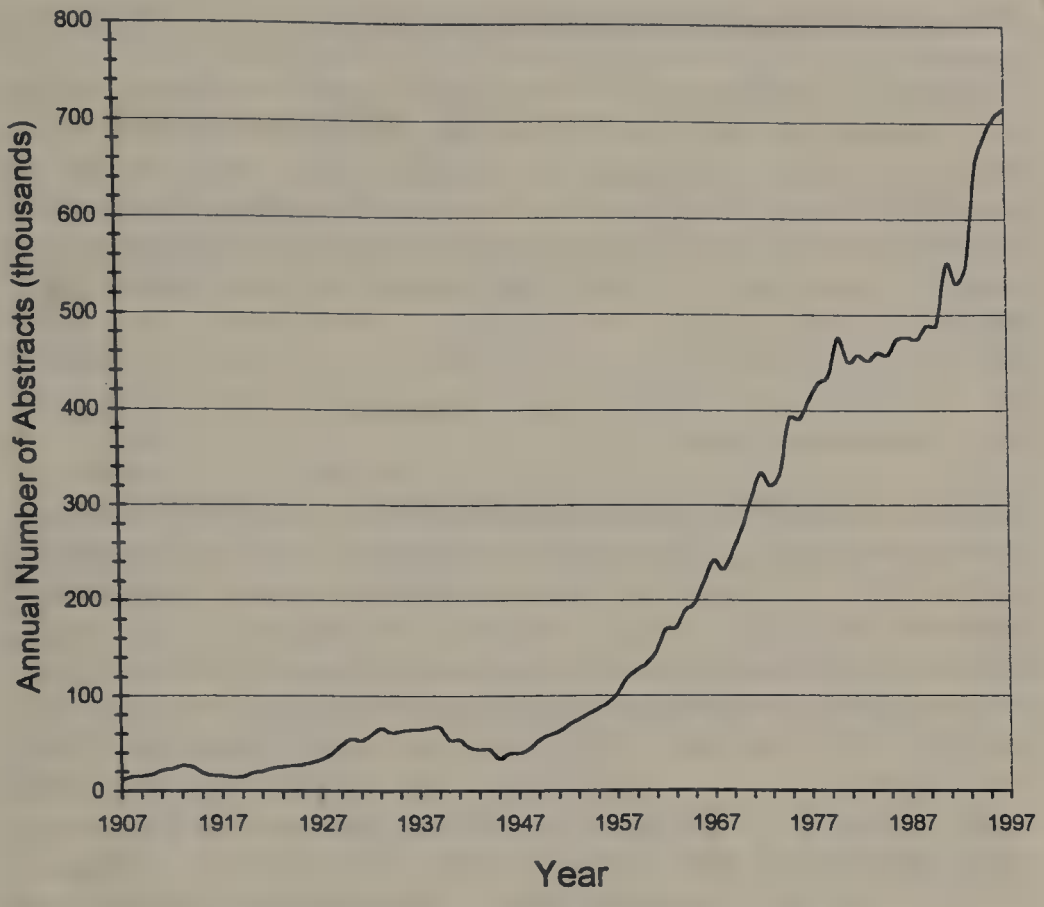


Figure 3.5 Growth of the chemical literature abstracted in *CA*.

Not only the sheer number of documents but also the amount of work done by the editorial staff has grown tremendously. In the beginning, *CA* was essentially an abstracting service, with the abstracts provided by volunteers world-wide, and only the indexing performed by the editorial staff. The shift away from volunteer abstractors to in-house staff in the 1960s resulted from the need for greater timeliness in abstracting and indexing and the introduction of computer processing of the data.

Another reason for the growth in staff has been the increasing importance of the indexing effort. It is the subject indexing that usually allows the searchers to identify the existence of abstracts (and thus of documents) of interest to them. Most of the CAS intellectual analysis effort is devoted to indexing. There are chemical substance name, chemical formula, general subject, author name, and patent indexes. *The CA Twelfth Collective Index*, covering the 5-year period 1987–1991, is listed in the *Guinness Book of World Records* as the world's longest index. It consisted of 115 books, having almost 216 000 pages, referencing over 3.0 million documents in chemistry. The *CA Thirteenth Collective Index*, covering

1992–1996, published in 1997 with 150 books is expected to replace that record!

In preparing its information services, CAS receives and monitors more than 8000 scientific journals or magazines published all over the world. Books, conference proceedings, patents, university dissertations, and government reports are also covered. Almost 80% of the 700 000 documents abstracted in 1997 are individually authored papers or articles from journals. The next largest category of documents is patents, amounting to about 120 000 documents.

3.8 CAS registry system

The ever-increasing growth in the number of documents to be abstracted and the corresponding number of chemical substances to be indexed was a strong driving force in the development by CAS of a computerized registry of chemical substances. Prior to 1965, every time a chemical substance was selected from the literature for *CA* indexing, with the exception of some 2500 common substances, its structure was drawn by hand on a separate special sheet of paper and then named. There was no easy way for the indexers to determine whether a compound was new or had been reported in the literature and indexed previously in *CA*. Even a manual comparison with compounds previously indexed could only be done after the naming was completed. The structure diagrams of many compounds were drawn and the compounds named over and over again. Thus, there was a strong need for a system that would allow CAS staff to recognize previously encountered substances and retrieve their index names for reuse.

The CAS Chemical Registry System grew out of a concept suggested in the late 1950s by G. Malcolm Dyson, Research Director at CAS. The initial 'registry' was an experimental file of some 14 000 organic fluorine-containing compounds that had appeared in the *CA Formula Index*. These were coded with a linear notation originally devised by Dyson and subsequently adopted by IUPAC; chemical nomenclature had been rejected at an early stage as the basis for the registry because of the likelihood of continual change. Each compound was assigned a 'register number'. The file was later expanded to include all of the ring systems from the second edition of *The Ring Index*.

CAS research and development staff soon recognized the limitations of the linear notations as the basis for a registry system, particularly if one wanted to search for chemical substances in the registry on the basis of their substructural features. As an alternative to the linear notations, CAS staff began to experiment with connection tables in which the arrangements of atoms and bonds in structural diagrams are reduced to table

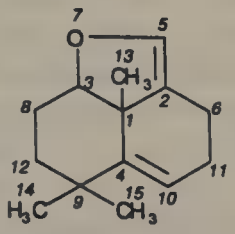
form (Figure 3.6). Harry L. Morgan, of the CAS staff, building on work done by D. J. Gluck at Du Pont, perfected an algorithm that translates two-dimensional structural diagrams into a tabular form that can be manipulated and searched by computer. Except for stereochemistry, the resulting connection table uniquely describes the bond attachments and the elemental makeup.

In theory, the Morgan algorithm generates all possible connection tables for a structure and, based on a series of hierarchical rankings, selects one table as the preferred table for the structure. The algorithm assures that the same connection table is generated for the same substance, no matter how or when the substance is entered again into the system after its first registration. This algorithm became the basis of the CAS Registry, and was also adopted by Du Pont for its registry system.

In late 1964 an experimental Chemical Registry System was established and used to support production of the new CAS publication *Chemical-Biological Activities*. After about 6 months, the Registry System, by then no longer considered experimental, was extended to cover most of the substances indexed in *CA* beginning with the volumes published in 1965. Underlying research and development for continued improvement of the Registry was funded in part by the National Science Foundation from 1965 into the early 1970s.

Node no.	Atom	Connected to	Bond type
1	C	-	-
2	C	1	*1
3	C	1	*1
4	C	1	*1
5	C	2	*2
6	C	2	*1
7	C	3	*1
8	C	3	*1
9	C	4	*1
10	C	4	*2
11	C	6	*1
12	C	8	*1
13	C	1	-1
14	C	9	-1
15	C	9	-1
Ring closure	-	5-7, 9-12, 10-11	*1

1=single bond *=cyclic bond
2=double bond -=non-cyclic bond



Registration translates the two-dimensional diagram for a substance into a tabular form that can be searched by computer. The connection table (left) for the 10 millionth substance, CAS Registry Number 125417-03-0, describes the bond attachment and elemental makeup.

Figure 3.6 Example of CAS chemical structure connection table. (Reproduced with permission from the American Chemical Society.)

Over the 30-year history of the CAS Registry, three distinct successive versions were developed, each resulting in handling more classes of chemical substances and providing better representation, search, and retrieval capabilities.

The initial version of the Registry System, now known as Registry I, was limited to the registration of fully defined organic compounds. Because of the anticipated comprehensive nature of the Registry database, the system was designed to register substances as specifically as possible. Thus, the connection table based on the Morgan algorithm was supplemented by additional data such as abnormal mass, abnormal valence, charge, and hydrogen count. Alternating cyclic single- and double-bond systems were 'normalized'; that is, they were recognized by the system as being equivalent and stored in identical machine-language representations. Stereochemistry was represented by an associated text descriptor.

Compounds identified only by name or molecular formula were manually registered; that is, they were added to the database and assigned a Registry Number on the basis of an intellectual determination of uniqueness without the benefit of an associated structural record.

The second version of the CAS Registry System, Registry II, was introduced in 1968. In this version, machine registration was extended to include inorganic substances, coordination compounds, polymers, mixtures, alloys, and certain incompletely defined substances (i.e. substances in which the positions of certain substituents or groups are either unknown or unspecified). The stereochemistry text descriptors were completely reorganized and made more systematic.

Many tautomeric systems (i.e. substances in which migration of a hydrogen atom results in two or more structures that are in equilibrium) were normalized. The CAS Chemical Registry System handles the problem by algorithmically recognizing certain tautomeric and alternating bond structures, replacing the explicit single and double bonds with special normalized bonds, and associating the migrating tautomeric hydrogen atoms with groups of atoms rather than just single atoms.

With the advent of Registry II, every substance indexed for *CA* was processed through the Registry System via name or structure input. If the substance name used by the author of the original report matched a name already present on the Registry Nomenclature File, the CAS Registry Number was retrieved, together with the *CA* index name and molecular summation formula, for use in the *CA Chemical Substance Index* and the *CA Formula Index*, respectively. Figure 3.7 shows several entries from a recent *CA Chemical Substance Index* with the *CA* index names and CAS Registry Numbers.

If the author's substance name did not match or if no name was available for input, the structure diagram was drawn and input for structure

Benzoic acid
——, **2-acetyl-** *[577-56-0]*
Friedlaender reaction of acetyltropolones, 86602g
process for producing (quinolin-2-yl)benzoic acids, P 58340b
ethyl ester *[103935-10-0]*, prepn. of ketoaryl carboxylic esters using phenyliodoso diacetate, *pr* 275360n
——, **3-acetyl-** *[586-42-5]*
prepn. of heterocyclylthiocarbapenems, P 10476y
methyl ester *[21860-07-1]*, prepn. of heterocyclylthiocarbapenems, *pr* P 10476y
——, **4-acetyl-** *[586-89-0]*
monomer starting material; phase transfer catalysis in prepn. of poly(amido-carbonates) and poly(amido-thiocarbonates), 59291y
methyl ester *[3609-53-8]*, substituent effects in gas-phase basicities of benzaldehydes, acetophenones, and Me benzoates, 167211f
——, **2-(acetylamino)-** *[89-52-1]*
for prepn. of rhodium(III) aryl amide complex, 130765p
——, **2-acetyl-6-amino-** *[103440-81-9]*
reactant in prepn. of pyrimidinyl salicylate deriv. herbicides, 161008y
——, **4-(acetylamino)-** *[556-08-01]*
biochem. manuf. of benzyl alcs. from benzoic acids, P 8734n
prepn. of procainamide metabolites, 247328q
prepn. of substituted benzene derivs. useful as neuraminidase-inhibiting antiviral agents, *pr* P 328309m

Figure 3.7 Examples of entries from the *CA Chemical Substance Index*, Volume 125 (July–December 1996). (Copyright American Chemical Society.)

matching. If a match was then obtained, the CAS Registry Number was retrieved, together with the index data. If no match occurred, the structure was added to the Registry database, the substance was assigned a new CAS Registry Number and, subsequently, a new *CA* index name was generated and recorded. This procedure, which attempts to first identify a substance on the basis of its name or structure and forwards only new substances for naming, avoids much redundant and costly intellectual effort.

The most recent version of the Registry System, Registry III, became operational in 1973. This refinement introduced major changes in the Registry records designed to make the system more effective in its support of *CA* index name generation and in the computer-based structure output operation. Registry III created a four-component chemical record consisting of (a) the connection table topology which involved separate segments for the cyclic and acyclic portions, (b) a text descriptor defining the stereochemical characteristics, (c) an isotopic labeling component, and (d) a derivative component that included salt, charge, and tautomer information. The basic algorithms for registration, however, remained unchanged.

Within the Registry System each unique, individual substance is assigned a separate CAS Registry Number (see Table 3.4 for examples). Different positional and stereochemical isomers have distinct numbers. Similarly, salts of acids and bases have CAS Registry Numbers that are distinct from those of the acids and bases themselves.

The CAS Registry Number itself is an automatically assigned serial number that contains no chemical intelligence. It was simply designed to function as a machine address within the associated Registry System files. It designates only one substance, and thus provides a means of linking

Table 3.4 Examples of CAS Registry Number specificity

Common Name*	CAS Registry Number
Xylene (unspecified isomer)	1330-20-7
o-Xylene	95-47-6
m-Xylene	108-38-3
p-Xylene	106-42-3
Epinephrine (unspecified stereoisomer or racemate)	329-65-7
d-Epinephrine	150-05-0
l-Epinephrine	51-43-4
l-Epinephrine hydrochloride	55-31-2
l-Epinephrine bitartrate	51-42-3
Ethylenediaminetetraacetic acid	60-00-4
monosodium salt	17421-79-3
disodium salt	139-33-3
trisodium salt	150-38-9
tetrasodium salt	64-02-8
sodium salt (unspecified ratio)	7379-28-4

*Note that these are common names and are not the *CA* Index Names used for these substances.

often unrecognized synonymous names with the description of molecular structure. A format was developed using hyphens to make the numbers easier to read and to identify. A Registry Number includes up to nine digits, which are separated into three parts by hyphens. The first part, starting from the left, has up to six digits, the second part has two digits, and the final part is a single check digit to verify the validity of the total number (e.g. 7732-18-5 is the CAS Registry Number for water).

The CAS Registry Number is written in general form as $N_i \ldots\ldots N_4N_3$-N_2N_1-R where R represents the check digit and N_i through N_1 represent the individual sequential numbers. The check digit is derived from the following formula.

$$\frac{iN_i + \ldots + 4N_4 + 3N_{3+} + 2N_2 + 1N_1}{10} = Q + \frac{R}{10}$$

where Q represents a digit that is discarded.

For example, water has the total Registry Number 7732-18-5, the validity of which is checked by means of the following formula.

$$\frac{(6\times 7) + (5\times 7) + (4\times 3) + (3\times 2) + (2\times 1) + (1\times 8)}{10} = \frac{105}{10} = 10 + \frac{5}{10}$$

Hence, Q is 10 (disregarded) and R (the check digit) is 5.

The Registry Structure File connection table (Figure 3.6) is a detailed inventory of the atoms and bonds that comprise the two-dimensional representation of a substance's structure. Within the connection table the computer identifies each ring system and replaces the ring portion of

the connection table with a ring system identifier. This separate identification of specific ring systems in the machine structure record allows the system to provide certain substance name generation support and makes it possible to readily recreate structure diagrams algorithmically from the structure record.

Stereochemical (or three-dimensional) representation has been an integral part of the CAS Registry System since its inception in 1965. Connection tables usually describe only the two-dimensional graph of a chemical structure. The third-dimensional detail may be added by means of a verbalized stereochemical descriptor associated with the connection table or by incorporation into the connection table itself. CAS chose to record the stereochemical information as a text-based descriptor. The stereochemical description included as a part of the *CA* substance index name has been a reflection of that descriptor.

This text-based approach to identifying stereoisomers served CAS internal needs for support of indexing and registration, but it did not serve external user needs for stereospecific structure display or search. To provide these additional capabilities, CAS began algorithmically adding specific atom and bond stereochemistry directly to the connection table in 1992 to make such information more amenable to computer handling and manipulation. Stereochemical parity data has now been incorporated into more than three-quarters of the stereospecific Registry connection table records.

The increasing utility of the Registry database in substance searching led to user requests for CAS to provide structure-based access to the earlier chemical literature. In 1984, CAS began to register the substances indexed in *CA* prior to 1965. This registration was accomplished by optically scanning the printed formula indexes to *CA* and converting the systematic names to connection tables using algorithms developed by CAS in the late 1960s and early 1970s. Registration of substances from the Sixth and Seventh Collective Indexes (1957–1966) was completed in 1990, with 690 000 new registrations added to the database.

There has been a rapid growth in the volume of biotechnology information in recent years. About one-third of the abstracts that now appear in *CA* are in the 19 biochemistry sections. To support evolving user needs in this important area, CAS has provided since 1993 what is probably the largest biosequence database in the world. Included are about 465 000 protein sequences indexed from the world's journals and patents and 2 million polynucleotide sequences from the same literature and the GenBank File, a US National Institutes of Health genetic sequence database. Because of their size, structure records for these large biomolecules consist of sequence notations using codes for the amino acid and nucleotide base units rather than using the typical atom-bond connection table records. A feature that distinguishes the CAS sequence portion of

the Registry database from other sequence databases is the presence of a large number of chemically modified sequences and sequences drawn from the patent literature.

In 1962, Dyson estimated that a registry system of all known compounds would require a capacity to handle 2.0 to 2.5 million compounds, of which 1.8 million would be organic compounds. Dyson's estimate was substantially higher than the figures commonly accepted at that time. As it turned out, these projections still significantly underestimated the growth of the file. Annual additions grew from an average of 262 000 in the first 5 years to an average of more than 1 million in each of the most recent 3 years. By the end of 1997, the CAS Registry File contained more than 17 million substances. The Registry System has withstood the test of time in being able to handle this large growth. The growth of the CAS Registry File is shown in Figure 3.8.

The Registry database also contains over 24 million names, both *CA* Index Names and the many trivial, commercial, systematic and semi-systematic names, and acronyms that CAS staff have encountered during indexing.

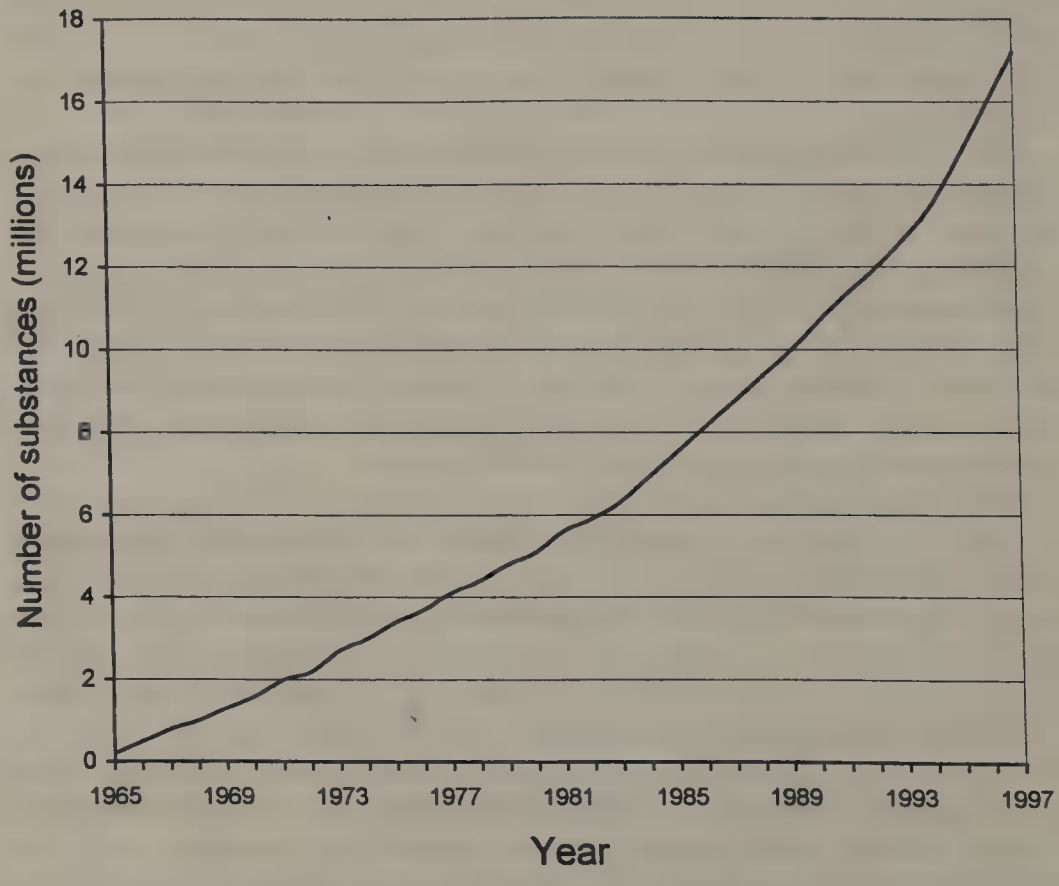


Figure 3.8 Growth of the CAS Chemical Registry file.

Table 3.5 CAS Chemical Registry database contents (31 December 1997)

Type of Substance	Number	%
Machine registered	14 740 878	85.5
Manually registered	2 499 612	14.5
Total	17 240 490	
Stereospecific compounds	3 810 102	22.1
Biosequences	2 480 156	14.4
Coordination compounds	1 091 375	6.3
Polymers	688 453	4.0
Alloys	570 973	3.3
Minerals	7 940	0.05

Today the CAS Chemical Registry database is the largest collection of information on naturally occurring and synthetic chemical substances in the world, including elements, atomic particles, organic compounds, inorganic compounds, polymers, coordination compounds, alloys, biosequences, and minerals. Table 3.5 provides a summary of some of the classes of chemical substances in the Registry database. Figure 3.9 shows a representative example of the range of information recorded in the Registry database for a given substance.

In addition to their inclusion in the CAS databases, CAS Registry Numbers are now used for substance identification in many public and private databases. Currently, the cluster of databases on STN International containing Registry Numbers includes 7 CAS databases and 55 non-CAS databases. Among the latter are such diverse files as AIDSLINE, BIOSIS, Design Institute for Physical Property Data File (DIPPR), MEDLINE, Natural Products Alert (NAPRALERT), Pharmaceutical News Index (PNI), and Plastics Materials Selection Database (PLASPEC).

Many handbooks, guides, and other reference works include CAS Registry Numbers and provide special Registry Number indexes that allow the reader immediate access to the proper place in the text without having to first identify the full name of the substance, which may differ from handbook to handbook resulting in unavoidable different alphabetical arrangements. Among these works are *CRC Handbook of Chemistry and Physics*, *Kirk-Othmer Encyclopedia of Chemical Technology*, *The Merck Index* published by Merck and Co., the *USP Dictionary of USAN and International Drug Names* published by the US Pharmacopoeial Convention, the *National Institute of Occupational Safety and Health (NIOSH) Registry of Toxic Effects of Chemical Substances*, and the proposed and recommended International Non-proprietary Names lists published by the World Health Organization.

CAS Registry Numbers are also widely used as standard identifiers for chemical substances in many governmental regulatory agency commercial

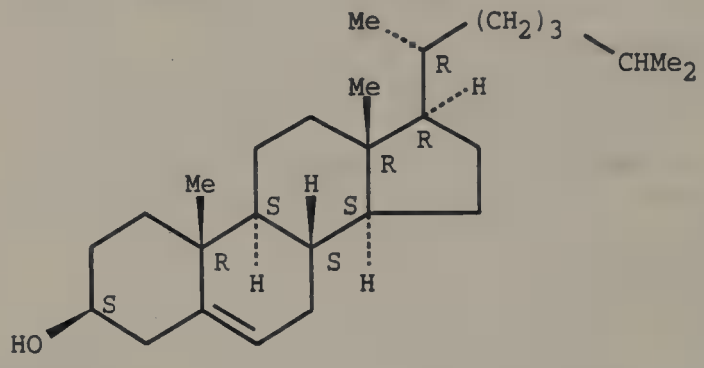
CAS Registry Number: 57-88-5

CA Index Name: Cholest-5-en-3-ol (3β)-

Synonyms: Cholesterol (8CI)
(-)-Cholesterol
Δ5-Cholesten-3β-ol
3β-Hydroxycholest-5-ene
5:6-Cholesten-3β-ol
Cholest-5-en-3β-ol
Cholesterin
Cholesteryl alcohol

Molecular formula: C27 H46 O

Structure (stored as connection table):



Ring System Data:

Elemental Analysis	Elemental Sequence	Size of the Rings	Ring System Formula	Ring Identifier (RID)	RID Occurrence
C5-C6-C6-C6	C5-C6-C6-C6	5-6-6-6	C17	4432.3.5	1

Figure 3.9 Example of CAS Chemical Registry file information. (Copyright American Chemical Society.)

chemical inventories such as the Toxic Substances Control Act (TSCA) Inventory in the U.S., the Ministry of International Trade and Industry (MITI) Inventory in Japan, the European Inventory of Existing Commercial Chemical Substances (EINECS), and the Canadian Domestic and Non-Domestic Substances Lists (DSL/NDSL). CAS Registry Numbers can often be found as part of the ingredients listings on many commercial products.

CAS Registry Numbers have become surrogates in many cases for chemical names because of their specificity, conciseness, and international acceptance. Today, whenever a chemical substance is sold, transported, imported, exported, reported to a regulatory agency, or disposed of, a CAS Registry Number is probably involved.

3.9 Computer-Supported Name Generation Systems

Because of the importance of index names as precise identifiers of chemical substances and as reliable access points in the printed indexes, CAS places great emphasis on assuring the accuracy of the names assigned by its nomenclature specialists. In addition, because of the large volume of newly encountered substances that require the generation of names, CAS has sought to make the process more efficient by adding computer support for the name assignment process and where possible by providing automatic assignment of names.

In the early days of *CA*, each indexer was expected to know the principles of chemical nomenclature and to assign the preferred index names to all chemical substances as they were indexed. With the introduction of the Registry System and the ability to match chemical substances being indexed with ones already on file by names or structures, it became possible to limit the intellectual derivation of names to just the newly encountered substances and thereby to centralize the nomenclature function within a group of staff specifically skilled in chemical nomenclature. The nomenclature experts could also be specialists in various areas of chemistry such as alkaloids, carbohydrates, peptides, steroids, alloys, coordination compounds, and polymers. This organization helped to assure the quality of the index names which, once assigned, could be stored on the Registry database and retrieved again and again as needed.

Additionally, CAS developed automated nomenclature editing procedures to validate the intellectually assigned substance names. These can be divided into three groups: automatic error detection and correction, automatic error detection with manual correction, and nomenclature translation. The first group involves the detection of errors that can be corrected unambiguously. These include errors in format such as capitalization, italicization, and super- and subscripting of data.

The second group involves the detection of errors and the display of diagnostic messages intended to alert the nomenclature editor to possible problems that require manual intervention for their resolution. These include the presence of unbalanced enclosing marks in name segments, punctuation errors that might indicate the omission of data or other problems (e.g. 2-chloro-methyl-), atomic masses that are inconsistent for the elements in question (e.g. mercury-^{97}Hg), inconsistent ring names (e.g. bicyclo[4.2.1]octane), and disagreement between the number of locants and the immediately following multiplying terms (e.g. 2,4-trichloro-).

One of the most powerful error detection programs developed by CAS is nomenclature translation. This is a program that converts a systematic chemical substance name to an atom-bond connection table and then compares that table with the connection table already on file for the substance being named. If the name is not translated or if the two structure

records do not correspond, the name must be reviewed and the name (or structure) record may have to be corrected. If the name is translated and the two structure records are identical, then the name is considered to be a fully accurate description.

With the introduction of Registry III in late 1973 came major adjustments in the Registry records to make the system more effective in its support of the CAS index name generation function. Of particular importance for support of name derivation is the ability of the system to recognize various types of partial structure matches between new substances and substances already on the file. Each kind of partial match permits a corresponding type of nomenclature support.

The three types of partial match are: (1) a match at the two-dimensional level but differing in some level of greater detail such as stereochemistry or isotopic labeling, (2) a match of the ring system with one already on file and (3) a match of a component substance with one already known (e.g. a match of one or more known monomers which are components of a new copolymer).

In the first case, there is a strong likelihood that any substance on file having the same topology as the new substance will have a name very similar to the new name that must be derived. Retrieval of the name for the related substance will simplify the intellectual assignment of the new substance's name since generally only minor alterations will be needed. Derivation of ring names can be one of the more challenging aspects of naming new substances, and therefore the matching of a ring system with one already on file, the second type of partial match, and the retrieval of that ring system's name can considerably simplify the naming process.

Many substances indexed and registered by CAS consist of two or more separately structured components. The principal classes of substances handled in this way are molecular addition compounds, copolymers, and mixtures. With the advent of Registry III, these multicomponent substances were registered by using the Registry Numbers of their components. The ability to provide automated support for the naming of these component-based substances arises from the fact that their index names are based largely on the names of the components. If some of the components of the new substance are already on file, the program is able to retrieve those names to support the manual index name assignment. If all of the components are already on file, the program is able to derive the index name(s) by automatically assembling the names of all of the components, including the generation of uninverted names for the name modification and the appropriate phrase (e.g. 'compd. with' or 'polymer with').

CAS has recently taken the next step in extending computer support to the naming of new chemical substances with the development of the Algorithmic Nomenclature Generation System (ANGS). ANGS generates

the unique CAS index name from connection table information for many organic and inorganic substances. The system implements the CAS nomenclature rules in a series of expert systems, that is a series of computer programs that consider a set of knowledge, reason, and then recommend a course of action. The systems examine the structure, identify the substructures and principal functional groups to be named, generate ring-system names, generate names for the substructures, generate stereodescriptors, and combine the resulting substructural names into complete substance names. Multiple viewpoints are employed when alternative names must be considered, for example, in tautomeric substances.

The objective of the ANGS effort is to represent all CAS nomenclature rules, to automate the production and timeliness of *CA* index names, and to improve the consistency of the names. In 1997, ANGS identified and named general organics, stereochemical isomers, tabular inorganics, coordination compounds, amino acids and peptides, carbohydrates, simple, fused, bridged fused, and von Baeyer ring systems and natural products with stereochemistry. About 60% of all CAS connection table-based registrations were being named by ANGS and efforts continue to extend the scope and utility of ANGS.

3.10 Challenges for the future

The challenge to CAS for the future is to be able to respond quickly to the changing nomenclature needs in rapidly evolving areas of science and technology such as biotechnology and materials science where substances of increasing complexity are being produced and for which there are often no established nomenclature standards. An indexing organization like CAS is faced with the need to index a document promptly to assure the most timely access for its users. While nomenclature commissions and committees have some time to consider, debate, and vet the options, indexing organizations must move more quickly.

Fortunately, CAS has a long history and considerable breadth of nomenclature experience that enable it to respond to these challenges in a timely and highly competent matter. In considering the options, CAS always looks to the future and tries to anticipate where the particular area of science is headed. CAS realizes it should not implement nomenclature practices that just serve the moment but are not extensible to a broader family of compounds than the immediate case. There is always a new example tomorrow of somewhat greater complexity, and CAS must be prepared.

Bibliography

Dittmar, P. G., Stobaugh, R. E., and Watson, C. E. (1976) The Chemical Abstracts Service Chemical Registry System. I. General design. *Journal of Chemical Information and Computer Sciences*, **16(2)**, 111–121.

Donaldson, N., Powell, W. H., Rowlett, R. J., Jr, White, R. W. and Yorka, K. V. (1974) Chemical Abstracts index names for chemical substances in the Ninth Collective Period (1972–1976). *Journal of Chemical Documentation*, **14(1)**, 3–15.

Giles, P. M., Jr and Metanomski, W. V. (1993) The history of chemical substance nomenclature at Chemical Abstracts Service (CAS), in *Organic Chemistry: Its language and Its State of the Art* (ed. M. V. Kisakürek) Basel: Verlag Helvetica Chimica Acta; Weinheim: VCH, pp. 173–196.

Leiter, D. P., Jr, Morgan, H. L. and Stobaugh, R. E. (1965) Installation and operation of a registry for chemical compounds. *Journal of Chemical Documentation*, **5(4)**, 238–242.

Patterson, A. M., Capell, L. T. and Walker, D. F. (1960) *The Ring Index*, 2nd edn, American Chemical Society, Washington.

Stobaugh, R. E. (1988) The Chemical Abstracts Service Chemical Registry System. 11. Substance-related statistics: Update and additions. *Journal of Chemical Information and Computer Sciences*, **28(4)**, 180–187.

Vander Stouw, G. G., Gustafson, C., Rule, J. D. and Watson, C. E. (1976) The Chemical Abstracts Service Chemical Registry System. IV. Use of the Registry System to support the preparation of index nomenclature. *Journal of Chemical Information and Computer Sciences*, **16(4)**, 213–218.

Vander Stouw, G. G. (1975). Computer programs for editing and validation of chemical names. *Journal of Chemical Information and Computer Sciences*, **15(4)**, 232–236.

Weisgerber, D. W. (1997). Chemical Abstracts Chemical Registry System: History, scope, and impacts. *Journal of the American Society for Information Science*, **48(4)**, 349–360.

4 IUPAC nomenclature part 1, organic

K. J. THURLOW

4.1 Introduction

Man: Hello my boy. And what is your dog's name?
Boy: I don't know. We call him Rover.

(New Scientist 3 October 1974)

Chemical names of some sort have been used for a long time, but it was not until the nineteenth century that serious attempts were made to produce a logical system of chemical nomenclature. The Geneva Conference [1] of 1892 was set up to examine nomenclature of organic chemicals and attempt to produce unified rules. Although it was impossible to produce rules for everything, many of the guidelines produced then are still in use today. The International Union of Pure and Applied Chemistry (IUPAC) was formed in 1919, and in 1923 set up a group to consider nomenclature of organic chemicals. As time passed, more and more aspects of nomenclature were considered and recommendations were published for naming chemicals of all kinds. A major problem is that organic chemistry is viable, in its true sense. Just when you think you have rules covering every possibility, somebody finds an entirely new chemical structure. A recent example is the discovery of fullerenes – how do you decide which carbon atom is number two? Work is constantly in progress to cover new eventualities and to tidy up old problems.

The Chemical Abstracts Service (CAS) and Association for Science Education (ASE) have both used IUPAC recommendations, but have refined them for their own use and needs. IUPAC recommendations are subject to extended review in an attempt to gain consensus. This can be a lengthy process. Recommendations on 'phane' nomenclature (of which more anon) have been in the pipeline for some years. CAS registers thousands of new chemicals every year and cannot wait, so have produced their own names in a number of areas, although the general principles are largely the same. An important difference is that CAS aims for the unique name, clearly important for indexing purposes, whereas IUPAC rules frequently permit equally valid alternatives. Work is continuing to provide IUPAC recommendations for a 'preferred' name, and it is expected that these recommendations will be published in the near future. ASE are teaching students about chemical structures, so have become more systematic than IUPAC. IUPAC still use acetic acid and pyrogallol

as permitted names, but ASE [2] prefer the more systematic approach, and use ethanoic acid and benzene-1,2,3-triol. These names tell you more about the structure, an obvious advantage for teaching purposes, but any student moving on to industry will soon need to learn the more traditional names. There is a place for trivial names, but the oft repeated suggestion, even in apparently serious scientific magazines, that you can get by using **only** trivial names is misguided.

The Laboratory of the Government Chemist (LGC) was formed in 1842 and first became involved in nomenclature in the First World War when it was realized that the dyestuff industry was in enemy hands. A list of essential chemicals was compiled so that they might be protected from full import duty. LGC continues to this day to advise Government departments on chemical matters. Interest in nomenclature continued and the Chemical Nomenclature Advisory Service (CNAS) was formed in 1976 and the role has expanded. A number of bodies outside Government, including the British Standards Institution (BSI) and the International Organization for Standardization (ISO) receive input from CNAS, as well as an increasing number of private companies, especially since LGC was privatized in 1996. The entry of the United Kingdom into the European Community (EC) saw LGC providing names for the European Customs Inventory of Chemical Substances (ECICS). These ran from number 10001 (the start) to number 26591. This was referred to in CNAS as the 'Line of Death' because subsequent names were taken from any source without much checking. This speeded things up, but produced duplications, inconsistencies and mistakes. More recently, the EC has been combing ECICS for errors and has corrected many of them, as well as deleting items no longer of interest. Unfortunately, some numbers have been reused, so the 'Line of Death' has become blurred.

CNAS has traditionally provided IUPAC names, although obviously some knowledge of other systems is necessary. If a customer wants an IUPAC name, we are always delighted to see a CAS number to help identify the structure correctly, which is not always possible from the name the customer has given! We have had considerable involvement with ISO names for pesticides and International Non-Proprietary Names (INN) for pharmaceutical products. It is important with names provided for lists that a consistent approach is adopted. As mentioned earlier, IUPAC does allow alternative names, so we try to give the same type of name to a customer, particularly if it is part of a list. IUPAC rules do not always cover the structure you are considering. CNAS has compiled a massive set of rationales over the years, which complement and expand on IUPAC recommendations. It is hoped that these ideas are logical, but every nomenclator has his own ideas and is normally prepared to defend them to the end of time. Indeed, nomenclators cannot even agree on whether they are nomenclators or nomenclaturists. These two chapters on IUPAC

nomenclature will give an idea on how to produce names and will incorporate some of the LGC rationales. These have been devised over many years and every member of CNAS has contributed to their formulation.

4.2 Organic names

It is impossible to provide a complete coverage of organic nomenclature in such a short space. Most IUPAC organic nomenclature is based on IUPAC's *Nomenclature of Organic Chemistry* (1979) [3] known as the 'Blue Book', because of the colour of its cover. The 1993 *Guide to IUPAC Nomenclature of Organic Compounds* [4] (also with a blue cover) expands on the Blue Book, but does not supersede it. It is generally known as the Guide. Biochemical recommendations are covered in the 'Compendium of Biochemical Nomenclature' [5] (1992), known as the 'White Book'. Other rules and recommendations are published from time to time, usually in *Pure and Applied Chemistry*.

4.2.1 Where do we begin?

Nomenclature is not easy. It takes years to become an expert and, even then, you suddenly find that you are presented with a structure that appears not to be covered in the books. But it should be possible to deal with most of the structures. A former head of CNAS, E. W. Godly, produced a book [6] with this laudable aim, using flow diagrams.

Let us start with a saturated acyclic hydrocarbon (Figure 4.1). Saturated means that there are no double bonds, acyclic that there are no rings, and hydrocarbon shows that only hydrogen and carbon are present.

$$CH_3–CH_2–CH_2–CH_2–CH_2–CH_3$$

Figure 4.1 Hexane.

'Hex' means six carbons, and 'ane' tells you that it is saturated.

Cyclohexane similarly means a cyclic saturated hydrocarbon, normally shown as below in Figure 4.2:

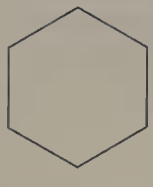


Figure 4.2 Cyclohexane.

A long list of multiplying prefixes is given in the 'Blue Book.' The first four hydrocarbons are not systematically named, but thereafter a classical education helps identify the number of carbons present. Dodecane has 12 carbons, octadecane 18, etc. Occasionally there are suggestions that the earlier alkanes should be named systematically, but it is unlikely that the scientific community would take any notice if methane, ethane, propane and butane were replaced with ane, diane, triane and tetrane, so the former names will undoubtedly survive.

4.2.2 How do we decide on a name?

Even simple compounds usually have other groups attached. These fall into two categories, those that are named only as prefixes, and those named as prefixes or suffixes (see Figure 4.3 and Figure 4.4). Figure 4.5 shows the decreasing priority of principal groups. These figures are taken from the 'Blue Book.'

Figure 4.6 shows the schematic approach to naming all organic compounds. Thus, the normal procedure is:

1. Identify Principal Group (PG) if any.
2. Identify Parent Structure.
3. Collect several Principal Groups together if possible.
4. Identify other characteristic groups directly attached to parent

Characteristic group	*Prefix*	*Cf. Rule*	*Page*
–Br	Bromo	C–102.1	144
–Cl	Chloro	C–102.1	144
–ClO	Chlorosyl	C–106.2	146
$-ClO_2$	Chloryl	C–106.2	146
$-ClO_3$	Perchloryl	C–106.2	146
–F	Fluoro	C–102.1	144
–I	Iodo	C–102.1	144
–IO	Iodosyl	C–106.1	146
$-IO_2$	Iodyl (replacing iodoxy)	C–106.1	146
$-I(OH)_2$	Dihydroxyiodo	C–106.3	146
$-IX_2$	X may be halogen or a radical, and the prefix names are dihalogenoiodo, *etc.*, or, for radicals, patterned on diacetoxyiodo	C–106.3	146
$=N_2$	Diazo	C–931.4	290
$-N_3$	Azido	C–941.1	291
–NO	Nitroso	C–851.1	275
$-NO_2$	Nitro	C–852.1	275
=N(O)OH	*aci*-Nitro	C–852.1	275
–OR	R-oxy	C–205.1	154
–SR	R-thio (similarly R-seleno and R-telluro)	C–514.1	213

Figure 4.3 List of groups used only as prefixes.

Class	*Formula**	*Prefix*	*Suffix*	*Cf. Rule*
Cations		-onio -onia	-onium	C–82 B–6
Carboxylic acid	–COOH –(C)OOH	Carboxy –	-carboxylic acid -oic acid	C–401
Sulfonic acid	$-SO_3H$	Sulfo	-sulfonic acid	C–641
Salts	–COOM –(C)OOM	– –	Metal ... carboxylate Metal ... oate	C–461
Esters	–COOR –(C)OOR	R-oxycarbonyl –	R ... carboxylate R ... oate	C–463
Acid halides	–CO–Halogen –(C)O–Halogen	Haloformyl –	-carbonyl halide -oyl halide	C–481
Amides	$-CO-NH_2$ $-(C)O-NH_2$	Carbamoyl –	-carboxamide -amide	C–821 C–822
Amidines	$-C(=NH)-NH_2$ $-(C)(=NH)-NH_2$	Amidino –	-carboxamidine -amidine	C–951
Nitriles	–C≡N –(C)≡N	Cyano –	-carbonitrile -nitrile	C–832
Aldehydes	–CHO –(C)HO	Formyl Oxo	-carbaldehyde -al	C–301 C–301
Ketones	>(C)=O	Oxo	-one	C–311
Alcohols	–OH	Hydroxy	-ol	C–201
Phenols	–OH	Hydroxy	-ol	C–202
Thiols	–SH	Mercapto	-thiol	C–511
Hydroperoxides	–O–OH	Hydroperoxy	–	C–218
Amines	$-NH_3$	Amino	-amine	C–812
Imines	=NH	Imino	-imine	C–815
Ethers	–OR	R-oxy	–	C–211
Sulphides	–SR	R-thio	–	C–514
Peroxides	–O–OR	R-dioxy	–	C–218

Figure 4.4 Group names used as prefixes and suffixes.

1. 'Onium and similar cations (see Subsection C–0.8)
2. Acids: in the order COOH, C(=O)OOH, then successively their S and Se derivatives, followed by sulfonic, sulfonic acids, etc.
3. Derivatives of acids: in the order anhydrides, esters, acyl halides, amides, hydrazides, imides, amidines, etc.
4. Nitriles (cyanides), then isocyanides
5. Aldehydes, then successively their S and Se analogues; then their derivatives
6. Ketones, then their analogues and derivatives, in the same order as for aldehydes
7. Alcohols and phenols; then their S, Se and Te analogues; then neutral esters of alcohols and phenols with inorganic acids, except hydrogen halides, in the same order
8. Hydroperoxides
9. Amines; then imines, hydrazines, etc.
10. Ethers; then successively their S and Se analogues
11. Peroxides

Figure 4.5 Decreasing order of priority as principal group.

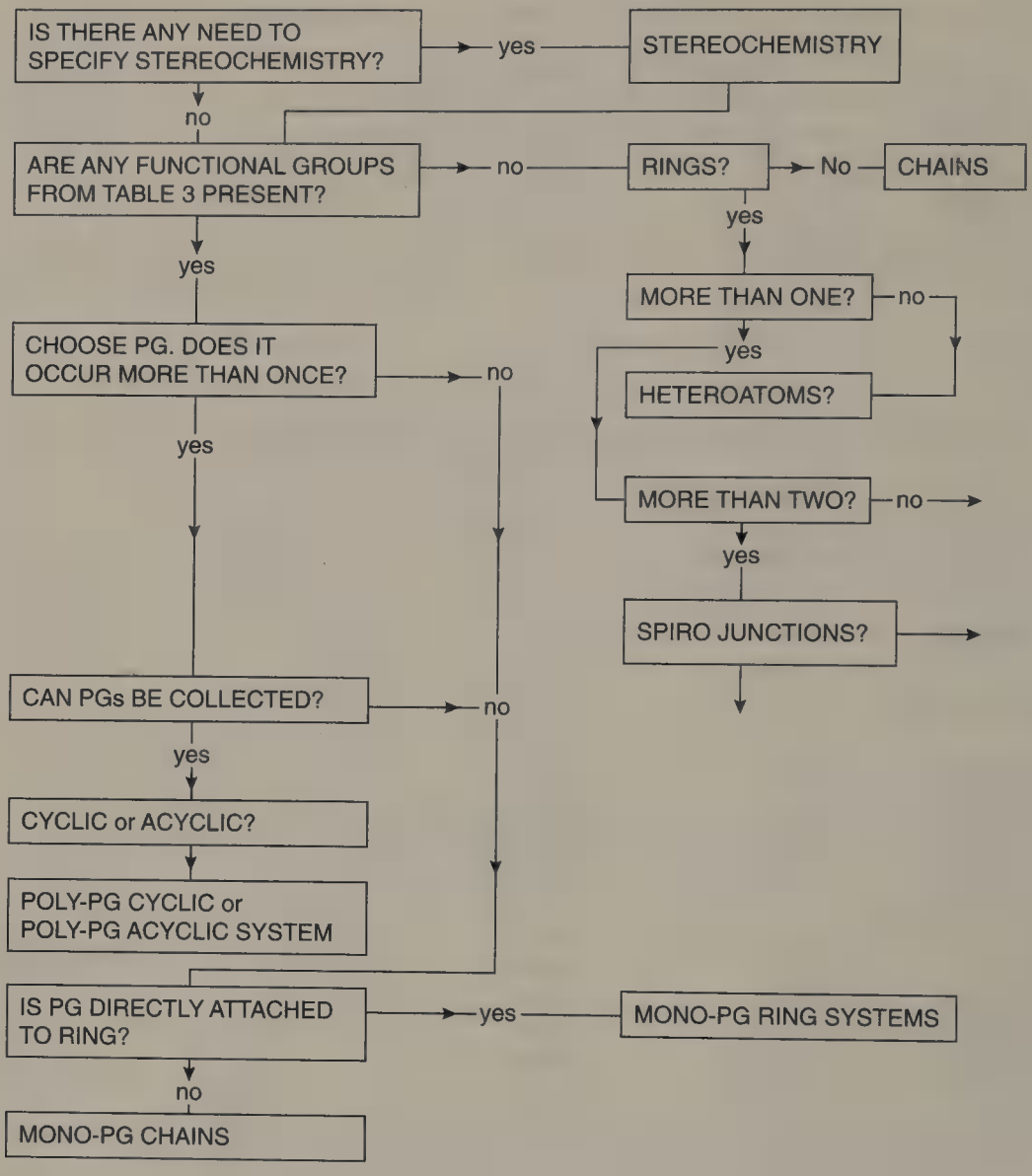


Figure 4.6 How do we decide on a name?

5. Identify other groups attached to parent; e.g. simple radicals, composite radicals, etc.
6. Assemble components into complete name, using alphabetical order for substitutive prefixes accompanied by appropriate locants.
7. Add stereochemical descriptors if necessary.

Consider this structure (Figure 4.7):

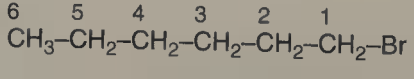


Figure 4.7 1-Bromohexane.

This can be called hexyl bromide in radicofunctional nomenclature, but substitutive nomenclature is normally preferred. Using the above scheme:

Find Principal Group (Figure 4.5). NONE! – Br can only be used as a prefix.
Find Parent Structure. Six carbons in a line = hexane.

This is the longest possible chain of carbons. So this is bromohexane, but this is still ambiguous, as the name does not tell us where the Br is attached. It needs a locant to tell us this. The chain is numbered from one end to the other, giving the lowest number to the carbon with the attached group. Therefore, the correct name is 1-bromohexane, not 6-bromohexane. This principle of lowest locant applies throughout, although seniority of attachments comes into play with more complicated examples. Consider the structure (Figure 4.8):

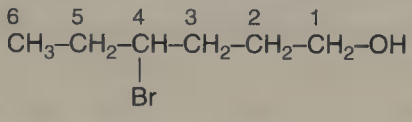


Figure 4.8 4-Bromohexan-1-ol.

Find Principal Group – -OH
Find Parent Structure – six carbons in a line = hexane

In this case, the -OH is most important, so we number from that end of the chain and find the name 4-bromohexan-1-ol. The change in the ending from '-ane' to '-anol' reflects the presence of the alcohol group. Note the positioning of the locants. The 'Blue Book' uses the style '1-hexanol', but CNAS has used 'hexan-1-ol' for many years, and the Guide also adopts this approach. It clarifies matters if locants are as close as possible to the part of the name to which they refer. Note that the 'e' is elided in '-an-1-ol', as another vowel follows the 'e', although a number splits the two parts of the name.

If we add another -OH group to the hexane, the name would change significantly (Figure 4.9).

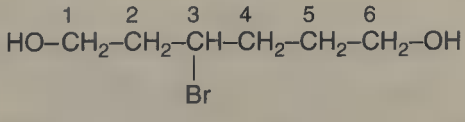


Figure 4.9 3-Bromohexane-1,6-diol.

Here we have a choice of which way to number the chain, as there is an -OH on each end. Either could be '1'. The 'tie-break' is that the bromine takes a lower locant if the structure is numbered as shown. This

is 3-bromohexane-1,6-diol. It would be incorrect to say 4-bromohexane-1,6-diol. Note that the elided 'e' has returned, as it is now followed by a consonant.

The next example offers a choice of Principal Groups (Figure 4.10):

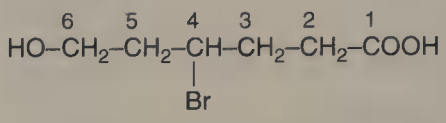


Figure 4.10 4-Bromo-6-hydroxyhexanoic acid.

Both -OH and -COOH can be named as suffixes, but Figure 4.5 shows us that the -COOH is senior. The -OH must be named as a prefix, because you are only allowed one type of suffix. Again we have six carbons, but note this time that the carbon in the -COOH is included in the total carbon count. That carbon becomes number 1, so this compound is called 4-bromo-6-hydroxyhexanoic acid. This raises several important points. The '-ane' ending is replaced by '-anoic acid', and this tells you that the acid group is on position '1'. The -OH group is now a prefix, so it becomes '6-hydroxy'. The two prefixes are cited in alphabetical order. 'B' for bromo comes before 'H' for hydroxy. There are no gaps in the name except before 'acid'.

Similar considerations apply to cyclic compounds. Figures 4.11a and 4.11b are examples.

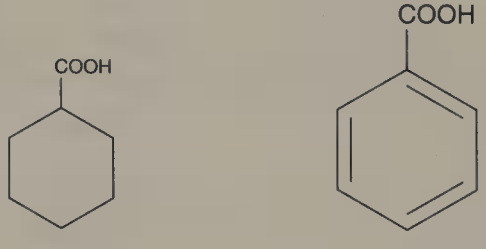


Figure 4.11 (a) Cyclohexanecarboxylic acid. (b) Benzoic acid.

For Figure 4.11a, the principal group is again the -COOH, and the parent structure is cyclohexane. Here the -COOH is additional; it does not replace a CH_3, so the name becomes cyclohexanecarboxylic acid. A locant is not necessary, because all the possible points of attachment are equal. Figure 4.11b could be called benzenecarboxylic acid, but traditionalists will be pleased to know that benzoic acid has been retained for this structure.

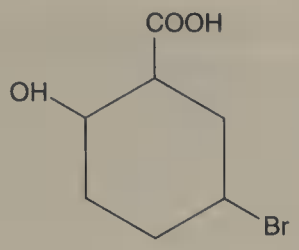


Figure 4.12 5-Bromo-2-hydroxycyclohexanecarboxylic acid.

Figure 4.12 continues the theme of substitution. The Principal Group is -COOH, which is senior to -OH. This structure is called: 5-bromo-2-hydroxycyclohexanecarboxylic acid, **not** 3-bromo-6-hydroxycyclohexanecarboxylic acid, **nor** 2-hydroxy-5-bromocyclohexanecarboxylic acid. The -COOH is senior and takes position '1', but this is understood so you do not have to cite it. CNAS takes the attitude that if you can omit a locant or bracket **without ambiguity**, then do so. After assigning position '1', the other numbers are inserted so as to give all the substituents lowest locants, hence '2,5' is lower than '3,6'. Then the prefixes are put in alphabetical order.

However, locants must be used if there is a danger of ambiguity. You cannot call Figure 4.13a hexabromocyclohexane, because it could equally apply to Figure 4.13b.

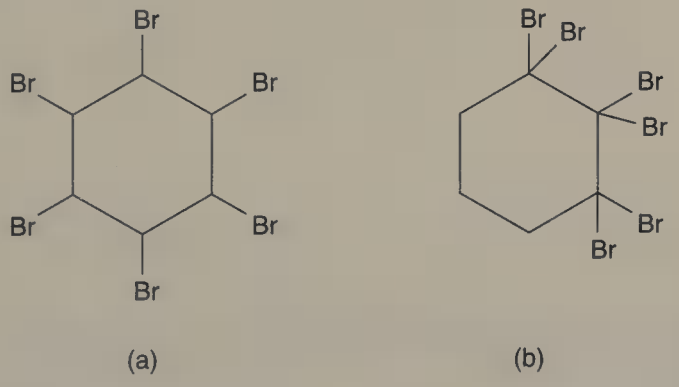


Figure 4.13 (a) 1,2,3,4,5,6-Hexabromocyclohexane. (b) 1,1,2,2,3,3-Hexabromocyclohexane.

So far we have looked at chains and rings, but only on their own. Suppose we have a mixture like Figure 4.14.

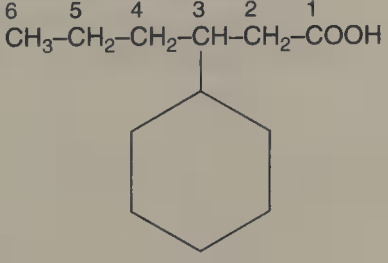


Figure 4.14 3-Cyclohexylhexanoic acid.

Again the Principal Group is -COOH and it is attached to the chain, not the ring. Hence the chain becomes the parent structure, and we have a substituted hexanoic acid. Radicals are formed by replacing the '-ane' ending with '-yl', so this structure is named 3-cyclohexylhexanoic acid.

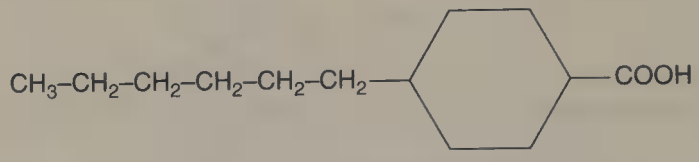


Figure 4.15 4-Hexylcyclohexanecarboxylic acid

Figure 4.15 above illustrates the opposite case. The ring is now the parent structure and the acid group is at position '1'. Hence, the name is 4-hexylcyclohexanecarboxylic acid. The hexane is numbered so as to give the lowest locant to the point of attachment. So far we have been making fairly gentle progress through the field of organic nomenclature, but this field is littered with rabbit holes, waiting to wrench the ankle of the unwary. One such rabbit has been at work in Figure 4.16.

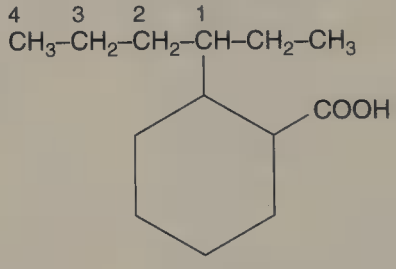


Figure 4.16 2-(1-Ethylbutyl)cyclohexanecarboxylic acid.

The parent is cyclohexanecarboxylic acid, and a six-membered chain is attached to position '2' on the ring. However, this time the point of attachment is not the end of the chain, it is somewhere in the middle. Does this make a difference? Well, it depends which book you use. The 'Blue Book' stresses (Rule A-1.2) that for univalent radicals, the carbon atom with free valence is numbered as 1. Take the longest carbon chain, with the point of attachment as '1', and you reach four carbons if you go to the left, or three if you go to the right. Four beats three, so the radical is 'butyl'. This is substituted itself at position '1', so the radical becomes 1-ethylbutyl. This shows there is an 'ethyl' on the 'butyl' at position '1'.

However, the Guide (rule R-2.5) does also permit the use here of 'hexan-3-yl'. Many traditionalists threw their arms up in horror when they saw that, but one can see it has the advantages that it is shorter and simpler. Compare the two possibilities:

2-(1-ethylbutyl)cyclohexanecarboxylic acid
2-hexan-3-ylcyclohexanecarboxylic acid

CNAS uses the former for the sake of continuity. This name introduces the use of parentheses. The parentheses show that everything within them is on the '2' position of the cyclohexane. Figure 4.17 is a more complex example.

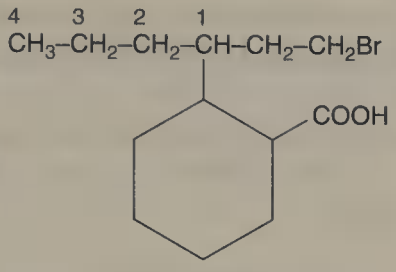


Figure 4.17 2-[1-(2-Bromoethyl)butyl]cyclohexanecarboxylic acid.

The (2-bromoethyl) is all on the '1' position of the butyl, giving [1-(2-bromoethyl)butyl], then all of that is on the '2' position of the cyclohexane as before. Of course this can be complicated further.

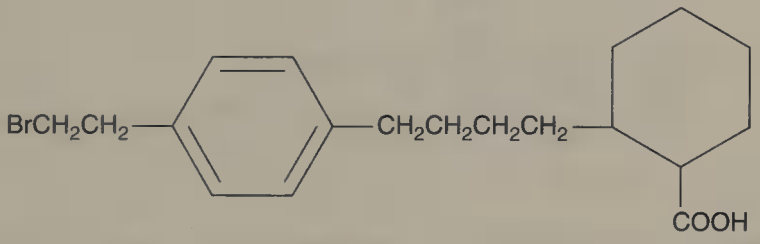


Figure 4.18 2-{4-[4-(2-Bromoethyl)phenyl]butyl}cyclohexanecarboxylic acid.

In Figure 4.18 the (2-bromoethyl) is all on the 4 position of the phenyl, giving [4-(2-bromoethyl)phenyl]. This is all on the 4 position of the butyl, giving {4-[4-(2-bromoethyl)phenyl]butyl} . This is all on the 2 position of the cyclohexane. Note the nesting order of the brackets, which is {[()]}. If necessary, this is repeated *ad infinitum*, leading to {[({[()]})]}. This differs from the CAS system, which uses only ( ) and [], so you would have [[[()]]], etc.

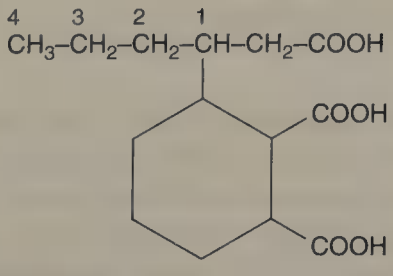


Figure 4.19 3-(1-Carboxymethylbutyl)cyclohexane-1,2-dicarboxylic acid.

In Figure 4.19 the principal group is carboxylic acid, which appears on both the chain and the ring. As the ring has more carboxylic groups, it is regarded as senior, and the carboxylic group on the chain is named as a prefix. You may wonder what happens if both chain and ring have an equal number of senior groups. The rules do not cover this possibility, so CNAS recommends making a decision based on the relative sizes of the ring and chain, or if you are dealing (e.g.) with a whole series of cyclohexanecarboxylic acids with alkyl substitution, then continue the series in a similar manner.

It has already been stated that substituents are cited in alphabetical order. This is true, but the definition of alphabetical order can vary. Consider Figures 4.20–4.22.

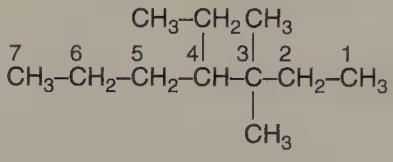


Figure 4.20 4-Ethyl-3,3-dimethylheptane ('E' before 'M').

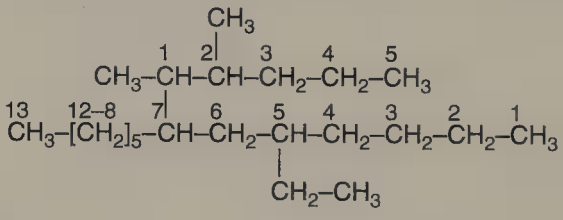


Figure 4.21 7-(1,2-Dimethylpentyl)-5-ethyltridecane ('D' before 'E').

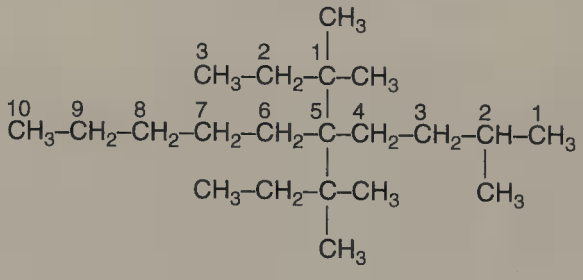


Figure 4.22 5,5-Bis(1,1-dimethylpropyl)-2-methyldecane ('D' before 'M').

These demonstrate this and also the use of 'bis' rather than 'di' as a multiplying prefix. In Figure 4.20, the substituents are the **simple** radicals, 'ethyl' and 'dimethyl', but the alphabetization is decided first, 'E' before 'M', then the multiplying prefixes inserted. Figure 4.21 includes the **complex** radical '2-methylpentyl', which is considered to start with the first letter of its complete name, 'M'. The same process applies in Figure 4.22.

In addition, we see that 'di' (tri, tetra, etc.) are used as multiplying prefixes for simple radicals, but 'bis' (tris, tetrakis, etc.) are used if multiplying complex radicals.

4.2.3 Unsaturated compounds

So far we have only considered saturated compounds, but similar procedures apply to unsaturated hydrocarbons. Instead of the '-ane' ending for saturated hydrocarbons, '-ene' means double bond(s), and '-yne' means triple bond(s) (Figures 4.23–4.25).

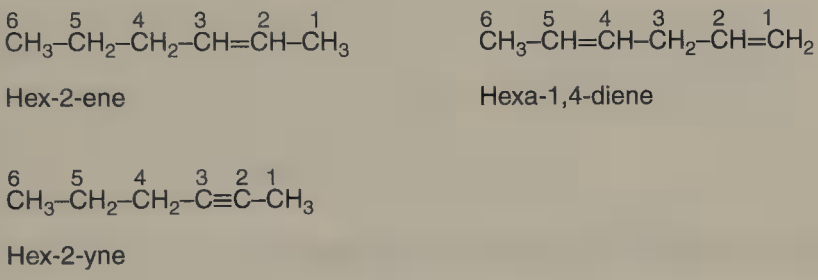


Figure 4.23

Note the elision of the 'a' when it precedes 'e' in Figure 4.23.

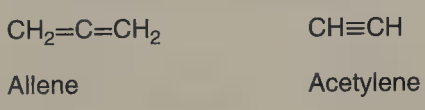


Figure 4.24

Figure 4.24 shows permitted trivials, but note that ethylene may now only be used for $-CH_2-CH_2-$, not $CH_2=CH_2$.

Figure 4.25 shows mixed double and triple bonds.

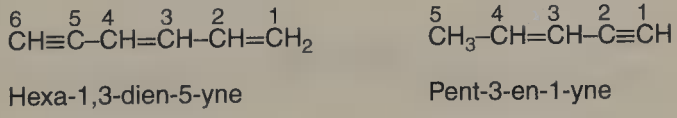


Figure 4.25

Multiple bonds take the lowest possible numbers, but when there is a choice, the lowest number goes to the type of bond that appears more, or if that is equal, to the double bond, as 'ene' is alphabetically before 'yne'.

Radicals are treated as follows:

1. Take chain with most double and triple bonds.
2. Take chain with most carbon atoms.
3. Take chain with most double bonds.

Again, the radical occupies position 1 (Figures 4.26 and 4.27).

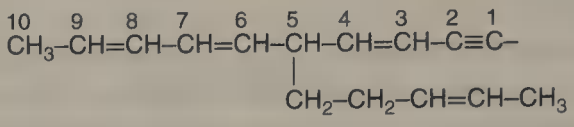


Figure 4.26 5-(Pent-3-enyl)deca-3,6,8-trien-1-ynyl.

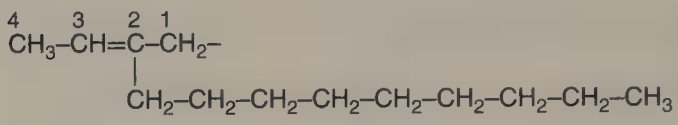


Figure 4.27 2-Nonylbut-2-enyl.

But if you remove the double bond (Figure 4.28):

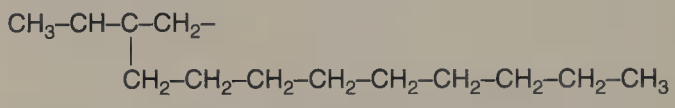


Figure 4.28 2-Ethylundecyl.

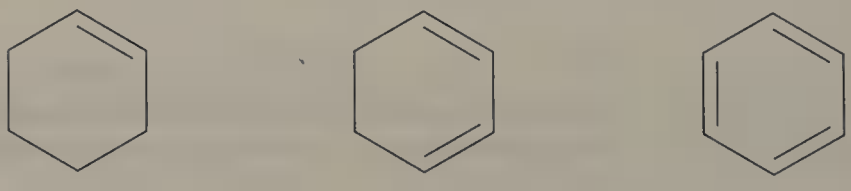
CAS prefers to use the longest chain in each case, so they would call Figure 4.27 2-ethenylundecyl. IUPAC are considering using the same approach.

Multivalent radicals are formed as below (Figure 4.29).

–CH_2–CH_2–CH_2–CH_2–CH_2–CH_2	hexamethylene (old) or hexane-1,6-diyl (new)
≡C–CH_2–CH_2–C≡	butanediylidyne
≡C–CH_2–CH_2–C=	butan-1-yliden-4-ylidyne

Figure 4.29

4.2.4 Unsaturated cyclic compounds



Cyclohexene Cyclohexa-1,3-diene Benzene

Figure 4.30

In Figure 4.30 the double bonds in benzene are shown in the usual stylized way; a circle inside the ring may be a more accurate representation. One magazine recently reported the isolation of 'cyclohexa-1,3,5-triene', but it was not clear if this were serious or not.

Figure 4.31 below shows radicals from unsaturated cyclic compounds. Note that where possible, the radical is position 1, or failing that, as low as possible.

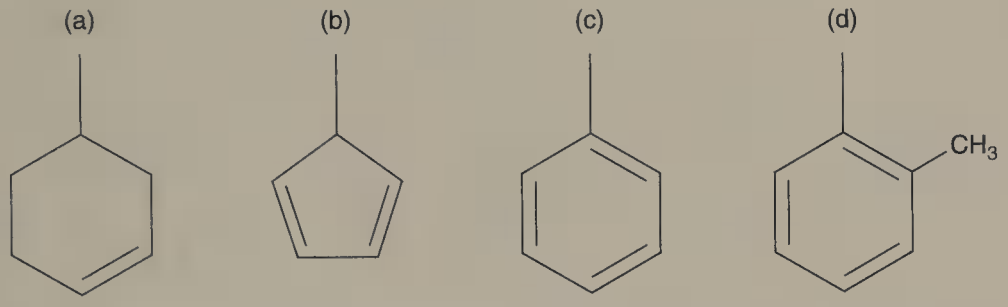


Figure 4.31 (a) Cyclohex-3-en-1-yl; (b) cyclopenta-2,4-dien-1-yl; (c) phenyl; (d) *o*-tolyl.

4.2.5 Multivalent radicals from unsaturated cyclic compounds

See Figure 4.32 below.

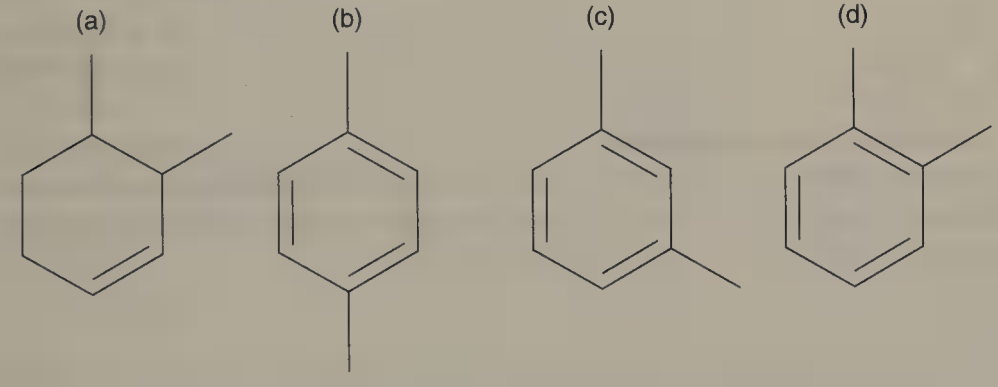


Figure 4.32 (a) Cyclohex-3-en-1,2-ylene; (b) *p*-phenylene; (c) *m*-phenylene; (d) *o*-phenylene.

Again, the radicals take the lowest possible locants. '*Ortho*', '*meta*' and '*para*' are used for convenience, but numbers can be used instead. Indeed, numbers are sometimes clearer.

4.2.6 Fusion

A large number of cyclic systems are rather complicated. Luckily, IUPAC allows many trivial names, else, life would be very difficult. Obviously it

is not possible to have trivial names for every structure, so it is necessary to fuse certain permitted trivially named structures. These must be unsaturated systems. If any ring is saturated, it may only be fused if it is assumed to be unsaturated, then 'hydro' prefixes used afterwards to reflect the saturation. If 1,2,3,4-tetrahydronaphthalene were regarded as a fusion compound, you would have to treat it as benzene fused to benzene, and then add the hydro prefixes (Figures 4.33a and 4.33b).

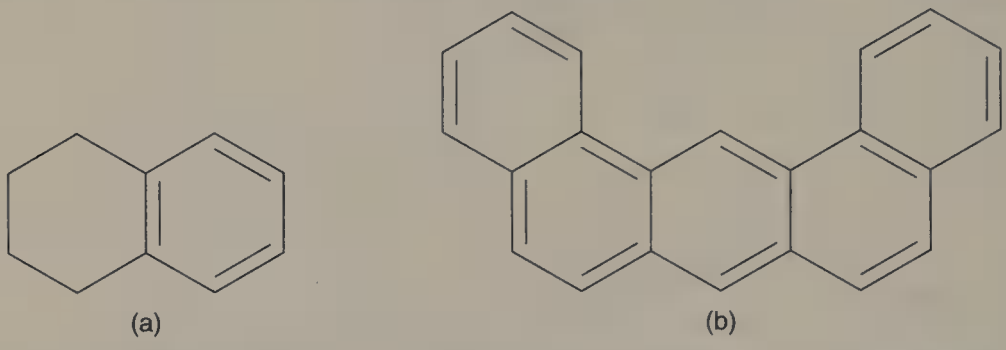


Figure 4.33 (a) 1,2,3,4-Tetrahydronaphthalene; (b) dibenzo[*a,j*]anthracene.

Imagine the problems in naming dibenzo[*a,j*]anthracene if you were not allowed to fuse anything bigger than benzene. 'Benzo' indicates that a benzene has been fused on to the anthracene. A brief explanation of this example is that you select the most senior component, here the anthracene and add the two benzenes, one at each end. However, you need locants to establish where the benzenes fit. The smaller components use numbers as usual (when necessary), but the larger component uses italicized letters to identify the sides, rather than numbers to identify the atoms. Number one atom in anthracene is at the top of the right hand ring, then it numbers clockwise round the first ring. So the side 1-2 is called '*a*', then the next one '*b*', etc. Hence, the right-hand benzene is on side '*a*' and the left-hand one is on side '*j*'.

Once the rings are fused, they have to be renumbered to take account of the entire structure. The fused system has to be oriented correctly before the numbering can be assigned. Figure 34a shows this orientation.

You put the greatest number of rings in the horizontal row, then the maximum number of rings in the upper right quadrant, then the minimum number of rings in the lower left quadrant. Then number from the most anti-clockwise position of the uppermost ring farthest to the right. However, atoms common to two or more rings are not given fresh numbers. So the correct orientation in Figure 4.34a is numbered 1, 2, 3, 3a, 4, 5, 5a, etc.

Consider Figure 4.34b which shows another aspect of numbering. Note the numbering in the square brackets, which is based on the original

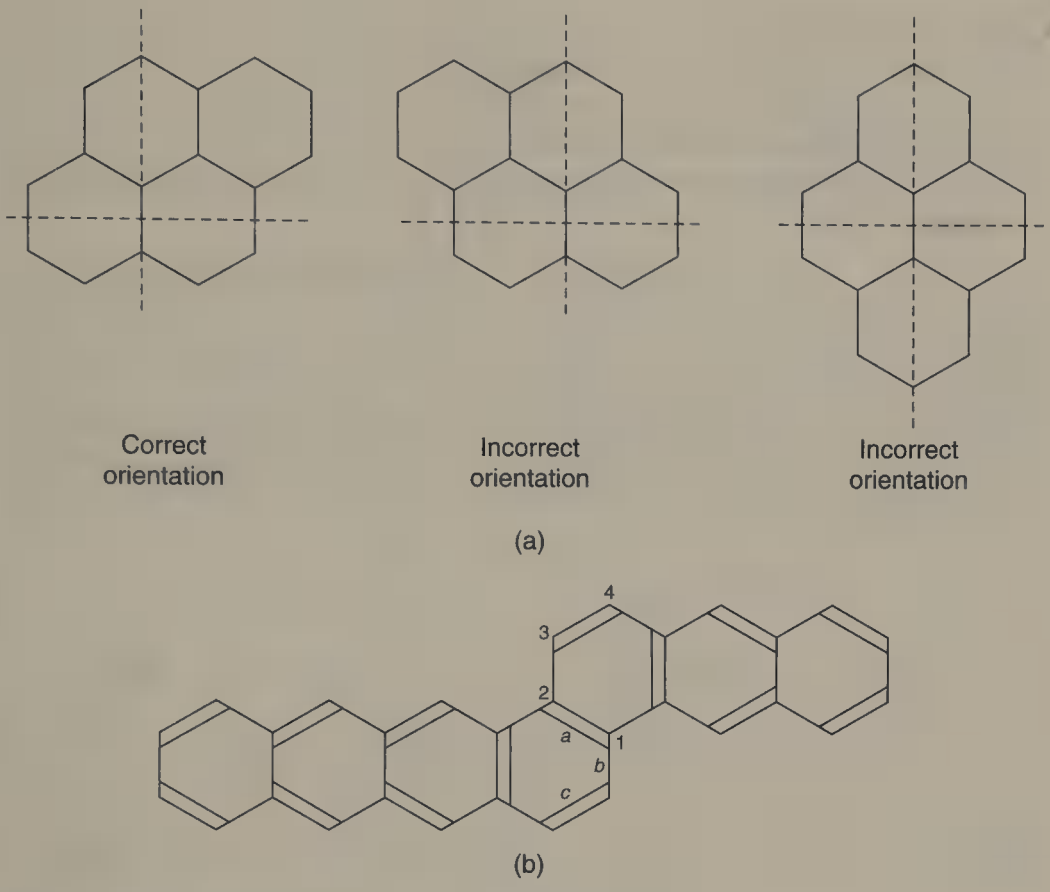


Figure 4.34

numbering of the individual components, rather than the numbering of the fused system. More detail may be found in the section on fused heterocyclic systems later in this chapter.

4.2.7 Bridges

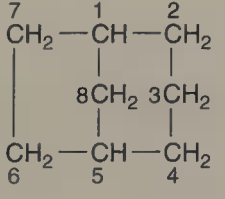


Figure 4.35 Bicyclo[3.2.1]octane.

The basis of the name bicyclo[3.2.1]octane (Figure 4.35) is quite simple. 'Bicyclo' tells you that there are two rings. 'Octane' means there are eight carbons, with no double bonds. Numbering is important:

1. Start at bridgehead.
2 Number by longest path to second bridgehead.
3. Now take longer remaining path to first bridgehead.
4. Take shorter path between bridgeheads.
5. The square bracket shows the number of carbons between the bridgeheads.

In the above example, which shows the correct numbering, start from position 1. There are three routes to position 5, the longest containing 3 carbons, (numbers 2, 3 and 4), which gives the '3' in the square bracket. Having reached position 5, the longer route back to position 1, is via carbons 6 and 7, hence the '2' in the square bracket. This then leaves the remaining route between positions 1 and 5, via carbon 8, hence the '1' in the bracket.

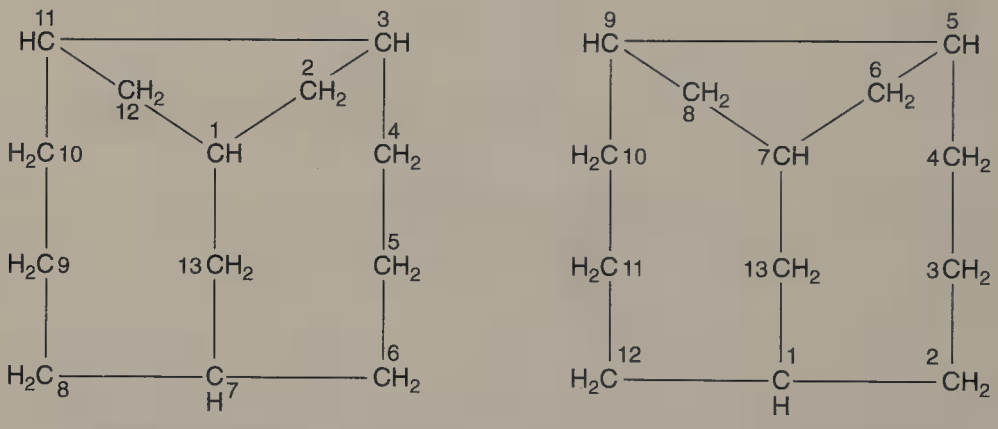


Figure 4.36 Tricyclo[5.5.1.$0^{3,11}$]tridecane **not** Tricyclo[5.5.1.$0^{5,9}$]tridecane.

Figure 4.36 shows a more complicated example.

Here there are three rings, with two apparently equal ways of numbering the structure. However, the last bridge is between carbons 3 and 11, or 5 and 9. The first choice is preferred, because 3 is lower than 5. This continues the theme of preferring the lower (or lowest) locant. Superscripts are used to show the locants of the final bridgeheads. A useful check is that the number of digits in the bracket exceeds the number of rings by one. The '$0^{3,11}$'counts as one digit.

4.2.8 Spiro compounds

In Figure 4.37, one atom is shared between two rings. Numbering starts at the atom next to the spiro junction.

Figure 4.37 Spiro[4.5]decane.

1. Start in the smaller ring – go round it.
2. Go through spiro junction.
3. Go round the next ring.

The expression [4.5] means there are four carbons in the first ring then five in the second ring, excluding the spiro atom. The decane shows there are ten carbons in all. This can be extended, as in Figure 4.38.

Figure 4.38 Dispiro[4.1.5.2]tetradecane.

This is numbered from one end to another, starting at the smaller end, and taking the shorter route round the rings. Much more complicated examples are under consideration by IUPAC.

4.2.9 Hydrocarbon ring assemblies

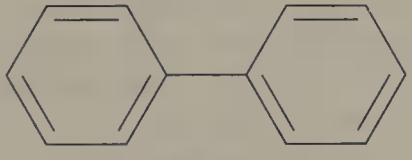


Figure 4.39 Biphenyl.

This is biphenyl, not phenylbenzene. Although one can see the logic of the name, radicals do sound a bit odd. For example (Figure 4.40)

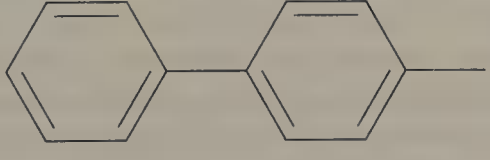


Figure 4.40 Biphenyl-4-yl.

This is biphenyl-4-yl. Numbering starts where the rings are joined. One ring is numbered 1,2,3,4,5,6 and the other 1′,2′,3′,4′,5′,6′. The latter locants are spoken of as 'one prime, two prime . . . etc'. If only one ring is substituted or has an attachment, unprimed locants are preferred. It is possible to collect more rings.

Figure 4.41 1,1′:4′,1″-terphenyl.

Figure 4.41 depicts 1,1′:4′,1″-terphenyl. The third ring is numbered with 'double primes'. This can be extended further, and we will look at more assembly names later.

4.2.10 Heterocyclic systems

So far we have only considered hydrocarbons, but suppose atoms other than carbon appear in rings or chains. There are many trivial names permitted, like pyridine, pyrimidine, etc. Hantzsch–Widman [7] rules cover cyclic compounds containing atoms other than carbon. Various prefixes are used for the replacing atoms, and appropriate suffixes indicate ring size. Full details may be found in the 'Blue Book'. Consider Figure 4.42.

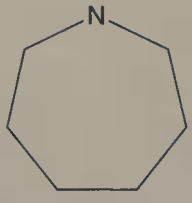
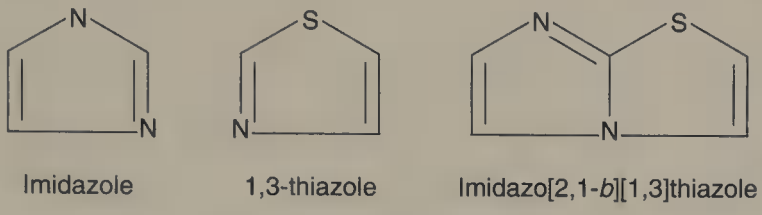


Figure 4.42 Azepane.

This is called azepane, taking 'aza' meaning nitrogen and '-epane' meaning a 7-membered saturated ring. The heteroatom takes position 1, and any radicals the lowest possible number after that. The author has encountered this structure several times under the name 'hexamethyleneimine'. This was quite baffling at first. Hexamethylene should have meant a chain of six CH_2s with a bond at each end and 'imine' means $=NH_2$. All became clear when the components were drawn. Joining up the bonds revealed the structure. It was intended to represent a cyclic compound, although there were no clues to this in the name. Hexamethyleneimine was swiftly

consigned to our 'Abominations' file. One of the great nomenclatorial arts is deducing structures from awful names.

We briefly looked at fused systems earlier. Heterocyclic systems may be fused in a similar manner, but with one important difference. Consider Figure 4.43.



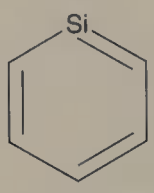
Imidazole 1,3-thiazole Imidazo[2,1-*b*][1,3]thiazole

Figure 4.43

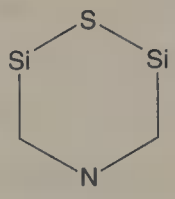
Imidazole is a permitted trivial name. Locants are used to show the position of the heteroatoms. Sulfur is senior to nitrogen, so takes position 1. 'Thia-' means sulfur, and we have already encountered 'aza'. CNAS recommends 1,3-thiazole, although the 'Blue Book' allows thiazole; 1,2-thiazole is called isothiazole in the 'Blue Book.' The trouble with this is that if you see 'thiazole' in a name, you do not know if the person naming it has forgotten the locants. Including the locants removes ambiguity. This deals with the individual names, but what happens when we fuse them? First, you must decide which component is senior. The base component should contain nitrogen, or failing that, a heteroatom as high as possible in the list shown in the Hantzsch–Widman paper [7]. Both components here contain nitrogen, both have only one ring, of the same size, and have the same number of heteroatoms. But 1,3-thiazole has two different heteroatoms, so it is senior. Now comes the important difference from fused hydrocarbon systems. When these two components are fused, a nitrogen is common to both rings. The final structure is named so that the nitrogen appears in both rings. When naming these systems, you have to imagine the components as separate entities. The numbering in the square brackets shows the bond that is fused. Imidazole is numbered with the nitrogens as '1' and '3'. The fused bond takes lowest locants, so 1,2 is preferred to 2,3. As the 1,3-thiazole is senior, the bond from sulfur going anticlockwise is 'a', and the next one 'b'. As this ring is being numbered anticlockwise, the imidazole is numbered clockwise. If you imagine the bonds as paths, with people walking round each ring, they would be walking alongside each other along the fused bond. Numbering in the square brackets again reflects the numbering before fusion. Note the final numbering of the compound. The S is 1; then going clockwise, the numbers are 2, 3, 4 (the N), 5, 6, 7 (the N), 7a. The fused heteroatom is numbered differently from the fused carbon. Fusion nomenclature can become very complicated.

4.2.11 Replacement nomenclature

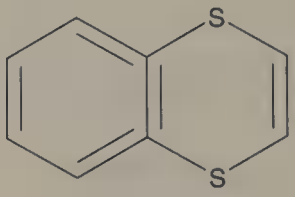
It is sometimes useful to name compounds with heteroatoms by naming them as the appropriate hydrocarbon, then replacing carbons. See examples in Figure 4.44. Locant 1 goes to the atom highest in the table. The hydrogens are omitted.



silabenzene



1-thia-4-aza-2,6-disilacyclohexane



1,4-dithianaphthalene

Figure 4.44

This system can be used also with acyclic structures, but we do not recommend doing this unless there are at least four heteroatoms. For example: CH_3-CH_2-NH-CH_2-CH_3 is diethylamine, not 3-azapentane.

4.2.12 Seniority of chains

There is frequently a choice of numbering chains. Consider Figure 4.45.

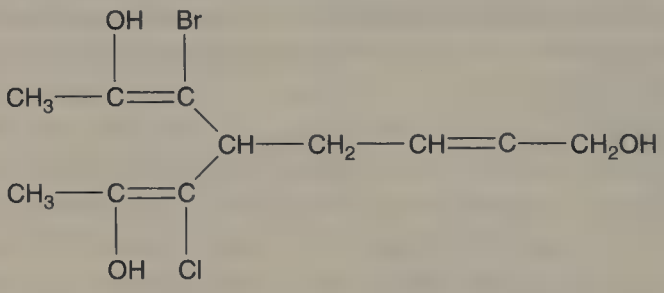
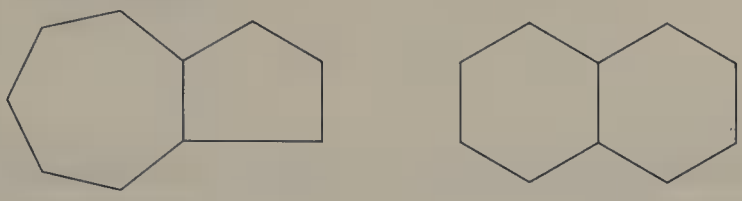


Figure 4.45 6-Bromo-5-(1-chloro-2-hydroxyprop-1-enyl)octa-2,6-diene-1,7-diol.

The senior chain will contain the maximum number of the senior suffix, in this case the -OH group. There are three different chains with two OH groups, so they are equal. Each chain has two double bonds, still equal. The 'U'-shaped one on the left has only seven carbons, whereas the others have eight, so we can eliminate the former. Both of the remaining chains have OH groups at positions 1,7 – still equal. The bromine and chlorine both appear at position 6, so that is equal. However, bromine is earlier in alphabetical order so that wins.

4.2.13 Seniority of rings

The first rule is that all heterocycles are senior to all carbocycles. Heterocyclic seniority is determined as before. After that, take the largest number of rings, then the largest individual ring. Hence, in Figure 4.46, the left-hand system wins.



7 + 5 6 + 6

Figure 4.46

After that, take the largest number of atoms common to the rings, so naphthalene beats spiro[5.5]undecane, etc. Choice after that depends on a variety of lowest locants for different features and also degree of unsaturation. The 'Blue Book' has full details.

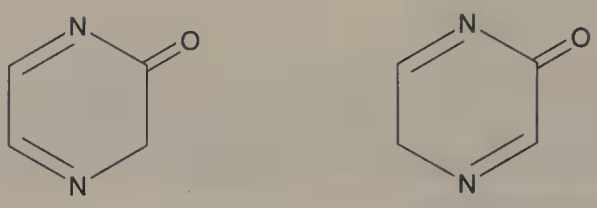
4.2.14 Indicated hydrogen

This particular feature causes a lot of confusion. The basic premise is that hydrogens go in pairs, so 1,2-dihydronaphthalene shows that a double bond is missing from the structure. Hydro prefixes always go in pairs. Dihydro, tetrahydro, hexahydro, etc. are all fine, showing one, two and three missing double bonds, but trihydro, pentahydro, etc. are wrong. This is easy enough, but suppose you have pyran (Figure 4.47).

2*H*-pyran 4*H*-pyran

Figure 4.47

There is no way you can fit in three double bonds, because the oxygen is divalent, so there is one carbon that is saturated. This position is shown by the italic *H* with locant to show its position. The oxygen is number 1. The system really comes into play with ketones.



pyrazin-2(3*H*)-one pyrazin-2(5*H*)-one

Figure 4.48

The hetero atoms are numbered 1 and 4, and the ketone group then takes the lowest possible number. There is obviously no double bond at position 2, so the indicated hydrogen shows where the double bond is not present.

4.2.15 Other suffixes

CH_3-CH_2-CH_2-CH_2-CH_2-CN is hexanenitrile, as it is considered to derive from hexanoic acid, but where the acid has an established trivial name, the '-ic acid' is replaced by '-onitrile'. Hence: CH_3-CH_2-CN is propiononitrile, **not** propionitrile, which is a common mistake. Aldehydes are named in a similar fashion. CH_3-CH_2-CH_2-CH_2-CH_2-CHO is hexanal. ('-al' replaces '-oic acid'). CH_3-CH_2-CHO is propionaldehyde ('-aldehyde' replaces trivially named '-ic acid').

4.2.16 Summary

1. Find Principal Group.
2. Find Parent Structure.
3. Use lowest numbers possible.
4. Use alphabetical order.
5. These basic principles cover even the most complex structures.

We have already looked at some simple names and at a wide variety of parent structures. In the next chapter, we put it all together.

References

1. *Organic Chemistry: Its Language and its State of the Art* (1993) ed. M. Volkan Kisakürek, VHCA.
2. *Chemical Nomenclature, Symbols and Terminology* (1979), Association for Science Education.
3. IUPAC (1979) *Nomenclature of Organic Chemistry*, Pergamon (Blue Book).
4. IUPAC (1993) *A Guide to IUPAC Nomenclature of Organic Compounds*, Recommendations 1993, Blackwell Scientific (The Guide).
5. IUBMB (1992) *Biochemical Nomenclature and Related Documents*, A Compendium, 2nd edn, Portland Press (White Book).
6. Godly, E. W. (1995) *Naming Organic Compounds* – A Systematic Instruction Manual, 2nd edn, Ellis Horwood.
7. (1983) *Pure and Applied Chemistry* **55 (2)**, 409–416.

5 IUPAC nomenclature part 2, organic, inorganic and others

K. J. THURLOW

5.1 Specific examples

Now we have looked at the basic rules of organic nomenclature, it is time to put it all together.

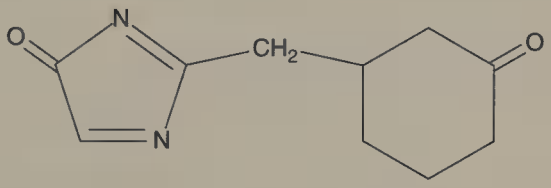


Figure 5.1 2-(3-Oxocyclohexylmethyl)imidazol-4-one.

In Figure 5.1 the principal group is =O, but there is one on each ring. The heterocyclic ring is senior, so the other =O group is named as a prefix. Imidazole is numbered so that the nitrogens are 1 and 3. If you start with the nitrogen at the top, the other ring is attached at 2 and the =O at 6. Starting with the other nitrogen, the attachments are at 2 and 4. 1,2,3,4 beats 1,2,3,6 so the name becomes 2-(3-oxocyclohexylmethyl)imidazol-4-one **not** 2-(3-oxocyclohexylmethyl)imidazol-6-one. We do not need to indicate hydrogen, because it is obvious where the double bonds are.

The next example is a fairly large spiro compound (Figure 5.2).

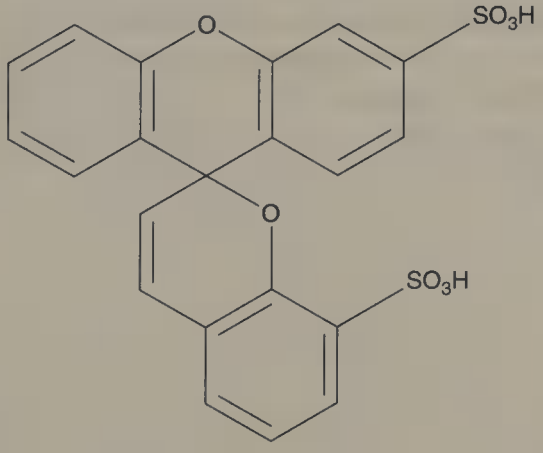


Figure 5.2 Spiro[chromene-2,9′-xanthene]-3′,8-disulfonic acid.

Sulfonic acid (-SO_3H) is the principal group and you would expect the component with three rings (xanthene) to be senior to the component with two rings (chromene). However, the spiro junction allows you to collect the two ring systems as one parent structure. The spiro junction takes the lowest possible number and the second component has 'primed' numbers. Hence, Spiro[chromene-2,9′-xanthene]-3′,8-disulfonic acid. Chromene's numbering starts with the 'O' as 1 and continues anticlockwise round the rings, so spiroatom is 2, then 3, 4, 4a, 5, 6, 7, 8 (the SO_3H), 8a. Xanthene's numbering is an exception to the normal rule. Starting at the bottom of the right hand ring is 1′ and continues anticlockwise 2′, 3′ (the SO_3H), 4′, 4a′, 10 (the O), 10a′, 5′, 6′, 7′, 8′, 8a′, 9′ (the spiroatom), 9a′. It should be noted that terms like clockwise and anticlockwise are not absolute, but refer to the way the structures are depicted.

5.1.1 Assembly names

First one you cannot assemble (Figure 5.3):

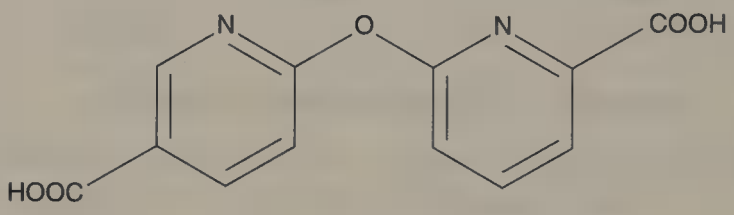


Figure 5.3 6-(5-Carboxy-3-pyridyloxy)pyridine-2-carboxylic acid.

Each pyridine ring has a carboxylic acid, so it would be nice to collect them. Unfortunately, one ring is pyridine-2-carboxylic acid and the other one is pyridine-3-carboxylic acid, which are not identical. If they were both benzoic acid, you could collect them. The lower locant for the principal group is senior, so this structure is named: 6-(5-carboxy-3-pyridyloxy)pyridine-2-carboxylic acid **not** 6-(6-carboxy-2-pyridyloxy)pyridine-3-carboxylic acid. 3-Pyridyl is the contracted form of pyridin-3-yl, so the locant precedes the contraction. Pyrid-3-yl is incorrect. A similar example of a contracted form is where 2-naphthyl is used instead of naphthalen-2-yl.

Now an example you can assemble (Figure 5.4).

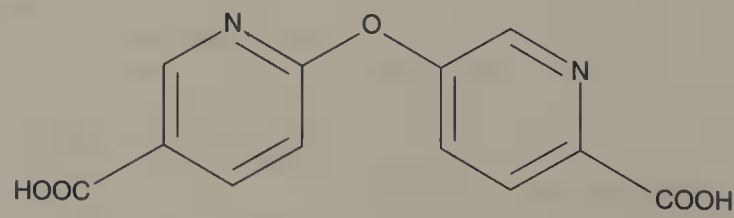


Figure 5.4 5,6′-Oxydinicotinic acid.

Figure 5.4 shows a case where the parent structure and principal group are identical. Pyridine-3-carboxylic acid has the trivial name nicotinic acid. This structure is called 5,6′-oxydinicotinic acid. The unprimed number is lower. If the two rings were linked directly, without any intervening atoms (Figure 5.5) then the name would be 5,6′-binicotinic acid. Note that in this case 'bi' is used instead of 'di'.

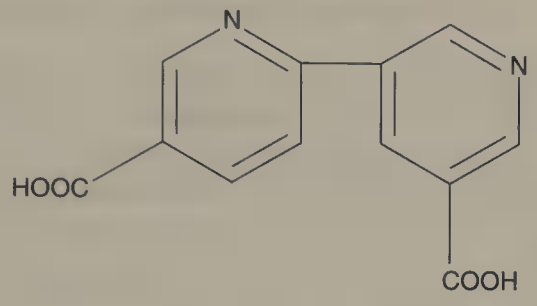


Figure 5.5 5,6′-Binicotinic acid.

A slightly more difficult example (Figure 5.6):

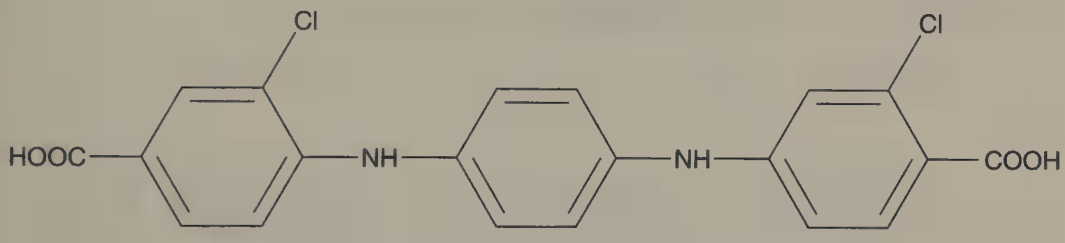


Figure 5.6 2,3′-Dichloro-4,4′-*p*-phenylenediiminodibenzoic acid.

Here we see another assembly. There are two benzoic acids, one at each end, so we may collect them, although the chloro-substitution is at different positions. As the 'chloro's are prefixes, they do not interfere with the basic symmetry. One benzoic acid is numbered normally, the other has primed locants, the choice being made depending on lower locants at first point of difference. An essential point is that the centre of the molecule is also symmetrical.

There are some interesting points here. '-NH-' is called 'imino' in this case, because the structure is being named in both directions at once. If you started at one end and worked towards the other, then it would be 'amino'. The other point is that the 'chloro's appear before the benzoic acids are collected. This would apply even if the 'chloro's were both at position '2' (say). You would say '2,2′-dichloro . . .', not '. . . bis(2-chlorobenzoic acid)'.

The lack of symmetry in Figure 5.7 prevents the assembly name:

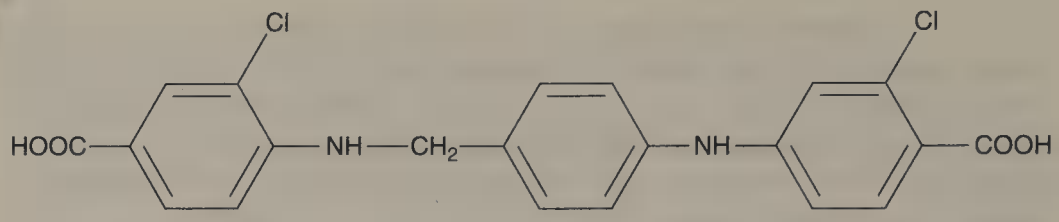


Figure 5.7 4-[4-(4-Carboxy-3-chloroanilino)benzylamino]-3-chlorobenzoic acid.

This looks similar – again there are two benzoic acid groups, but the central part is not symmetrical. We have to select one benzoic acid as parent and the other one as prefix. Starting at the left-hand end, it is 4-[4-(4-carboxy-2-chloroanilinomethyl)anilino]-2-chlorobenzoic acid. Starting at the right-hand end, it is 4-[4-(4-carboxy-3-chloroanilino)benzylamino]-3-chlorobenzoic acid. Note how the numbering has changed. The chlorines are at '2' or '3' depending which way you travel. Note that C_6H_5-NH- is anilino and CH_3-C_6H_4-NH- is toluidino. The 'Guide' does not allow toluidino when the methyl group is substituted. The second name is preferred as it is earlier in the alphabet.

Alphabetical order can cause difficulties (Figure 5.8):

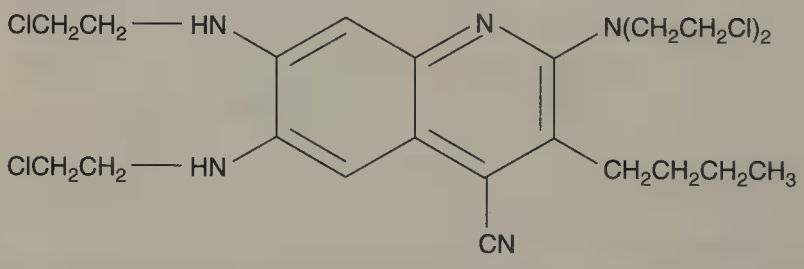


Figure 5.8 2-[Bis(2-chloroethyl)amino]-3-butyl-6,7-bis(2-chloroethylamino)quinoline-4-carbonitrile.

The principal group is carbonitrile and the parent structure is quinoline, another permitted trivial name. So you have quinoline-4-carbonitrile and all you have to do is insert the various substituents in alphabetical order. We list the substituents:

(i) 2-[bis(2-chloroethyl)amino]
(ii) 3-butyl
(iii) 6,7-bis(2-chloroethylamino)

Now we spot a problem with the alphabetical order. What is the difference between (i) and (iii)? Actually there is a subtle difference in the meaning of 'bis', rather like the difference between simple and complex radicals discussed in Figures 4.20–4.22. Substituent (i) refers to a complex radical in one position, so the name is considered to start with the 'b' of 'bis'. Everything is on position '2'. Substituent (iii) refers to complex radicals

also, but this time to a radical that appears in two different places. This radical is considered to begin with the 'c' of 'chloro'. The radical is on positions '6' and '7'.

This does look confusing. When CNAS staff give talks on nomenclature, there is invariably some head scratching when this example is considered.

5.1.2 Special systems

The 'White Book' [1] covers a wide variety of naming systems for biochemical structures, including amino acids, sugars, steroids, vitamins, etc. There are many acceptable trivial names, which helps considerably. More detail may be found in the chapter on natural products in this book.

5.1.3 Phane nomenclature

'Phane' nomenclature has been under development for some time now. The premise is that it is awkward to name even relatively simple structures like Figure 5.9:

$CH_3–C_6H_4–CH_2–C_6H_4–CH_2–C_6H_4–CH_3$

Figure 5.9

You end up with a rather long and messy name. Phane nomenclature is analogous to replacement nomenclature, and this structure would be called something like: 2,4,6-tribenzenaheptaphane. This pretends that the benzene rings are carbon atoms, and the whole 'chain' is seven atoms long. '2,4,6-tribenzena' shows that the 'atoms' at positions 2, 4 and 6 are actually benzene. The 'hepta' reflects that there are seven 'atoms', and the 'phane' that some are not real atoms, but something more complicated. The difficulty arises when you need to insert locants for the connection between the benzenes and the methyl groups. If it is all '*para*' it is easy enough, but suppose it isn't? IUPAC are discussing ways of overcoming this problem.

5.1.4 Fullerenes

Readers will be aware of the excitement caused by the discovery of fullerenes. 'Buckminsterfullerene' (C_{60}) is spherical. So, how do you number it? Where is position 2? CAS and IUPAC have been discussing this and a preliminary survey has now been published [2]. The systematic name for 'buckminsterfullerene' is [60-I_h]fullerene, the 'I_h' indicating the point group symmetry, which may be discarded if there is no ambiguity, leaving [60]fullerene. There is still plenty of work to be done, not only

on fullerenes, but also on the bowl-shaped compounds which are now coming into prominence. Another article shows how to make paper models of fullerenes [3].

5.1.5 Ions

Figure 5.10 shows various ions:

$[C_3H_9]^+$	$CH_3–CH_2–C^+H_2$	$CH_3–C^+H–CH_3$	$CH_3–\bar{C}H–CH_3$
propanium	propylium	propan-2-ylium	propan-2-ide

Figure 5.10

Propanium is a general name for this cation, where you do not know the location of the charge. Propylium is the truncated version of propan-1-ylium. This and propan-2-ylium are the names where the position of the charge is established. Similarly, the last example shows the '-ide' ending used for anions.

Compounds with two ionic centres are named as shown in Figure 5.11.

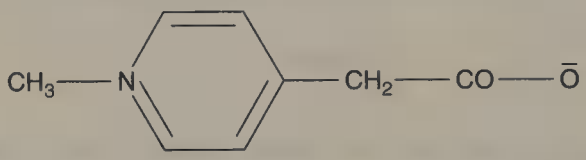


Figure 5.11 (1-Methylpyridin-1-ium-4-yl) acetate.

'-1-ium' tells you the positive charge is on position 1 of the pyridine, and the '-4-yl' shows where the normal radical is.

5.2 Stereochemistry

> . . . and should not I have pity on Nineveh, that great city; wherein are more than six score thousand persons that cannot discern between their right hand and their left hand;
>
> *Jonah 4:11*

This is very important and can be rather complicated. *Cis–trans* isomerism is relatively simple. Take the case where there is one double bond, which we take as the reference plane (Figure 5.12).

According to the Sequence Rule, devised by Cahn, Ingold and Prelog [4–7], atoms are put in order of seniority, so here I > Br > Cl, as they have decreasing atomic number. So, in Figure 5.12a, Br >Cl on atom 1, and I > Br on atom 2. Both senior atoms are below the double bond,

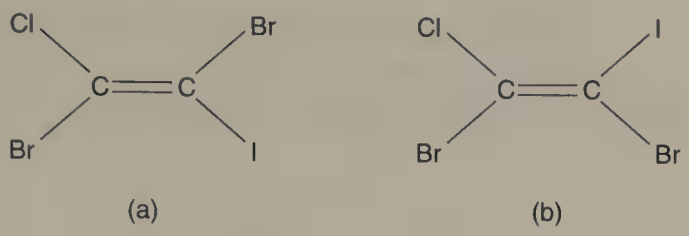


Figure 5.12 (a) (*Z*)-1,2-dibromo-1-chloro-2-iodoethene. (b) (*E*)-1,2-dibromo-1-chloro-2-iodoethene.

so this is (*Z*)-1,2-dibromo-1-chloro-2-iodoethene. (*Z*) for the German *zusammen* (same); this is '*cis*'.

In Figure 5.12b, the same seniority applies, but this time the senior atoms are opposite sides of the double bond, so it is (*E*)-1,2-dibromo-1-chloro-2-iodoethene. (*E*) for the German *entgegen* (opposite); this is '*trans*'.

5.2.1 Chirality

The Sequence Rule is used for specifying the absolute molecular chirality (handedness) of a compound, using the letters (*R*) for the Latin *rectus* (right) and (*S*) for the Latin *sinister* (left). The simplest case of a chiral centre is to have a carbon atom, with four different ligands, which are attached tetrahedrally. The Cahn–Ingold–Prelog papers deal with this, as does the 'Blue Book', but the simplest general description was written by Isaac Asimov [8]. IUPAC are considering clarification of the Sequence Rule, as there is disagreement over some aspects of its use.

It will be seen that there is no asymmetry if two of the ligands are the same. The differing ligands are normally called a, b, c, and d, and are cited in decreasing atomic number. Hence 'a' is most senior, and 'd' is least senior. Consider the simple case in Figure 5.13:

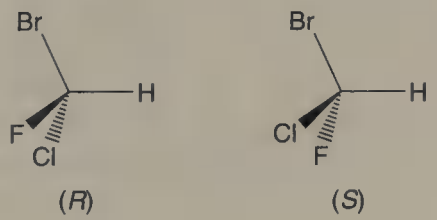


Figure 5.13 (a) *R* and (b) *S* enantiomers.

Dotted lines imply that the bond is beneath the plane of the paper, and solid lines (or wedges) are above the plane of the paper. Once the seniority has been established, you orientate the molecule and view it from the side opposite 'd'. You are left with three bonds, apparently planar at angles of 120° to each other. Trace a path from 'a' to 'b' to 'c'. If it goes clockwise,

the stereochemistry is (*R*), if anticlockwise, (*S*). So in the above examples, Br > Cl > F is clockwise in Figure 5.13a (*R*), but anticlockwise in Figure 5.13b (*S*). However, it is rarely that simple. Consider Figure 5.14.

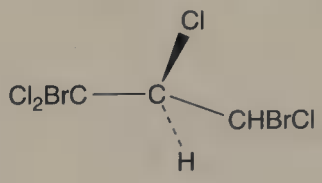


Figure 5.14

The central asymmetric carbon has four attachments. The Cl is clearly senior hence 'a', and the H is clearly least senior, hence 'd', but the other two atoms are both C and there is no decision. We now have to look at the attachments to these carbons. The one on the left has (Br, Cl, Cl), the one on the right (Br, Cl, H). These are cited in decreasing seniority. Now compare the most senior atom on each side. Br vs. Br is a tie, so we look at the next atom. Cl vs. Cl is also a tie, so we look at the last atom. Cl > H. At last, a decision. As Cl beats H, then the left-hand carbon is 'b' and the other one 'c'. If you look down the page towards the H, you will see this is (*S*). This might be a bit difficult to imagine. CNAS staff have found that the easiest way to visualize this is to have a small model of the molecule. We have a central 'carbon' with different coloured plastic bonds (to represent a, b, c and d) radiating from it tetrahedrally. If this is picked up by the 'd' bond, it may be twisted until you are looking towards 'd', and the clockwise or anticlockwise rotation is easily deduced.

Large volumes may be written on stereochemistry, but we will consider just one more example, albeit a very complicated one. The material in Figure 5.15 was produced by a former CNAS expert.

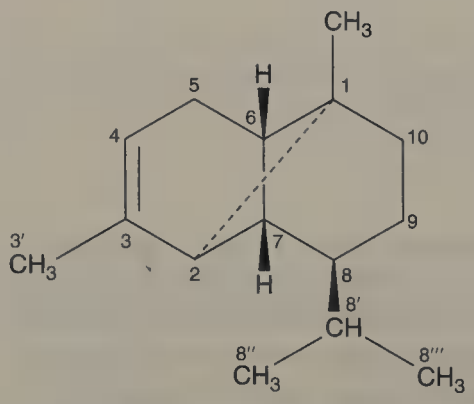


Figure 5.15 (1*R*,2*S*,6*S*,7*S*,8*S*)-8-isopropyl-1,3-dimethyltricyclo[4.4.0.$0^{2,7}$]dec-3-ene(-)-α-copaene.

Select the first chiral centre, which is C_1, and arrange the connected atoms in Set 1:

$C_1 : C_{1'}$
$\quad : C_2$
$\quad : C_6$
$\quad : C_{10}$

All the connected atoms are the same, so no decision is possible yet. We must look at all of the atoms attached to these carbons, working away from the chiral centre. These are tabulated as follows in Set 2:

$C_{1'}$: H H H
C_2 : C_3 C_7 H
C_6 : C_5 C_7 H
C_{10} : C_9 H H

Compare the most senior atoms attached to each atom from Set 1. Now we see that C_1' is least senior, hence 'd', as $C_3 = C_5 = C_9 > H$. There is no decision for the others. Now look at the second most senior atoms. C_{10} is next least senior atom, hence 'c', as $C_7 = C_7 > H$. There is still no decision for the last two unordered atoms, so we look at the final (least senior) atoms. They are both H, so we need to go to Set 3. $C_{1'}$ and C_{10} have already been dealt with so we can forget them.

C_2) C_3 :$C_{3'}$ C_4 (C_4) – 3 C
$\quad C_7$:C_6 C_8 H – 2 C 1H

C_6) C_5 :C_4 H H – 1C 2H
$\quad C_7$:C_2 C_8 H – 2C 1H

Note that the two C_7 sets are different. This is because you must move away from the atom you have just visited. The (C_4) indicates a 'phantom' atom, as a double bond is involved. The subsequent attachments to C_4 count only once; the attachments to the phantom C_4 count as nothing.

First, we must consider the more senior carbon from each set. So for C_2, the first set (C_3) is senior, as there are 3 carbons. For C_6, the second set (C_7) is senior. Again you compare the individual atoms from C_3 and C_7, but the difference is obvious as the former has CCC, which beats CCH.

$C_{3'} = C_2$, $C_4 = C_8$, (C_4) > H

So C_2 is 'a', C_6 is 'b', C_{10} is 'c' and $C_{1'}$ is 'd'. The stereodescriptor for carbon atom C_1 is 1*R*.

If readers wish to test their stereochemistry skills, they may care to try all the other chiral centres in this example. The full name is: (1*R*,2*S*,6*S*,7*S*,8*S*)-8-isopropyl-1,3-dimethyltricyclo[4.4.0.$0^{2,7}$]dec-3-ene. It is unusual to encounter examples as difficult as this!

5.3 Deducing the structure from the name

2-bromo-4-{3-[3-chloro-5-(3-hydroxy-2-iodopropyl)anilino]-5-fluorophenoxy} benzoic acid

What does this tell us?

(A) On the benzoic acid, we have	bromo	at position 2
	{. . . . .}	at position 4
(B) On the phenoxy group, we have	[. . . . .]	at position 3
	fluoro	at position 5
(C) On the anilino group, we have	chloro	at position 3
	(. . . . .)	at position 5
(D) On the propyl group, we have	hydroxy	at position 3
	iodo	at position 2

Hence, the structure is as Figure 5.16:

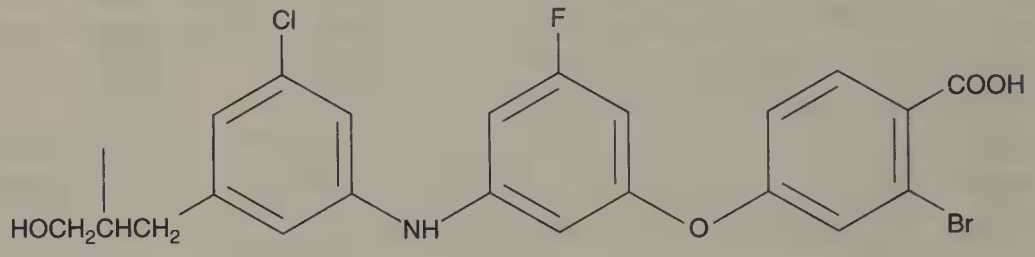


Figure 5.16 2-Bromo-4-{3-[3-chloro-5-(3-hydroxy-2-iodopropyl)anilino]-5-fluorophenoxy) benzoic acid.

This has been a fairly selective wander through organic nomenclature; it is impossible to cover everything, but it is hoped that at least some of it will be useful.

5.4 Inorganic names

It is not intended to cover this subject in as much depth as the organic names, but there are some important points. The elements are the basic building blocks, and are dealt with elsewhere in this book. See 'Red Book' [9] for detailed coverage. It may be useful to comment briefly on sulfur. It is still common to encounter people who are antagonistic to this spelling, because it is 'American'. In fact, it is spelled with an 'f' nearly everywhere, and there is no etymological justification for 'ph'. The 'ph' does not denote a Greek usage, unlike phosphorus. It seems logical to go with the majority viewpoint, and readers can take comfort from the fact that there is a town called 'Sulphur Springs' in Texas.

5.4.1 Simple inorganic names

NaCl is sodium chloride. Note the word endings. The cation ends -ium, just as in the organic examples cited in Figure 5.10. Of course, this is more a case of the organic nomenclature imitating inorganic names, but at least it is consistent. Similarly, the anion ends -ide. This does not apply to all anions, as the -ate ending is also common. All electropositive parts are placed before all electronegative parts.

Multiplying prefixes are used as necessary. So Fe_3O_4 is triiron tetraoxide – note that the 'a' is not elided, and Fe_2O_3 is diiron trioxide. The prefixes may be omitted if there is no ambiguity. $CaCl_2$ is called calcium chloride.

5.4.2 Alphabetical order

Formulas and names are expressed in alphabetical order, even if these differ. $KMgCl_3$ is magnesium potassium chloride. Obviously, there will be differences in different languages also. Note the spaces. Multiplying prefixes are treated in a similar manner to organic names. 'Triiron' comes before 'sodium'. Hydrogen is an exception, being placed last of the electropositive parts.

5.4.3 Oxidation and charge numbers

> The naming of Cats is a difficult matter
> It isn't one of your holiday games
> At first you may think I'm as mad as a hatter
> When I tell you a cat must have THREE DIFFERENT NAMES
>
> *The Naming of Cats – T. S. Eliot*

Oxidation (Stock) numbers show the oxidation state, given as a Roman numeral in brackets immediately after the appropriate element. $FeCl_3$ would be iron(III) chloride. The number is assumed positive unless otherwise specified.

Charge (Ewens–Bassett) numbers show ionic charge. $FeCl_3$ would be Iron(3+) chloride. Positive or negative signs are used. It will be seen that this gives quite a choice of names. Consider our old friend $CuSO_4$. This may be called:

Copper sulfate
Copper(II) sulfate
Copper(2+) sulfate
Copper tetraoxosulfate

The first of these is most common and anyone reading it will understand what is intended. If there is a doubt – it could be copper(I) sulfate – it is hoped that the writer would draw attention to it. It has to be said that

it is not normally a good idea to assume writers know what they are doing. The second and third names give more structural information, due to the oxidation and charge numbers. The fourth name gives even more information. It may seem confusing that sulfate usually means 'SO_4', but here means 'S'. However, the 'tetraoxo' is a clue that something funny is going on. Cats have only three different names, chemicals can have many more. The above systems can also be used for acid names. Looking at the analogous H_2SO_4, normally known as sulfuric acid, this may be called dihydrogen tetraoxosulfate, or tetraoxosulfuric acid. This can be quite useful, as it removes all doubt about the structure. Sulfurous acid would be trioxosulfuric acid. The use of systematic names like these is a handy learning tool. This form of name is ideal for complicated structures, and we will see more of it later.

5.4.4 Hydrides

Inorganic hydrides are named much as organic ones are. Just as methane is the (trivial) name for carbon surrounded by the maximum number of hydrogens, then BH_3 is borane, AsH_3 is arsane and SnH_4 is stannane. This can be extended. NH_3 is ammonia but may be called azane. H_2O is water but may be called oxidane. H_2S is hydrogen sulfide but may be called sulfane. The traditional names would be used most of the time, and particularly for the unsubstituted compounds. One former CNAS colleague labelled the office water container oxidane, which puzzled some visitors to the area. One even complained about him keeping chemicals in the office.

You would expect 'oxane' to be the name used for H_2O, but this is already in use in Hantzch–Widman nomenclature, for the saturated six-membered heterocyclic ring with one oxygen.

Extensions to these names are analogous with organic nomenclature. $SiH_3SiH_2SiH_3$ is trisilane, and H_2O_2 may be called dioxidane. It has to be said that the phrase 'dioxidane blonde' has a certain ring to it. Unsaturation is dealt with in the usual way. $NH_2{=}NH_2$ is diazene. Cyclic compounds are similar. The structure of cyclopentazane is easy to deduce – a saturated ring of five nitrogens.

5.4.5 Coordination compounds

So far, things have been pretty straightforward, but coordination names can be a real nightmare. Let us start with a couple of simple examples. Special names are used for various ligands: CO is carbonyl, NH_3 is ammine, H_2O is aqua. Hence, $[Fe(CO)_5]$ is pentacarbonyliron.

Consider $K_4[Fe(CN)_6]$. This is traditionally called potassium ferrocyanide. IUPAC allows three names for this:

potassium hexacyanoferrate(II)
potassium hexacyanoferrate(4-)
tetrapotassium hexacyanoferrate

The 'ate' at the end indicates the anionic part of the structure. In the first name, the (II) is Stock notation, where a Roman numeral indicates the oxidation state of the central atom. The four potassiums contribute +4 and the six cyano groups -6, which requires +2 to balance it out. The second name shows an alternate procedure, whereby the Arabic numeral, in this case '4-', shows the overall charge of the coordination entity. Both these styles require some thought to either create or decipher the name. Perhaps the easiest system is shown in the last name, which clearly tells you there are four potassiums, six cyano groups, and an iron.

This case is also relatively simple, but sometimes we need to specify which part of a ligand is attached to a central atom. Consider the structure in Figure 5.17:

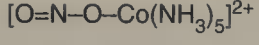


Figure 5.17

The name for this has to take into account the cobalt–oxygen bond. If you do not specify this in the name, it could equally apply to a complex with a cobalt–nitrogen bond. The kappa convention is used to avoid ambiguity. The above structure is named: pentaamminenitrito-κ*O*-cobalt(III) ion. 'Penta' means five and 'ammine' is the preferred name for ammonia as a ligand name, as mentioned earlier, although you can use the systematic 'azane'. Anionic ligands end in 'o', so 'nitrite' becomes 'nitrito'. Similarly, 'ethanol' would become 'ethanolato' and 'acetate' would become 'acetato'. The' κ*O*' shows that the oxygen in the nitrite is the ligating atom, i.e. it attaches to the Co. This is a fairly simple case.

Consider Figure 5.18, which shows an extension of the kappa convention.

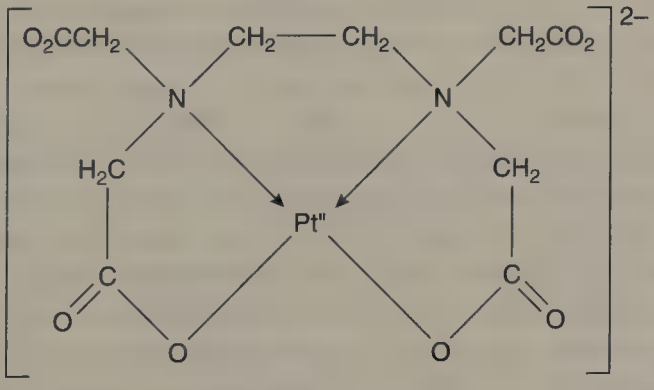


Figure 5.18

This would be named: [(ethane-1,2-diylnitrilo-κ^2N,N')tetraacetato-κ^2O,O'']platinate(2-). The superscripts on the κ show the number of ligating groups involved. Here there are two nitrogens and two oxygens as the points of attachment. The *N* and *N′* indicate that we are talking about two different nitrogens, not two bonds to the same one. The use of *O* and *O″* is similar, in this case showing that the oxygens come from different acetate groups. 'κ^2' shows that only two of the acetates are attached to the platinum. This is a tricky example, but it could be worse.

Another problem arises when the ligand acts as a bridge, instead of being attached only at one end. Consider Figure 5.19.

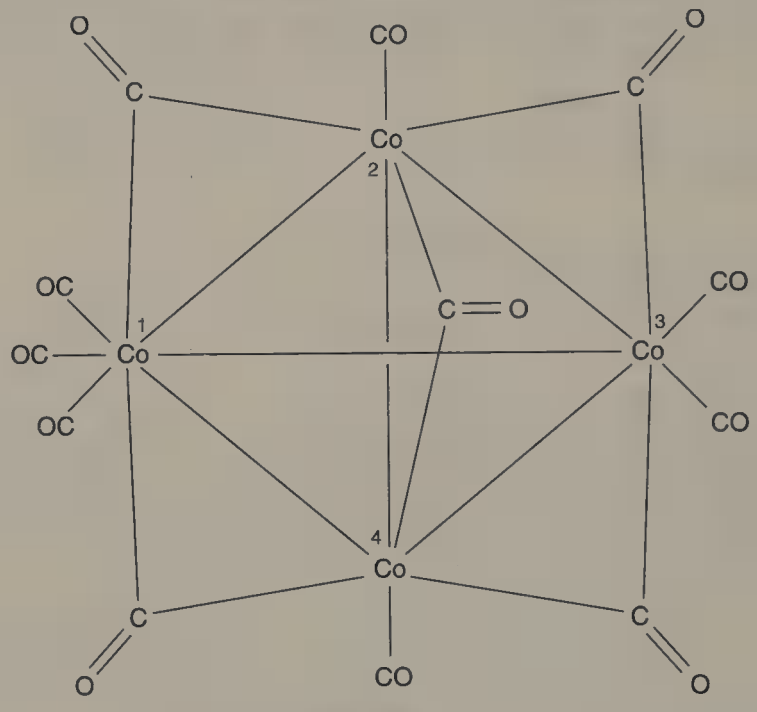


Figure 5.19

This is named: penta-μ-carbonyl-1:2κ^2C;1:4κ^2C;2:3κ^2C;2:4κ^2C;3:4κ^2C-heptacarbonyl-1κ^3C,2κ*C*,3κ^2C,4κ*C*-*tetrahedro*-tetracobalt(4 *Co-Co*)

Even experienced nomenclators greet such names without great enthusiasm. The 'penta-μ' shows that five carbonyls are acting as bridges, the first between cobalts 1 and 2, hence the '1:2', then the next between cobalts 1 and 4, etc. Kappa is used as before, so the carbon is the ligating atom. Then the seven carbonyls are distributed so there are three at cobalt 1, hence the '1κ^3C', one at cobalt 2, two at cobalt 3, and one at cobalt 4. The *tetrahedro* describes the structural relationship of the cobalts, and the (4 *Co-Co*) shows that there are four metal–metal bonds.

Sometimes you do not know how the metal attaches to the ligand and it may even attach to a number of atoms in the ligand. Figure 5.20 is an example of this.

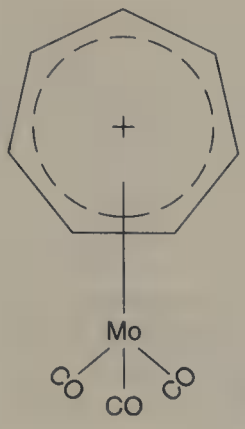


Figure 5.20

IUPAC goes Greek again and uses what is called 'hapto' nomenclature, designated by the Greek letter etha (η). So we have: tricarbonyl(η^7-cycloheptatrienylium)molybdenum(1+). The η^7 means that the central atom attaches to seven ligating atoms in the cycloheptatriene, and the 'ium' shows the positive charge. If only four (say) of the atoms were ligating, and the position of these is known, locants can be used instead of superscripts. Consider Figure 5.21.

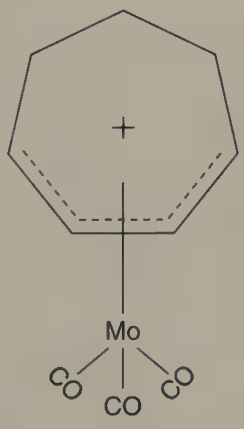


Figure 5.21

You could call this: tricarbonyl(1,2,3,4-η-cycloheptatrienylium)molybdenum(1+).

It may appear that these systems are rather confusing, but it is difficult to see how the names could be expressed any more easily, and at least the Greek letters are an immediate warning that you are in for a bout of head-scratching.

I did publish information on these last few examples elsewhere and received one complaint that it was too complicated. Well, nomenclature

is complicated. Luckily, examples like these are not too common and I hope the above information helps readers recognize roughly what the names mean. It might be best to seek help if you actually want to name such a structure.

5.4.6 Porphyrins, etc.

Researchers are taking an increasing interest in metal coordination complexes with porphyrin analogues. There are recommended IUPAC trivial names for many of the analogues. For example, Figure 5.22, is known as phthalocyanine, which is considerably shorter than its systematic name.

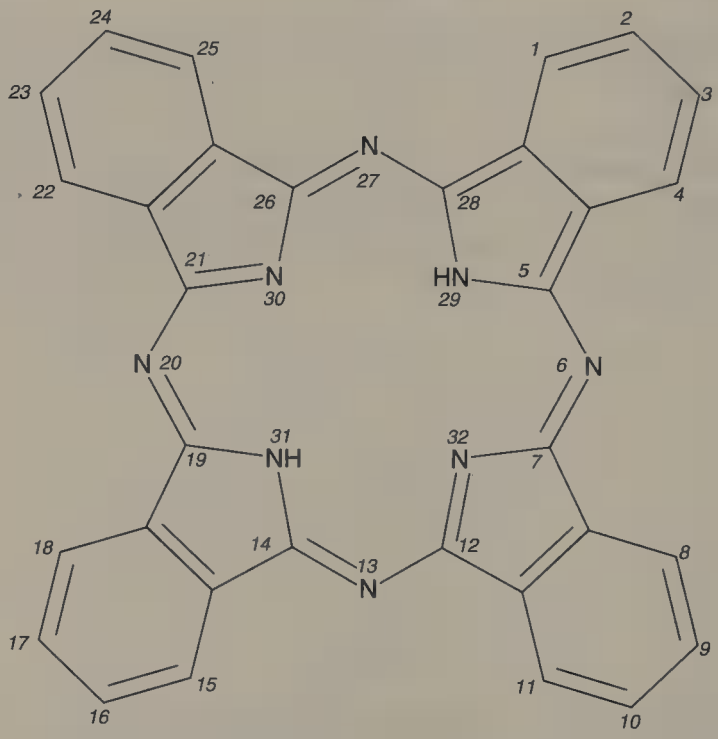


Figure 5.22 Phthalocyanine.

Note the unusual numbering of the central nitrogen atoms, which you would expect to go clockwise. It does not matter normally, as you rarely find substitution on only one of those nitrogens. As for metal complexes, if you insert a magnesium bonded to the four central nitrogens, it is simply called: phthalocyaninatomagnesium. There are many other specialized naming systems for complicated structures, some of which will be mentioned elsewhere in this book.

5.4.7 General problems

I alluded in the previous chapter to the 'Abominations File', which contains examples of bad nomenclature. Some examples are obviously bad. A European Communities document refers to 'methylene', which should be $-CH_2-$, as: "actual methylene, that is to say raw methyl alcohol . . .". This came as a bit of a shock. Other errors are more subtle. Ted Godly has mentioned earlier the difficulties with spoken names. An LGC colleague was puzzled recently to be asked for the 'oxylene' content of a sample. It took some time to realize they meant '*o*-xylene'. Some are subtler again. Someone wanted some 'phenylethylamine'. But what did he want?

(i) Ph-NH-CH_2CH_3?
(ii) Ph-CH_2-CH_2-NH_2?
(iii) $CH_3CH(Ph)NH_2$?

Names like isopropanol are not recommended because isopropane does not exist; isopropyl alcohol is acceptable.

However, the favourite entry remains 'Homberg's Phosphorus' which, as every schoolchild knows, is calcium chloride.

5.4.8 Summary

I have tried to cover the basic points of nomenclature in these two chapters, along with more complicated examples. I hope this may have cleared up some problem areas, but equally I recognize it may have alerted readers to problems they did not know existed. It would be impossible to cover every aspect of IUPAC nomenclature in a couple of chapters. Other areas are covered in depth elsewhere in the book. Even so, there are inevitably omissions.

Nomenclature is a controversial subject and until IUPAC produces its 'Preferred Name' book, there will always be arguments about which particular form of name is right. Even then, there will doubtless be areas that are open to interpretation, and the process of expansion and revision is never ending. CNAS has advised on nomenclature for many years, and has provided a huge number of IUPAC names. We try to adopt a consistent approach. If a customer has a long list of names derived from the Blue Book, we would normally prefer to continue giving names in the same style to that customer. These names may not always agree with the more recent 'Guide', but if readers are used to seeing a substituted 'toluidino', the insertion of a 'methylanilino' may cause confusion. If you change one, you should change them all. We have also made logical extensions to the rules. If 'acetanilide' is permitted for 'phenylacetamide', why not 'acetotoluidide' and 'acetoxylidide', with appropriate locants? It makes life easier and life in nomenclature is rarely easy. Consider the next structure(Figure 5.23), taken from *The Pesticide Manual* [10]:

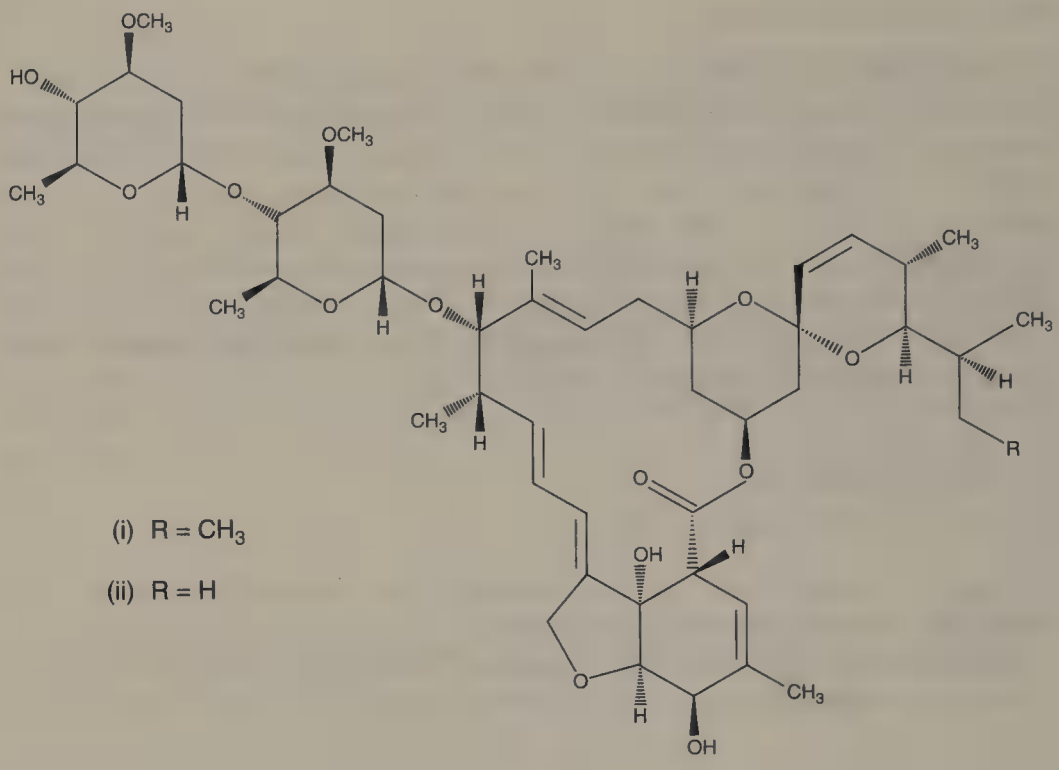


Figure 5.23 Abamectin (ISO) (INN). Structure taken from the *Pesticide Manual*, copyright British Crop Protection Council, with permission.

This has the systematic name: (10*E*,14*E*,16*E*,22*Z*)-(1*R*,4*S*,5′*S*,6*S*,8*R*,12*S*,13*S*,20*R*,21*R*,24*S*)-6′[(*S*)-*sec*-butyl]-21,24-dihydroxy-5′,11,13,22-tetramethyl-2-oxo-3,7,19-trioxatetracyclo[15.6.1.1^{4,8}.0^{20,24}]pentacosa-10,14,16,2-tetraene-6-spiro-2′-(5′,6′-dihydro-2′*H*-pyran)-12-yl 2,6-dideoxy-4-*O*-(2,6-dideoxy-3-*O*-methyl-α-L-*arabino*-hexapyranosyl)-3-*O*-methyl-α-L-*arabino*-hexapyranoside mixture with (10*E*,14*E*,16*E*,22*Z*)-(1*R*,4*S*,5′*S*,6*S*,8*R*,12*S*,13*S*,20*R*,21*R*,24*S*)-21,24-dihydroxy-6′-isopropyl-5′,11,13,22-tetramethyl-2-oxo-3,7,19-trioxatetracyclo[15.6.1.1^{4,8}.0^{20,24}]pentacosa-10,14,16,22-tetraene-6-spiro-2′-(5′,6′-dihydro-2′*H*-pyran)-12-yl 2,6-dideoxy-4-*O*-(2,6-dideoxy-3-*O*-methyl-α-L-*arabino*-hexapyranosyl)-3-*O*-methyl-α-L-*arabino*-hexapyranoside 4:1.

Well, what else would you call it? It does raise some useful points. The stereodescriptors (*E* and *Z*) are cited before (*R* and *S*). The '=' sign at the end of the line means the name continues without a break and the small capital 'L' and 'D' are special symbols in sugar names. The 4:1 at the end shows the ratios of the two components. It will come as no surprise that it takes quite a while to arrive at a name for abamectin.

This completes our look at IUPAC names. CNAS have always believed that the most important rule of nomenclature is the one that is not written.

Rule Zero: Whatever name you use, be sure ALL your audience knows what you mean.

References

1. *Biochemical Nomenclature and Related Documents* (1992) 2nd edn, prepared by C. Liébecq, Portland Press.
2. Godly. E. W. and Taylor, R. (1997) Nomenclature and terminology of fullerenes: a preliminary study. *Pure and Applied Chemistry*, **69 (7),** 1411–1434.
3. Beaton, J. M. (1995) *Journal of Chemical Education*, October **72 (10)**, 863–869.
4. Cahn, R. S., Ingold, C. K. and Prelog, V. (1966) *Angewandte Chemie International Edition* **5**, 385.
5. Cahn, R. S., Ingold, C. K. and Prelog, V. (1966) *Angewandte Chemie* **78**, 413 (in German).
6. Cahn, R. S. and Ingold, C. K. (1951) *Journal of the Chemical Society (London)*, 612.
7. Cahn, R. S., Ingold, C. K. and Prelog, V. (1956) *Experientia*, **12**, 81.
8. Asimov, I. (1975) *The Left Hand of the Electron*, Panther, St Albans, UK.
9. IUPAC (1990) *IUPAC Nomenclature of Inorganic Chemistry*, Recommendations 1990, ed. G. J. Leigh, Blackwell Scientific Publications.
10. British Crop Protection Council (1997) *The Pesticide Manual*, 11th edn, ed. C. D. S. Tomlin.

6 Nomenclature for polymer chemistry

A. D. JENKINS

6.1 Introduction

Although anyone is at liberty to invent a system of chemical nomenclature, and some organizations have done so to suit their own purposes, it is generally recognized that the standard chemical nomenclature is that developed and recommended by the International Union of Pure and Applied Chemistry (IUPAC). In 1919, IUPAC inherited responsibility for chemical nomenclature from the International Union of Chemical Societies (which had identified one of its goals as the development of the 'nomenclature of inorganic and organic chemistry'), building upon work that had been done on the formulation of names for organic and inorganic compounds for some 30 years or more. Readers interested in the progress of nomenclature during this period, and the history of IUPAC, might wish to consult a number of excellent books on these topics [1–3]. A review of the problems in devising nomenclature for polymers may also be of interest [4].

IUPAC work on nomenclature specific to polymer chemistry commenced as recently as 1950. At that time, IUPAC had not yet elevated macromolecular science to divisional status but it had founded a Commission on Macromolecules, and this body then set up a Sub-Commission on Nomenclature, which published its first report in 1952 [5]. Ten years later, it issued a further report [6] but the long-term systematic development of nomenclature in the polymer field really commenced after the Commission had become a Division and the Sub-Commission had become the Commission on Macromolecular Nomenclature, in 1968. This latter Commission will hereafter be referred to simply as 'the Commission'. In 1991, the principal recommendations of this Commission were published in collected form as a *Compendium of Macromolecular Nomenclature* [7], the so-called 'Purple Book' (it was unfortunately published with a mauve cover), which contains many of the papers referred to below; it has been translated into several other languages.

Any experienced chemist will have some idea of the difficulties encountered in attempting to compose concise, but accurate and universally acceptable definitions of technical terms such as are met with in the nomenclature of chemical compounds. The polymer chemist is beset by all of these but, in addition, he has the task of coping with the fact that

the molecules of interest to him do not conform to the general rule of chemistry that all molecules of a particular substance are identical with respect to composition and, moreover, that a great variety of types of molecular architecture are encountered.

To begin with, polymer molecules are large, far outside the range of molecular weight (molar mass) met in general chemistry. In general, a chemist would regard a molecule as quite extraordinarily big if it had a molar mass of, say, 1500 g mol^{-1} (the molar mass of the very large and complicated molecule of Vitamin B_{12}, for example, is 1330 g mol^{-1}); a polymer chemist would regard a polymer molecule as exceptionally small if it had a molar mass as low as 5000 g mol^{-1}. The great majority of synthetic polymer molecules have molar masses in the range 10^4 to 10^8 g mol^{-1}; however, the real problem arises not from sheer size but from the fact that the size is not uniform, covering a wide distribution which it may or may not be possible to define. In fact, the lack of identity of structure inherent in polymers has led, at times, to their classification as 'chemicals' being challenged but such pedantry is no longer encountered, and it is now accepted that they are chemicals of a special type.

The fact that, in a sample of a given polymer, pure in all other respects, the size of the molecules varies over a wide range, makes it extremely difficult even to define the word 'polymer'. To take a very common example, poly(vinyl chloride) has a structure often written as $[CH_2CH(Cl)]_n$; the implication is that the molecule contains a large number (denoted by the subscript '*n*') of -$CH_2CH(Cl)$- units, giving the molecule the approximate molecular formula C_2H_3Cl, but an approximation is certainly involved because there must be groups of some kind at both ends of the molecule to satisfy the loose valencies (represented by the '-' in the structural formula); the assumption is almost always made that both these end-groups are so insignificant a part of the structure that they can be left out of account for nomenclature purposes.

6.2 The Stipulation

The qualification (about the insignificance of end-groups) made at the end of the previous paragraph is so typical of the approximations encountered in trying to make useful statements about polymers that it seems appropriate at this point to quote from one of the foundation documents issued by the Commission. In the introduction to its document on stereochemical definitions [8], one finds the following paragraph, which will be referred to below as 'The Stipulation'.

> In order to present clear concepts it is necessary that idealized definitions be adopted but it is recognized that the realities of polymer science must be faced. Deviations from ideality arise with polymers at both molecular and bulk levels

in ways that have no parallel with the small molecules of organic or inorganic chemistry. Although such deviations are not explicitly taken into account in the definitions below, the nomenclature recommended can usefully be applied to the predominant structural features of real polymer molecules, if necessary with self-explanatory, if imprecise, qualifications, such as 'almost completely isotactic' or 'highly syndiotactic'. Although such expressions lack the rigour beloved by the purist, every experienced polymer scientist knows that communication in this discipline is impossible without them.

The sentiments expressed in the previous paragraph could profitably be repeated in every publication purporting to describe polymers.

6.3 Polymers and polymer molecules: basic definitions of terms

Before proceeding, one very important distinction must be made clear. In polymer science it is necessary to distinguish carefully between the **substance** (the **polymer**) and the **molecules** of which it is composed; the latter are known as **macromolecules** (or **polymer molecules**). In modern nomenclature, there will be one set of definitions for the polymer and a parallel set for the corresponding macromolecule (and indeed a third set for oligomers – low-molecular-weight polymers – but these will not be dealt with here). It has been suggested that the term **polymer** is redundant [9] but this is not the view of the IUPAC Commission [10].

The two definitions, **polymer** and **macromolecule**, are absolutely fundamental to all subsequently formulated definitions of terms to describe the structure or properties of polymers. In fact, the Commission did not manage to define 'macromolecule' at all in its first attempt at composing basic definitions of terms relating to polymers [11] but this omission was rectified in the recent revision [12] entitled 'Glossary of Basic Terms in Polymer Science'.

The 1974 structure-based definition of 'polymer' is remarkable in being based on the properties of the material: thus a polymer is "a substance composed of molecules characterized by the multiple repetition of one or more species of atoms or groups of atoms linked to each other in amounts sufficient to provide a set of properties that do not vary markedly with the addition or removal of one or a few of the constitutional units."

Perhaps the use of the term 'structure-based' should be discussed before proceeding. It is possible to look at a polymer from the point of view of **either** its structure **or** the identity of the monomer(s) from which it was made, actually or conceptually: the former approach requires that a sufficiently detailed analysis of the structure can be performed to enable a name based on structure to be formulated with confidence, while the latter merely needs a knowledge of the components of the reaction mixture. To take an example, if styrene is heated in the absence of oxygen to a

temperature above 100°C, it will polymerize, and we may reasonably call the product 'polystyrene'; that will naturally be its **source-based name**. If the structure of the polymer can be assumed or, better, be shown to be $\text{+CH}(\Phi)\text{CH}_2\text{+}_n$, where Φ denotes C_6H_5, then we may call it poly(1-phenylethene), and this will be its **structure-based name**. The use of the structure-based name implies that the structure is **almost** perfectly regular, ideally with every alternate carbon atom in the backbone substituted by a phenyl group, and with a total absence of groups corresponding to phenyl substitution on adjacent atoms along the chain – 'almost' because of The Stipulation.

As structure-based nomenclature relies on a precise knowledge of the detailed molecular architecture, the development of new analytical techniques may alter the degree of precision with which structures are known, and hence the structure-based name of a polymer may have to change to accommodate the heightened understanding of its constitution. This fact may seem to point to a preference for the more permanent source-based names but here we encounter the complication that a certain monomer may give rise to polymers with different structures when reacted in different ways, thus giving rise to ambiguity. For example, the monomer isoprene can react in the 1,2-, the 3,4- or the 1,4- mode and, in the last case, the residual double bond may have either the *cis* or the *trans* configuration; it is not satisfactory simply to label all these products with the source-based name 'polyisoprene', especially as the properties of the various polymers differ greatly, only the *cis*-1,4- variety having the characteristic properties of rubber. Obviously, the structure-based name conveys more information about the detailed nature of the macromolecule but, inevitably, the mere fact that it encodes the elements of structure entails a measure of complication, so that it may well be cumbersome and unsuited to verbal communication. Whether the source-based or structure-based name is to be used must depend on the nature and purpose of the communication, and it may even be advisable to use both. Generic nomenclature, which adds a prefix to describe the chemical nature of the linkage between adjacent monomer units, is under development to ease this awkward situation.

The 1996 structure-based definition [12] of 'macromolecule' is simpler in form than that of 'polymer'; it reads as follows: "A molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetition of units derived, actually or conceptually, from molecules of low relative molecular mass." If this definition had been available in 1974, it might have been thought reasonable to define a polymer as "a substance composed of macromolecules", however, the 1996 definition contributes a valuable codicil to the definition in the form of the note: "In many cases, especially for synthetic polymers, a molecule can be regarded as having a high relative molecular mass if the addition or

removal of one or a few units has a negligible effect on the molecular properties." In IUPAC nomenclature, the type of unit referred to in the definition of macromolecule is properly known as a **constitutional repeating unit**, usually abbreviated to CRU.

The purist will still object that the definitions of both 'polymer' and 'macromolecule' contain arbitrary adjectives (like 'high', 'few' and 'negligible') but one can only point to The Stipulation, and ask him to try to do better.

In the text above, polymer formulae have been written so as to imply that there is a linear sequence of the units (monomer residues) in the structure: the simplest polymers are indeed linear in nature but many are branched, sometimes in quite complicated fashion. Thus, polymer chains, that can be thought of as having been formed independently, may become linked through mutual side-chains or cross-links; this type of polymer is actually a network, sometimes called an 'infinite molecule' because one single molecule fills the entire container. Worse still, from the point of view of nomenclature, there can be two (or more) networks present which interpenetrate one another. Of course, if a polymer molecule is branched, even just once, there must be an element in the structure (the branch point or junction point) which is different from the remainder of the units. The challenge to develop nomenclature to cope with all these ramifications is formidable but considerable strides have been made towards that end (section 6.9).

6.4 Regular single-strand, quasi-single-strand, and double-strand polymers

After a long struggle leading to its first recommendations for basic terms in 1974, the Commission moved rapidly to deal with the nomenclature of regular single-strand organic polymers [13]. (The adjective 'regular' has a very specific meaning in this context: in a regular polymer, the structure of the molecules can be represented by a single sequential arrangement of a plurality of CRUs of a single type.) It is interesting to note that the seeds of the 1996 definition of 'macromolecule' were present in this document, which states the fundamental principle that "This nomenclature system rests upon the selection of a preferred *constitutional repeating unit* (CRU) of which the polymer is a multiple . . .".

The name of the selected class of polymers (regular single-strand) corresponds to what are often called linear homopolymers, the least complicated form of polymer structure. Thus, the structure-based name of poly(vinyl chloride) is poly(1-chloroethene) [N.B. In 1976, it was poly(1-chloroethylene)] but it can happen that the CRU is not identical to the monomer unit; polyethylene, usually represented as $[CH_2CH_2]$, is actually named

poly(methylene), because the **smallest possible** CRU is chosen as the basis of the name, and in this case that is CH_2. In proceeding with this exercise, the Commission faced the requirement that the organic parts of the polymer names had to be consistent with the agreed nomenclature of organic chemistry, as embodied in the IUPAC recommendations; the challenge was then to devise rules for naming polymers which were both in line with organic nomenclature and which described the polymeric nature of the structures. This 1976 document covered many types of regular organic polymers with remarkable success but a revision is currently being carried out which takes account of the developments in the nomenclature of organic chemistry during the intervening twenty-odd years; the revised recommendations are likely to be published in late 1998.

The next development in geometrical structural sophistication led to definitions for regular quasi-single-strand polymers [14], molecules in which inorganic or coordination units cause loops or rings to be incorporated into the polymer chain without destroying the essentially linear structure; in this endeavour, it was necessary to maintain consistency with the IUPAC recommendations on the nomenclature of inorganic chemistry. A stronger deviation from the single-strand concept is involved when ladder or spiro structures are present, and this situation had to wait until 1993 for a resolution, see below. It is necessary to consider how to define expressions such as 'single-strand', 'quasi-single-strand', and 'double-strand'; they fall into the category of things that seem easily to be recognized in a common-sense way, yet defy simple concise definition. The Commission defines 'single-strand' as a structure containing CRUs, each of which has just two terminal units, each composed of a single atom, 'quasi-single-strand' as one in which CRUs are joined through a single atom at one terminal and through two atoms at the other terminal, while 'double-strand' applies to structures comprising an uninterrupted sequence of rings, with adjacent rings having one atom in common (in a spiro polymer) or two atoms in common (in a ladder polymer). All these varieties are exemplified in the document.

In all these cases, the basic procedure for arriving at the correct systematic name for a polymer is the same:

(i) **identify** the preferred CRU;
(ii) **orient** the CRU;
(iii) **name** the CRU; and
(iv) **name** the polymer.

The selection of the **preferred** CRU may require the application of seniority rules to determine which of a number of possible CRUs should provide the basis of the name; if this step were omitted, it would often be possible to name a single polymer in several different ways, adding to (rather than relieving) the confusion.

6.5 Polymer formulae

Of course, communication in print is not always based on the use of written names because a more immediate impact is often gained by a formula. However, just as it is possible (in the absence of rules) to compose various names for a polymer, so a formula can be written in a variety of ways, unless guidelines are observed. To meet this need, the Commission tackled the formulation of procedures for the graphic representation of polymer structures (i.e. chemical formulae for polymers) with the result published in 1994 [15]. This document presents rules for formulae compatible with the nomenclature documents cited above for regular polymers, and it also provides for irregular polymers, copolymers, and star polymers. The reader may find this document particularly useful because it presents many examples of formulae together with both structure-based and source-based names.

There are actually eight guidelines governing the construction of an unambiguous graphical representation of a macromolecule; they are as follows.

(i) The established usage of organic, inorganic and polymer nomenclature will be followed.
(ii) The order of citation of constitutional units is arbitrary.
(iii) Dashes representing chemical bonds may be omitted, except at the ends of units.
(iv) Side-groups or substituents are set between enclosing marks.
(v) Enclosing marks with subscript letters denote a multiplicity of the units specified.
(vi) Subscripts *n*, *p*, *q*, *r*, etc. denote multiplicities of polymer sequences whereas *a*, *b*, *c*, etc. denote multiplicities of oligomer sequences.
(vii) End-groups can be attached to terminal units, outside the enclosing marks.
(viii) Mass fractions, mole fractions, molar masses, and degrees of polymerization can be indicated in parentheses after the formula.

6.6 Copolymers

Copolymers provide the nomenclature enthusiast with a new challenge, not only because there must inevitably be more than one CRU present but also because there are several different ways in which the CRUs can be arranged. There are organized structures, block and graft copolymers, in which long sequences of units of each type are maintained, and there is the much more common variety resulting from the reaction of a mixture of monomers, where the order in which monomer units enter the structure

is partly governed by chance. It will mostly be assumed in this discussion, for the sake of simplicity, that only two monomers are involved in any given copolymer structure but the recommended nomenclature extends to more elaborate materials, e.g. star polymers.

In attempting to construct a source-based nomenclature for copolymers, it is easy enough to state the names of the source monomers, the problem is to describe the nature of the resulting copolymer structure, if it is known. When the common 'bucket-chemistry' approach is employed to make a copolymer, there is little or no regularity in the sequential arrangement of monomer units in a copolymer molecule. In very rare cases, the monomer units may find themselves in a truly random order (i.e. one governed by chance alone) but almost invariably some element of chemical preference is also involved; for this reason, the Commission preferred [16] to name such copolymers 'statistical', although extant text-books will usually describe them (incorrectly) as 'random'. Very occasionally, the monomer units enter the copolymer structure in strictly alternating fashion, giving rise to 'alternating' copolymers. (In systems involving more than two monomers, parallel 'periodic' structures are possible.) In addition, there are the block and graft varieties. The Commission decided to represent a binary copolymer by a name in which the names of the source monomers (say, A and B) are linked by an italicized connective denoting the particular type of arrangement. The connective '*co*' was employed to indicate that the manner in which the monomer units are arranged in the structure is unknown. If A and B represent the names of the monomers involved, the seven basic types of copolymer (unknown, statistical, random, alternating, periodic, block, graft) would thus have names like the following: poly(A-*co*-B), poly(A-*stat*-B), poly(A-*ran*-B), poly(A-*alt*-B), poly(A-*per*-B-*per*-C), poly(A-*block*-B) and poly(A-*graft*-B). An alternative system was also presented by the Commission, in which the seven basic types would have names such as: copoly(A/B), *stat*-copoly(A/B), *ran*-copoly(A/B), *alt*-copoly(A/B), *per*-copoly(A/B/C), *block*-copoly(A/B) and, finally, *graft*-copoly(A/B).

The reader may be wondering whether the problem, mentioned earlier, that a single monomer may polymerize in different ways, can be addressed through the medium of copolymer nomenclature. Suppose, for example, that isoprene reacts to provide a polymer structure containing monomer residues in both the 1,4- and 1,2- modes, or suppose that simple poly(vinyl acetate) is partially hydrolysed, or again suppose that poly(propene oxide) polymerizes to yield a structure containing both *R*- and *S*- units; are these materials to be regarded as copolymers? The Commission preferred the solution of calling such materials 'pseudo-copolymers', and noted that they could then be named like true copolymers.

A particular nomenclature problem arises with certain polymers of the familiar simple polyester type, e.g. the product of reaction between

ethylene glycol and terephthalic acid. This polymer can be made from these two components directly, suggesting the source-based name poly[(ethylene glycol)-*alt*-(terephthalic acid)]. It can also be made by self-condensation of the partial ester $HOCH_2CH_2O.CO.C_6H_4COOH$, or alternatively of the ester bis(2-hydroxyethyl) terephthalate, in both cases suggesting a source-based name after the fashion of homopolymers, e.g. poly[bis(2-hydroxyethyl)terephthalate]. Regardless of the preferred source-based name, the structure-based name is poly(oxyethylenoxyterephthaloyl). For all such materials, where it is possible to visualize two initial reactants undergoing preliminary 1:1 reaction to form an 'implicit monomer', the homopolymerization of which would yield the actual product, the Commission felt that the single-strand structure-based (homopolymer) nomenclature would then be suitable but the presence of a third component would certainly necessitate copolymer nomenclature.

6.7 Irregular single-strand organic polymers

For single-strand polymers that do not conform to the definition 'regular', the nomenclature scheme has been extended in a way that provides an alternative structure-based formulation for the names of copolymers [17].

'Irregular', in this context, implies that more than one type of CRU is involved in the structure or that CRUs of a single type are not all connected in a unique directional sense. The first of these two classes embraces copolymers, for which a source-based system of nomenclature had been published in 1985, as described above. The second class is exemplified by structures in which 'head-to-head' pairs of units appear, in addition to the customary 'head-to-tail' arrangement. This occurs, for example, in the homopolymerization of 1-chloroethene (vinyl chloride), where a typical segment in a chain may be as shown below:

$$-CH_2-CH(Cl)-CH_2-CH(Cl)-CH_2-CH(Cl)-CH(Cl)-CH_2-CH_2-CH(Cl)-$$

Again, buta-1,3-diene polymerization may give rise to a product in which the monomer units are incorporated in both the 1,2- and 1,4- modes, thus:

$$-CH_2-CH=CH-CH_2-CH_2-CH=CH-CH_2-CH_2-CH=CH_2CH-CH_2-CH=CH-CH_2-$$

The procedure for naming an irregular single-strand organic polymer is the following.

(i) **Write** the structure.
(ii) **Select and orient** the minimum number of constitutional units necessary to represent that structure.

(iii) **Check** by combining the selected constitutional units in all possible ways to form a polymer chain. Incorrect constitutional units will give chain segments that do not correspond to the structure as written in (i).
(iv) **Name** the constitutional units according to the rules of organic and structure-based polymer nomenclature.
(v) **Write** the name as specified in Rules 1 to 7 of the recommendations.

To take one of the simplest examples, a typical segment of a molecule of a statistical copolymer of 1-phenylethene (styrene) and 1-chloroethene (vinyl chloride) might have the following structure

$$-CH_2-CH(Cl)-CH_2-CH(\Phi)-CH_2-CH(Cl)-CH_2-CH(Cl)-CH_2-CH(Cl)-CH_2-CH\Phi)-$$

and it would be named poly(1-chloroethylene/1-phenylethylene).

6.8 Double-strand organic polymers

After dealing with regular and irregular single-strand organic polymers, it was appropriate to extend the range of polymer nomenclature to encompass their double-strand counterparts, which have either spiro or ladder character [18]. As mentioned above, a spiro macromolecule is one consisting of an uninterrupted sequence of rings, with adjacent rings having only one atom in common. By contrast, a ladder macromolecule is one consisting of an uninterrupted sequence of rings, with adjacent rings having two or more atoms in common.

The rules to be followed in constructing the name of such a polymer are identical to those for the regular single-strand variety but more care is required in selecting, orienting and naming the preferred constitutional repeating unit; the necessary procedures are embodied in rules in the document.

6.9 Non-linear polymers

Extension of the system of polymer nomenclature to branched and cyclic macromolecules, networks, and macromolecular assemblies held together by non-covalent forces (blends, networks and complexes) was accomplished in a document published in 1997 [19]. After listing a number of relevant definitions, which also appeared in the Glossary of Basic Terms in Polymer Science [12], the rules of nomenclature for non-linear macromolecules and systems thereof are laid down, the non-linear nature being denoted by italicized prefixes and connectives added to the name of the corresponding linear material. Some self-explanatory examples of the use of prefixes are: *branch*-polystyrene, *star*-polystyrene, *cyclo*-poly(dimethylsiloxane)

and *net*-polybutadiene. The connectives include *net* (network), μ-*net* (micro-network) and *v* (crosslink), giving rise to such names as: poly (methyl methacrylate)-*net*-poly(ethylene oxide), and polystyrene-*v*-divinylbenzene.

For assemblies of macromolecules, examples include: *ipn* (interpenetrating polymer network), *sipn* (semi-interpenetrating polymer network), *blend* and *compl* (complex).

6.10 Stereochemistry

Since the recognition of stereoisomerism in vinyl polymers in the early 1950s, it has become particularly important to be able to specify the stereochemical features of polymers. In the case of vinyl polymers, two distinct types of stereoisomerism may have to be taken into account, sometimes in combination. Thus, the tactic nature of the polymer (isotactic, syndiotactic, atactic) may need to be specified and, in the case of diene polymerization, the configuration of residual double bonds (*cis*, *trans*) may have to be made explicit.

The nomenclature devised by the Commission [8] introduced some new concepts relating to the microstructure of polymer chains; it dealt with stereoblock polymers, in which the blocks differ only in the stereochemical arrangement within the blocks, with sequences of units, with the relative configurations of consecutive units, and with the conformations of polymer molecules.

6.11 Individual macromolecules, their assemblies and dilute polymer solutions

It has been remarked in the introduction that polymers typically comprise a collection of molecules of assorted lengths; in order to characterize such a substance, it is necessary to use some form of average molar mass, a problem that does not arise in dealing with 'small-molecule' chemistry. The difficulty encountered here is that there are various ways to take an average and, moreover, different ways are appropriate to different circumstances. There are also many characteristics of polymer molecules that are of no interest in ordinary chemistry, parameters such as: root-mean-square end-to-end distance, radius of gyration, persistence length and thermodynamically equivalent sphere. The same is true of the characteristic features of dilute solutions, notably: viscosity, theta solution, excluded volume and expansion factor.

Although all these terms have been current for a long time, usage has not been consistent nor, in some cases, clearly defined. The Commission

brought all this material into good order in a document [20] which can claim to be historic in at least one respect, namely, that it outlawed the ridiculous (internally contradictory) description 'monodisperse' for a polymer consisting of molecules uniform with respect to relative molar mass.

6.12 Crystalline polymers and liquid-crystal polymers

In the same year (1989), the Commission also introduced a system of nomenclature for crystalline polymers [21] and the process of molecular crystallization. Every effort was made to rely on the 'Basic Definitions of Terms Relating to Polymers' of 1974, and to conform to the generally accepted usage of the study of the crystalline state.

A more formidable task was to be encountered later after the explosion of work on liquid-crystal polymers in the 1980s. Because the whole subject of the liquid-crystalline state had undergone tremendous expansion in recent years, both experimentally and theoretically, it was deemed desirable to produce a document on liquid-crystallinity, covering materials of both high and low relative molar mass. This has reached an advanced stage of drafting (in consultation with the International Liquid Crystal Society) and will shortly be distributed for public comment, with a view to finalization and publication in 1998/1999.

6.13 Polymerization reactions

Almost any textbook of polymer chemistry will divide polymerization reactions into two classes: condensation and addition processes. Polymers are then classed as condensation polymers or addition polymers, accordingly. In recent years, there has been a tendency to think rather in terms of the more detailed mechanism, and so to class reactions as either chain-growth or step-growth, depending (respectively) on whether or not the process follows a kinetic chain type of mechanism. In chain-growth reactions, every propagation involves monomer whereas, in step-growth reactions, molecules of all sizes can engage in the growth step.

With increasing knowledge of different types of polymerization processes and their mechanisms, it has become clear that a slightly more elaborate classification is necessary, and the Commission has published recommendations [22] which cover all presently known polymerization reactions.

In the new system, the classes of reactions are:

(i) **chain polymerization**, essentially consisting of kinetic chains involving monomer at each step;

(ii) **condensative chain polymerization**, which is really a sub-group of (i) with the additional qualification that a small molecule is eliminated at each step (as in e.g. the polymerization of *N*-carbonic anhydrides);
(iii) **polyaddition** where growth proceeds by addition reactions between molecules of all sizes and
(iv) **polycondensation** where growth proceeds by condensation reactions between molecules of all sizes.

6.14 Degradation and ageing

The useful lifetime of polymers clearly limits their applications; the change in properties following oxidation, hydrolysis, etc. is a source of concern to both manufacturers and users, so it is appropriate that there should be agreed terminology for the description of these and related processes. This issue was addressed by the IUPAC Commission in a document published in 1996 [23].

6.15 Multi-phase polymer systems

As materials science progresses, so the systems in which polymers are exploited commercially become more complex. Polymers are now frequently encountered in mixtures with one another (blends) or in combination with a non-polymeric substance (composites). Terminology for these materials has become necessary, and the IUPAC Commission is actively engaged on drafting recommendations that will probably be published early in 1999. One of the most contentious terms in this field is 'polymer alloy'. Not everyone sees this term as a synonym for 'polymer blend' but the definition that appears likely to be agreed is "A macroscopically homogeneous polymeric material that is comprised of either a compatible polymer blend, a miscible polymer blend, or a multi-phase copolymer". This definition begs the question of the meaning of 'miscible'; does it imply miscible under all known conditions? The answer is probably 'no' but this sort of issue illustrates well the difficulties inherent in drafting definitions of technical terms.

6.16 Mechanical properties

The growing importance of the materials aspects of polymers has caused attention to become focussed on nomenclature for their mechanical properties. A set of recommendations [24] has been prepared, dealing with bulk polymers and their concentrated solutions, with particular emphasis on the elastic and viscous properties of isotropic materials.

6.17 Current and future projects

The Commission is currently working on documents dealing with terminology concerned with the Kinetics and Thermodynamics of Polymerization, Generic Nomenclature for Polymers, and Asymmetric Polymerization. In the near future, it expects to become seriously engaged in drafting recommendations for Thermal Properties of Polymers, Ceramics, Cyclic Macromolecules, Hyperbranched Macromolecules, and Membranes. Further ahead, studies may well embrace: Dielectric, Optical, and Acoustic Properties of Polymers; Gels; Chromatography of Polymers; Ion-Containing Polymers; Polymerization in Dispersed Systems; and the Functionalization of Polymers. Clearly, there will be no shortage of topics requiring attention for many years to come.

6.18 Conclusion

The struggle to reach international agreement on the most basic terms used to describe polymers and macromolecules was protracted but, once this task had been accomplished, it proved possible to make much more rapid progress in polymer nomenclature by proceeding stepwise, starting with the most simple structures, and gradually elaborating on them. It must be realized that this was possible only because the Commission was looking ahead all the time, anticipating the direction to be taken by future projects.

Since the Commission can only meet once a year, usually for 4 or 5 days, it has to use meeting time with the utmost efficiency, and much must be achieved by circulation of drafts for comment during the intervening months. Moreover, some five or six main projects are always under consideration simultaneously, working parties being convened to prepare the way for discussion by the full Commission.

Although the construction of a system of nomenclature by the Commission may seem to make very slow progress, it takes the average chemist a great deal longer to absorb the recommendations and put them into practice. It is a most remarkable thing that a scientist will spend months and years on his research, only to commit his results to paper in a way which may be misleading, ambiguous or downright unintelligible because he cannot be bothered to take the minimal trouble to use the proper nomenclature, which it has taken a lot of conscientious people years of labour to produce.

Acknowledgements

Grateful thanks are due to the John Benjamins Publishing Co. and the International Union of Pure and Applied Chemistry for permission to quote extensively from material published previously in *Terminology* and *Pure and Applied Chemistry*.

References

1. Crosland, M. P. (1978) *Historical Studies in the Language of Chemistry*, 2nd edn, Dover Publications, New York.
2. Verkade, P. E. (1985) *A History of the Nomenclature of Organic Chemistry*, D. Reidel Publishing Company, Dordrecht.
3. Fennell, R. W. (1994) *History of IUPAC 1919–1987*, Blackwell Science Ltd., Oxford.
4. Jenkins, A. D. (1995) Problems in composing definitions of terms for polymer chemistry. *Terminology*, **2**, 351–364.
5. Huggins, M. L., Natta, G., Desreux, V. and Mark, H. (1952) Report on nomenclature in the field of macromolecules. *Journal of Polymer Science*, **8**, 257–277.
6. Huggins, M. L., Natta, G., Desreux, V. and Mark, H. (1962) Report on nomenclature dealing with steric regularity in high polymers. *Journal of Polymer Science*, **56**, 153–161.
7. Metanomski, W. V. (1991) *Compendium of Macromolecular Nomenclature*, Blackwell Scientific Publications, Oxford.
8. Commission on Macromolecular Nomenclature of IUPAC (1981) Stereochemical definitions and notations relating to polymers (1980). *Pure and Applied Chemistry*, **53**, 733–752.
9. Staudinger-Woit, M. (1996) Macromolecule and polymer. *Terminology*, **3**, 343–347.
10. Jenkins, A. D. (1997) *Terminology*, **4(1)**, 157–159.
11. Commission on Macromolecular Nomenclature of IUPAC (1974) Basic definitions of terms relating to polymers (1974). *Pure and Applied Chemistry*, **40**, 477–491.
12. Commission on Macromolecular Nomenclature of IUPAC (1996) Glossary of basic terms in polymer science (1996). *Pure and Applied Chemistry*, **68**, 2287–2311.
13. Commission on Macromolecular Nomenclature of IUPAC (1976) Nomenclature of regular single-strand organic polymers (1975). *Pure and Applied Chemistry*, **48**, 373–385.
14. Commission on Macromolecular Nomenclature of IUPAC (1985) Nomenclature for regular single-strand and quasi-single-strand inorganic and coordination polymers (1984). *Pure and Applied Chemistry*, **57**, 149–168.
15. Commission on Macromolecular Nomenclature of IUPAC (1994) Graphic representations (chemical formulae) of macromolecules (1994). *Pure and Applied Chemistry*, **66**, 2469–2482.
16. Commission on Macromolecular Nomenclature of IUPAC (1985) Source-based nomenclature for copolymers (1985). *Pure and Applied Chemistry*, **57**, 1427–1440.
17. Commission on Macromolecular Nomenclature of IUPAC (1994) Structure-based nomenclature for irregular single-strand organic polymers (1994). *Pure and Applied Chemistry*, **66**, 873–889.
18. Commission on Macromolecular Nomenclature of IUPAC (1993) Nomenclature of regular double-strand (ladder and spiro) organic polymers *Pure and Applied Chemistry*, **65**, 1561–1580.
19. Commission on Macromolecular Nomenclature of IUPAC (1997) Source-based nomenclature for non-linear macromolecules and macromolecular assemblies (1997). *Pure and Applied Chemistry*, **69**(12), 2511–2521.
20. Commission on Macromolecular Nomenclature of IUPAC (1989) Definitions of terms relating to individual macromolecules, their assemblies, and dilute solutions (1988). *Pure and Applied Chemistry*, **61**, 211–241.
21. Commission on Macromolecular Nomenclature of IUPAC (1989) Definitions of terms relating to crystalline polymers (1988). *Pure and Applied Chemistry*, **61**, 769–785.

22. Commission on Macromolecular Nomenclature of IUPAC (1994) Basic classification and definitions of polymerisation reactions (1994). *Pure and Applied Chemistry*, **66**, 2483–2486.
23. Commission on Macromolecular Nomenclature of IUPAC (1996) Definitions of terms relating to degradation, aging, and related chemical transformations of polymers (1996). *Pure and Applied Chemistry*, **68**, 2313–2323.
24. Commission on Macromolecular Nomenclature of IUPAC (1998). Definitions of terms relating to the non-ultimate mechanical properties of polymers (1998). *Pure and Applied Chemistry*, **70**, in press.

7 Natural Products

J. BUCKINGHAM

7.1 Introduction

The history of natural products is virtually the history of organic chemistry itself. Even at the end of the Twentieth Century, Nature still surprises us by the revelation of hitherto unknown structural types of molecule, and so it is no surprise that the nomenclature of natural products encompasses and exemplifies many of the problems that apply to organic compounds generally, often in an empirical and somewhat chaotic form. To add to the difficulties imposed by the sheer structural diversity of natural product molecules, there exist some genuine philosophical differences between different 'user groups' as to how the nomenclature should be organized and systematized, except in the case of structurally rather homogeneous classes such as carbohydrates and lipids.

At the time of writing, there is no published version Section F of the IUPAC Organic Chemistry rules *Natural Products and Related Compounds* although several versions of a draft have been circulated for comments, the first as long ago as 1976. Successive drafts have recognized that the nomenclature of natural products has suffered from much confusion.

In the chapter which follows, I will first review the general principles of natural product nomenclature, with special reference to the practical difficulties associated with any body such as IUPAC attempting to evolve a uniform template, and in section 7.7 will then give a brief review of the nomenclature of the different types of natural products, paying particular attention to groups that illustrate general points of good or bad practice, or particular difficulties. The chapter as a whole will concentrate on points of general principle, practical problems of nomenclature and 'grey areas', rather than on groups that are nomenclaturally stable and well codified. It will be impossible within the scope of a single chapter to deal with every ramifying subgroup, and it will be similarly impossible to review the whole range without making some controversial statements. However, it is hoped that this review will at least serve to point the reader towards good practice, and that the references cited will enable him or her to obtain a view of how to name new or existing natural products without alienating **everyone**.

An extensive detailed guide to the subgroups of natural products dealing with their biogenesis, structure and stereochemistry is published as the

introduction to Volume 7 of the print edition of the *Dictionary of Natural Products (DNP)* (1992), and is now available in an updated version as a read-only Acrobat file accessible from the introductory screen of the DNP CD-ROM version. Readers of this chapter are referred to this source for a more detailed exposition of the individual natural product subgroups and for detailed numbering and nomenclature schemes for each individual subgroup. Further important information can be gleaned from the CAS Index Guide and other CAS sections.

The number of known and documented natural products is currently 100 000+, taking a relatively traditional view of what constitutes a natural product and therefore excluding the majority of oligomeric biomolecules, i.e. polypeptides, oligosaccharides and polynucleosides. These are profoundly important from a biochemical point of view but they do not show the constitutional diversity that makes the secondary metabolites so interesting chemically and causes the complications in naming them.

This number is relatively small compared with the number of known organic compounds (currently about 15 million). The universe of possible natural products is constrained by biogenetic considerations; in other words, for a natural product to be biosynthesized by any organism, that organism must have the enzyme systems to assemble it, and enzymes, while diverse and extremely inventive, do not have all of the powers of the synthetic chemist. The nomenclature of natural products is influenced by these biosynthetic possibilities; it has evolved to fit the structural types that are actually found in nature. In domains where diversity rules to a high degree and structure of the next isolated natural product is difficult to predict, attempts at nomenclature regularization tend to break down; for example, the whole class of polyketide-derived natural products is structurally extremely diverse and resistant to systematization. Therefore, new microbial products are usually given trivial names by the researchers, then left to the tender mercies of the specialists, such as those at *Chemical Abstracts*, to tidy up by providing them with fully systematic IUPAC names or semi-systematic alternatives. However, for the majority of groups and subgroups, the subject specialists try to bridge the gap between the trivial name and the fully systematic name by evolving nomenclature that ties together the known structures in a way that makes it possible to rationalize and classify them meaningfully. This is an important process; classifying and naming things is an important part of the way in which science evolves, and a good nomenclature scheme enables logical connections to be made between the diverse structures.

Natural product nomenclature should in general always follow the fundamental IUPAC nomenclature principles, such as those governing the order of citation of substituents in a name, however complex the molecule being considered. In general the majority of workers in the field, and journal editors, are familiar with these although at the biological

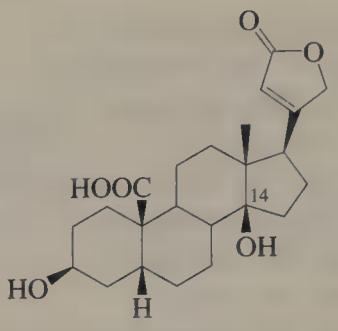


Figure 7.1 3β,14β-Dihydroxy-5β-card-20-enolid-19-oic acid (cannogeninic acid). Non-standard nomenclature of a lactone in natural product chemistry.

'fringes' a considerable number of trivial errors are made by those who may not consider themselves primarily as chemists.

There is one major departure from recommended IUPAC nomenclature rules that is made by almost all those publishing in natural products; a fundamental IUPAC principle is that each name shall have only one principal group. This is broken by the usual naming of molecules containing both a lactone and carboxyl substituent, which are rather common, because of the sheer inconvenience of naming them by strict IUPAC rules. An example is shown in Figure 7.1.

7.2 Trivial names

The first natural products were isolated, wholly or partially purified and at least to some extent characterized, well before the emergence of any well-defined chemical bonding/structure theory in the second half of the nineteenth century. Not surprisingly, the first natural products to attract attention were those with powerful biological properties, (particularly alkaloids, some of which are easy to isolate), and whilst many plant or animal extracts proved to be intractable for a long time (for example, the cardioactive digitalis from *Digitalis purpurea*), others yielded pure substances at an early date. The table below summarizes some early natural product isolations.

Isolate	Worker(s)	Year
Morphine	Derosne, Séguin	1803–1804
Narcotine	Robiquet	1817
Strychnine	Pelletier and Caventou	1818
Brucine	Pelletier	1820
Coniine	Giesecke	1827

Obviously, such substances could only be given **trivial names**, in other words names that imply nothing about the structure, and the practice has continued to the present day of first allocating such a trivial name when a new natural product is isolated. A large number of organic compounds of simple structure whose names have passed into the toolkit of systematic nomenclature were first encountered from natural sources, particularly if the term natural source is given a wide connotation to include sources such as coal tar. The names of, for example, methane and naphthalene are thus in fact trivial names by origin [1] and the boundary today between systematic and trivial names is thus to some extent one of consensus. When a new natural product of rather simple structure and easily memorable name is isolated, such as 2-methylquinoline from the urine of the red fox [2], there is no point in allocating it a trivial name in addition to its short and convenient systematic name. Another reason why aromatic natural products in particular tend to be referred to by their systematic names is that they are biogenetically diverse; structurally similar aromatic compounds may arise in different organisms by different biogenetic pathways (terpenoid, polyketide or shikimate) and there is less scope for emphasizing biogenetic relationships by nomenclature and numbering schemes.

The usual convention is that a trivial name is derived from the Latin binomial for the species from which the natural product is isolated, with the suffix giving some indication of the chemical nature of the compound. Thus, to borrow the terminology used in the IUPAC draft, the hypothetical species *Paradigma exemplare* (family Biespeiliae) might give rise to the basic (alkaloidal) natural products Beilspeiline, Paradigmine or Exemplarine, or the alcohols Beispeilol, Paradigmol or Exemplarol. Deviations from this model are not uncommon, for example the baleabuxidines from *Buxus balearica* [3] and the aflatoxins from *Aspergillus flavus* demonstrate derivation from both parts of the Latin binomial; the number of natural products isolated from intensively studied genera such as *Streptomyces* would make it almost impossible to devise enough different trivial names from this generic.

There are many total exceptions to the recommendation that trivial names derive from binomial names. In some cases the species may be unidentified, or the authors may choose to derive the compound name from a colloquial name for the species. In other cases the natural product may have been isolated from a crude drug of undefined or complex composition. In yet other cases there may be a complete divergence from this time-honoured custom of starting from the biological source (Figure 7.2). The third example given is non-representative and is generally regarded as an aberration.

There is a divergence of opinion between different workers concerning the ongoing function of trivial names. One set of proposals has it that as

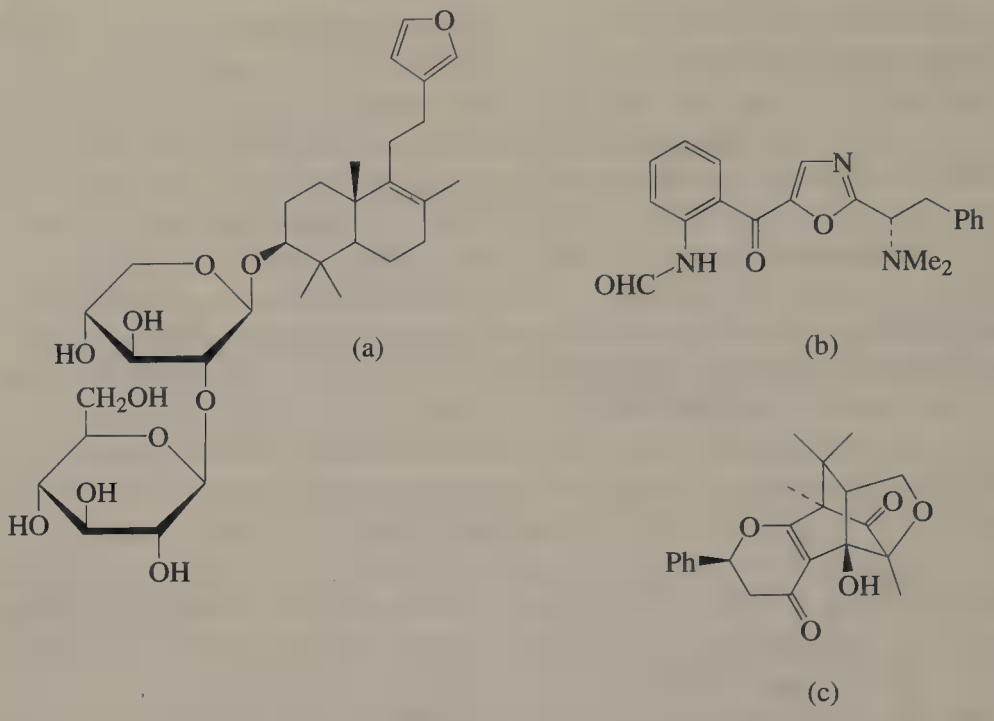


Figure 7.2 Non-standard derivation of trivial names. (a) Baiyunoside (from the Chinese drug Bai-Yun-Shen) [4]. (b) Almazole A (from an unidentified red seaweed collected at Almadies, Senegal) [5]. (c) Louisfieserone (named after Louis Fieser, a well-known organic chemist) [6].

soon as functional groups begin to be identified in the molecule, the trivial name should be modified, e.g. *Paradigmune* → *Paradigmol* to show an OH group, and then when the structure is determined fully a systematic or semi-systematic name should be allocated and the trivial name discontinued.

In the opinion of many dealing with the natural products literature, however, this approach can only cause much confusion in the literature. Since many natural products now being characterized have increasingly complex structures which are more or less intractable in terms of their systematic nomenclature, the trivial name continues to play a vital role in providing an unambiguous 'handle' for the compound which survives any amount of structural uncertainty and changes in nomenclature, and furthermore is meaningful to scientists who are not organic chemists. Accordingly, I and others have proposed an alternative set of recommendations for promulgation by IUPAC, in which the continued use of trivial names is encouraged, together with recommendations about their correct derivation and usage, and the trivial name is carried forward unchanged as the structure is revealed in greater detail.

Another fact that should not be lost sight of is that even in an era of powerful structure determination methods, revisions of previously published

natural product structures are not uncommon. A proportion of the structures currently in the literature, especially for the groups where nature is prolific and researchers typically publish many structures, are undoubtedly wrong [7]. Structure revision poses no particular problems if the trivial name has continued in use; the natural product name now maps onto a new systematic name. An example is provided by the flavonoid Hortensin, (Figure 7.3) which was assigned the structure (a) when first isolated in 1989 [8]; this was revised to (b) in 1992 [9] then again to (c) in 1995 [10].

However, were the trivial name to be discarded, the situation becomes something like 'the natural product formerly considered to be 3,4′-dihydroxy-6,7-dimethoxyflavone, was between 1992 and 1994 regarded as 4′,5-dihydroxy-6,7-dimethoxyflavone but is now known to be 5,7-dihydroxy-4′,6-dimethoxyflavone', with considerable scope for confusion between the natural product and the compounds that really do have these structures, if they are known [11].

7.2.1 Good practice in the assignment of trivial names

The following good practice for the assignment of trivial names should be followed.

(1) A trivial name should be allocated to any natural product that appears to be new. Names for groups of closely related natural products isolated from the same source are traditionally allocated by variations of the alphabetical suffix, e.g. Strychnine, Strychnicine, Strychnidine, or by assigning alphanumeric suffixes, as in Gibberellin A13 [12]. The assignment of suffixes should, for reasons described below, preferably not

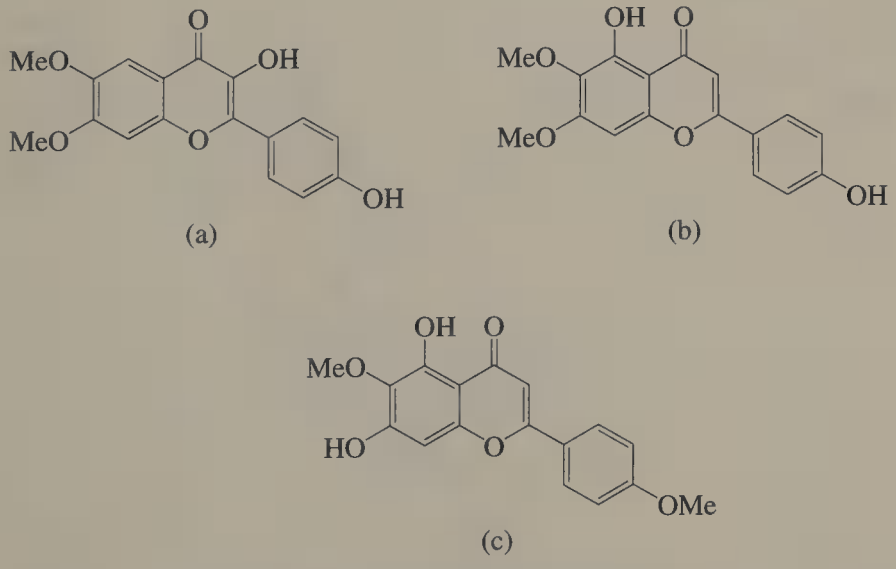


Figure 7.3 Hortensin structures.

attempt to convey structural information. They are usually allocated in order of characterization, but there are exceptions; for instance, in the names of the alkaloids Usambaridines/Dihydrousambaridines Br, Vi and Vc, the suffixes refer to their colour reactions [13]. An example of a series where the suffixes convey structural information is the series of compounds from *Mammea africana* and *Mammea americana* where the suffix in, for example, Mammea A/AA describes the two side-chains attached to the coumarin nucleus [14]. Practice differs as to whether to allocate a trivial name to every new natural product, however straightforwardly derived from a known parent. Where the new natural product is a simple acetate or glycoside of a known product, it is frequently just named, for example, as '*O*-acetyloleanolic acid' or 'luteolin 3-*O*-glucoside'. However, even this practice has its drawbacks; application of this method of nomenclature to more profound modifications of the structures arrives at the undesirable 'semi-trivial' names described below.

(2) Once assigned, the trivial name should not be altered. Retrospective amendments are sometimes made by the discoverers in order to regularize some point of chemistry, but this is almost invariably counterproductive [15].

(3) A search of the literature should be carried out to ensure that the proposed name (**and structure**!) is indeed new. This seems obvious, but there are for example three Orientins, four Obtusin(e)s and no less than six Odoratin(e)s in the literature. Of the latter, one is a piperidine alkaloid, one a sesquiterpene, and one a nortriterpenoid, but the other three, (a)–(c), are all flavonoids and two of them are isoflavones of different structure isolated by different workers from the same plant, *Dipteryx odorata* (Figure 7.4) [16–18].

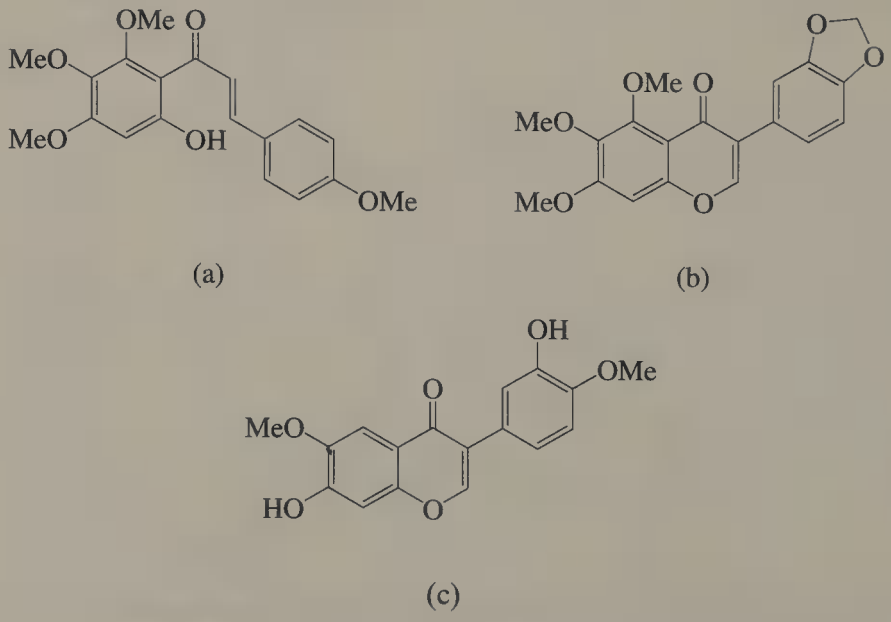


Figure 7.4 Odoratin(e) flavonoid structures.

(4) The name should be assigned at the time of publication, not at the point of isolation. A few authors cause an undesirable proliferation of names by the practice of assigning them to all compounds isolated from a natural source before going on in the same paper to show that a number of them are identical with natural products already known. Such practice (as well as the apparent failure to search the literature diligently to ensure that a compound has not been isolated before) partly accounts for the fact that Oleanolic acid, for example, has at last 16 other obsolete trivial names.

7.3 'Semi-trivial' names; a class that should be discouraged

Many authors prefer to tie a newly isolated natural product to one already known by the expedient of naming it in a way that shows the relationship. In very simple cases such as 3-*O*-acetyloleanolic acid referred to above this is relatively harmless, but pushing the technique further is undesirable and can cause confusion in the literature. Three examples are given in Figure 7.5.

The reasons for potential confusion are at least four:

(1) The semi-trivial name selected by the authors may be an inefficient and roundabout way of describing the structure, since it may be referred by them to a trivially named compound that is not very closely related. Different authors may thus name closely related structures by different routes.
(2) The stem component may itself have more than one trivial name.

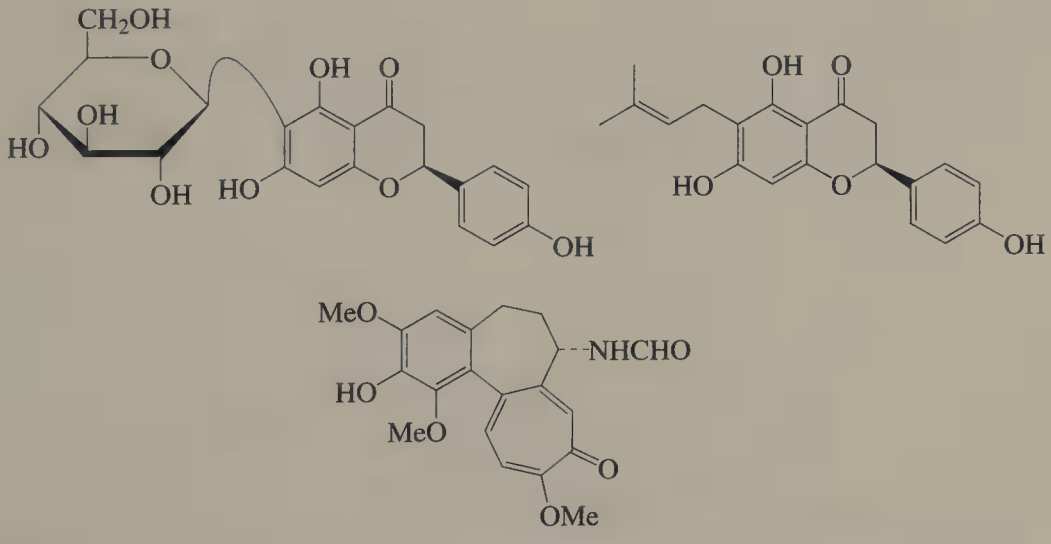
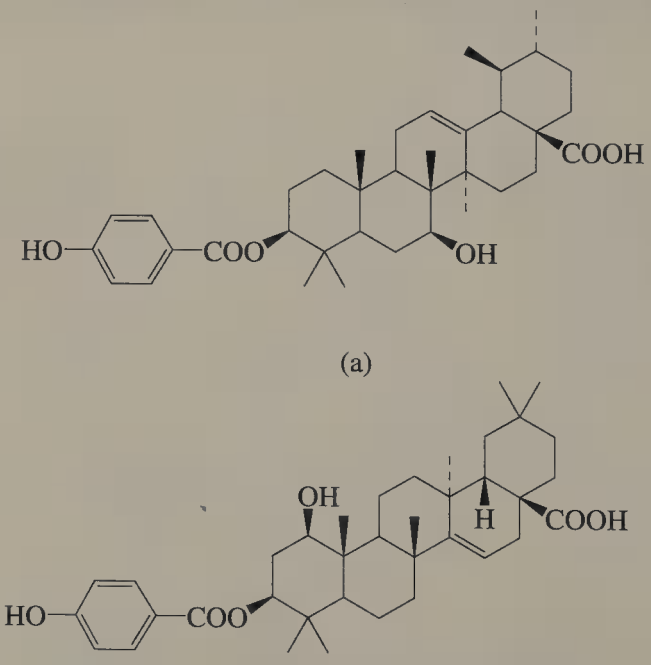


Figure 7.5 Some semi-trivial names. (1) 6-*C*-Glucosylnaringenin. (2) 6-Prenylnaringenin. (3) 2-Demethyl-*N*-deacetyl-*N*-formylcolchicine.

(3) There is no standardization between different names for the modifying radicals or operands. Even 3-*O*-acetyloleanolic acid can also be called oleanolic acid 3-acetate, but for the examples given above the possibilities for variation are much greater. In Figure 7.5(1) they include 6-glucosylnaringenin, 6-glucopyranosylnaringenin and 6-*O*-β-D-glucopyranosylnaringenin (explicit specification of structural features at different levels of precision; colloquially, glucosyl residues are normally considered to be β-D-pyranosyl unless stated otherwise), as well as naringenin 3-glucoside, naringenin 3-glucopyranoside and naringenin 3-*O*-β-D-glucopyranoside. The substituent in Figure 7.5(2) has been variously called γ,γ-dimethylallyl, 3,3-dimethylacrylyl, isopentenyl, isoprenyl (most unfortunately) and the systematic 3-methyl-2-butenyl as well as the now more generally accepted prenyl; furthermore, the trivial name for Figure 7.5(3) allocated by the authors violates good IUPAC practice in the alphabetical ordering of the substituents (deacetyl should precede demethyl); there are many such annoying trivial errors and variants in the literature for this sort of name.

(4) Probably the most serious drawback is the possibility for confusion caused by structure revisions. A good example is provided by a terpenoid $C_{37}H_{52}O_6$ isolated from *Maprouana africana* (Figure 7.6). The original workers [19] assigned the structure (a) and called it '7β-hydroxymaprounic acid 3-*p*-hydroxybenzoate'. In 1989 the structure of maprounic acid was revised from an ursane to a taraxerene [20] and



(a)

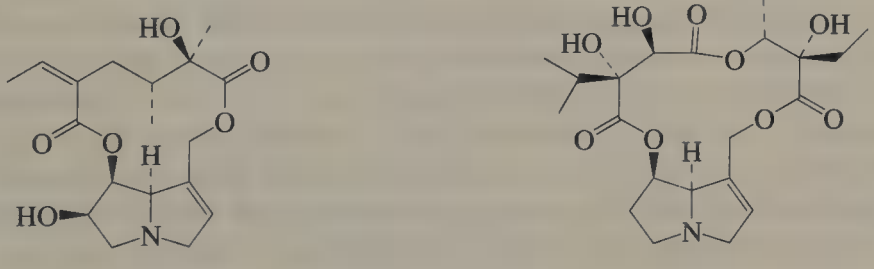
Figure 7.6 Terpenoid isolated from *Maprouana africana*.

was found to be identical with the previously known aleuritolic acid. In 1995, the structure of the C_{37} compound was reinvestigated and was shown not only to have the taraxerene skeleton, but to have the OH group on C-1 not C-7 (structure (b)). Therefore the authors renamed it '1β-hydroxyaleuritolic acid 3-*p*-hydroxybenzoate' [21] It would have been much simpler to have given the compound a trivial name when first isolated and the authors would then merely have to report that the structure of 'maprounemic acid' (for example) had been revised from 3β,7β-dihydroxy-14-ursen-29-oic acid 3-*p*-hydroxybenzoate to 1β,7β-dihydroxy-14-taraxeren-29-oic acid 3-*p*-hydroxybenzoate.

The attitude of IUPAC and CAS to this type of name should be noted. IUPAC sanctions the use of **functional parents**, which are defined as 'structures that have certain terminal hetero atoms or groups, such as are found in carbohydrates, amino acids and nucleosides'. However, functional parents should preferably not be used in the open literature outside well-understood and IUPAC-sanctioned domains such as the three mentioned.

CAS uses names derived from functional parents for complex natural products that cannot readily be named systematically, and for which no semi-systematic skeleton (see below) is in use. For example, among the macrocyclic pyrrolizidine alkaloids (Figure 7.7), the majority are named semi-systematically from the key parents Senecionan and Crotalanan; thus Anacrotine (a) is 6,12-dihydroxysenecionan-11,16-dione. But Parsonsianine (b), which falls into the rare and recently discovered Parsonsine subgroup, is named by CAS as 16-hydroxy-22-norparsonsine.

The tendency in recent years has been for CAS to introduce many more functional parents even for relatively small molecules, so that it is essential to consult the Index Guide in detail to determine CAS indexing policy for a particular group. This point can be appreciated by comparing the CAS index names for biennin C and bicyclomycin (Figure 7.8). The former is indexed under the cumbersome systematic Spiro[furo[2′,3′:5,6]cyclohepta[1,2-*c*]pyran-3(2*H*),7′(8′*H*)-[6,8*a*]methano[8*aH*]benzo[4,5]cyclohepta



(a) (b)

Figure 7.7 Macrocyclic pyrrolizidine alkaloid structures.

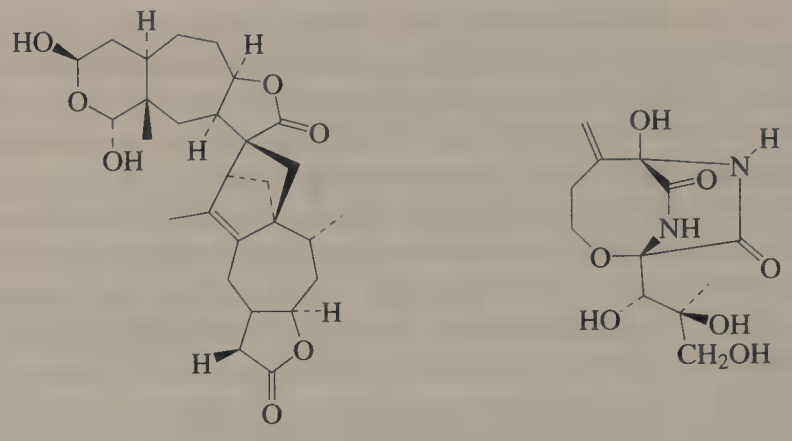


Figure 7.8 Structures of biennin C and bicyclomycin.

[1,2-*b*]furan]-2,2′(3′*H*)-dione, 3*a*,3′*a*,4,4′4*a*,5,6′,7,8,8*a*,9,9′10,10′,10*a*,10′*a*-hexadecahydro-5,7-dihydroxy-3′4*a*,5′,9,9′-pentamethyl, [3*R*-[3α(3′*S**, 3′*aR**,6′*S**,8′α*S**,9′*R**,10′*aR**),3*a*α,4*a*β,5α,7β,8*a*α,9α,10*a*α]] whereas the CAS name for the smaller molecule (b) is bicyclomycin.

The conclusion has to be that in the hands of nomenclature experts 'semi-trivial' nomenclature is a necessary evil, but that it should be used cautiously if at all in the primary literature. Authors should distinguish between stating colloquially that 'Compound X was found to have the structure of 6-prenylnaringenin' and formally naming it as such.

7.4 Biogenetic numbering and the natural product specialists' semi-systematic schemes

For many groups of natural products there is a bifurcation in the path between that followed by nomenclature specialists, such as those working at CAS, and the subject specialists. The former group have the need to work in accordance with IUPAC principles to produce rational and consistent names for the whole gamut of chemical structures using systematic or semi-systematic nomenclature, whereas the subject specialists seek a scheme giving the greatest possible degree of consistency and understandability within the domain of their particular group of natural products, the membership of which, as already noted, is determined by biogenetic constraints. The practice of the two groups overlaps in the semi-systematic area, but even here there are many cases where CAS have needed to 'crystallize' the nomenclature/numbering of a particular skeleton in a way that the subject specialists at the time or subsequently consider undesirable, for example biogenetically misleading [22]. IUPAC are currently examining proposals for the regulation of this kind of semi-systematic numbering.

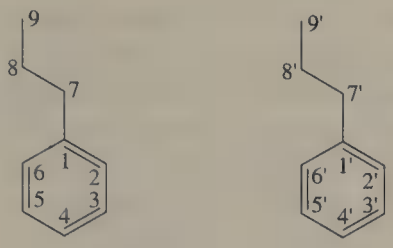


Figure 7.9 Lignan fragments.

To see how this works in practice, let us consider the nomenclature of a medium-sized group of plant natural products, the lignans [23]. Biogenetically these are derived from the oxidative coupling of two (or more) C-9 cinnamate-derived moieties. In order to think about and classify them, the subject specialists number the two C-9 fragments as shown in Figure 7.9; note that as this stage there is no question of 'correct' or 'incorrect' nomenclature because we are talking about the numbering scheme for a notional framework without any functional groups or bond orders.

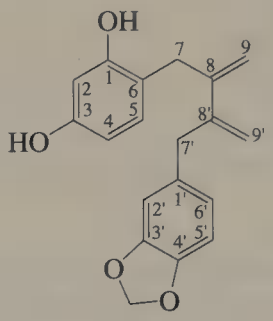
The most widespread and earliest-studied examples, the lignans proper, have the two units joined by an 8,8′ bond (also called a ββ- bond) followed by various cyclizations, but in more recent years a wide variety of so-called neolignans have been isolated in which virtually every other biogenetically reasonable type of dimerization is known, including dimers formed through oxygen bridges. (The neolignan skeletons known up to 1977 are tabulated by Gottlieb [24] but many more types have since been discovered). Combining these various possible dimerizations with a large number of secondary cyclizations and other processes produces an extensive and superficially rather heterogeneous class of natural products, which numbered about 1500 members up to 1996 (figures from the *Dictionary of Natural Products* database).

A small representative selection of lignans and neolignans Anolignan A, Liovil and Americanin D is shown in Figure 7.10 together with their systematic names and numbering according to current *Chemical Abstracts* practice, and their 'lignan-specialist' names and numbering according to the scheme developed by Moss [25] and now recommended by IUPAC.

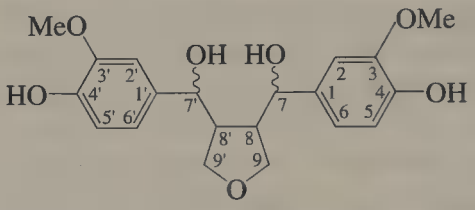
From these examples, it can readily be seen that the systematic names give no indication of the underlying structural/biogenetic relationships within the group and that using systematic nomenclature, the numbering varies irregularly and disguises the biogenetic and other relationships. The response of the active workers in such fields, as with the lignans, is to evolve a scheme that emphasizes these.

There is no ready solution to this 'conflict' between the requirement of the natural product worker to rationalize relationships within and between

Systematic (CAS) name	*Lignan name*
1,3-Benzenediol, 4-[3-(1,3-benzodioxol-5-ylmethyl)-2-methylene-3-butenyl]-	2,4-Dihydroxy-3′,4′-methylenedioxy-8,8′-ligna-8,8′-diene



3,4-Furandimethanol, tetrahydro-α,α′-bis(4-hydroxy-3-methoxyphenyl)-	4,4′,7,7′-Tetrahydroxy-3,3′-dimethoxy-8,8′-epoxylignan



2-Propenal, 3-[2-(3,4-dihydroxyphenyl)-2,3-dihydro-7-hydroxy-3(hydroxymethyl)-5-benzofuranyl)	3,8′-Epoxy-3′4′,5,9-tetrahydroxy-3,8′-lign-7-en-9-al

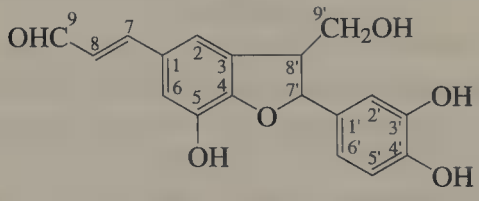


Figure 7.10 Some lignan structures.

different natural product groups, and the urge of the nomenclature experts to standardize nomenclature across a wider canvas. Once it is realized that both types of name fulfil a purpose, the 'conflict' merely becomes an inconvenience.

To summarize, a given natural product may have at least three names:

(a) A trivial name
(b) A fully systematic name
(c) A semi-systematic name emphasizing biogenetic and/or structural relationships with other natural products.

For a particular natural product, (a) and/or (c) may be absent but owing to factors such as differing conventions or simultaneous publication may

also be duplicated or multiply occurring. For a particular group of natural products, *Chemical Abstracts* may index under names falling into any of the classes but in the case of (c) their scheme may not be widely accepted.

In the information age, this trifurcation causes less strain on the memory-banks of the average chemist than might have been the case in the past, because readily searched information systems are there to provide a mapping between these different sets of names.

7.5 Systematic nomenclature

The systematic nomenclature of natural product types is not considered in detail in this chapter; it raises no general aspects that are not covered by other chapters in this book although of course many problems of systematic nomenclature arise when individual natural products or groups are looked at in detail. Many natural products are named systematically in *Chemical Abstracts* and it can be difficult even for a nomenclature specialist to remember or predict which groups are named systematically in CAS and which semi-systematically.

In consulting CAS it is important always to bear in mind that what to the natural product chemist appears a minor structural change between compounds of similar biogenesis and essential ring structure frequently causes profound changes in their CAS indexing. The common methylenedioxy substituent is a frequent cause of 'losing' compounds in CAS since according to the systematic scheme the methylenedioxy function becomes part of the defining ring system. Another major cause is the 'uncovering' of a principal group such as phenolic OH in one of a set of homologous compounds. An example is Dauricine in Figure 7.36(c).

7.6 IUPAC semi-systematic names

Semi-systematic names are normally derived by the creation of a semi-systematic parent. It is variations in the choice and construction of semi-systematic parents that cause many of the problems with natural product nomenclature, and it is perhaps unfortunate that IUPAC in this field tends towards the descriptive rather than the prescriptive, thus allowing considerable variation between different authors in the primary literature and other bodies such as CAS. However, the heterogeneity in structure between different new natural products, and the need to publish names quickly, makes it difficult to see how the whole system, as opposed to retrospective improvement of individual groups, could be made so as to allow timely publication of consistent names by authors around the world.

It should also be remarked that in order to be able to choose a fundamental parent structure it should reflect the basic skeleton that is common to most compounds in that class, which implies that there is already a 'class' awaiting nomenclature rather than a single new natural product of unique structural type. However, it is certainly the intention of IUPAC that authors in the primary literature should be free to bring in new fundamental semi-systematic parents when it seems appropriate.

The majority of semi-systematic skeletons in use are hydrocarbons, and although a limited number of hydroaromatic skeletons such as spirostan and cardenolide have become well-sanctioned by custom, heterocyclic skeletons are in general not desirable and the introduction of new ones should be avoided. Note that Spirostan, since it is not a hydrocarbon, does not have a terminal -e (Figure 7.11).

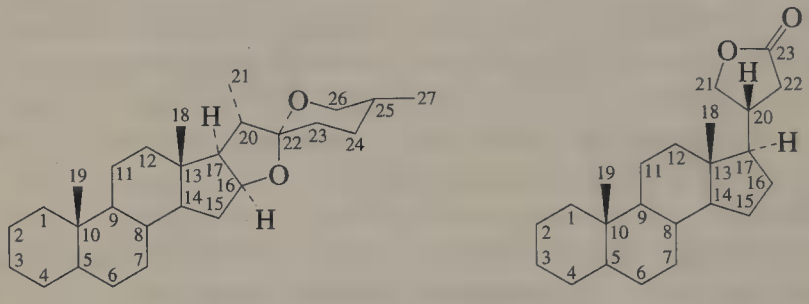


Figure 7.11 Spirostan and cardanolide parents.

7.6.1 Modification of fundamental parent structures

If a natural product contains one of the established fundamental parent skeletons, it can be straightforwardly named using the stem name modified by the usual conventions of organic nomenclature; for example, Veratramine = 14,15,16,17-Tetrahydroveratraman-3,23-diol. In many cases the skeleton will be similar to, but not identical with, one of the established parent skeletons in which case there are well-established and IUPAC-approved modifiers (operands) for changes to the skeleton. The majority of these are straightforward and are listed in Table 7.1. Modifiers can refer either to operations on certain numbered carbon (or hetero-) atoms of the skeleton, or in some cases can refer to operations on a certain ring, lettered A,B.... In each case a convention needs to have been established concerning the skeleton numbering or ring lettering. Some of the parent skeletons have well-established ring lettering conventions (e.g. steroids), others do not.

These prefixes may be used in combination to effect more far-reaching modifications of a starting structure. There may be more than one combination for producing the required result, in which case the preference should

Table 7.1 Skeletal modifications of fundamental parent structures

Operand	Operation	Notes
nor-	Removal of an unsubstituted skeletal atom (from a ring or acyclic part)	Removal of a heteroatom is allowed. A locant no. is required; the highest possible number should be chosen for removal. In CAS usage this may be a ring descriptor (e.g. *A*-Noroleanane). The plural of nor- is dinor-.
homo-	Addition of a methylene group between two skeletal atoms (in a ring or acyclic portion)	Locant no. of the introduced carbon must be given (e.g. 10*a*-homotaxane). In CAS usage may be a ring descriptor (e.g. *B,C*-Dihomocholestane). Special notation is used for homologation at a 'bond connector' position, e.g. in 8(9)*a*-homo-5α-androstane.
cyclo-	Creation of an additional ring by means of a direct link between any two atoms of a parent structure	It may be necessary to specify the configuration, as in 16β*H*-1,16-cyclocorynan. Formation of two rings is indicated by the plural prefix dicyclo-.
seco-	Cleavage of a (saturated or unsaturated) ring bond	The original numbering is retained
apo-	Removal of all of the side-chain beyond the indicated locant	Carotenoids only (see below)
X(Y→Z) abeo	Migration of one end of a single bond (locant X) from Y to Z.	Numbering and stereochemistry of the starting skeleton is retained. Introduction of further stereodescriptors may be necessary.
retro-	Shift by one position of all double and single bonds of a conjugated polyene system	Carotenoids only. Infrequent use. The prefix retro- has also been used to indicate 9β,10α-steroids but this usage is not recommended. It is also used to denote peptides with a reversed amino-acid sequence.
des-	Removal of a terminal ring	Numbering and stereochemistry are retained

be for using the smallest possible number of prefixes, or if more than one name is possible using the same number of prefixes, by favouring homo/nor over cyclo/seco. The order of citation of multiple prefixes is chosen so as to avoid invalid operations; for example, 8(14)*a*-Seco-8(14)*a*-homopregnane not 8(14)*a*-Homo-8(14)*a*-secopregnane because the homo-operation cannot validly be carried out on 8(14)*a*-secopregnane. Subject to this limitation, two alternative protocols are in use for the order of the modifying prefixes; either alphabetically ignoring multipliers, according to normal IUPAC nomenclature rules, or alphabetically within two groups with the bond rearrangement operators abeo, cyclo, retro, seco- preceding removal/addition operators apo, de, des, homo, nor. In either case, these prefixes are preceded by any replacement 'a' prefixes (oxa, aza, etc.) if present.

Since certain of the parent skeletons used in natural product structures are heterocycles not carbocycles, the rules allow use of the 'a' modifiers even when the starting structure is already a heterocycle, and also sanction

the use of 'carba' to denote replacement of a heteroatom in the parent structure by carbon. (If the starting heteroatom is unnumbered, the new carbon atom is numbered using an 'a' suffix to the immediately adjacent lower numbered skeletal atom). The notation can also be used to denote replacement of one heteroatom by another, as in 1-Thiaergoline (replacement of N by S). These prefixes form a non-detachable part of the stem name, for example 5,6-dihydroxy-5,6-dihydro-10′-apo-β,y-caroten-10′-al is correct, not 10′-apo .....

It is possible to further elaborate these widely accepted semi-systematic parent skeleton names by using the techniques of aromatic/heterocyclic fusion nomenclature in order to describe variant structures (whether new natural products or semi-synthetic compounds). An example taken at random from the CAS Index Guide is Estr-1-eno[2,1-*d*]thiazole to describe steroids having an additional thiazole ring fused to C1–C2. However, the construction and correct numbering of such hybrid skeletons is complex and clear IUPAC guidance has not yet been published. Authors introducing new parent structures of this type run the risk of deviating from the name chosen by CAS, Beilstein and other bodies, and in general their promiscuous introduction should be discouraged.

Because the semi-systematic natural product parents are at various levels of unsaturation, several kinds of modifier are necessary to denote bond order modification.

(i) If the parent structure is fully saturated (ending in -an, -ane or -anine), dehydrogenation is indicated by replacement of these endings by -en,, -adien, yne, etc., as in Pregn-4-en-20-yne. Note that the semi-synthetic parent does not have to be fully saturated, only saturated in the relevant portion as in 1*H*-Coryn-16-ene (Figure 7.12a). Note also the potential for confusion in the case of 5-Conenine, derived from the saturated parent which is called Conanine (Figure 7.12b).

(ii) The introduction of additional unsaturation into a parent structure not ending in -an, -ane or -anine or conversion of a double bond to triple are denoted by the prefix dehydro- accompanied by a multiplying prefix and appropriate locants, e.g. 9,10-Didehydroergoline (Figure 7.12c).

(iii) The dehydro- operand may encompass rearrangement of double bond position, as in the example of 3,8-didehydrosenecionan (Figure 7.12d; cf. Figure 7.7).

7.7 Stereochemical considerations

The range of stereochemical variation shown by natural products of different type is extensive and a full discussion of the implications of this

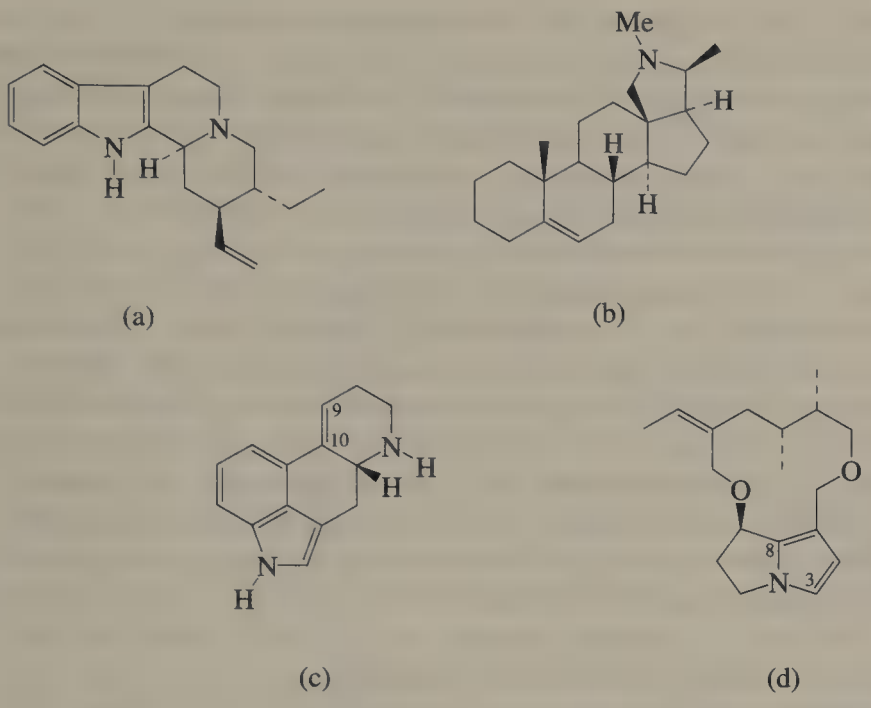


Figure 7.12 Unsaturation in semi-systematic parents.

in their nomenclature would be out of place. Here we touch upon some general points.

(1) Allocation of a trivial name to a natural product traditionally implies a certain relative stereochemistry in the case of molecules with two or more chiral centres (or other stereogenic features), even though at the time of allocation this relative configuration may not be known. Known diastereomers are frequently distinguished by use of the stereoeprefixes *epi-*, *allo-* and *epiallo-* as well as the more historical prefixes *pseudo*(ψ-) and others; practice has varied as to whether these prefixes are italicized as in *pseudo*-ephedrine or form an unitalicized part of the stem name as in pseudoephedrine; the latter is now more accepted.

(2) There is no general rule as to whether a trivial name once allocated is applicable to both enantiomers of the molecule and the racemate, or whether it strictly refers only to the enantiomer isolated. To take an example, the (-) enantiomer of the compound shown in Figure 7.13 was isolated from *Myoporium laetum* (1925) and named Ngaione, whilst the (+) enantiomer was isolated from sweet potatoes infected with

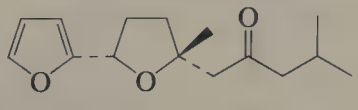


Figure 7.13 Ngaione/Ipomemarone ((-)-enantiomer illustrated).

Ceratostomella fimbriata (1943) and named Ipomeamarone [26]. There is no general agreement as to whether one of these names (presumably by chronological precedence) should supersede the other or whether both names should continue in use. I prefer the former solution, on the grounds that under the second scheme the racemate has two equally valid names. However, Ngaione/Ipomemarone is an example where the workers in the field seem to have taken by consent the second path.

(3) A fundamental parent structure should include as much stereochemistry as possible that is common to the relevant class of natural products, and the name of the fundamental parent implies the absolute configuration at all chiral centres as well as double bond configurations, except in some cases where certain chiral centres are variable and have to be specified. If a compound has the reverse configuration at one or more stereogenic centres, this is specified. Thus, in the name (5α,9α)-pregnan-3α-ol, the three alphas perform different functions; the 3α- because it specifies the configuration of a substituent, the 9α- because it is the unusual configuration for the C-9 centre in pregnane, and the 5α- because the pregnane stereoparent denotes unspecified configuration at C-5.

(4) Configurations at stereocentres in planar or quasi-planar ring systems are denoted by the symbols α, β- and ξ-(for unknown). This implies an agreed orientation for the ring system. Where there is any doubt about the correct orientation of the ring system, where the stereogenic centre is attached to a chain that is free to rotate (cf. Vitamin D derivatives below), or where there is a bridged ring system that may make the display of the correct absolute configuration and/or its accurate description by the α,β-system unreliable, the Cahn–Ingold–Prelog (*R,S*)-system should be preferred. A good example of potential uncertainty caused by bridged rings is provided by diterpene alkaloids containing the atidane skeleton when substituted at C-13, C-14, C-19 or C-20; the α,β-system is ambiguous and structure diagrams in the literature are frequently unclear, therefore the (*R,S*)-system should be used to specify these configurations (Figure 7.14).

(5) Configurational inversion at one chiral centre is denoted by the prefix *epi-* (separable italicized prefix) or epi- (inseparable prefix) plus a

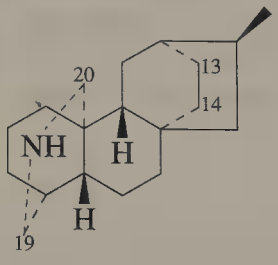


Figure 7.14 Atidane skeleton.

locant. The latter form seems to be becoming more usual. The plural of epi- is diepi-.

(6). There are a number of cases in which different parent skeletal names are used where the only variation between them is stereochemical (either epimerism, enantiomerism, or dehydrogenation leading to fewer centres of chirality). Some of these are listed in Table 7.2.

(7) Care needs to be taken in retrieving *Chemical Abstracts* information for compounds substituted at a gem-dimethyl site which is not a stereogenic centre in the parent skeleton. Usual practice among natural products chemists is to differentiate between such methyl groups in the numbering scheme, the α-Me invariably being assigned the lower number. CAS practice is to treat the two methyl groups when unsubstituted as indistinguishable, assigning the substituent to the lower-numbered methyl group and adding a stereochemical descriptor to indicate the stereochemistry at the newly generated chiral centre (Figure 7.15). Kauranes showing this phenomenon are particularly tricky (Figure 7.16).

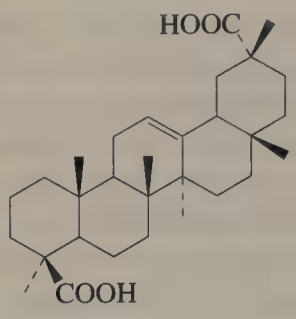
7.7.1 The ent-*convention*

A stereochemical convention introduced by Klyne and employed only for natural products (mostly terpenoids but a few others, e.g. catechins) is the *ent-* convention. Use of this prefix implies inversion of the configuration at *all* chiral centres in the molecule so that, for example *ent*-6α,13*R*-dihydroxy-5β-labdan-15-al is the enantiomer of 6α,13*R*-dihydroxy-5β-labdan-15-al.

Although the convention is relatively easy to apply for particular compound provided the user is clear about the rule, it is very easy to make mistakes when discussing transformations, e.g. to forget that to go from *ent*-5β-Labdan-3α-ol to *ent*-5β-Labdane-3α,13*R*-diol requires addi-

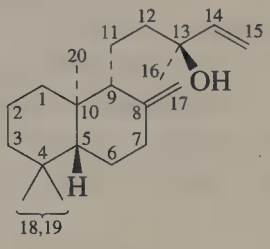
Table 7.2 Examples of the application of more than one name to constitutionally identical or almost identical carbon skeletons

Sesquiterpenoids	Acoranes–Alaskanes	enantiomerism
	Cadinanes–Muurolanes–Bulgaranes– Amorphanes	epimerism at two stereogenic centres
	Cadinanes(etc.)–Calamenenes– Cadalenes	dehydrogenation
	Bicyclogermacranes–Lepidozanes	epimerism
Diterpenoids	Clerodanes–Neoclerodanes (kolavanes)	enantiomerism (obsolescent)
	Pimaranes–Isopimaranes	enantiomerism
	Kauranes–Phyllocladanes	epimerism
	Bifloranes–Serrulatanes	dehydrogenation
Triterpenoids	Ursanes–Taraxastanes	epimerism
Steroids	Ergostanes–campestanes	epimerism
	Stigmastanes–poriferastanes	epimerism



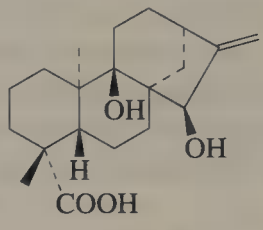
Olean-12-ene-24,29-dioic acid =
Olean-12-ene-23,29-dioic acid
(4β,29α)(CAS)

Figure 7.15 Nomenclature of an oleanane substituted at the *gem*-dimethyl groups.



(a)

ent-8(17),14-Labdadien-13*R*-ol



(b)

ent-9,15α-Dihydroxykaur-16-en-19-oic acid
= Kaur-16-en-18-oic acid, 9,15-
dihydroxy (4α,15β)(CAS name)

Figure 7.16 Application of the ent-convention.

tion of a 13*S*-hydroxy group, and there are numerous errors in the literature. The situation is particularly confused in the case of Kauranes (and some other smaller diterpenoid groups) where CAS (presumably on the grounds that nearly all naturally occurring kauranes are in fact *ent*-kauranes) calls *ent*-kaurane kaurane (Figure 7.16) The terpene chemists' name *ent*-kaurane for this skeleton is based on stereochemical dishomogeneity at the key chiral centre C-10 with other terpenoids of the normal or 10β-series [27].

In the *Dictionary of Natural Products* data structure, all stereoisomeric forms, including *ent*-forms, of a given constitutional structure are collected together as stereoisomeric variants within the same entry. The entry names do not contain any *ent*-prefixes . . . Thus, for example the entry for 8(17),14-Labdadien-13-ol contains information on the four variants (13*R*)-form (Manool), (13*S*)-form (13-Epimanool), (*ent*-13*R*)-form (*ent*-Manool) and (*ent*-13*S*)-form (*ent*-13-Epimanool). This considerably reduces the risk of confusion.

7.8 Review of natural product classes

7.8.1 Aliphatic compounds; lipids

Two local conventions are in use for specifying fatty acid structure and stereochemistry. The first of these is a shorthand notation for describing straight-chain fatty acid structures and takes the form A:B(C), where A indicates the chain length, B represents the number of centres of unsaturation and C the position and configuration of unsaturation. Thus, oleic (*cis*-9-octadecenoic) acid is 18:1(9*Z*).

The second convention uses the symbol ω*x* or *n*-*x* to denote the indicate the position of the centre of unsaturation closest to the tail end of the molecule, it being assumed that all other double bonds are methylene-interrupted (1,4,7 . . . . . unsaturation) and *Z*-. According to this convention, arachidonic (all-*cis*-5,8,11,14-eicosatetraenoic) acid is 20:4(ω3) or 20:4(n-3); this a clear case of the form of a convention being driven by the biosynthetic possibilities, since all of the nutritionally significant natural fatty acids are methylene-interrupted all-*Z* polyunsaturated.

The systematic nomenclature of the fatty acids is easy, but the nomenclature of the glycerides is more complex. CAS names all of these systematically, which means that they become indexed at the carboxylic acid of highest precedence and not as glycerol (1,2,3-propanetriol) derivatives. Lipid chemists generally name them as glycerol derivatives but in two different ways, either as glycerol alkanoates or as alkanoylglycerols, as illustrated in the examples below (Figure 7.17). Whichever method is used, if two or three alkanoyl residues are present they should be ordered alphabetically in accordance with good IUPAC practice (Figure 7.17b); there are many incorrect names in the literature.

Figure 7.17(c) also illustrates the Stereospecific Numbering (*sn*-) convention for glyceride configurations. By convention, the Fischer projection for the lipid is drawn with the C-2 *OH* to the left so that the enantiomer of 17(c) is *sn*-glycerol 3-butanoate 2-hexadecanoate 1-octadecanoate. This convention has the advantage of preserving the numbering through minor transitions involving changes to the C-1 and C-3 R groups which might cause configurational inversion according to the Cahn–Ingold–Prelog system. However, for describing the absolute configurations of individual glycerides, the (*R*,*S*)- convention is probably clearer and more foolproof.

7.8.2 Aromatic, heteroaromatic and reduced aromatic systems

As already stated, a detailed review of these will not be given. Trivial names are in widespread use among natural products chemists for a number of the simpler heteroaromatic nuclei (e.g. chromene for 2*H*-1-benzopyran,

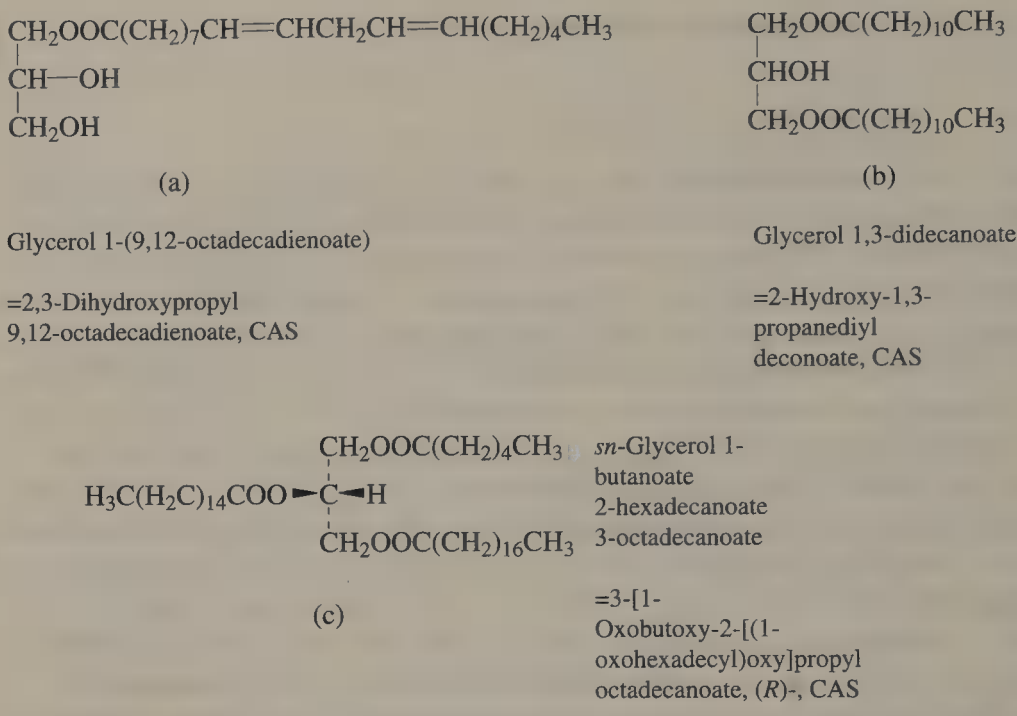


Figure 7.17 Glyceride nomenclature.

coumarin for 2*H*-1-benzopyran-2-one), but since the numbering does not differ between the trivial and systematic forms there is little room for difficulty.

Shikimic acid and quinic acid deserve particular comment. Two conventions for the ring numbering of quinic acid are in almost equal usage despite IUPAC/IUB recommendations going back many years. C-3 can be defined as the carbon atom bearing the OH group either *cis*- (Beilstein) or *trans*- (IUPAC, CAS) to the carboxyl group (Figure 7.18) [28]. With shikimic acid and its stereoisomers the position should be clearer; the numbering should be determined by the position of the double bond, but unfortunately the opposite numbering is sometimes encountered. Biogenetically speaking, formal dehydration of quinic acid can of course go

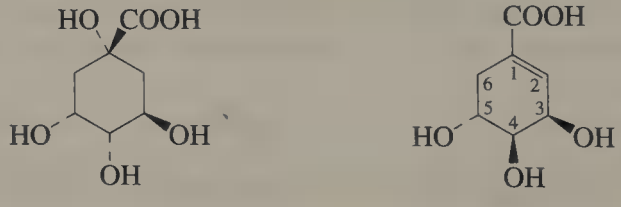
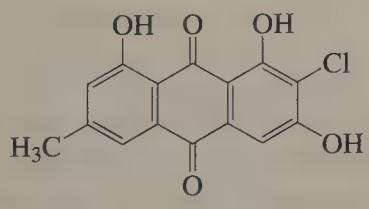


Figure 7.18 Shikimic and quinic acids.

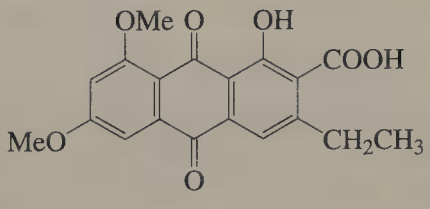
either way round the ring although in practice dehydration takes place towards the *trans*- OH group [29].

Among the naturally occurring polycyclic aromatic systems, natural product chemists are guilty of two departures from IUPAC good practice in the naming and numbering of the anthraquinones as a class. The majority of natural anthraquinones have a C-2 alkyl substituent (but not all; for example, anthraquinone itself is a natural product) and some but not all workers number simple anthraquinones so as to preserve C-2 as the alkyl group. An example is 2-chloro-1,3,8-trihydroxy-6-methylanthraquinone (Figure 7.19a) which has been called 6-chloro-4,5,7-trihydroxy-2-methyl ..... This is dying out and should be discouraged. A more permissible practice is to treat anthraquinone as a stem name so as to maintain nomenclatural homogeneity even for carboxyl-substituted compounds so that, for example, Austrocorticinic acid (Figure 7.19(b) becomes 3-ethyl-1-hydroxy-6,8-dimethoxyanthraquinone-2-carboxylic acid rather than the strictly correct IUPAC/CAS name 3-ethyl-9,10-dihydro-1-hydroxy-6,8-dimethoxy-9,10-dioxo-2-anthracenecarboxylic acid.

The nomenclature and numbering of the important anthracycline group requires care (Figure 7.20). Most antibiotic workers preserve the underlying biogenetic/structural relationships by treating them all as 5,12-naphthacenediones. The parent naphthacenedione nucleus can be numbered in either of two directions owing to symmetry and the choice is made by making C-9 the locant of the attached alkyl group. This numbering is shown in Figure 7.20(a) for Cytorhodin A. This molecule is a stereoparent in CAS having the numbering illustrated (the numbering of this stereoparent in CAS conforms with the biogenetic numbering usually found in the literature but is not according to IUPAC numbering principles since there is an OH group on C-4 which could be numbered C-1). However, many anthracyclines contain a dominant carboxyl-related substituent as illustrated in Figure 7.20(b) for Pyrromycin. These compounds CAS names systematically as derivatives of 6,11-dioxo-1-naphthacenecarboxylic acid, so that C-10 becomes C-1. However, there are also anthracyclines which lack this



(a)



(b)

Figure 7.19 Anthraquinones.

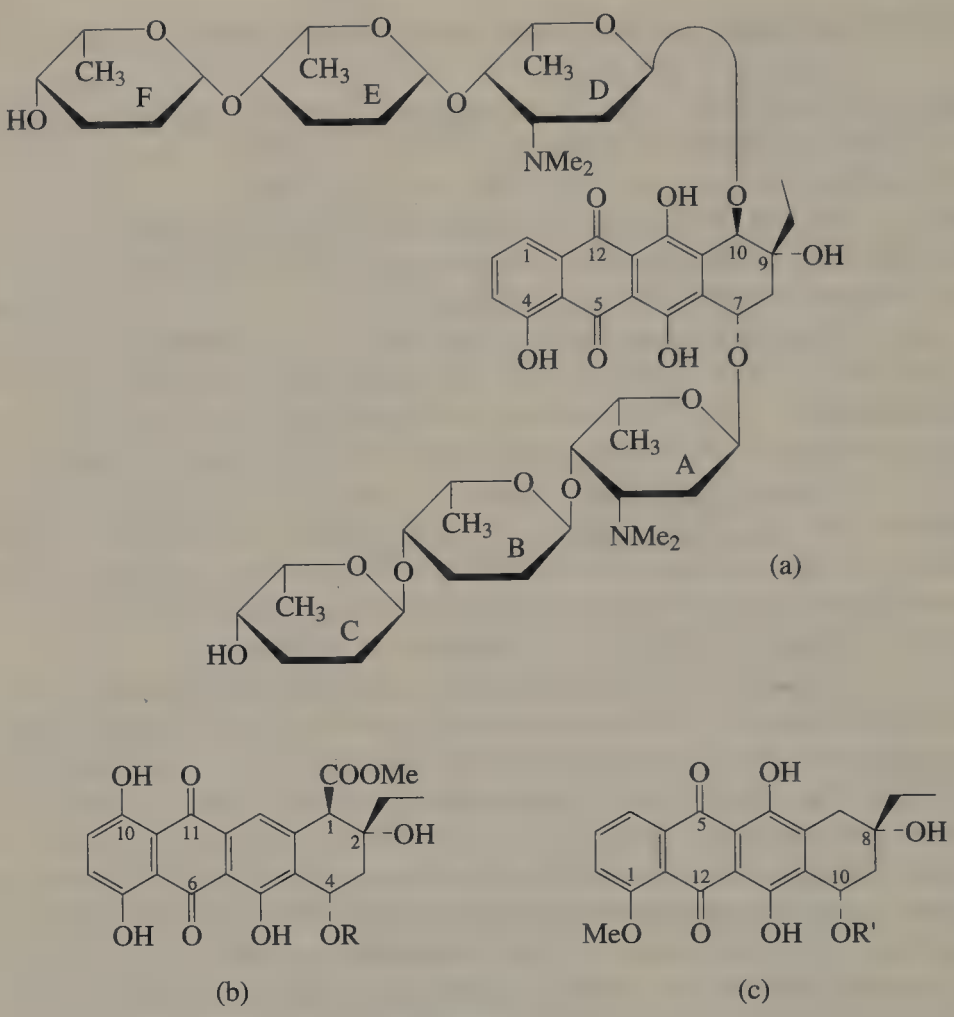
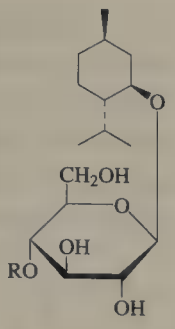


Figure 7.20 CAS numbering of anthracyclines.

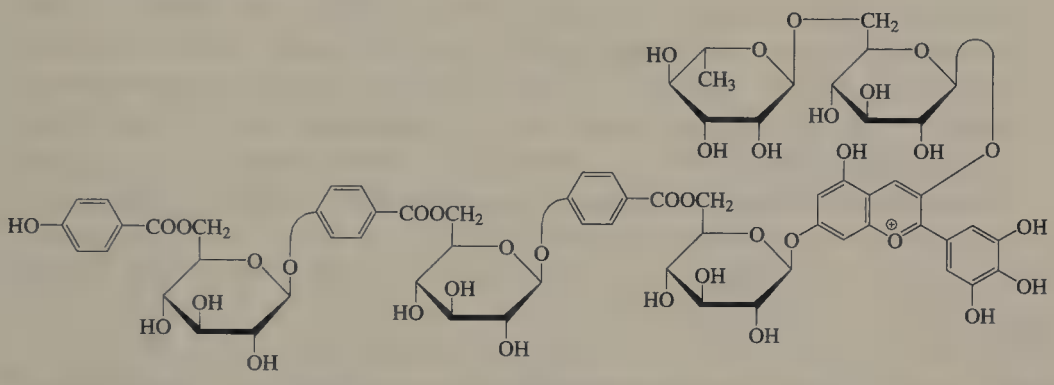
dominant group and are named systematically by CAS; an example is Feudomycin A (Figure 7.20(c)) which is named as a naphthacenedione with the numbering going in the opposite direction to the biogenetic scheme.

7.8.3 Carbohydrates; glycosides

The nomenclature of carbohydrates themselves is well established and will not be covered here [30]. For the natural products worker, more complex issues arise in the nomenclature and numbering of glycosides and especially complex glycosides, i.e. those in which carbohydrate residues are intermingled with non-carbohydrate fragments. The isolation of new types of these has to some extent outstripped available IUPAC/IUB guidance but certain points can be made about usual practice (Figure 7.21).



(a) R = H
(b) = *p*—OH—C_6H_4CH=CHCO—



(c)

Figure 7.21 Naming and numbering of glycosides.

1. The IUPAC/IUB rules describe three methods of naming glycosides, as exemplified by the three possible names (excluding stereodescriptors) for 23(a): menthyl *O*-β-D-glucopyranoside, 3-β-D-glucopyranosyloxy-*p*-menthane and *O*-β-D-glucopyranosylmenthol. Of these alternatives, the former is the more tractable when naming complex glycosides. This can be seen even for the fairly simple example Figure 7.20(b), which according to the first method is menthyl *O*-[4-*O*-(*p*-hydroxycinnamoyl)-β-D-glucopyranoside] but which becomes less convenient by the second and third methods.
2. It is common to simplify glycoside names by the omission of the *O*-prefixes (glycosides being understood to be *O*-glycosides unless otherwise stated) and to omit the ol→yl modifier, so that Figure 7.20(b) becomes menthol [4-(*p*-hydroxycinnamoyl)-β-D-glucopyranoside].
3. Special care is needed in the naming of glycosides of glycuronic acids. The IUPAC carbohydrate rules [27] should be consulted.
4. An increasing number of complex glycosides are being isolated which contain alternating carbohydrate and acylating groups. An example is

Campanin (Figure 7.21(c)) where the C-7 substituent in CAS requires multiple nesting brackets making it difficult to deduce the structure from the name; 7-[[6-*O*-[4-[[6-*O*-[4-[[6-*O*-(4-hydroxybenzoyl)-β-D-glucopyranosyl.... A simpler convention which is being introduced into the *Dictionary of Natural Products* is to call the substituent 7-*O*-[4-hydroxybenzoyl(→6)-β-D-glucopyranosyl(1→4)-4-hydroxybenzoyl ... (note absence of locant before the first arrow).

5. At least three conventions are in use for ring locant positions in glycosides (refer to Figure 7.20(a)) The dimethylamino substituent in the ring labelled A of Cytorhodin A can be referred to as 3′ with substituents on the other rings being distinguished by the use of multiple primes; but this becomes unwieldy when it becomes necessary to label the rings in order to indicate which ring the single primes refer to, which the double primes, etc. The *Chemical Abstracts* convention of using capital letter ring identifiers as superscripts so that the NMe_2 substituent becomes 3^A is becoming much more widely used. A third convention (IUPAC) involving the use of roman figures has not found wide acceptance.

7.8.4 Tannins

These natural products, combining carbohydrate moieties linked in a variety of modes to aromatic residues the simplest of which is galloyl (3,4,5-trihydroxybenzoyl) and to various residues at an intermediate state of dehydrogenation, are difficult to name. The various residues encountered have well-accepted trivial names (hexahydroxydiphenoyl, flavogallonoyl, gallagyl, elaeocarpusinoyl, chebuloyl, dehydrochebuloyl, brevifoloyl, trilloyl, dehydrodigalloyl, sanguisorboyl, valoneoyl, macaranoyl, tergalloyl, euphorbinoyl), and new ones continue to be characterized and named. However, the description and consistent numbering of the complete structures of the tannins resulting from the assembly of these subunits is currently something of a black art. There are no semi-systematic stereoparents currently in use for complete molecule skeletons as opposed to subunits. One representative example is shown in Figure 7.22. This has the trivial name Agrimonic acid A, the *Chemical Abstracts* name α-D-Glucopyranose, cyclic 2,3;4,6-bis(4,4′,5,5′6,6′-hexahydroxy-[1,1′-biphenyl]-2,2′-dicarboxylate) 1-[3-(6-carboxy-2,3,4-trihydroxyphenoxy)-4,5-dihydroxybenzoate], stereoisomer, and the 'colloquial' name 1-*O*-Dehydrodigalloyl-2,3:4,6-bis(hexahydroxydiphenoyl)-α-D-glucopyranose.

7.8.5 Terpenoids

As every student knows, this large class of natural products is derived by oligomerization of two or more prenyl (isoprenyl) residues according to the

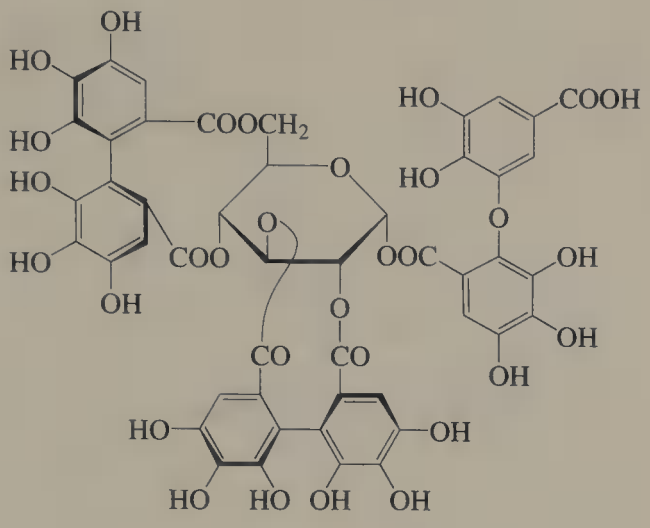


Figure 7.22 A representative tannin.

biogenetic isoprene rule that encompasses the possibility of rearrangements during biosynthesis. Although ideally the underlying relationships within the whole kingdom of terpenoids would be brought out by uniformities of nomenclature and, particularly, numbering, there are many irregularities caused notably by numbering schemes becoming established on the basis of unknown biosynthesis or biosynthetic schemes later shown to be erroneous. In the discussion which follows it will clearly impossible to cover every subgroup, of which DNP lists about 340, but the main groups will be covered together with any other groups that exemplify useful general points.

In general, the nomenclature and numbering of terpenoid skeletons bear little or no relationship to systematic nomenclature. *Chemical Abstracts* used a biogenetic nomenclature scheme similar to that described below for the smaller terpenoids until 1973; at the commencement of the 9th collective index period the switch was made to systematic nomenclature for most groups. However, the triterpenoids continue to be named semi-systematically.

Acyclic terpenoids Regular acyclic terpenoids from C_{10}-C_{30} including farnesanes, phytanes and squalanes, have a chain of $4n/5$ carbons substituted by $n/5$ equally spaced one-carbon substituents. Examination of their biogenesis shows that the biosynthetic numbering scheme begins at the end not bearing the proximal C_1 group, and therefore is the reverse of the systematic numbering for the parent hydrocarbons. However, the systematic numbering is influenced by the substitution pattern. Thus, for example whereas farnesane (Figure 7.23(a)) is 2,6,10-trimethyldodecane, 4-farnesanol (Figure 7.23(b)) is 3,7,11-trimethyl-4-dodecanol and 14-farnesanol (Figure 7.23(c)) is 6-methyl-2-(4-methyl-2-pentyl)-1-octanol.

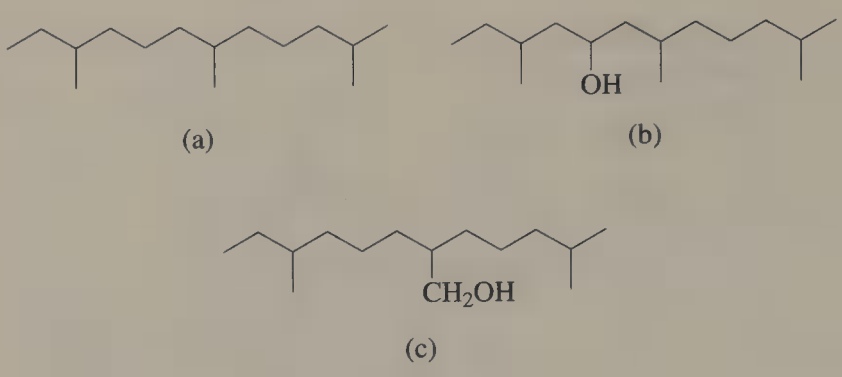


Figure 7.23 Acyclic terpenoids.

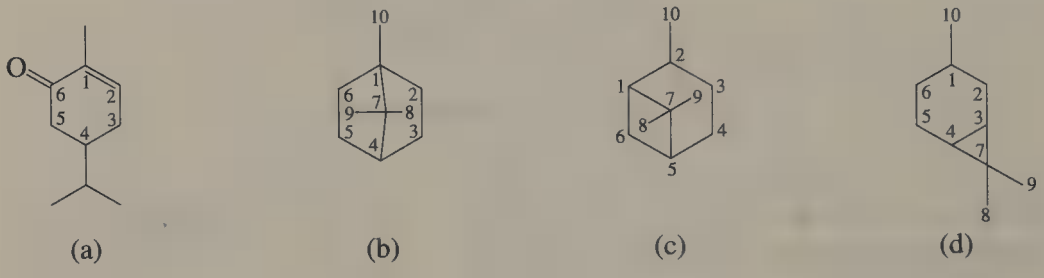


Figure 7.24 Some monoterpenoid skeletons.

Mono-, bi- and tricyclic monoterpenoids Figure 7.24 shows the generally accepted biogenetic numbering scheme for some of the commonest monoterpenoid types. There are several deviations from consistent biogenetic numbering and indeed for the bicyclic camphane and pinane skeletons (Figure 7.24 (b)–(c)) the modern schemes are based on the von Baeyer bicyclic systematic scheme rather than any presumed biogenetic analogies.

Figure 7.24(a) also demonstrates the IUPAC principle that numbering should be chosen to avoid compound locants if possible. Thus, Figure 7.24(a) is *p*-menth-1-en-6-one not *p*-menth-1(6)-en-2-one.

Sesquiterpenoids, diterpenoids and sesterterpenoids The majority of the 130 or so sesquiterpenoid groups formed by cyclizations and/or rearrangements of the farnesane system (see above) are regular in their numbering and will not be described further. The biogenesis of some groups (e.g. Brasilanes, Herbertanes) is currently unknown. Special cases among the sesquiterpenoids include the following:

(1) Abscisic acid and its relatives are 6,11-cyclofarnesanes but a non-farnesane numbering system has become well established (Figure 7.25(a)).
(2) The germacranes and cembranes(Figures 25b–c)represent examples of skeletons in which nomenclatural confusion may be caused by

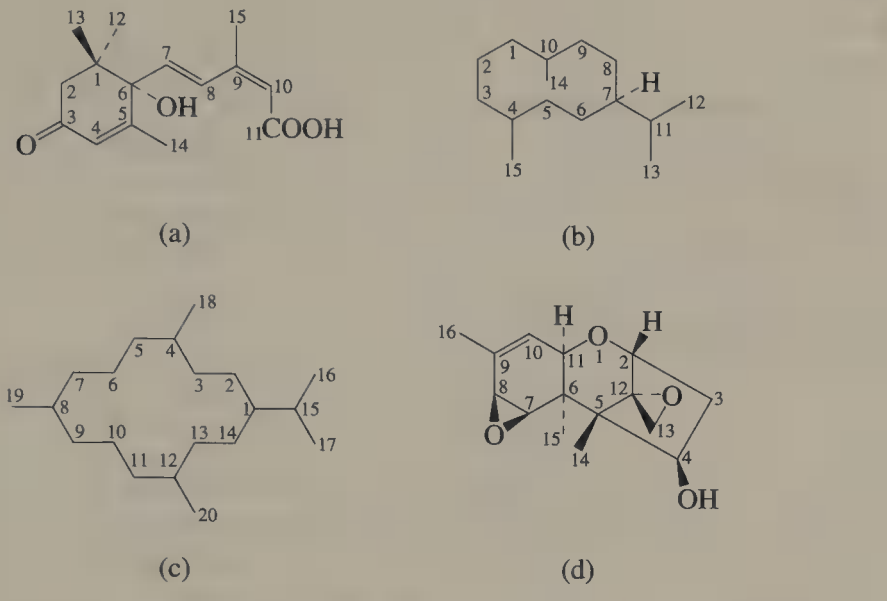


Figure 7.25 Numbering sesqui- and diterpenoids.

symmetry considerations, since the bare skeleton can in each case be numbered in either of two directions (although in the case of the germacranes (Figure 7.25(b)) this may not be immediately evident from the way in which they are conventionally drawn). In such cases, the modern convention of natural products workers is to impose a preferred direction of numbering so as to bring out biogenetic homogeneity. For germacranes, the convention is to orient with H-7 in the α-configuration as shown, which imposes a preferred numbering except when there is a double bond at C-7 destroying this centre of chirality (rare). In the case of cembranes, the convention is to choose between C-7 and C-9 by assigning C-7 to a carbon bearing a double bond or equivalent (e.g. epoxide). Of course, in systematic nomenclature (cyclodecane and cyclotetradecane, respectively) numbering may be in either direction. Similar putative symmetry in a 14-membered ring is shown by the Cericerane group of sesterterpenoids.

Particular problems associated with stereochemistry of the kauranes and related groups have already been referred to in section 7.7 (Figure 7.16b). The majority of other diterpenoid skeletons are fairly straightforward. CAS now uses few other diterpenoid skeletons as stereoparents; for example labdanes (named as such in the 8th Collective Index and earlier) are now named systematically based on decahydro-1,1,4*a*,6-tetramethyl-5-(3-methylpentyl)naphthalene for the parent labdane skeleton. The usual numbering is shown in Figure 7.16(a).

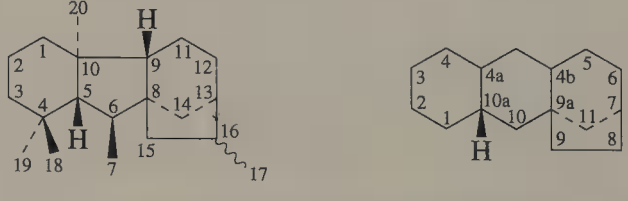
The biologically important trichothecane (scirpane) group is normally numbered using the IUPAC heterocyclic stereoparent trichothecane, the

numbering of which is shown in Figure 7.25(d). CAS uses the same scheme although others have appeared in the literature. Since nearly all of the naturally occurring trichothecanes have the 12,13α-epoxy function illustrated, they are frequently named as scirpanes, scirpane = 12,13-epoxytrichothecane, so that the compound illustrated (Crotocol) is 7,8; 12,13-diepoxytrichothec-9-en-4-ol or 7,8-epoxyscirp-9-en-4-ol. There are numerous macrocyclic analogues and these are named in CAS from stereoparents, e.g. Verrucarin A; the situation is analogous to the pyrrolizidine alkaloids described above.

The gibberellane group also requires special attention. Biogenetically it is well established that gibberellins are derived from *ent*-kauranes, which leads to the numbering scheme shown in Figure 7.26(a); many of the natural gibberellins are in fact 19-nor C_{19} compounds, so the numbering scheme is discontinuous. CAS names them from the C_{15} stereoparent Gibbane, which has a completely different numbering scheme (Figure 7.26(b)). Similar divergence occurs between the CAS numbering scheme and the terpenoid numbering scheme in the case of the quassinoid nortriterpenoids based on the heterocyclic picrasane skeleton. In CAS usage the ring oxygen is numbered (0–17) eliminating the biogenetic analogy with other triterpenoids.

Triterpenoids For the purposes of nomenclature discussion, the triterpenoids can be divided into three main groups; the tetracyclic triterpenoids containing a steroid-like nucleus and forming a biogenetic continuum with the steroids proper; the pentacyclic triterpenoids; and the nortriterpenoids. All three groups show a strong divergence between CAS and biogenetic nomenclature practice.

The first group is exemplified by lanosterol, which follows the standard and universal steroid numbering scheme with the additional methyl groups, lost during steroid biogenesis, numbered 28, 29, 30 (according to an earlier scheme they were numbered 31, 30, 32, respectively, by analogy with stigmastanes). For this group, CAS uses lanostane and dammarane as stereoparents but not the other generally recognized stereoparents. This



(a) (b)

Figure 7.26 Gibberellin numbering.

Table 7.3 Nomenclature of common tetracyclic triterpenoid skeletons

	Biogenetic Nomenclature	CAS Nomenclature
Tetracyclic triterpenoids		
	Protostane	Dammarane, (8α, 9β, 13α, 14β)
	Fusidane	29-Nordammarane, (4α, 8α, 13α, 14β)
	Cycloartane	9,19-Cyclolanostane
	Cucurbitane	19-Norlanostane, 19-methyl, (9β, 10α)
	Euphane	Lanostane, (13α, 14β, 17α)
	Tirucallane	Lanostane, (13α, 14β, 17α, 20S)
	Apotirucallane	Cholestane, 4,4,8-trimethyl, (13α, 17α, 20*S*)
Pentacyclic triterpenoids		
	Taraxerane	*D*-Friedooleanane
	Multiflorane	*D*:*C*-Friedooleanane
	Glutinane	*D*:*B*-Friedooleanane
	Friedelane	*D*:*A*-Friedooleanane
	Baccharane	*D*:*B*-Friedo-18,19-secolupane
	Bauerane	*D*:*C*-Friedoursane
	Hopane	*A*′-Neogammacerane
	Neohopane	*B*′*A*′-Neogammacerane
	Fernane	*D*:*C*-Friedo-*B*′:*A*′-neogammacerane
	Adianane	*D*:*B*-Friedo-*B*′,*A*′-neogammacerane
	Arborinane	*D*:*C*-Friedo-*B*′:*A*′-neogammacerane
	Stictane	21,21-Dimethyl-29,30-dinorgammacerane
	Serratane	*C*(14*a*)-Homo-27-norgammacerane
	Onocerane	8,4-Secogammacerane

is illustrated in Table 7.3. A similar situation pertains among the pentacyclic triterpenoids, where CAS bases the majority of known skeletons on the common ones (oleanane, lupane, neogammacerane).

Carotenoids The nomenclature of this intensively studied class of tetraterpenoid compounds is unique and is based on the stem name 'carotene' modified by two Greek prefixes to denote the cyclization/unsaturation of the end groups. The structures of these end groups are shown in Figure 7.27. α-, δ-, η- and ξ-carotenes are trivial names (e.g. α-carotene ≡ β,ε-carotene)

This system is clear and unambiguous with regard to the structures of the parent carotenes, but divergences in practice arise in the modification of these parents in the naming of their many hydro- and hydrated derivatives. IUPAC rules treat 'hydro' prefixes in carotenoid names as a non-detachable prefix to the carotene part; many authors as well as CAS treat the hydro- as detachable. Another source of variation is the disallowance by CAS of parents such as β-caroten-6-ol which are incapable of existence. This is illustrated by the molecule shown in Figure 7.28 which is 5,6-dihydro-β,β-caroten-6-ol according to IUPAC nomenclature, but in CAS is 5,6-dihydro-6-hydroxy-β,β-carotene.

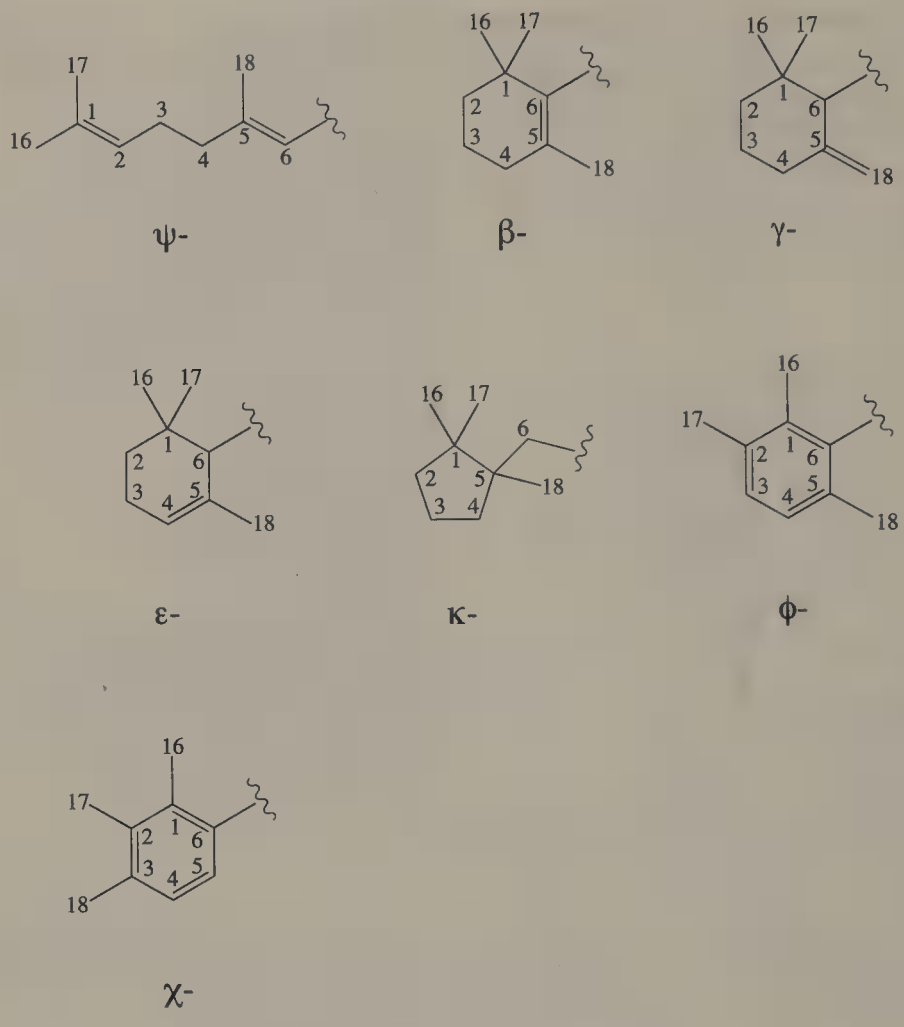


Figure 7.27 Carotenoid end groups.

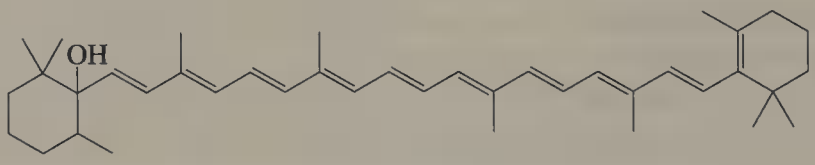


Figure 7.28 An example of carotenoid nomenclature.

A clear and full explanation of CAS nomenclature policy for carotenoids is given in the CAS Index Guide. Carotenoid numbering is retained for carotenoid-derived natural products (apocarotenoids) in which the chain is formally shortened from one or both ends. The prefix apo- is used to indicate that all of the molecule beyond the carbon with that locant has been removed. It is not necessary to give a Greek end-group designation if the apo-locant is greater than 5.

7.8.6 *Steroids*

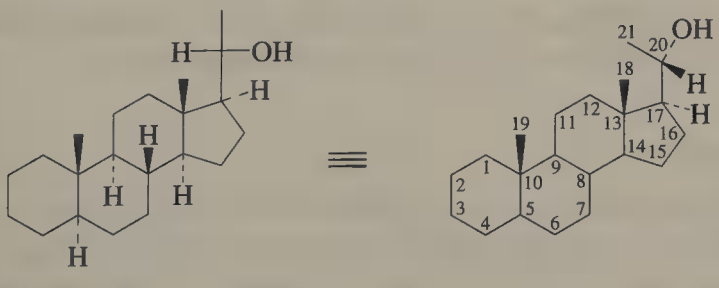
The nomenclature and numbering of the standard steroid skeletons based on the cyclopenta[*a*]phenanthrene nucleus is too well known to bear repetition here. However, a number of complicating factors of which those expecting to use the steroid literature effectively should be aware, are described.

The standard steroid stereoparents define the configurations at C-8(β-), C-9(α-), C-10(β-), C-13(β-), and C-14(α-), as shown in Figure 7.29 for 5α-pregnan-20-ol, and the C-8, C-9 and C-14 hydrogens are usually omitted from formulae. The configuration at C-5 needs to be specified. Prior to 1967, 5α- and 5β- steroids were indexed separately in *Chemical Abstracts*.

Configurations of substituents in the side-chain were formerly also indicated by α- or β- according to the Fieser convention in which the side-chain was drawn in Fischer projection with the highest-numbered locant at the top. This convention has now been replaced by the use of Cahn–Ingold–Prelog (*R,S*) descriptors (Figure 7.29).

Complications in labelling configurations also occur in secosteroids such as the vitamins D in which ring A loses its stereochemical anchoring relative to the rest of the molecule. The sequence rule is again recommended but is unfortunately not always used in the literature (or in CAS), resulting in frequent ambiguities (Figure 7.30). Subscripts 2 and 3 are used in this series to denote secosteroids belonging to the Vitamin D_2 (Ergocalciferol) and Vitamin D_3 (Cholecalciferol) series, which have ergostane and cholestane side-chains, respectively.

The numbering of estrane (oestrane) steroids is another example of the application of IUPAC principles concerning compound locants. Estranes with ring A aromatic are numbered as 1,3,5(10)-trienes in preference to 1(10),2,4-trienes, but in the case of the equine estrogens with naphthalenic unsaturation in rings A and B it is possible to avoid the use of compound locants altogether, so that they become estra-1,3,5,7,9-pentaenes (Figure 7.31).



20α 20*S*-

Figure 7.29 Fieser and Cahn–Ingold–Prelog descriptors for Pregnan-20-ol.

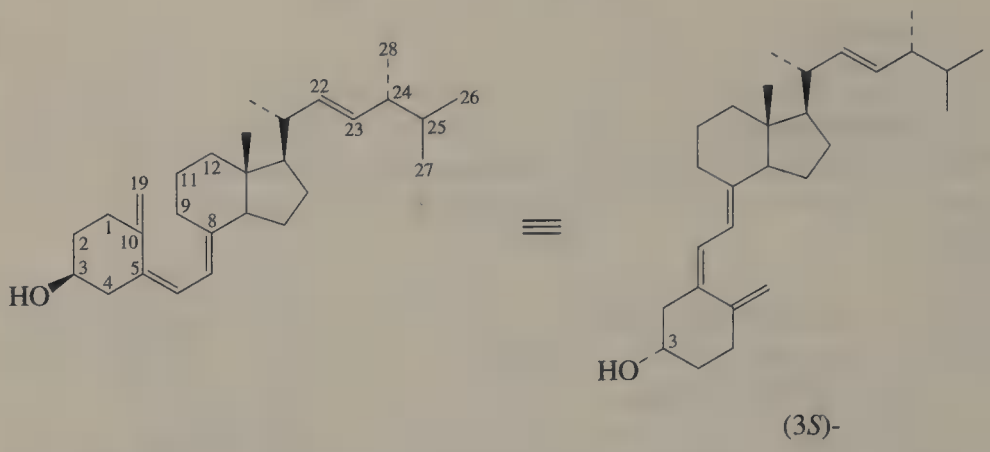


Figure 7.30 Numbering and stereochemistry of ergocalciferol.

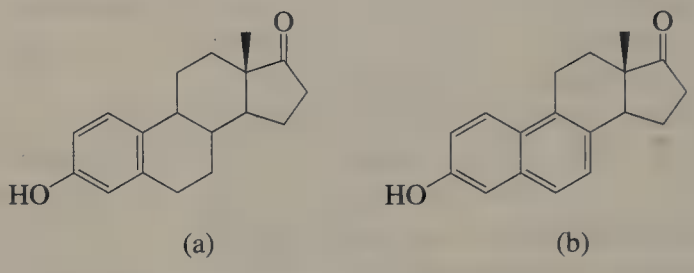


Figure 7.31 Nomenclature of estratrienes and estrapentaenes. (a) Estrone (Oestrone) =3-hydroxyestra-1,3,5(10)-trien-17-one. (b) 3-Hydroxyestra-1,3,5,7,9-pentaen-17-one.

The cardanolide skeleton has already been illustrated (Figure 7.1). Almost all natural cardanolides have the 5β,14β- configuration. The 5-configuration needs to be explicitly stated, but there has been a change in convention concerning the specification of the 14-configuration for cardanolides as well as the related bufanolides. Prior to the IUPAC–IUB 1989 recommendations, 14α- was assumed as for other steroids, but the two stereoparents are now defined as having 14β-configuration.

Some complications concerning the numbering of side-chains in the tetracyclic triterpenoids and the overlapping class of sterols have already been described. Steroidal alkaloids are considered below.

7.8.7 Flavonoids

The simplest and most numerous of this large group of plant products are based on the flavan and isoflavan nuclei which in systematic nomenclature are respectively dihydro-2- and 3-phenyl-2*H*-1-benzopyrans. Flavan and isoflavan are IUPAC-sanctioned semi-systematic parent names. (The alternatives 2 or 3-phenylchromone should be avoided.) Numbering of the simple flavonoids causes no problems; the phenyl ring locants are always

primed. There are however two conventions for ordering the locant numbers; for example, zapotin is 2′,5,6,6′- or 5,6,2′,6′-tetramethoxyflavone. The former is recommended as being more consistent with electronic indexing.

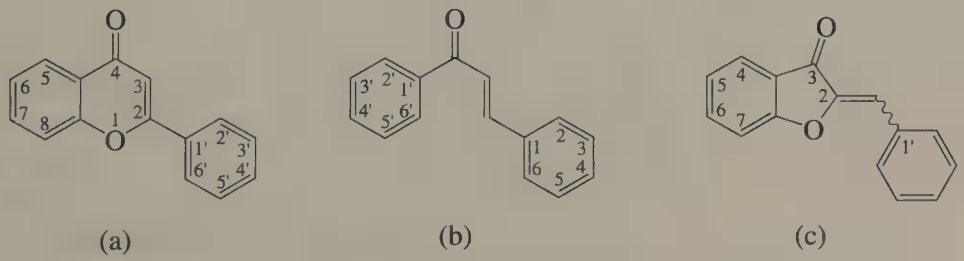
Chemical Abstracts allowed flavone and isoflavone as semi-trivial parents until 1973, but now names these compounds systematically. From the natural product worker's perspective this is unfortunate, since not only does it separate closely related flavonoids widely in the indexes and mix up the flavones with the isoflavones, but considerable mental agility is needed to work out and look up the systematic names without making trivial errors. Thus, to take a very simple example, 2′,4′,5-trihydroxy-7-methoxyflavone is 2-(2,4-dihydroxyphenyl)-5-hydroxy-7-methoxy-4*H*-1-benzopyran-4-one but the isomeric 2′,5,7-trihydroxy-4′-methoxyflavone is 5,7-dihydroxy-2-(2-hydroxy-4-methoxyphenyl)-4*H*-1-benzopyran-4-one.

Other classes of flavonoid whose nomenclature and numbering follows on readily from that of the flavones are the flavanones (2,3-dihydroflavones), flavan-3-ols, flavan-3,4-diols, flavans, anthocyanidins (2-phenylbenzopyryliums or flavyliums), dihydroflavonols and flavonols (3-hydroxyflavones). The semi-systematic parent name flavonol is in widespread use for the latter group so that, for example, 3,4′,5,7-tetrahydroxyflavone can also be called 4′,5,7-trihydroxyflavonol.

The complexities of naming the many glycosides and prenylated homologues of the simple flavonoids have already been referred to.

In the plant, flavones are formed biogenetically by cyclization of chalcones which also undergo bioreduction to dihydrochalcones. Chalcones can likewise be named semi-systematically in which case the chalcone skeleton is numbered as shown in Figure 7.32(b) or systematically as 2-propen-1-ones in which case ring C retains the same numbering (primed) but ring A changes its numbering so that 5,6,7,8- become 2″,3″,4″,5″. An alternative cyclization of chalcones in the plant produces the aurones, which are numbered according to their systematic nomenclature as 2-Benzylidene (*phenylmethylene in CAS*)-3(2*H*)-benzofuranones (Figure 7.32c).

There are three other main groups of modified flavonoids of fairly widespread occurrence. Two of these, the pterocarpans or 6*a*,11*a*-dihydro-



(a) (b) (c)

Figure 7.32 Flavone, chalcone and aurone skeletons

6*H*-benzofuro[3,2-*c*][1]-benzopyrans and the coumestans or 6*H*-benzofuro[3,2-*c*][1]benzopyran-6-ones are regular in that the widely accepted numbering schemes mirror the IUPAC/CAS systematic numbering. This is not true for the third group, the rotenoids, where several numbering schemes can be found in the literature and the most widely used scheme differs from that found in CAS. There is no general agreement on the numbering of the cyclized prenyl function. There are also many miscellaneous flavonoids containing additional rings produced by prenyl group cyclization or from other biosynthetic routes. The nomenclature of these is confused since they are not always given trivial names and their literature names are often unsatisfactory attempts at semi-systematic nomenclature, using fused ring parent names such as pyrano[. . .]flavone.

Oligomeric (tannin) flavonoids such as arecatannin (Figure 7.33) are now being characterized with increasing frequency. Systematic nomenclature (in this case based on hexahydro[4,8′;4′,6″-ter-2*H*-1-benzopyran]) is unrevealing and they are often named semi-trivially; epicatechin-(4β→8)epicatechin(4β→6)catechin. A more systematic variant that has the merit of describing the oligomeric relationships, which we have been using for entry names in the DNP, is 3,3′,4′,5,7-pentahydroxyflavan(4→8)-3,3′,4′,5,7-pentahydroxyflavan(4→6)-3,3′,4′,5,7-pentahydroxyflavan.

7.8.8 Lignans

The main features of lignan nomenclature have already been dealt with (section 7.4) and there is no escape from the fact that most lignans will

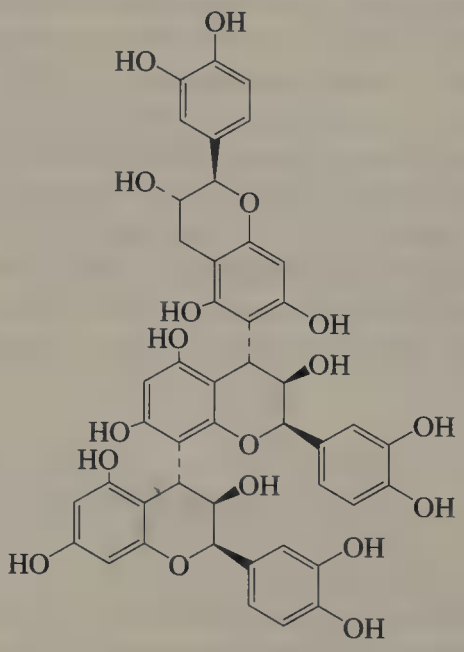


Figure 7.33 Arecatannin.

conveniently have a semi-systematic Freudenberg/Moss name according to the scheme outlined in that section, and a systematic name as well as usually a trivial name. Oligomeric lignans and hybrid lignans such as flavonolignans are now being discovered but are not very numerous and are structurally diverse.

7.8.9 Amino acids and peptides

The 20 protein-occurring parent amino acids are nomenclaturally straightforward, as are the majority of the 200 or more non-protein α-amino acids that occur in plants. It is however important to remember that CAS indexes the common α-amino acids and their substituted derivatives under the amino-acid name; for example 2-amino-3-methylhexanoic acid is indexed as 3-methylnorleucine.

There is an IUPAC–IUB 3-letter code for all of the protein amino acids and also for many of the commoner non-protein amino acids, and more recently one-letter codes to assist in computer indexing of peptides have also been introduced. The codes can be found in many reference sources [31].

The L-configuration (which corresponds to *S*- for all of the common amino acids except cysteine and cystine) is understood for all peptide residues unless marked as D-.

One fairly important class of amino-acid derivatives for which uniformity of nomenclature is sorely lacking is the dimeric amino-acid anhydrides (diketopiperazines) [32].This nucleus occurs not only in the simple representatives, which crop up frequently in the biochemical literature, but also in antibiotic/alkaloidal systems such as bicyclomycin and the ergot alkaloids. A simple representative is shown in Figure 7.34 and is worth considering in a little more detail because it exemplifies in miniature the trivial problems of natural product nomenclature. It is often called leucylalanine (or leucylalanyl) anhydride but is equally alanylleucine (or alanylleucyl) anhydride. The latter should take precedence by the rule of alphabetical sequencing but in practice the two names occur with similar frequency. Other workers prefer cyclo(alanylleucine) or cyclo(alanylleucyl) but here again cyclo(leucylalanine) or cyclo(leucylalanyl) are structurally equally valid. Recourse can then be made to systematic nomenclature in which case it is 3-isobutyl-6-methyl-2,5-piperazinedione, except that isobutyl is no longer valid in CAS. Conversion to the *Chemical Abstracts* name results in a change in alphabetical ordering (and numbering!) so that the CAS name is 3-methyl-6-(2-methylpropyl)-2,5-piperazinedione. Added to this, some authors refer to the parent nucleus as diketopiperazine or dioxopiperazine, while yet others use amino-acid abbreviations, as in cyclo(AlaLeu), with the cyclo- prefix either italicized or unitalicized.

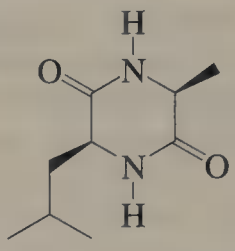


Figure 7.34 A simple diketopiperazine.

7.8.10 Alkaloids

This section will be considerably shorter than would be justified by the importance of alkaloids as natural products. To a first approximation – and it is a very rough division – in terms of their nomenclature and structure (as opposed to their biogenetic origin which is the way they are usually classified) the many groups of alkaloids can be divided into two large categories; those such as the benzylisoquinolines and berberines where the members of the group are structurally and nomenclaturally fairly homogeneous (often based on a heteroaromatic nucleus) and those where the individual members show a great variety of carbon skeleton based either on variations in the later stages of biosynthesis, or because they are acetate-derived resulting in many alternative chain cyclization modes. In the latter groups, the preferred nomenclature and numbering where possible should follow biogenetic considerations as for the terpenoids. The largest of these 'skeletally diverse' groups is the very diverse class of indole alkaloids, further considered below. Other alkaloid groups which are particularly large, or which present special points of interest, are also discussed briefly here.

Chemical Abstracts makes extensive use of semi-systematic parents (Cytochalasan and Securinan are examples). However, it also makes extensive use of functional parents, e.g. Vincaleucoblastine, Parsonsine (mentioned above, Figure 7.7) and Decaline, while many alkaloids (not always those of simplest structure), are named systematically.

Tropane alkaloids

The nomenclature and numbering of these is regular and based on the tropane (8-methyl-8-azabicyclo[3.2.1]octane) skeleton. Most of the naturally occurring alkaloids are in fact esters of hydroxytropanes and are thus scattered in the CAS indexes under the name of the esterifying acid.

Pyrrolizidine alkaloids The alkaloids in this group are esters of pyrrolizidine bases such as retronecine and its stereoisomers, and may be divided into two subgroups, the simple esters in which one or more OH group of the pyrrolizidine moiety is acylated, and macrocyclic diesters

such as monocrotaline. The nomenclature of the latter group, which is semi-systematic, is actually the simpler and has been described above in section 7.3. The systematic names of the simple ester alkaloids are, like the tropanes, dominated by the various acyl groups, but there is a further complication due to the two numbering systems of the pyrrolizidine nucleus (the tertiary carbon atom can be numbered 7a, as in CAS, or 8).

Sparteine alkaloids A point of interest of the sparteine nucleus (an IUPAC stereoparent but systematically dodecahydro-7,14-methano-2H,6H-dipyrido[1,2-a;1′,2′-e][1,5]-diazocine and named as such in CAS) is its two-fold rotation-reflection symmetry axis (Figure 7.35), which requires extra care in numbering and especially in the choice of stereodescriptors. Members of this lysine-derived subgroup having fewer or more rings do not have this symmetry property.

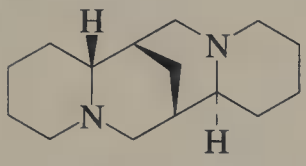


Figure 7.35 Sparteine.

Benzylisoquinoline alkaloids The nomenclature and numbering of uncyclized members of this very large group is straightforward, as exemplified by reticuline (Figure 7.36a). Divergence occurs in modified members such as cularine (Figure 7.36b), for which the systematic name (CAS) is 2,3,12,12*a*-tetrahydro-6,9,10-trimethoxy-1-methyl-1*H*-[1]-benzoxepino[2,3,4-*ij*]isoquinoline with consequent change in numbering. The same kind of divergency occurs among the bisbenzylisoquinoline alkaloids, of which there are at least 28 different structural subgroups. Where two or more oxygen bridges are present, semi-systematic parents may be in use, e.g. Berbaman (IUPAC, CAS), but alkaloids with only one bridge, e.g. Dauricine (Figure 7.36c), are named systematically by CAS (stem name = phenol).

Aporphine alkaloids The simple members of this large group are based on the aporphine skeleton, which in systematic parlance is 5,6,6*a*,7-Tetrahydro-6-methyl-4*H*-dibenzo[*d,e*] quinoline, with identical numbering (e.g. Tuduranine, Figure 7.37(a), 10-hydroxy-1,2-dimethoxyaporphine). There are numerous groups of modified aporphines with the potential for divergence between numbering schemes. This biogenetic continuum even goes as far as the aristolochic acids, in which the nitrogen has become oxidized and exocyclic, although there appears to be no attempt to extend

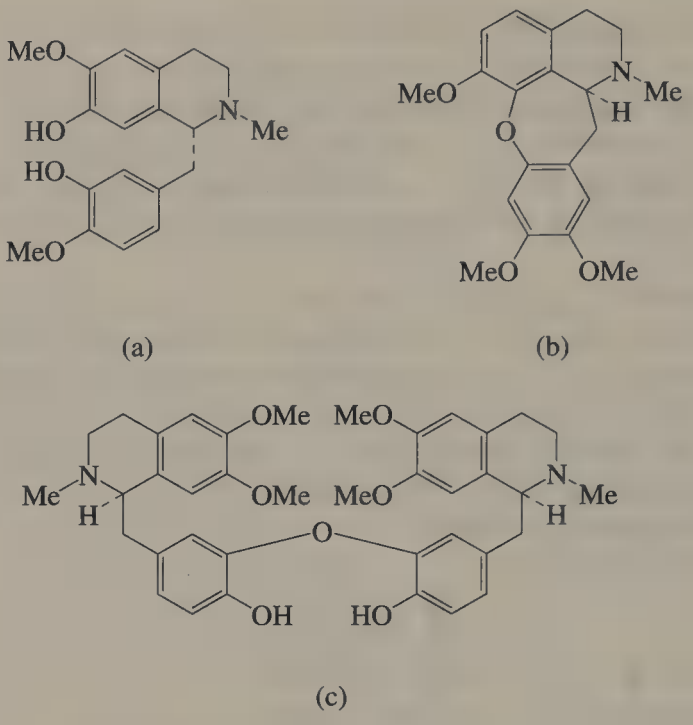


Figure 7.36 Benzylisoquinoline alkaloids.

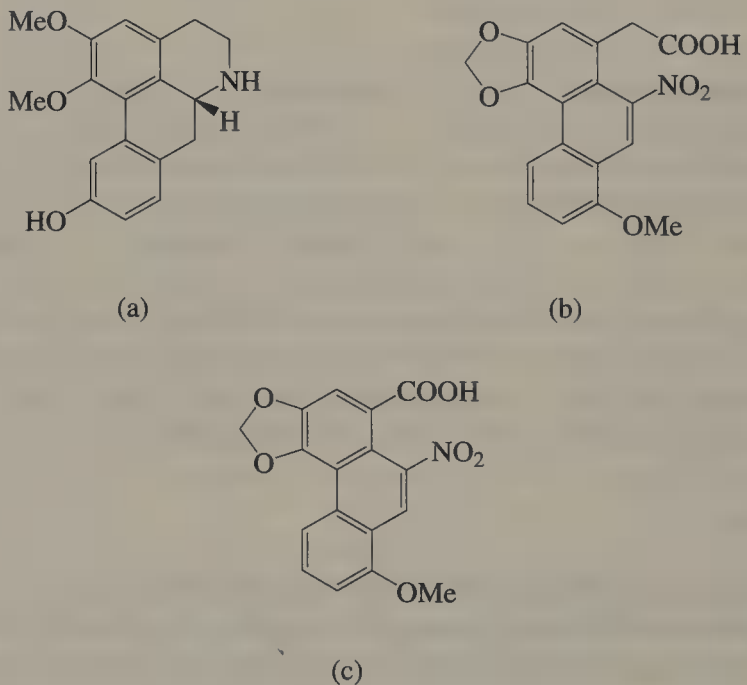


Figure 7.37 Aporphines and aristolochic acids.

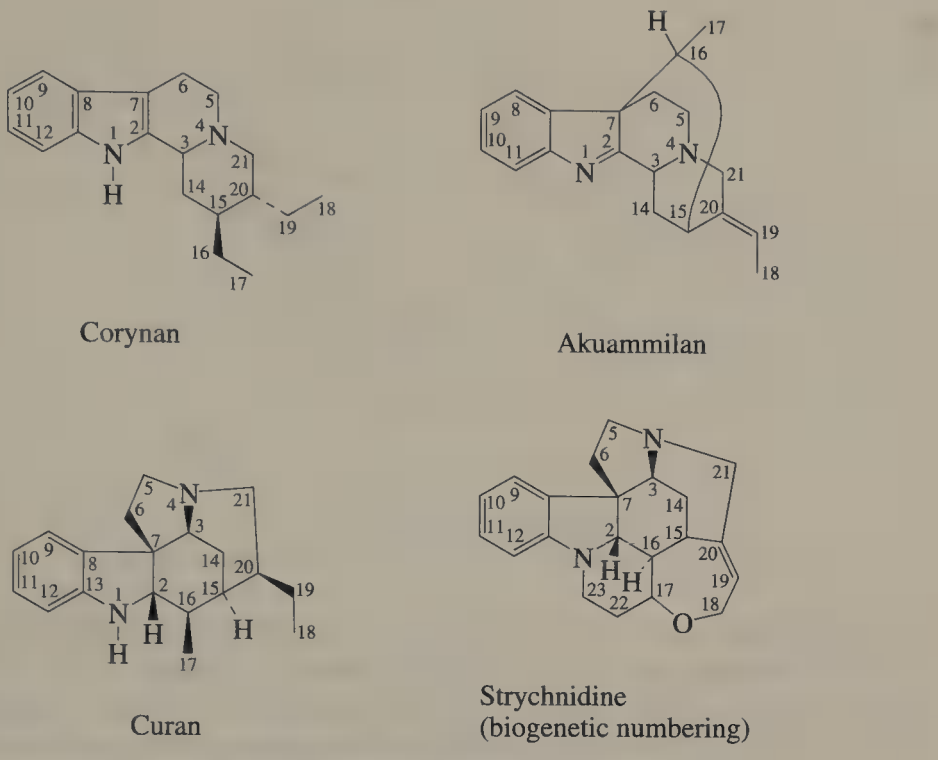


Figure 7.38 Indole alkaloid skeleton.

biogenetic numbering into what has now become a methylenedioxyphenanthrene (or in CAS usage, a phenanthro[3,4-*d*]-1,3-dioxole), for example Debilic acid, Figure 7.37(b).

Indole alkaloids

Figure 7.38 illustrates the homogeneity of the preferred biogenetic numbering scheme in just a few illustrative skeletons from this vast group. Note that in several cases (e.g. curan, strychnidine) this does not agree with the CAS numbering, which unfortunately corresponds neither with current biosynthetic knowledge nor with the IUPAC systematic numbering for these ring systems. For example in strychnidine, CAS numbering, 9, 10, 11, 12 become 1, 2, 3, 4, respectively.

Steroidal alkaloids

Some of these are simple aminopregnane derivatives and can be named as such, but there are several groups in which a cholestane-type side chain has become cyclized with incorporation of nitrogen, and in the case of the cevane skeleton illustrated in Figure 7.39(a), accompanied by a C-nor-D-homo rearrangement. A local convention applies in the case of the large Buxus group of alkaloids, illustrated by the example Baleabuxidine F illustrated in Figure 7.39(b). The suffix letter denotes the substitution

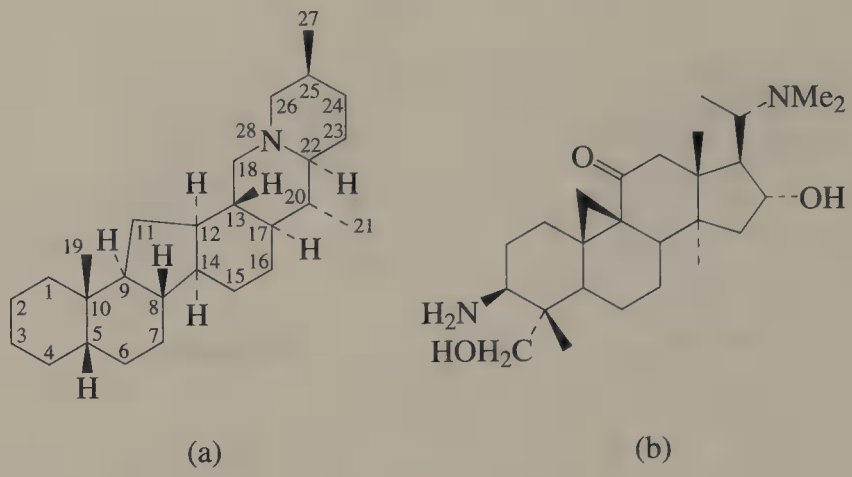


Figure 7.39 Steroidal alkaloid structures.

pattern of the one or two nitrogen atoms of the alkaloids; thus I means N^{20},N^{20}-dimethylation in an alkaloid having Ns at C-3 and C-20. The convention has been well adopted, although it is a good example of what was described in section 7.3 as undesirable semi-trivial nomenclature, i.e. the suffixes would have to be changed if the structures were revised, which fortunately does not appear to have happened to any great extent. The nomenclature of this whole group is however beset by synonymy and semi-trivial nomenclature. A number of structures have also been published which seem to be incorrect since they ignore the revision of the C-4 configuration in these alkaloids that was published about 20 years ago [33].

7.8.11 Polypyrroles

The polypyrroles are a numerically limited class of natural products centred on the haems, the chlorophylls, the bilins and Vitamin B12. These are all derived biogenetically from a common tetrapyrrolic intermediate, Uroporphyrinogen III (often abbreviated to Uro'gen III). An important feature of virtually all natural tetrapyrroles is that the acetate and propanoate side-chains in one of the rings are exchanged relative to the other three rings; this has the effect of lowering the symmetry characteristics of the skeleton so that all four rings are unique and they are always referred to as rings A–D as shown in Figure 7.40.

The standard system of numbering according to IUPAC–IUB for the tetrapyrrole nucleus is also shown in the Uro'gen formula. The carbon atoms are numbered 1–20 and the nitrogens 21–24. This has replaced the older scheme due to H. Fischer in which the outer pyrrole carbons were numbered 1–8 and the four bridging carbons were lettered α, β, γ, δ. The same numbering system is carried over into the open-chain Bilin group

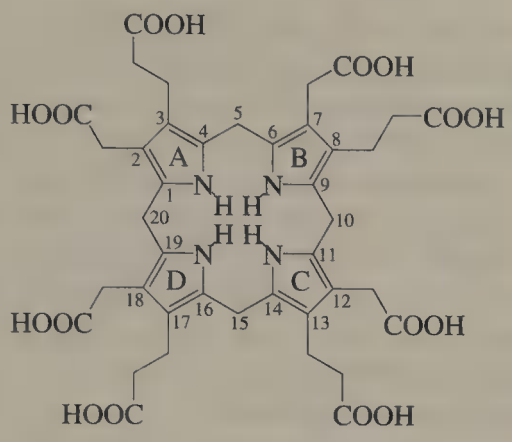


Figure 7.40 Uroporphyrinogen III.

which lack one of the nitrogens; to preserve parity with the cyclic analogues this is considered to be N-20.

Unfortunately, although the porphyrin nucleus is numbered identically by IUPAC–IUB and by CAS, the IUPAC–IUB numbering always starts on the A ring as shown for Uro'gen III, whereas CAS numbers compounds using systematic principles so as to give the lowest locants for the propanoate chains: thus Uro'gen III is 3,8,13,17-Tetrakis(carboxymethyl)-5,10,15,20,22,24-hexahydro-21*H*,23*H*-porphine-2,7,12,18-tetrapropanoic acid and the numbering begins at the position labelled 14 in the diagram and goes in the opposite direction. The suffix III in the name for Uro'gen III denotes its regioisomeric structure with respect to the disposition of the side-chains. The structures of the isomeric Urogens (or rather the uroporphyrins to which they readily oxidize) are given by the roman figure suffixes and the 15 regioisomers of the derived protoporphyrins by a similar series of conventional suffixes I–XV. The only natural protoporphyrin is Type IX and if the suffixes are omitted from compounds of this type, series III or IX, respectively, are implied.

The chlorophylls and Vitamin B12 analogues, derived by elaboration of tetrapyrrole precursors, are chemically and structurally more complex, but do not raise any major points of general principle in their nomenclature that have not already been covered.

7.9 Conclusion

Natural products embrace a very wide range of structural types. The needs of the specialist in various areas means that a large number of conventions have arisen. The chemist who works mostly or exclusively in a particular area should soon fall in with fellow experts to adopt good practice if certain basic principles described in this chapter are followed. It is

more difficult for those engaged in, for example, wide-ranging isolation work. Here the best advice must be that unless you are sure how to name a new natural product accurately using systematic or semi-systematic nomenclature, avoid doing so and do not spurn the use of trivial names coupled with clear structure diagrams. This does not prevent you from additionally **describing** the structure using the tools of systematic or semi-systematic nomenclature, but allows you to stop short of allocating a complete name that may cause future problems in documentation or, worse, actual misunderstandings of the science involved. It is hoped that this chapter will also have helped those who need to consult various parts of the literature on an occasional basis.

Notes and References

1. This can be appreciated by asking oneself the hypothetical question what would happen if, by some extraordinary combination of circumstances, the structure of methane were to be found to be CH_5 not CH_4. The substance found in natural gas, etc. would retain the name methane and the textbooks would report that, before 1997, methane was thought to be CH_4 but is now known to have a different structure. However, if a similar thing were to happen to hexane, we would have to say that the substance with a boiling point of 68.95°C that was formerly thought to be hexane has now been shown to be something else, and that the real hexane has some other set of properties; in other words, the name stays with the **structure** of six carbons in a straight chain. This scenario is far-fetched, but analogous situations arise with some frequency in natural product chemistry.
2. Whitten, W. K. *et al.* (1980), *J. Chem. Ecol.*, **6**, 49.
3. Herlem-Gaulier, D. *et al.* (1968), *Bull. Soc. Chim. Fr.*, 763.
4. Tanaka, T. *et al.* (1985), *Chem. Pharm. Bull.*, **33**, 4275.
5. N'Diaye, I *et al.* (1994), *Tetrahedron Lett.*, **35**, 4827–4830.
6. Dominguez, X.A. *et al.* (1980), *Phytochemistry*, **19**, 1262.
7. A search of the *Dictionary of Natural Products* database in 1997 revealed that at least 500 natural products, or 0.4% of the known total, were mentioned as having had their structures revised (including stereochemical revisions). This does not include historical revisions during the classical period of structure determination, or other cases where the DNP entry does not mention that the now-accepted structure is a revision. The current rate of publication of structure revisions of less well-known natural products would suggest that the proportion of incorrect structures in the literature, including those for which corrections have not yet been published, is at least 1%.
8. Bunyapraphatsara, N *et al.* (1989), *Phytochemistry*, **55**, 1555.
9. Nair, A. G. *et al.* (1992), *Phytochemistry*, **31**, 671.
10. Hase, T. *et al.*,(1995), *Phytochemistry*, **40**, 2.
11. This is without even taking into account whether the semi-systematic nucleus 'flavone' is allowable. Substitution of the up-to-date CAS alternatives for the flavone names given in this statement results in intractable complexity.
12. If the isolating authors fail to allocate a name and do not assign a structure, subsequent workers have no alternative than to refer to the compound as 'Bloggs' alkaloid VII' or something similar. This obviously has the potential for confusion; for example, the *Dictionary of Natural Products* database lists 11 'Alkaloid As' and this includes only those that have still not yet been assigned a structure. There are many more Alkaloid As in the older literature corresponding to alkaloids whose structure was subsequently determined. The situation is further complicated by the fact that other authors mentioning the substance may refer to it differently, for example as 'Base A' instead of 'Alkaloid A'.

13. Coune, C. *et al.* (1978) *J. Pharm. Belg.*, ***33***, 11, 284.
14. Crombie, L. *et al.* (1967), *J. Chem. Soc (C)*, 2553.
15. An example is provided by Bryostatin 3. The original isolate was assigned the incorrect configuration at C-20 and when this was realized it was renamed 20-Epibryostatin 3; the name Bryostatin 3 was reassigned to its newly isolated 20-epimer. (Schaufelberger, D. E. *et al.* (1991), *J. Org. Chem.*, **56**, 2895.)
16. Bose, P. K. *et al.* (1973), *Phytochemistry*, **12**, 667–668.
17. Nakano, T. *et al.* (1979), *J. Chem. Soc., Perkin Trans. 1*, 2107.
18. Hayashi, T. *et al.* (1974), *Phytochemistry*, **13**, 1943.
19. Wani, M. C. *et al.* (1983), *J. Nat. Prod.*, **46**, 537.
20. McPhail, A. T., *et al.,* (1989), *J. Nat. Prod.*, **52**, 212.
21. Chaudhuri, S. K. *et al.* (1995), *J. Nat. Prod.*, **58**, 1.
22. It has to be admitted that considerable confusion can be caused by the indiscriminate coining of numbering schemes in some areas. Thus, for the Isocomane group of sesquiterpenoids there have been several schemes in simultaneous use, including no less than four different ones from the Bohlmann group.
23. Ayres, D. C. and Loike, J. D. (1990) *Lignans; Chemical, Biological and Clinical Properties*, Cambridge University Press.
24. Gottlieb, O. R. (1978), *Fortschr. Chem. Org. Naturst.*, **35**, 1.
25. Moss, G. P., personal communication.
26. Kubota, T. *et al.* (1958), *J. Chem. Soc.*, 3667.
27. In addition the members of the smaller phyllocladane group which are epimeric with the kauranes at C-5, C-9 and C-10 are named by CAS as kauranes.
28. Scholz-Böttcher, B. M. *et al.* (1991), *Annalen*, 1029–1036.
29. Haslam, E. *et al.,* (1971), *J. Chem. Soc. (C),* 1489–1495.
30. An updated set of IUPAC carbohydrate nomenclature recommendations has recently been published (*Pure Appl. Chem* (1996) **68**, 1919–2008), also available as an Acrobat file on the CD-ROM disc of the *Dictionary of Carbohydrates* (Chapman & Hall, 1997).
31. Rhodes, P. H.(1995), *The Organic Chemist's Desk Reference*, Chapman & Hall.
32. Sammes, P. G. (1975), *Fortschr. Chem. Org. Naturst.*, **32**, 51–107.
33. Sangare, M. *et al.* (1975), *Tetrahedron Lett.*, 1791.

8 Trivial nomenclature: the INN and ISO systems

R. B. TRIGG

8.1 Introduction

Two industries above all others possess an insatiable appetite for new chemical and biochemical substances: pharmaceuticals and agrochemicals. Consequently, nomenclature systems are required in order that scientists, commercial workers, health professionals and regulators working with these materials can communicate with each other in a precise and concise manner. It frequently happens that different groups of specialist workers elaborate colloquial forms of nomenclature that are meaningful within a given work area but which would be seen as mindless jargon by workers outside that area.

To meet the requirement for easy communication two parallel nomenclature systems have been developed to serve these two communities:

- International Nonproprietary Names (INN)
- International Organization for Standardization of Common Names for Pesticides and other Agrochemicals (ISO Names)

8.2 The INN System

8.2.1 Background

With the rapid expansion of pharmaceutical research after the Second World War, it soon became apparent that some form of international co-ordination was necessary to harmonize the assignment of generic names to substances being produced for screening for biological activity. Without centralized administration, there was a likelihood that the instances of the same substance being assigned different names by different agencies would increase. Different names for 4-acetamidophenol had already been assigned, paracetamol in Europe and acetaminophen in the USA, and it was clearly not in the best interests of health professionals or patients for such divergences to continue.

Accordingly, in 1950, representatives from France, UK and the USA got together under the auspices of the World Health Organization (WHO)

and the INN system was born. The original concept is still in place today. It consists of the selection and publication of nonproprietary names initially as **proposed** INNs and, after a due period of time for public comment, re-publication as **recommended** INNs.

The first list of proposed INN was published in 1953 [1]. It consisted largely of a round-up of names in use at that time (e.g. phenylbutazone, phentolamine) and the first list of recommended INN followed 2 years later [2]. The creators of the system had the foresight to anticipate that the selected names might well be in conflict either with registered trade marks or with other generic names and that a rigorous search procedure would be both costly to administer and subjective in its assessments. Better, they reasoned, to place the onus on those with proprietary interests to protect to examine proposed INN and lodge objections with WHO in cases where such interests were compromised. This concept remains one of the fundamental concepts of the current system.

Lists of proposed INN are published twice yearly in *WHO Drug Information*. A period of 4 months is allowed for the lodging of an objection by any party with a *bona fide* interest in doing so. Objections are reviewed by the INN secretariat and if found valid are formally lodged. A name subject to a formal objection remains as a proposed INN. Names to which no objections are raised proceed to become Recommended INN about 1 year after initial publication. Lists of Recommended INN are also published in *WHO Drug Information*.

About 7000 INN have been adopted, a cumulative list of which was last published in 1996 [3]. At the time of writing 78 lists of proposed INN and 39 lists of recommended INN have been published. The cumulative list provides much useful information for each name but regrettably not the chemical definition nor graphic formulae, details of which must be sought in the proposed INN list in which the name was originally published.

8.2.2 Secretariat mechanism

The INN system is administered by a secretariat serviced by the Division of Drug Management and Policies of WHO. It is advised by a Panel of Consultants drawn from members of the *WHO Expert Advisory Panel on the International Pharmacopoeia and Pharmaceutical Preparations*, a pool of experts appointed to advise the Organization on pharmacopoeial and related matters. Consultants are appointed on account of their expertise in the field, irrespective of nationality. The currently serving consultants are from the USA, UK, Japan, France, Spain, Nigeria, Poland and Indonesia. The business is conducted largely by correspondence supplemented by an annual Consultation held in Geneva at the end of April each year.

The INN process begins when a manufacturer starts clinical trials with a potential new medicinal substance. At this point, a generic name will

be required to facilitate communication between the health professionals concerned with the drug's development. Hitherto, the substance is likely to have been known by a laboratory code number or by a colloquial name of some kind. Manufacturers are discouraged from premature internal use of an informal name for fear of it either not complying with the increasingly rigid guidelines that underpin the selection process or of its being in conflict with an existing name. It is even more important not to use a name in the scientific literature unless its has been adopted as an INN since if the name is subsequently rejected or is substantially modified there is no easy way of expunging the name from the scientific record.

Adoption of an INN at too early a stage in the drug's development leads to wastage of the name if the drug subsequently fails to fulfil its expectations and is withdrawn from trial. With an assurance that the drug is in clinical trial, the manufacturer applies to WHO for the adoption of an INN using the form shown in Figure 8.1. The applicant is invited to put up proposals for the name and is required to provide the systematic chemical name, graphic formula and molecular formula, in the case of an organic substance, or in the case of a biological substance which nowadays is most likely to be a product of recombinant technology, a scientifically rigorous definition which should be either biologically descriptive or, in the case of a polypeptide or glycoprotein, given in terms of its amino acid sequence. For all substances the applicant is expected to provide the Chemical Abstracts Registry Number (or take steps to procure one), details of any code numbers by which the substance may be known in the scientific literature, an indication of the likely therapeutic application and most importantly, a concise statement of the pharmacological mode of action. Provision of supporting pharmacological data is particularly vital when a mode of action is being claimed for which there are no previous INN examples.

The manufacturer's proposals are expected to comply with certain guiding principals (see section 8.2.3). Applications are screened initially by the secretariat to see that the guiding principles have been followed. Failure to observe this requirement will result in almost certain rejection. A careful study is carried out looking for relationships with existing substances, either through common fragments of chemical structure, similar target therapeutic applications or, increasingly importantly, a similar pharmacological mode of action. The secretariat maintains collections of reviews of substances related in these ways which can be drawn upon and attached to the applications when they are sent to the consultants for comment. In this way each application is supported by a package of data supplementing that provided by the manufacturer. It is not necessarily in the manufacturer's best commercial interest to make the relationship to a pre-existing substance too transparent but it is important that such links are revealed if a rational system of nomenclature is to be followed.

Request for an international nonproprietary name (INN)
Demande de dénomination commune internationale (DCI)

Authority or manufacturer:
Autorité ou fabricant:

Name of applicant / *nom du demandeur:*
Name of responsible officer / *nom du responsable:*
Address / *adresse:*

Telephone No/*No. de téléphone:* Fax No./*No. de fax*

We hereby request the World Health Organization to establish a free and unrestricted INN for the pharmaceutical substance described below.
L'OMS est priée de bien vouloir établir une DCI à usage libre pour la substance pharmaceutique en question.

For completion by WHO/
A remplir par l'OMS

No:
Date:
Copies forwarded:

Date:

Acknowledged:

SUGGESTED NAMES (in order of preference): 1.....
DENOMINATIONS PROPOSÉES
(par ordre de préférence) 2.....

3.....

CHEMICAL NAME OR DESCRIPTION (INCLUDING STEREOCHEMICAL INFORMATION):
NOM OU DESCRIPTION CHIMIQUE (Y COMPRIS DE L'INFORMATION SUR LA STÉRÉOCHIMIE)

GRAPHIC FORMULA:
FORMULE GRAPHIQUE

MOLECULAR FORMULA:
FORMULE BRUTE

CHEMICAL ABSTRACTS SERVICE (CAS) REGISTRY NUMBER:
NUMÉRO DU REGISTRE CAS

TRADE NAME (known or contemplated):
NOM COMMERCIAL (connu ou envisagé)

ANY OTHER NAME OR CODE:
AUTRE NOM OU CODE

PRINCIPAL THERAPEUTIC USES AND POSOLOGY; PHARMACOLOGICAL ACTION:
UTILITÉ THÉRAPEUTIQUE ET POSOLOGIE; ACTION PHARMACOLOGIQUE

For completion by WHO/
A remplir par l'OMS

Additional information may be given overleaf/Toute information complémentaire à fournir au verso.
Conditions and Explanatory Notes overleaf/Conditions et notes explicatives au verso.

Figure 8.1 The INN Application Form.

Batches of proposals are sent out to the consultants at regular intervals. Several of the consultants are secretaries of national nomenclature agencies and have the advantage of being able to call upon their own national committees and thereby seek a far broader spectrum of expert comment than they themselves could bring. This secondary level of expert comment means that it may be some weeks before the INN secretariat can collate all the replies into a consolidated comment document.

The operation focuses on two annual events – the annual consultations held at WHO headquarters in April each year and a meeting of the secretariat with a senior Consultant in November. From each of these two assemblies, a first draft of a list of proposed INN arises.

The draft list is sent to the consultants together with a copy of a consolidated comment document and a request for further comments by a specified date, around 6 weeks from the date the list was drafted. This gives the consultants an opportunity to review the selections made at the April meeting and check the availability of the drafted names in their countries or to react to the views of fellow consultants in the case of the November list. Further rounds of comment ensue in which the consultants react to the endeavours of the secretariat to find solutions to problems that may arise during the negotiation. Difficulties may involve conflict with existing names, both trade and generic, misleading connotations and problems due to transcription into other languages.

A deadline has to be drawn eventually in order to meet publication schedules when any name not agreed by all the consultants will be held over until the next list. An average length of time to this point would be 9 months which, in terms of international negotiations, is very fair indeed. During this negotiation period, a manufacturer will not necessarily be involved in the week-by-week developments. He may be asked to supply further data or explain an esoteric point but in general it's a case of 'no news is good news'. In cases where the manufacturer's INN proposal is not entirely appropriate the secretariat and consultants endeavour to incorporate a distinctive syllable from the proposal in the eventual name in the hope that the counterproposal will be acceptable to the manufacturer. The manufacturer is, of course, invited to comment on the drafted name and any adverse response therefrom is fed into the following round of post-draft commenting.

The reason for not informing manufacturers every step of the way is the volume of correspondence and high cost of telecommunications it would generate. Around 150 names are processed annually by a secretariat with limited resources. To ease this burden, manufacturers in the USA, Japan, France and the UK are asked to submit their applications not to WHO but to their national nomenclature agencies where the applications will be processed for national adoption and as INN concomitantly. These agencies are known respectively as the United States Adopted Names Council (USAN), Japanese Accepted Name Committee (JAN), French Pharmacopoeia Commission (Dénomination Commune Française) and the nomenclature committee of the British Pharmacopoeia Commission (British Approved Names – BAN). These agencies work to the same guiding principles as the INN system and in very close harmony with the INN secretariat and, because of more convenient geography, they are able to liaise closely with manufacturers on a day-by-day basis.

8.2.3 Guiding principles

Two guiding principles provide the key to the philosophy of the INN system:

"1. International Nonproprietary Names (INN) should be distinctive in sound and spelling. They should not be inconveniently long and should not be liable to confusion with names in common use."
"2. The INN of a substance belonging to a group of pharmacologically related substances should, where appropriate, show this relationship. Names that are likely to convey to a patient an anatomical, physiological, pathological or therapeutic suggestion should be avoided."

Principle 1 is largely common sense and few would argue over its provisions. INN are required to be distinctive from other INN, trade marks both registered and unregistered, and from other forms of trivial scientific names. The assessment of similarity of two names is a very subjective process although computer algorithms are available to aid the process.

Principle 2 requires the similarity of a substance to an existing one to be reflected in its name and it is the failure to comply with this requirement that causes the most anguish. It has given rise to the prefix–stem approach to drug nomenclature that is normally applied in assigning a name.

General rules A number of subsidiary guidelines support these two primary rules:

1. When the first example of a new class of pharmacologically active substances arises, its name should be designed such that further members of the series can be named in a related manner. In other words, think ahead – choose a stem to identify the new class and only then select the specific name.
2. When naming acids, use a one-word name so that its salts can be named simply by adding the name of the cation, e.g. diclofenac and diclofenac sodium.
3. Whenever possible, select the name for the active moiety. In the case of esters of alcohols and salts of acids this generally means assigning the name to the free acid. The name of the derivative can be obtained by using normal systematic nomenclature practices. Thus: methylprednisolone acetate, erythromycin stearate, phenytoin sodium.
4. Avoid isolated letters and numbers in INN whenever possible, particularly letters. Naturally occurring antibiotics are sometimes obtained from fermentation processes as mixtures of complex molecules with small variations in their chemical structure. In some cases, it is neither economically sensible nor even necessary from the activity standpoint to separate the components. In the case of gentamicin these closely

related substances are labelled as 'gentamicin A', 'gentamicin B', etc. It would scarcely be sensible to name them in any other way but for purposes of pharmaceutical commerce it is sufficient for the mixture to be described just as 'gentamicin', thus avoiding detached letters. The regulatory agency will concern itself with precisely what is meant by 'gentamicin' and a full-bodied definition will be given in the pharmacopoeias. It then sometimes happens that the major component of a mixture becomes commercially available in an essentially homogeneous state. In such a case the major component of neomycin, known as 'neomycin C', was assigned the completely different INN 'framycetin' but in another example, actinomycin D became known as 'dactinomycin'.

Polymers are a second group of materials where recourse to appended characters is deemed sensible. There are many examples and one of the best known is the family of polyethylene glycols known in the pharmaceutical field as 'macrogols'. The INN system provides a general definition of macrogol together with the requirement that the name is supplemented with a number indicative of the average molecular weight. Thus, we have macrogol 400, macrogol 6000, macrogol 30 000, etc. Other polymeric series treated in this manner include the dextrans and the dimeticones; in the latter case the number is indicative not of the average molecular weight but of the viscosity.

5. The spelling of the English form of an INN is designed so that the minimal amount of transcription is necessary to render it suitable for use in other languages using the Roman alphabet. Accordingly, use 'f' in preference to 'ph', 't' instead of 'th' and 'i' instead of 'y'. Additionally avoid the letters 'h', 'k' and 'w' whenever possible. Unfortunately this can result in names which to the British reader look strange and may be ambiguous in pronunciation; consider 'carbofenotion' rather than the more conventional 'carbophenothion'!

8.2.4 The 'prefix–stem' approach

In the early days of the INN system, names were often devised simply by selecting key fragments of the systematic chemical name, Thus, the hypoglycaemic substance 3-(2-chlorophenothiazin-10-yl)propyldimethylamine became known as 'chlorpromazine'. This system, while providing chemists with a generic name with which they could associate, produced names that meant nothing to physicians, and, moreover, several thousand names later, produced a large number of names beginning chlor-, meth-, bromo-, etc. Some further mileage has been wrung from this approach by modifying these prefixes to clo- (even lo-), met-, bro-, etc, but it is commonly accepted that this vein of name construction is exhausted. Occasionally it is still necessary to use chemical prefixes in cases where an existing substance has

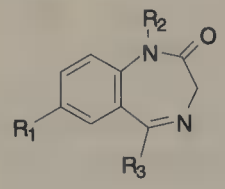
been modified simply by substituting, for example, a fluorine atom when it is usually sensible to relate the substances by using the prefix flu-.

The demise of the chemistry-based method of devising INN has been accompanied by a steep increase in the prefix–stem approach. This concept is not new. In the very early days of the INN system, if not before, it had become the practice to use the stem -cillin for semi-synthetic penicillins (ampicillin, amoxicillin) and -mycin for naturally occurring antibacterial substances obtained by fermentation process using strains of *Streptomyces* species (erythromycin, neomycin). The main advantage of this approach to drug nomenclature is that the resulting group of names with a common syllable, or group of syllables, is they carry an implicit message to health professionals. A physician meeting the name 'pivampicillin' for the first time will know instinctively that the substance is an antibiotic of the penicillin family and closely related to 'ampicillin'. Moreover, he will be alerted to the likelihood of allergic reactions to the substance in the case of patients known to be sensitive to other penicillin drugs.

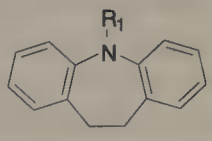
A common pattern found in the evolution of drug families is for an initial breakthrough into a new class of active substance to be followed by a period in which a series of substances is developed in an attempt to enhance the initial activity and at the same time minimize any adverse side-effects. It was therefore seen as logical to relate such a group of substances by a common stem in their names. The earliest groups of substances named in this manner usually exhibited common structural features as well as similar biological effects or therapeutic uses. For example, there is the -azepam group of hypnotic/anxiolytic substances based on 1,4-benzodiazepine (diazepam, nitrazepam), the -pramine group of tricyclic antidepressant substances based on dibenzazepine (imipramine, desipramine) and the well-known -olol group of beta-adrenoreceptor blocking agents all of which possess the 1-amino-3-arylpropan-2-ol skeleton (atenolol, propranolol) (see Figure 8.2).

Today the emphasis has shifted from chemistry to pharmacology, reflecting the huge increase in our understanding of drug biochemical reaction mechanisms. The physician need not concern himself with the chemical structure of the drugs he prescribes but he can be expected to have some feeling for the manner in which they exert their effects. We have 'agonists' and 'antagonists' of naturally produced biochemical substances, promoting or blocking their production in living systems and inhibitors of all manner of enzymes that react in human systems in an adverse manner. The natural material known as serotonin is the focus of much research interest at the present time and has spawned a series of serotonin receptor antagonists classified pharmacologically by the term '$5HT_3$', characterized by the stem -setron (e.g. granisetron, ondansetron) and used as antihypertensives, and a series of serotonin receptor agonists, classified as '$5HT_{1D}$', characterized by the stem -triptan (sumatriptan) and

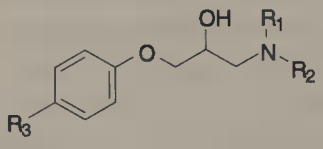
The essential structural features of:



a) the *-azepam* hypnotic/anxiolytics



b) the *-pramine* antidepressants



c) the *-olol* beta-adrenoceptor blockers

Figure 8.2

used as antiemetics. Substances assigned names using these stems can be expected to possess the same pharmacological action but may not possess a rigorously similar chemical parentage. The danger now for the INN system is to avoid being dragged too deeply into the increasing complexity of pharmacological classification. The physician is concerned largely with the biological effects shown by drugs and he would doubtless argue that from his viewpoint drugs with similar uses should have common stems – how their effects are brought about is not his prime concern.

The dilemma faced by the INN system in using the stem approach to naming substances is, therefore, to balance the needs of the physician for meaningful names without offending pharmacological taxonomy and without upsetting chemists too much by making bedfellows of substances with few, if any, structural elements in common.

Table 8.1 shows examples of stems in current use. A comprehensive document which is updated annually giving all the stems used in the creation of INN together with examples is available from WHO [4].

As for the prefix, its purpose is simply to distinguish one member of a series from the others in as distinctive a manner as possible. Inevitably, recourse is sometimes made to the chemical element or fragment that provides that distinction at the structural level but this results once more in an over-abundance of names beginning with flu-, met-, etc. The INN secretariat therefore encourages the use of totally abstract prefixes in order to minimize the number of names beginning with the all-important first syllable. Critics of the prefix–stem system argue that it results in names that look and sound alike but in its defence it can be argued that any system of rational scientific nomenclature results in similar names for

Table 8.1 Examples of stems used in INN construction

Stem	Definition
-anserin	serotonin receptor antagonists (mostly $5HT_2$) used as antihypertensives
-astine	antiasthmatics, antiallergics when not acting primarily as antihistamines
-azepam	substances of the diazepam group
-caine	local anaesthetics
cef-	antibiotics, derivatives of cephalosporanic acid
-cillin	antibiotics, derivatives of 6-aminopenicillanic acid
-conazole	systemic antifungals of the miconazole group
-dipine	calcium ion channel antagonists
-fibrate	substances of the clofibrate group
-floxacin	fluorine-containing antibacterial agents of the quinolone group
gli	sulphonamide hypoglycaemics
-grel	platelet aggregation inhibitors
-kalim	potassium channel activators
-mab	monoclonal antibodies
-metacin	anti-inflammatory substances of the indomethacin group
-mycin	antibiotics produced by *Streptomyces* strains
-nidazole	antiprotozoal substances of the metronidazole group
-olol	beta adrenoceptor antagonists
-oxacin	antibacterial agents of the quinolone group
-pafant	platelet-activating factor antagonists
-pril; -prilat	inhibitors of angiotensin-converting enzyme
-profen	anti-inflammatory substances of the ibuprofen group
prost	prostaglandins
-relin	hypophyseal hormone release-stimulating peptides
-sartan	angiotensin II receptor antagonists
-setron	serotonin receptor antagonists ($5HT_3$) used as hypertensives
-stat	enzyme inhibitors
-steine	substances of the acetylcysteine group
-tidine	H_2-receptor antagonists of the cimetidine group
-triptan	$5HT_{1D}$-serotonin receptor agonists
-vastatin	HMG CoA reductase inhibitors
-verine	spasmolytics with a papaverine-like action
-vir	antivirals

objects or organisms that show close similarities to each other, methanol and ethanol, for example, or, from the botanical world, *Heuchera* and *Heucherella*.

8.2.5 Stereochemistry

The ability of many substances to exist in two or more atomic configurations that are not superimposable on one another (stereoisomers) bedevils all forms of chemical nomenclature and very elegant solutions have been advanced to solve the problems raised. The INN approach to the challenge can scarcely be described as elegant, and is best regarded as pragmatic.

It starts from the premise that nonproprietary names for medicinal substances should not be encumbered with detached letters such as 'D' and 'L', or, more likely '*R*' and '*S*', the solution put forward many years

ago by Kahn and Ingold and described in depth elsewhere in this book. Solutions such as these are meaningful only to chemists while INN are intended to serve physicians whose needs are for names that are short and simple. However, to give quite different INN to two stereoisomers would not serve the medical profession well and would be deeply disturbing to chemists and pharmacologists. The latter group, in particular, is much concerned with the spatial geometry of molecules insofar as it plays a critical role in molecular interactions at receptor sites.

When a mixture of stereoisomers (a racemate) has already been assigned an INN and an INN is later required for one of the isomers, the solution has been to use the prefixes dextro- and levo- added to the existing INN. These prefixes are often shortened to dex- and lev-, according to needs dictated by the length of the existing name and overall euphony of the compound name. There are many examples: amfetamine, dexamfetamine, levamfetamine; propoxyphene, dextropropoxyphene, levopropoxyphene. The prefixes are used, of course, according to the direction of rotation of polarized light as it passes through a solution of the substance in a common solvent, usually water or dilute acid.

An advantage of this simplistic approach is that it provides equally for complex molecules exhibiting more than one centre of asymmetry (chiral centre). Regardless of the number of chiral centres a stereoisomer may contain it is most unlikely that the cumulative positive and negative effect of each chiral centre on polarized light will be an optical rotation of 0° and therefore the system will hold, especially as it will be very unlikely that more than two such isomers of any particular substance will prove to be of medicinal interest.

Life, however, is seldom that simple and there are several examples known where an acidic substance, for example, may exhibit 'dextro' rotation as a free acid but 'levo' rotation as a salt. This inversion of behaviour has no effect on the stereotaxonomy when the *R/S*-sytem is used in systematic chemical nomenclature but it does rather leave the INN system looking foolish if a sodium salt of a free acid (and the salt form is that in which the majority of acids will be used in medicine) has 'levo' rotation when the free acid form has 'dextro'. Regrettably, this effect is often not recorded until after a name has been assigned when it may be too late to change but, in cases where this behaviour has been observed and made known to the INN secretariat, steps will be taken to avoid the misfortune. One such example is the anti-inflammatory and racemic acid known as 'flurbiprofen'; because salts of this substance exhibit opposite effects on polarized light to that of the free acid, the *S*-isomer has been named 'esflurbiprofen' rather than 'levoflurbiprofen'. Similarly, the prefix *ar*- would be used for any *R*-isomer displaying this irregular property.

The case of a stereoisomer of a racemic substance already possessing an INN has been discussed but what of the position when the substance

being named is known to be a stereoisomer? Or that in which there is a need to assign a name for a racemic substance where a name has already been given to a stereoisomer thereof? The latter case would appear most unlikely but there has been indeed more than one example, that of ephedrine being the best-known case in point. Ephedrine is stereospecific and is one of four stereoisomers, there being two chiral centres in the molecule. Recently, medical interest arose in one of the two racemic pairs of optically inactive isomers whereupon the name 'racephedrine' was assigned; a simple solution, but effective nonetheless. In the example of an INN being required for a new substance known to be sterically active, no special account is taken. If it were, there would be a very large number of names all beginning dex- or lev-, and one of the primary rules of the INN system (names should be short and **distinct**) would be in danger of violation. Time enough, so it is argued, to worry about a prefix when a subsequent racemate or stereoisomer comes along.

8.2.6 Modified INN

With a few exceptions INN are always assigned to parent acids, bases and alcohols and not to their salts and esters. The idea behind this is to give the manufacturer the freedom to select among the salts and esters of biologically active acids and bases without the need to apply for an INN for each example. This saves WHO administrative resource and suits both parties. Such compound names are known as Modified International Non-proprietary Names (INNM) and it is this form of the INN rather than the INN itself that will feature in product labelling, documentation and in pharmacopoeial monograph titles.

8.2.7 Radicals and groups

To facilitate the INNM concept, names have been assigned to esterifying groups where there is no common name and the systematic name is inconveniently long. For example, the 1-acetoxyethyl ester of cefuroxime is known as 'cefuroxime axetil'. A representative list of these specially designed names for radicals and groups is shown in Table 8.2.

A similar approach is used for salts of bases where it is even more likely that the active moiety will reside in the free base. Basic substances are frequently present as hydrochlorides, tartrates, etc and the reason is that the salts are often chemically more stable than the corresponding free bases and will usually be far more soluble in body fluids, enhancing the bioavailability of the drug.

Unfortunately, the naming of esters is not quite so clear cut as these notes may initially suggest. No consistent policy has been used for esters of carboxylic acids. A dilemma exists in that, on the one hand, the

Table 8.2 Examples of radical and group names

Name	Chemical definition
acetonide	isopropylidene ether of a dihydric alcohol
aceturate	*N*-acetylglycinate
amsonate	4,4′-diaminostilbene-2,2′-disulphonate
axetil	l-acetoxyethyl
besilate	benzenesulphonate
camsilate	camphor-10-sulphonate
cilexetil	(*RS*)-1-[(cyclohexyloxy)carbonyloxy]ethyl
closilate	4-chlorobenzenesulphonate
cromesilate	6,7-dihydroxycoumarin-4-methanesulphonate
cipionate	β-cyclopentylpropionate
dipivoxil	(2,2-dimethyl-1-oxopropoxy)methyl
edisilate	ethane-1,2-disulphonate
embonate	4,4′-methylenebis(3-hydroxy-2-naphthoate)
enanthate	heptanoate
erbumine	*tert*-butylamine
esilate	ethanesulphonate
etabonate	ethoxycarbonyloxy
gluceptate	glucoheptonate
isetionate	2-hydroxyethanesulphonate
megallate	3,4,5-trimethoxybenzoate
mesilate	methanesulphonate
mofetil	2-(morpholino)ethyl
napadisilate	naphthalene-1,5-disulphonate
napsilate	naphthalene-2-sulphonate
pivalate	trimethylacetate
pivoxetil	1-(methoxy-2-methylpropionyloxy)ethyl
proxetil	1-[(isopropoxycarbonyl)oxylethyl
suleptanate	carbamoyl]heptanoyl
theoclate	8-chlorotheophyllinate
tosilate	toluene-4-sulphonate
troxundate	3,6,9-trioxaundecanoate
xinafoate	1-hydroxy-2-naphthoate

physician needs short, memorable names and, on the other, a rational approach to names for esters inevitably requires two-word names. The points in favour of the two-word name approach are firstly that the name of the parent acid will be shown and as this is likely to be the active principle, the physician is left in no doubt that the substance is an acid with which he is already familiar modified by esterification to impart chemical stability or to increase bioavailability through enhanced solubility in biolipids. Secondly, it avoids 'wasting' a name entirely different from that of the parent acid. Single-word names for esters may give the impression that an entirely new substance is to hand rather than a modification of an existing acid. Again, names have been created to serve as alkylating radicals in names for esters of acids alongside familiar radicals such as methyl and ethyl.

8.2.8 Recombinant biological substances

There are many natural biological products used in medicine. Materials such as insulin, growth hormone and blood coagulation factor VIII have been known and used for a long time and over the years have acquired a colloquial nomenclature that serves them very well. In many instances, they are mixtures of proteins or glycoproteins, often with one particular component predominating. As mixtures, these natural materials do not fulfil a basic requirement for an INN and accordingly few such products are known by an INN.

In contrast, the advent of substances obtained from recombinant biotechnology has created a requirement for a nomenclature system especially designed to accommodate the subtle inflections in molecular structure of which these materials are capable. In the case of biosynthetic proteins where the amino acid sequence is usually known, the normal guidelines of the INN system can be applied. Stems have been selected to identify substances with specific target actions with sub-groups established wherever necessary as, for example, in the blood coagulation factors (Table 8.3).

Glycosylated proteins present an added level of complication in that substances can have an identical amino acid sequence yet be glycosylated at different sites (which may or may not be known) with sugar residues that themselves may or may not be identified. Such analogues can be expected to display similar therapeutic activity yet possess a somewhat different biological profile from each other. It is therefore necessary to be able to distinguish between two or more such materials. This is achieved by using Greek letters spelt out as part of a two-word name: *epoetin alfa* and *epoetin beta*.

There are inevitably exceptions to every rule and the particularly complex interferon family is a case in point. Here, three main classes have been established, interferon alfa, interferon beta and interferon gamma. In this case the amino acid sequence is different in each case. Then, within each class, further biosynthetic variation in the chain termini is indicated by means of alphanumeric suffixes. Thus, we have interferon beta-1a, interferon beta-1b, interferon gamma-1a and interferon beta-1b. Commercial products containing each of these variants have been marketed in recent years.

However, the greatest challenge to the INN system has undoubtedly come from a group of materials known as monoclonal antibodies. These materials, some of which have been described in the popular scientific press as 'magic bullets', defy definition in simple biochemical terms but they are well-enough characterized to be worthy of INN assignment. They exhibit a wide range of therapeutic effects and are showing encouraging activity in the cancer field where their ability to seek out a particular physiological target is exploited.

Table 8.3 Stems used for products of recombinant biotechnology

Class	Stem	Example
erythropoietin derivatives	-poetin	epoetin alfa
blood coagulation factors	-cog	
factor VII	(-)eptacog	eptacog alfa
factor VIII	(-)octacog	
factor IX	(-)-nonacog	
colony stimulating factors (CSF)	-stim	
granulocyte CSF	-grastim	nartogastrim
granulocyte macrophage CSF	-gramostim	molgramostim
macrophage CSF	-mostim	
combination of above	-distim	milodistim
growth factors (GF)	-ermin	
epidermal GF	-dermin	murodermin
fibrinoblast	-fermin	ersofermin
tumour necrosis factor (TNF)	-nermin	
platelet-derived GF	-plermin	becaplermin
insulin-like GF	-sermin	
transforming GF	-termin	
growth hormones	som-	
bovine-type	-bove	sometribove
porcine-type	-por	somalapor
salmon-type	-salm	sometosalm
hormone-release stimulating factors	-relin	
growth hormone type	-morelin	
thyrotropin type	-tirelin	protirelin
hormone-release inhibiting peptdes	-relix	
interleukins	-kin	
type II-1	-nakin	pifonakin
type II-2	-leukin	aldesleukin
type II-3	-plestim	
type II-6	-exakin	
type II-8	-octakin	
type II-11	-elvekin	oprelvekin
interleukin receptor antagonists	-kinra	
type II-1	-nakinra	anakinra
pituitary hormones	-tropin	
follicle stimulating hormones	(-)follitropin	
luteinizing hormones	(-)lutropin	

Generic drug nomenclature has to be systematic as far as is possible; otherwise, it would be unable to convey the meaningful nuances to health professionals that is part of its value. The system that has evolved to serve the monoclonal antibodies is inevitably complex and results in names which seem quite extraordinary: consider 'abciximab', for example. The system is based on a compound stem comprising -'mab', to signify the general class, a second element (commonly a single letter) to indicate the source of the cell-line used in biosynthesis, and a third element to show the target disease or site of action. Finally, a prefix with no particular significance is added in an attempt to complete a name by which the material can be known. Table 8.4 sets out the stem system and gives some examples of the names produced.

Table 8.4 Stems used in monoclonal antibodies

General stem	Sub-stem for source of product		Sub-stem for disease or target	
-mab	human	-u-	cardiovascular	-ci(r)-
	rat	-a-	immunomodulator	-li(m)-
	hamster	-e-	infectious lesions	-le(s)-
	primate	-i-	viral	-vi(r)-
	mouse	-u-		
	chimeras	-xi-	colon	-co(l)-
	humanised	-zu-	testis	-go(T)-
			ovary	-go(v)-
			mammary	-ma(r)-
			melanoma	-me(r)-
			prostate	-pr(o)-
			miscellaneous	-tu(m)-

Examples:
abciximab; bicirumab; enlimomab; felivizumab; satumomab

Most of the names in this field produced so far appear strange to the English tongue but once the underlying concepts are appreciated the names become really meaningful. Nevertheless, it is to be expected that most of the commercial monoclonal antibodies are likely to be better known by their trade name. Unlike conventional synthetic compounds, which, once the patent has expired, may be freely manufactured by anyone in the almost certain knowledge that for all practical purposes they will be identical to the original substances (at which point the need for a non-proprietary designation such as an INN becomes increasingly necessary), products of recombinant biotechnology are likely to be company-specific. Because the products obtained will be very dependent on the precise conditions of manufacture, it is not difficult to understand why patent infringement litigation has bedevilled development of new products in this area. Moreover, in such cases, it can be argued, just one name can serve to identify the material. If the INN is going to be company-specific, it is effectively acting as a trade mark. Why burden health professionals with two names? Manufacturers can be expected to put that argument in reverse!

8.2.9 Mixtures

The terms of reference on which the INN system is based charges the WHO to concern itself with nonproprietary names for *substances*; this excludes formulated mixtures. However, many materials used in medicine are not pure, single substances and may well be mixtures of closely related substances, the separation of which is neither economic nor necessary from a efficacy standpoint. A medicinal substance will usually be used commercially at a purity level of about 98–99% and the impurities will

most likely be structurally related to the substance itself. It is not this form of 'mixture' that concerns us here but rather mixtures of structural isomers, close chemical homologues or stereo- or geometric isomers. These mixtures may be natural products of botanical or biological origin or materials derived from chemical synthesis that are processed in such a way that a product of consistent proportions and biological activity is obtained. In such cases INN are assigned secure in the knowledge that a rigorous definition can be provided. Examples include the antibiotics neomycin and gentamicin (homologues), the anti-infective agent halquinol (structural isomers) and the oestrogen receptor antagonist, clomifene (geometric isomers). In cases where the description of the mixture supplied at the time of application is ill-defined the usual fate is rejection.

Formulated mixtures are a different proposition. All requests for INN for such mixtures are firmly rejected but where a mixture of this kind has found favour with the medical fraternity and has been licensed by an appropriate authority it can be argued that a simple, single-word non-proprietary name should be established. It should not be necessary for health professionals to have to rely on a brand name to specify a medicinal product, especially as any one particular mixture may be known by more than one brand name.

In 1970, it was found that the antibacterial agents trimethoprim and sulfamethoxazole, in combination, exhibited an enhanced, synergistic effect when administered in the proportions 1 to 5. This discovery was developed jointly by the Wellcome and Roche companies and each had its own brand name for the mixture. In the UK, the nonproprietary descriptor co-trimoxazole was designated as a British Approved Name, the definition carefully specifying the 1:5 proportionality. Many years later it became politically expedient to create a set of names for other well-established products formulated as mixtures of fixed proportion. This device is particularly popular in the analgesic field where a combination of products can be expected to show an additive analgesic effect but with a lower likelihood of narcotic and other adverse reactions than from an equivalent dose of a single component. So that physicians would immediately recognize these names as names of combination products, each name began with the syllable co-. Thus a series of around 30 such names has been established in the UK, not only for analgesic combinations, but also for antibiotics (co-fluampicil: flucloxacillin + ampicillin, 1:1), antacids (co-magaldrox: magnesium and aluminium hydroxides in varying proportions) and diuretics (co-amilofruse: amiloride hydrochloride + furosemide, 1:8). Many of these names have become very familiar, co-proxamol (dextropropoxyphene hydrochloride + paracetamol, 1:10) perhaps being the best known. Nonproprietary names encourage the manufacture and marketing of pharmaceutical products in a generic form when the costs of brand name registration and support, and advertising are either not

applicable or minimal. The resulting savings in costs are reflected in lower pricing relative to branded forms of the same combination.

8.2.10 The trade mark interface and the protection of INN

While INN are assigned to medicinal substances, pharmaceutical trade marks are associated with medicinal products. Despite this difference, it is generally held that INN and trade marks should not be liable to confusion with each other. There are approximately 7500 INNs in existence but tens, if not hundreds, of thousands of trade marks registered throughout the world and in consequence it grows ever more difficult for manufacturers to find distinctive new trade marks. Brevity is of the essence. Think of Zantac, or Nurofen, or Aspro. Two-syllable marks are ideal, three more the norm, four very occasionally and five or more, scarcely ever. In contrast, an analysis of INN shows a massive peak at four-syllables in length, with three and five fairly common, two very seldom and six or more not entirely unknown.

From this, it might be deduced that INN and trade marks should sit comfortably alongside each other with little interference. However, this is not the case. There is an undercurrent of antagonism and mistrust between the two camps which is only kept in check by careful attention to the interface. Manufacturers have a commercial interest in encouraging the prescribing and dispensing of their products by means of trade marks. The value of a distinctive trade mark for a successful product is of paramount importance to a company. The quality and reputation of that product rubs off on other products from the same company. On the other hand the UK government, for example, has an obligation to its tax payers to keep a watchful eye on the cost of medicines dispensed under the national health service and this is best kept under some control by the use of medicines sold under, and prescribed using, an INN rather than a trade name – generic manufacture and prescribing.

INN are only attractive as a means of prescribing if they are kept to three or four syllables in length. To discourage their use some manufacturers prefer to see the length of the INN of their active substances substantially longer than the corresponding trade name. Proposals for new INN put forward by innovating companies are frequently six or even seven syllables long, even though the other requirements laid down in the guiding principles may have been followed. Such proposals are invariably rejected.

By focusing on three and four syllables as the optimum length of an INN, the system finds itself frequently in competition with trade mark creators for whom marks of this length are essential. However, the INN system is heavily constrained by its own rules. It does not possess the total freedom enjoyed by trade-mark creators but it is often found that this freedom is not exercised. There exists a school of thought that believes

there to be advantages to health professionals if a trade mark and the corresponding INN are not dissimilar to each other. Consider Paramol and paracetamol, for example (there are numerous other examples). The protagonists argue that a superficial similarity eases the memory burden for the prescriber and reduces the likelihood of the wrong product being dispensed. The antagonists reason that trade marks like Paramol are not distinctive if they resemble the INN and therefore do not meet an essential feature of a trade mark – to be distinctive. Moreover, if the names are to be similar to each other what purpose does the trade mark serve? Why not abandon the trade mark and just use the INN? This discussion is, and has been, the stuff of conferences and here is not the place to pursue this debate. However, there exists a more insidious aspect to this interface that we must explore.

In creating trade marks similar to the INN, which usually will have been established first, companies sometimes, unwittingly or otherwise, select a mark that incorporates the stem of the INN. Consider the trade mark 'Acepril' and the corresponding INN 'captopril'. This case is a good example of a serious concern held by the INN secretariat. The stem -pril is used in INN to indicate angiotensin-converting enzyme inhibitors and by using the stem in a trade mark the initial impression is conveyed to health professionals that Acepril is not a trade mark for captopril but rather an INN in its own right, particularly as the prefix ace- bears chemical structure connotations and has often been used as a prefix in INN. Furthermore, Acepril as a trade mark blocks the design of further INN in the -pril series. It is this latter point that so concerns the INN secretariat. The general issue has been discussed twice at the World Health Assembly (WHA) convened annually at WHO headquarters in Geneva, most recently in 1993 when Resolution WHA46.19 was passed. The Assembly thereby requested member states:

1. to enact rules or regulations, as necessary, to ensure that international nonproprietary names (or the equivalent nationally approved generic names) used in the labelling and advertising of pharmaceutical products are always displayed prominently;
2. to encourage manufacturers to rely on their corporate name and the international nonproprietary names, rather than on trade marks, to promote and market multi-source products introduced after patent expiration;
3. to develop policy guidelines on the use and protection of international nonproprietary names and to discourage the use of names derived from INN, and in particular names including established INN stems as trade marks.

Resolutions passed by the WHA and recommendations put forward by Expert Panels in WHO Technical Reports are just recommendations and carry no mandatory powers. It is the responsibility of individual member

states to take the necessary legislative action to enforce such recommendations. In the area of drug nomenclature in the UK it has always been the policy to regulate by persuasion rather than by obligation and there is no means in law of insisting that companies refrain from using trade names that incorporate an INN stem. However, the regulatory agency may well reject a licence application for a product bearing a trade name that uses an INN stem where the use of the product does not conform with the action or use implicit in the stem. The grounds for rejection would be that the product name would be misleading to health professionals, possible dangerously so. WHO takes a similar dim view of all trade marks that borrow too heavily from the INN regardless of whether the stem is used.

8.2.11 Computer support

The INN program is well supported by computerized services. One particular database focuses on the INN itself. For each INN the following fields have been established:

Serial number
INN – Latin
INN – English
INN – French
INN – Spanish
Proposed List No.
Recommended List No.
Molecular formulae
CAS Registry Number
Marketing Status
Vowel sequence

The vowel-sequence (e.g. paracetamol: 'a-a-e-a-o') field is particularly useful in that it provides a means by which 'sound-alike' comparisons can be made. It is a simple matter to seek out all names beginning with the same syllable but a change in the initial consonant, for example, will produce a name that is not so easy to pick up. Any name with such a minor difference would clearly constitute a conflict. An alphabetical sort on vowel-sequence field pulls together all names containing the same vowel sequence. Any pair of names containing the same vowel sequence may well be too close in handwriting or speech to each other even though the constituent syllables may be entirely different. Consider, for example, the INN 'labetalol' and the fictitious name 'fadelafod'. In poor handwriting, these names could appear very similar indeed and yet a syllable-based search would be unlikely to throw them up as a conflict. However they contain an identical vowel sequence, a-e-a-o, and, by using this search facility, the conflict would be revealed.

Another very useful analytical tool is provided by what the secretariat calls a 'letter-by-letter' breakdown. Each INN is recorded beginning with each letter in the name in turn:

paracetamol
aracetamol p
racetamol pa
acetamol par
cetamol para
etamol parac
amol paracet
mol paraceta
ol paracetam

In this way the 7500 INN list produces a table with over 70 000 records which, when sorted, brings all names containing key syllables together. In this way, the INN list can be supplied in a much more useful form as a word-processor file to interested parties without the need to supply the full database.

A second table in this INN database contains technical data such as the systematic chemical name, the molecular formula, the Chemical Abstracts Registry Number and, in an object field, the graphic formula. The data is held in each of the languages in which the INN lists are published (English, French and Spanish). Using database publishing software, such as Infopublisher/PageMaker or Corel Ventura, enables lists of Proposed and Recommended INN to be typeset within the secretariat office with considerable savings in time and cost of external publishing. Processing INN applications and publishing the lists in this way began 3 or 4 years ago.

8.2.12 Summary

This has been a summary account of the INN system that has been in operation for more than 40 years during which time it has served the pharmaceutical industry extremely well. The principles on which it is built have stood the test of time and have proved sufficiently flexible to meet the demands of most forms of molecular and biological innovation it has met. It should continue to provide a rational system of generic drug nomenclature well into the 21st century.

8.3 The ISO system

Many of the general principles of the INN system apply equally to the selection of names for pesticides and other agrochemicals but there are a

number of important differences. Rather than repeat much of the above discussion this section will focus on these differences.

8.3.1 Constitution

The adoption of names in this category is administered by the International Organization for Standardization (ISO) through one of its Technical Committees (ISO/TC 81) but in a markedly different way from that which WHO administers the INN system. ISO does not itself undertake the secretariat activities that play an essential role in co-ordinating international activity in pesticide nomenclature. Rather, the ISO/TC 81 administration is undertaken on its behalf by one of the member states of the ISO. At the present time, and indeed since the inception of the system in 1953, the secretariatship has been held by the UK in the form of the British Standards Organization (BSI). One of BSI's technical committees (currently styled AW/81) provides the hub of the whole system. ISO's role is two-fold: firstly, it convenes plenary meetings at appropriate intervals of all its member states who choose to participate in the function, and secondly, it publishes lists of selected names, known formally as ISO Common Names. The last cumulative list of ISO Common Names was published as ISO Standard 1750:1981; this has been supplemented by an Amendment in 1983, and by Addenda 1 and 2 also published in 1983. Addenda 3 to 8 have only been issued in draft form and Addendum 9 has not yet been issued.

Secretariat and committee activities Proposals for names for new substances with potential pesticidal, herbicidal or other form of agrochemical activity are put forward not directly by the innovating company but usually by the national standards organization in the country in which the company is headquartered. That standards organization is the regarded as the 'sponsor'. The application consists of the expected chemical data (structure, systematic name, molecular formula, CAS number) together with laboratory code numbers, trade name (if chosen) and, most importantly, the target activity. In addition, and this is a very significant difference from the INN system, the company, through its sponsor, is required to provide a copy of trade mark searches conducted in the trade mark registries of appropriate countries. This is to show that there are no immediately apparent conflicts with trade marks registered in relevant classes, normally the agrochemical and pharmaceutical classes. The latter includes toiletries and veterinary medicines.

Applications are not subjected to a vigorous technical screening by the secretariat. Instead, they are sent to BSI technical committee AW/81 for consideration at regularly held meetings. The committee is comprised of representatives of appropriate scientific and agrochemical trade associations

and from public service. Here, the proposals are scrutinized to see that the basic guidelines have been followed. These are laid down in another ISO Standard, namely ISO 257:1988 and are broadly similar to those underpinning the INN system with the exception of far less emphasis on the need to show similarity to existing substances through the use of common stems. BSI had previously issued a national version of the guiding principles as BS 1831:Part 1:1985.

The committee either tacitly agrees the proposal, recommends that further data be sought from the sponsor or rejects the application in which case the company is asked to put forward a revised proposal. The committee neither puts forward modifications to the proposal nor suggestions that it might wish to make. This is in stark contrast to the INN system where the secretariat provides a lead to its panel of experts when it foresees that a manufacturer's proposal is unlikely to prove unacceptable. But the difference here is that most of the INN work is done by correspondence between a panel of international experts and the co-ordinating secretariat, whereas the ISO system is able to function through the agency of regular meetings of the BSI AW/81 committee.

Trade-mark searches The initial trade mark search is central to the ISO system. While it might seem sensible for sponsors to supply evidence that a proposal is clear of such impediment, the requirement is something of a two-edged sword. The downside is that it inhibits the secretariat committee from making even the slightest adjustment to a proposal that might render it ultimately acceptable. A proposal acceptable to the committee can be sent on its way to the next hurdle, but if an amendment is required, it must be returned to the sponsor, from the sponsor to the company and back again. Even in this technological age of electronic mail and fax, these operations can take an inordinate length of time, resulting in considerable delay to the processing of the application. This inhibition centres, of course, on the likely invalidation of expensive trade mark searches that would arise if the original name was amended. In contrast, WHO takes the view that if you look hard enough in every country of the world you will find a conflict for every name and no INN would ever be published. It therefore chooses to place the onus on interested parties to scrutinize its lists of proposed INN and to raise objections if intellectual property rights have been compromised. The number of such objections is extremely small.

Public enquiry Given that a proposal is acceptable to the secretariat committee, it is then sent together with relevant technical details to those ISO member bodies that have elected to concern themselves with this aspect of the Organization's work. The process is known as 'public enquiry'. A period of 3 to 4 months is allowed for comment and during

this time the participating states will either arrange for one of their national standards committees to review the proposal or for an appropriate national expert to comment. Comments received from the member states are then examined by the secretariat BSI committee. Comments may take the form of criticism of the proposal on the grounds, for example, that a conflicting trade mark exists in one particular country or that an ambiguity in the systematic chemical name has been spotted. Any serious problem raised at this stage will necessitate referral back to the sponsor, with the possible need for the application to start afresh.

Draft addendum and letter ballot Once the participating countries have signified acceptance of a name, the secretariat adds it to the current draft addendum. The document is prepared bilingually, in English and French. It contains:

- the proposed common name in English, French and Russian with footnotes giving local spelling variations and countries where the ISO name is not acceptable,
- the systematic chemical name in English (IUPAC), French (UICPA) and according to Chemical Abstracts usage,
- the use, classified according to a system defined in ISO 257,
- sundry information as necessary.

The draft lists are then sent back to the participating states with the clear instruction that formal adoption will take place. This process is known as 'letter ballot'. If any state declares at this stage that a name is not acceptable in that particular country, then the name will be published with a footnote to the effect that it is not accepted in that country. Once the process of letter ballot is complete, the proposal may be considered to have become an ISO Common Name.

There is a degree of concern felt in some quarters at the length of time taken from initial application to publication as an ISO Common Name in an addendum or draft addendum, which may take several years. Products may be launched commercially well before negotiations for the Common Name are complete which is not a very satisfactory state of affairs. INN, on the other hand, generally take 6–12 months from application to publication and manufacturers are actively discouraged from using a name put forward as an INN proposal before publication is complete.

Guidelines Guidance to manufacturers in the form of ground rules to be followed is provided in ISO Standard 257: 1988.

Spelling and stems. Other than a requirement to avoid 'ph' in proposals (except when 'phenyl' is required as a radical name) no restrictions on spelling are imposed in contrast to INN where the letter

Table 8.5 Examples of ISO stems

Stem	Definition	Example
-alin	2,6-dinitroanilines	trifluralin
carb-, -carb- or -carb	carbamates and thiocarbamates	carbofuran
-conazole	fungicides and plant growth regulators based on imidazole or 1,2,4-triazole and containing an halogenated phenyl group	penconazole
-mectin	analogues of avermectin	abamectin
-meton	methoxy-substituted 1,3,5-triazines	secbumeton
-thrin	cyclopropanecarboxylic acids (pyrethrins)	permethrin
-uron	acyclic ureas in which one or both nitrogen atoms form part of a saturated ring system	linuron

sequences 'th' and 'ph' are rigorously forbidden and where 'y' used as a vowel is always replaced by 'i'. Again in contrast to the INN system, there has been much less reliance placed on the use of common stems to relate substances of similar structure and activity but in recent years several such series have been established, probably the best known being the -uron series of acyclic ureas and the -conazole series of antifungal agents (Table 8.5). The latter provides a sensible link with pharmaceutical antifungal agents containing the same basic structure and which have also been named as -conazoles. The pharmaceutical sector is in fact represented on the BSI secretariat committee in the hope of promoting common understanding of the two nomenclature systems and to ensure that where appropriate there is harmony between them.

Salts and esters. Harmony between the ISO and INN systems does exist in the approach to salts and esters. Usually the pesticide sector is concerned with acidic substances and the committee encourages manufacturers to specify which salt has been used in the formulation. Likewise, for esters, the manufacturer is required to specify the ester by adding its name as an hyphenated suffix, e.g. 'benzoylprop-ethyl'. As in the INN system, names have been adopted for ester groups and for alkyl radicals where a simple name is not available under the IUPAC system, simply to facilitate assigning two-word names to esters (Table 8.6).

Stereochemistry. Stereochemistry plays just as an important role in agrochemical reaction systems as it does in pharmaceuticals but the approach taken to distinguishing stereoisomers in ISO Common Names contrasts starkly with the method used in INNs. It will be recalled that the INN system relies on the use of the prefixes dextro- and levo- (sometimes truncated to dex- and lev-) affixed to an existing INN for the racemate according to the effect on polarized light of a solution of the stereoisomer in a common solvent. The ISO system uses a pair of suffixed letters, -P and

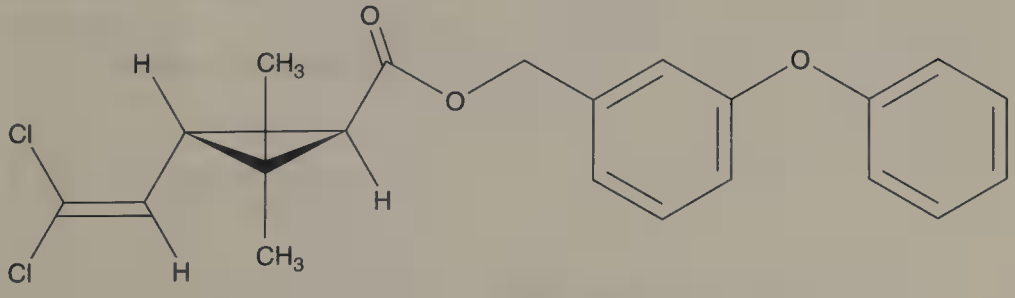
Table 8.6 Examples of names for ions and radicals used in ISO common names

Name	Chemical Definition
butometyl	2-butoxy-1-methylethyl
diclexine	dicyclohexylammonium
dimolamine	(2-hydroxyethyl)dimethylammonium
etotyl	2-ethoxyethyl
meptyl	1-methylheptyl
tefuryl	tetrahydrofurfuryl

-M or -PM together to achieve the same distinction. If a name is required for a stereochemical variant of an existing substance 'P' is used for a (+)-isomer, 'M' for a (–)-isomer and 'PM' for a racemate. In contrast to the INN prefixes, the P/M system is designed for substances for a single chiral centre. ISO 257 provides for more complex cases by allowing the creation of special local systems, as, for example, in the synthetic pyrethroids.

Pyrethroids and related compounds. Just as the INN has had to rise to a challenge, that presented by the complexity of the monoclonal antibodies for example, so the ISO system has had to devise a system of simple nomenclature to account for the complex stereochemistry that characterizes the family of synthetic pyrethroid insecticides.

Substitution at two positions in cyclopropane produces a *cis/trans* isomerism that gives rise to four stereoisomers. In the pyrethroids one of the substituents is an ester function, the radical of which commonly includes a chiral carbon atom and the other substituent is generally dichloro- or dibromovinyl but the possibility for side-chain *cis/trans* isomerism also arises. Thus, any one molecule can exist in a considerable number of steric variants. In practice, the early products in this field often consisted of mixtures of the cyclopropane-centred *cis/trans* isomers and the manufacturer when labelling products was required to specify the approximate proportions. Subsequently, other stereoisomers or mixtures of a substance to which an ISO name had already been assigned began to appear either as the result of different manufacturing processes or by design.



(1*R*,3*S*)- or (1*R*)-*trans* permethrin.

After a great deal of discussion at national and international level, including consultations with the Rothamsted chemists who had researched the area, a system was devised whereby Greek letters, spelt out, were used as prefixes. A new compound in the series receives a new name chosen in accordance with the principles laid down in ISO 257, in this case a name based on the stem '-thrin'. However, future compounds that differ only in their stereochemistry from a substance to which a name has already been assigned are given that name with a Greek prefix attached. The choice of Greek letters is restricted to: alpha, gamma, zeta, eta, theta, iota, kappa, lambda, sigma, tau, omega.

Each prefix is used once in a random manner. The manufacturer of a new isomer is to consult the BSI secretariat and choose from those prefixes which are available at any particular time, subject to the following condition. If a prefix has already been used to indicate a particular isomeric configuration or mixture, that prefix should be used with the same stereospecific significance in respect of any future 'root name'. Thus if, for example, the name omega-permethrin had been chosen for the (1*R*)-*cis*-isomer of permethrin, and if a name were subsequently required for the (1*R*)-*cis*-isomer of tetramethrin it would be omega-tetramethrin.

Summary This somewhat shorter account of the ISO system used for pesticide and agrochemicals nomenclature has drawn contrasts with the INN system by showing the ways in which the underlying philosophy and administration differ from each other. In reflecting on these differences, it must be borne in mind that the two systems serve markedly different user bases. On the one hand, there is the prescribing physician, stereotypically with very poor handwriting, and on the other the agricultural community. Both groups need different features in the names of the products they work with and in their own ways both the INN and ISO systems successfully meet those needs.

References

1. *Chron. Wld. Health Org.* **7**:299 (1953).
2. *Chron. Wld. Health Org.* **9**:185 (1955).
3. *International Nonproprietary Names (INN) for Pharmaceutical Substances (Lists 1-73 of Proposed INN and Lists 1-35 of Recommended INN)*, World Health Organization, Geneva, 1996.
4. Document PHARM S/NOM 15, WHO, Geneva.

9 Computer-generated chemical nomenclature

S. B. WALKER

Despite the advent of technology that has enabled chemical structures to be drawn graphically on a VDU screen, making the structure 'understood' via a connection table, and programs that can enable the structure to be searched both at the structural and sub-structural level, there is still a need for chemical names for a variety of purposes. Compounds need to be cited as text in chemical patents, and in other forms of technical literature such as research papers. They may be the invention themselves or the intermediates, reactants and reagents that are used in the preparation. Whilst these may well be exemplified with chemical structures, either drawn as full structures or as a Markush structure in a tabular form, they will also be referred to by a chemical name.

Other forms of technical literature such as data sheets are more likely to describe the chemical via a chemical name than by a chemical structure.

Chemical suppliers, producing and making available to research chemists a range of chemical intermediates in their catalogues, do so mainly by using a chemical name, often exemplified by some structures; few give structures for all the available molecules. Some form of structure display can be achieved via a linear structural diagram but this is mainly to save space and can be confusing and ambiguous.

Names are also very much used in commerce. The official names coined for existing and new pharmaceutical substances by the WHO (INN) and for pesticides (ISO) require chemical names as part of their definitions. Names are still very much used in customs work and by purchasing agents.

Metabolites and by-products of important commercial chemicals are more likely to be described by their chemical name than by a structure.

Thus, names are important and whilst structures have become easily accessible in such tools as chemical databases, there is still a need for chemical names and these are usually carried alongside the structure in such products. Names can help clarify what could still be uncertain in a chemical structure, e.g. the geometrical chemistry at a double bond which is invariably drawn as *trans* (E form) in a linear chain but may just be *cis* (Z form) or often unspecified. Any form of drawing, other than a linear structure, is likely to imply some stereochemistry.

There are a range of styles of names used in conjunction with chemicals. There are two official names widely used, these are IUPAC (International Union of Pure and Applied Chemistry) and CAS (Chemical Abstracts

Service), each carefully derived by a set of rules. Even so, there is still much scope for the production of the name and both systems are evolving to accommodate new chemistry (e.g. fullerenes) and to rationalize the basis of their rules.

Aspects such as the treatment of stereochemistry, salts, still provide problematic areas for these official systems, particularly if mixtures are involved.

Trivial nomenclature still has a strong grip on the industry. In IUPAC an amino benzene will no doubt be called an aniline but in CAS nomenclature it is a benzenamine. However, you are unlikely to see a commercial chemical sold as anything other than an aniline.

Chemical acronyms have been and are still used to describe a number of chemicals. These can create difficulties, as in the authors experience there is no control on their use and duplicates can and do occur, e.g. TEA which is used for triethanolamine, triethylaniline, triethylaluminium and possibly others.

There are currently three computer packages that can help with the creation of chemical names. All of these claim to be based on IUPAC rules. However, the names created may use the rules of IUPAC to produce a name but the final names often do not appear to provide the style of name that would be used in practise. The three packages are:

Autonom™ – produced by Beilstein Information Systems in Germany
ACD Name™ produced by Advanced Chemical Developments Inc. in Canada
Nomenclator and NamExpert produced by ChemInnovation Software Inc. in the USA.

The first two produce a name from a drawn structure. The latter will also produce a structure from an input chemical name (Nomenclator™), although only where this is a relatively simple molecule. In all instances an adequate drawing package is provided to enable the required input of the structure and the import of structures from other chemical drawing packages is also allowed, especially .MOL files from MDL Information Systems ISIS™ Draw software. The latter and CamSoft's software ChemDraw™ are probably the most widely used chemical structure drawing packages, although many more are available and the drawing portion of the ACD package (ACD ChemSketch) is easy to use.

A nice feature of the ACD Name package is that a full explanation of the IUPAC rules used to produce the name are given and as portions of the structure are sensed, they become highlighted and the place where they occur in the name appears in colour and is underlined.

At this stage, I would describe the packages as a good aid to the derivation of a name, especially for the more complex molecules. For those who have a good knowledge of nomenclature and who have been used to

naming compounds over many years, the systems will help with providing a backbone to work with. For the novice, then a reasonable quality name will be produced. It may not, however, be the practical name that most nomenclature 'experts' would use.

Areas where I have found such packages are weak would be where trivial nomenclature is still very much used, e.g. the naming of an acetophenone, an orthoformate, a complex heterocyclic ring system, a pinane, an adamantane, a crown ether or crab, a trifluoromethyl group, a chalcone, an acetanilide, a penicillin or penem, a steroid, a prostaglandin, an organosilane, etc.

Each package has its own particular oddities. The Beilstein Autonom tends to sprinkle the name with unnecessary hyphens and naturally names are derived of a style similar to those found in the Beilstein handbooks. With all packages there is a tendency to name large hetero containing rings systems as hetero substitued carbocyclics. ACD Name tends to name compounds along the style of CAS.

Some of these (e.g. a crown ether) may be described as not part of systematic nomenclature, but as an alternative system. However, with the usage that such names already receive, a nomenclature system should be able to derive names in these styles and should offer some choice of style.

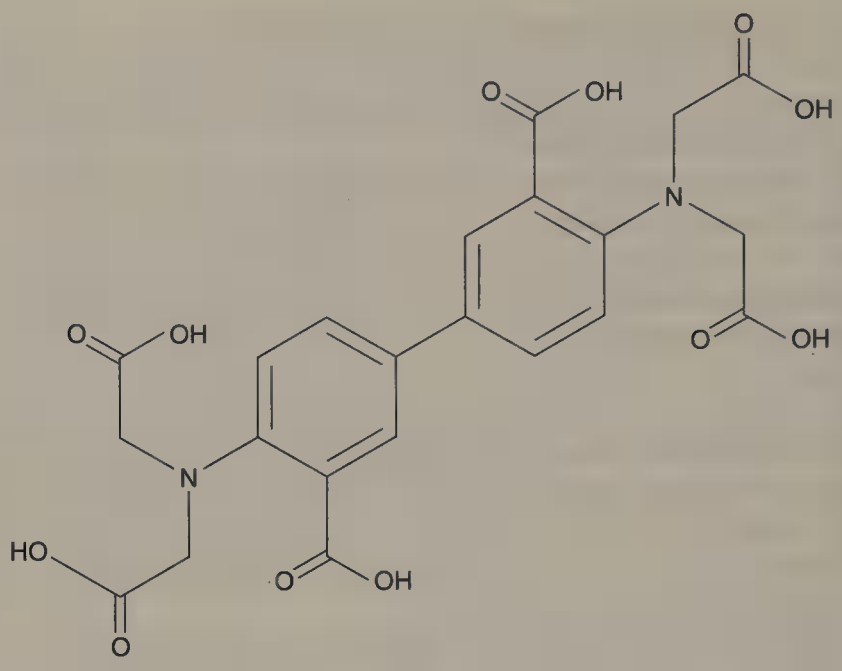
There is still scope for enhancements and improvements in the handling of stereochemistry, salts, mixtures, unspecified (often the drawing packages don't have any way of drawing these in the first place, (e.g. chlorodinitrobenzene).

Weak areas appear to be producing names of a style that is commercially used.

ACD Name does tend to produce a range of names from which the user can select the name closest to the style required. Stereochemistry of a structure is poor except in the ACD Name.

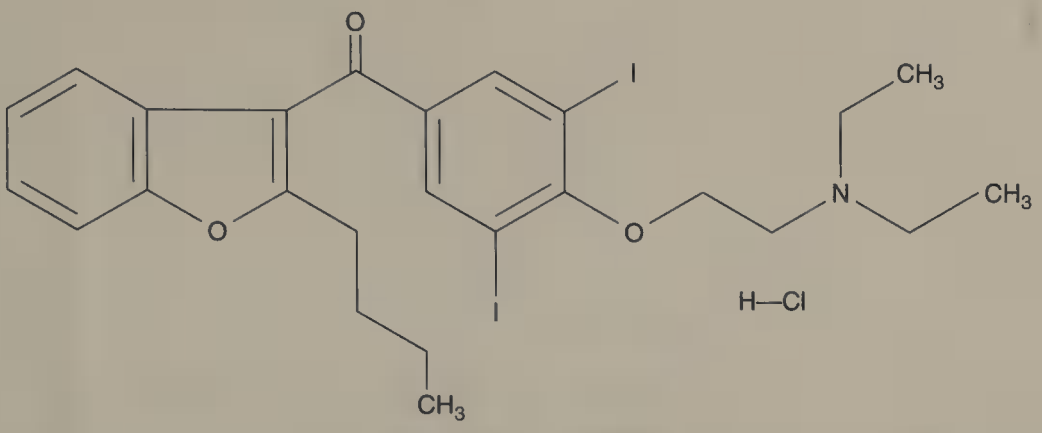
Obviously, all these packages are still at early stages of development and much more work needs to be done to make them suitable for all styles of usage. They are also relatively expensive.

Some examples of derived names and commercial names by which these substances are more generally known are given below.

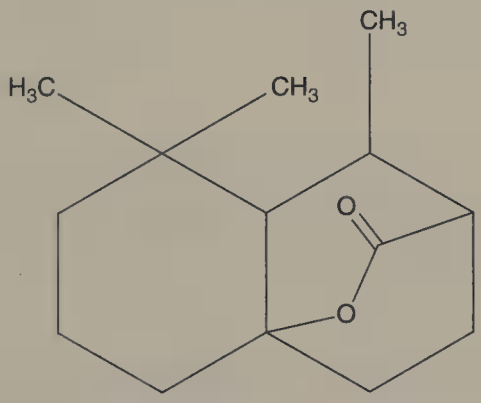


Beilstein Autonom name: 4,4′-Bis-(bis-carboxymethyl-amino)-biphenyl-3,3′-dicarboxylic acid
ACD Name: assemblies of cyclic systems are not supported in current version (2.51)
Commercial name: 3,3′-dicarboxybenzidine-*N*, *N*, *N*′,*N*′-tetraacetic acid

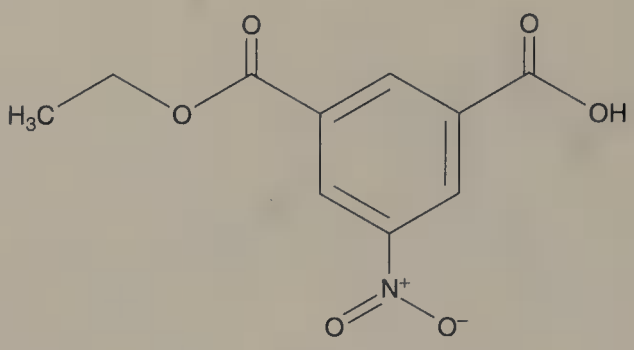
Beilstein name: 3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,10-Heptadecafluoro-dec-1-ene
ACD Name: 3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,10-heptadecafluoro-1-decene
Commercial names: 3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,10-heptadecafluoro-1-decene; 3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,10-heptadecafluorodec-1-ene; 1,1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8-heptadecafluoro-8-vinyloctane.



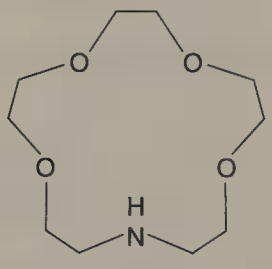
Beilstein Autonom name: (2-Butyl-benzofuran-3-yl)-[4-(2-diethylamino-ethoxy)-3,5-diiodo-phenyl]-methanone; hydrochloride
ACD Name: (2-butylbenzo[*b*]furan-3-yl)4-[2-(diethylamino)ethoxy]-3,5-diiodophenylmethanone
(ChemACD will not handle naming the salt and classes this as a mixture)



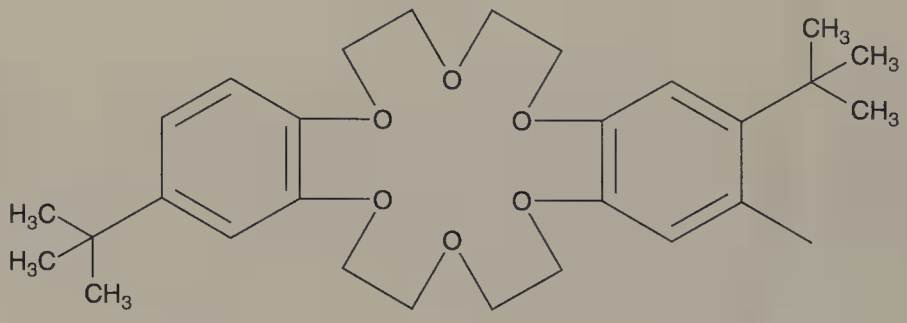
Beilstein Autonom name: 5,5,7-Trimethyl-10-oxa-tricyclo[6.2.2.$0^{1,6}$]dodecan-9-one
WARNING #7: Von Baeyer name suggested for a ring which is not yet in dictionary
ACD Name: 5,5,7-trimethyl-10-oxatricyclo[6.2.2.$0^{1,6}$]dodecan-9-one



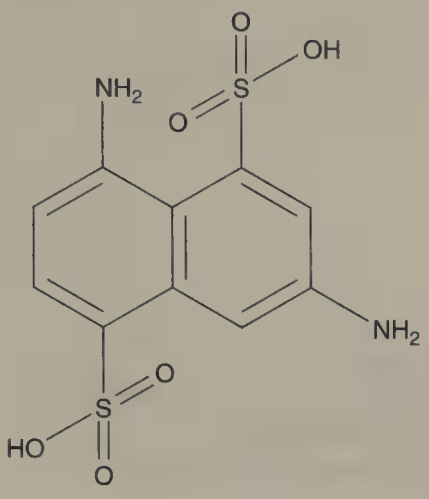
Beilstein Autonom name: 5-Nitro-isophthalic acid monoethyl ester
ACD Name: 3-(ethoxycarbonyl)-5-nitrobenzoic acid
Commercial names: monoethyl 5-nitro-isophthalate; ethyl hydrogen 5-Nitro-isophthalate.



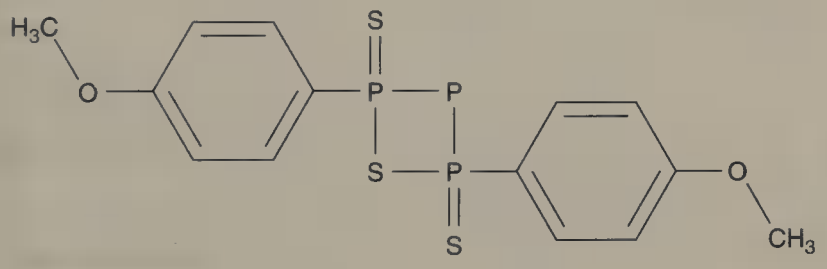
Beilstein Autonom name: 1,4,7,10-Tetraoxa-13-aza-cyclopentadecane
ACD Name: 1,4,7,10-tetraoxa-13-azacyclopentadecane
Commercial name: 1-aza-15-crown-5



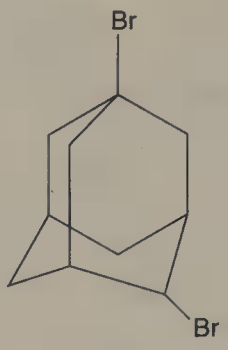
Beilstein Autonom name: 2,13-Di-*tert*-butyl-6,7,9,10,17,18,20,21-octa-hydro-5,8,11,16,19,22-hexaoxa-dibenzo[*a,j*]cyclooctadecene
WARNING #8: Replacement nomenclature used: fusion ring name might be better
ACD Name: 2,13-di(*tert*-butyl)-6,7,9,10,17,18,20,21-octahydrodibenzo-[*b,k*][1,4,7,10,13,16] hexaoxacyclooctadecine
Commercial name: 4,4′-di-*tert*-butyl-18-crown-6



Beilstein Autonom name: 3,8-Diamino-naphthalene-1,5-disulfonic acid
ACD Name: 3,8-diamino-1,5-naphthalenedisulfonic acid
Commercial names: 3,8-diamino-1,5-naphthalenedisulfonic acid; 3,8-diaminonaphthalene-1,5-disulfonic acid

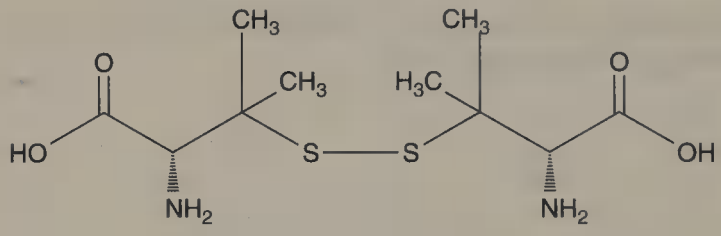


Beilstein Autonom name: 2,4-bis-(4-methoxy-phenyl)-[1,3,2,4]dithiadiphosphetane 2,4-disulfide
ACD Name: 2,4-di(4-methoxyphenyl)-1,3,2λ^5,4λ^5-dithiadiphosphetane-2,4-dithione
Commercial names: 2,4-bis(4-methoxyphenyl)-2,4-dithioxo-1,3,2,4-dithiadiphosphetane;
Lawesson's Reagent

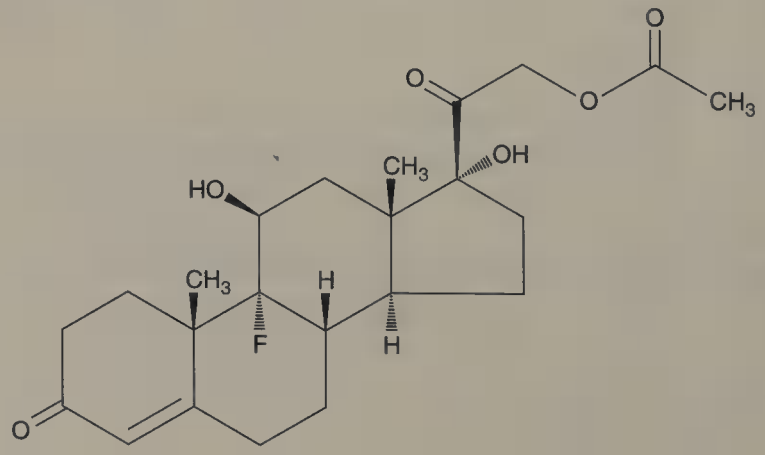


Beilstein Autonom name: 1,4-Dibromo-adamantane
ACD Name: 1,4-dibromoadamantane
Commercial name: 1,4-dibromoadamantane

It is interesting that both systems recognize an adamanatane ring structure and don't apply a bridged ring name to the compound.



Beilstein Autonom name: 2-Amino-3-(2-amino-2-carboxy-1,1-dimethyl-ethyldisulfanyl)-3-methyl-butyric acid
WARNING #1: Sterochemical information not coded into the name.
WARNING #3: Alphabetic order of prefixes ignored while selecting parent chain
ACD Name: (2*S*)-2-amino-3-[(2*R*)-2-amino-2-carboxy-1,1-dimethylethyl]-disulfanyl-3-methylbutanoic acid
Commercial names: 3,3′-dithio-bis-L-valine;
3,3,3′,3′-tetramethylcystine;
penicillamine disulfide;
3,3′-dithiodi-L-valine



Index

Chemical Nomenclature

Edited by K. J. Thurlow

Chemical nomenclature can be a complicated subject. As a result, most works on the subject are rather dry textbooks and primarily consist of sets of instructions on how to name chemicals. This practical book proves that chemical nomenclature can be interesting, not just a 'necessary evil'.

Written in a lively and engaging style by experts in their particular fields, this new book provides a general discussion on why good, clear nomenclature is needed. It introduces the reader to the various forms of nomenclature without reading like a textbook. Both 'systematic' and 'trivial' nomenclature systems are used widely (and interchangeably) in chemistry and this new book covers both areas. For example, systematic nomenclature in both the CAS and IUPAC styles is introduced. These systems have many similarities but important differences which the chemist should be aware of. Specialised naming systems are needed for polymers and natural products and these areas are covered in separate chapters. The naming of elements is a very topical subject at the moment and so this is included to ensure a comprehensive coverage.

Covering a wide range of topics in the area of nomenclature and acting as an introduction to a varied field, this book will be of interest to industrial chemists as well as students at senior undergraduate and postgraduate level.

K. J. Thurlow is Head of Chemical Nomenclature Advisory Service at the Laboratory of the Government Chemist.

ISBN 0-7514-0475-6

9 780751 404753 >

KLUWER ACADEMIC PUBLISHERS

DORDRECHT / BOSTON / LONDON