(6) Willett, P. The reaction indexing problem: a historical viewpoint. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: Aldershot, 1986.

(7) Harrison, J. M.; Lynch, M. F. Computer analysis of chemical reactions for storage and retrieval. *J. Chem. Soc. C* **1970**, *23*, 2082–2087.

(8) Lynch, M. F.; Willett, P. The automatic detection of chemical reaction sites. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 154–159.

(9) McGregor, J. J.; Willett, P. Use of a maximum common subgraph algorithm in the automatic identification of the ostensible bond changes occurring in chemical reactions. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 137–140.

(10) Arens, J. F. A formalism for the classification and design of organic reaction. I. The class of (±)ₙ reactions. *Recl. J. R. Neth. Chem. Soc.* **1979**, *98*, 155–161.

(11) Roberts, D. C. A systematic approach to the classification and nomenclature of reaction mechanisms. *J. Org. Chem.* **1978**, *43*, 1473–1480.

(12) Grethe, G.; Moock, T. E. Similarity searching in REACCS. A new tool for the synthetic chemist. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 511–520.

(13) Gordon, A. D. *Classification*; Chapman and Hall: London, 1981.

(14) Vickery, B. C. *Classification and indexing in science*, 3rd ed.; Butterworth: London, 1975.

(15) Littler, J. S. An approach to the linear representation of reaction mechanisms. *J. Org. Chem.* **1979**, *44*, 4657–4667.

(16) Guthrie, R. D. A suggestion for the revision of mechanistic designations. *J. Org. Chem.* **1975**, *40*, 402–407.

(17) Bonchev, D.; Kamenski, D.; Temkin, O. N. Graph theoretical classification and coding of chemical reactions with a linear mechanism. *J. Comput. Chem.* **1982**, *3*, 95–111.

(18) Johnson, A. P.; Cook, A. P. Automtic keyword generation for reaction searching. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: Aldershot, 1986.

(19) Schier, O.; et al. A system for the documentation of chemical reaction. *Angew. Chem. Int. Ed. Engl.* **1970**, *9*, 599–604.

(20) Bawden, D.; Wood, S. I. Design, implementation and evaluation of the CONTRAST reaction retrieval system. In *Modern approaches to chemical reaction searching*; Willett, P., Ed.; Gower: Aldershot, 1986.

(21) Zefirov, N. S. An approach to systematization and design of organic reactions. *Acc. Chem. Res.* **1987**, *20*, 237–243.

(22) Gelernter, H.; Rose, J. R.; Chen, C. Building and refining a knowledge base for synthetic organic chemistry via the methodology of inductive and deductive maching learning. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 492–504.

(23) Blurock, E. S. Computer-aided synthesis design at RISC-Linz: automatic extraction and use of reaction classes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 505–510.

(24) Hendrickson, J. B.; Miller, T. M. Reaction indexing for reaction databases. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 403–408.

(25) Ugi, I.; et al. Chemical similarity, chemical distance, and computer-assisted formalized reasoning by analogy. In *Concepts and applications of molecular similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990.

(26) Herges, R. Reaction planning: prediction of new organic reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 377–383.

(27) Brandt, J.; et al. A classification of reactions by electron shift parameters. Application of the Dugundji-Ugi model. *Chem. Scripta* **1981**, *18*, 53–60.

(28) Wilcox, C. S.; Levinson, R. A. A self-organized knowledge base for recall, design, and discovery in organic chemistry. *ACS Symp. Ser.* **1986**, *No. 306*, 209–230.

(29) Willett, P. *Similarity and clustering in chemical information systems*; Research Studies Press: Taunton, 1987.

(30) Adamson, G. W.; Bawden, D. Comparison of hierarchical cluster analysis techniques for the automatic classification of chemical structures. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 204–209.

(31) Willett, P.; Winterman, V.; Bawden, D. Implementation of nonhierarchic cluster analysis methods in chemical information systems. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 109–118.

(32) Bawden, D. DISCLOSE: an integrated set of multivariate display procedures for chemical and pharmaceutical data. *Anal. Chim. Acta* **1984**, *158*, 363–368.

AUTONOM: System for Computer Translation of Structural Diagrams into IUPAC-Compatible Names. 2. Nomenclature of Chains and Rings[†]

L. GOEBELS, A. J. LAWSON,* and J. L. WISNIEWSKI

Beilstein Institute, Varrentrappstrasse 40-42, D-6000 Frankfurt/Main 90, West Germany

The IUPAC recommendations for naming organic chains, carbocyles, and heterocycles are discussed with respect to the implementation in AUTONOM Version 1.0. The general criteria used in the nomenclature approach are given in detail, along with limitations and extensions of the current rules.

INTRODUCTION

The general subject of this series of papers[1] is AUTONOM, a fully automatic and practical computerized system for the generation of IUPAC systematic nomenclature direct from the input of structure diagrams of organic compounds. Our first paper outlined the general programming approach without significant detail of the actual nomenclature principles involved. This paper is concerned with some fundamental aspects of nomenclature, in particular with the implementation of the IUPAC recommendations of the standard work on this subject, the "Blue Book".[2]

This paper takes the form of a page-by-page documentation of those particular aspects of the Blue Book (Sections A and B) which are implemented in the algorithm, and also those which are not. This systematic approach is necessary, since the IUPAC recommendations (generally regarded as "rules") were consciously formulated to allow considerable freedom in their application, and in many cases are not fully defined to their logical conslusion. Thus, any given structure does not necessarily relate to one uniquely correct name, which is simultaneously the greatest strength (flexibility in passive usage)

and greatest weakness (imprecision in active usage) of this language. We have attempted to follow the recommendations as closely as possible. Nevertheless, we are aware that several "dialects" already exist in wide usage in the major reporting media and reference works, which differ slightly not only in style but also deviate from the IUPAC recommendations in some fundamental aspects. AUTONOM Version 1.0 does not cater to all these dialects, but we feel we cannot ignore them since they reflect current usage. Nomenclature experts will recognize in the following the elements of the dialect of IUPAC as used in the Beilstein Information System, which in our opinion follows "pure" IUPAC very closely indeed. We have taken pains in this paper to draw attention to those aspects which differ from other major IUPAC dialects, such as that practiced by Chemical Abstracts Services. It is our hope that ultimately these small variations may be resolved and that our work may contribute in achieving a consensus.

We have chosen to concentrate in the first phase on single component, uncharged organic species of less than 100 heavy atoms (atoms other than H), neglecting for the moment salts, mixtures, polymers, alloys, coordination compounds, and inorganic materials.

AUTONOM is not intended to produce "the" uniquely correct name for structures; on the other hand, the names

---

[†] For Mike Lynch, pioneer and gentleman.
* Author to whom correspondence should be addressed.

AUTONOM. 2

J. Chem. Inf. Comput. Sci., Vol. 31, No. 2, 1991 **217**

produced by any particular version of the program should obviously be reproducible and acceptable in practice. Analysis of the results produced from the current version (1.0) on the basis defined above shows a general success rate (measured by the production of a name on a non-steric basis) of well over 70% of the structures being currently published. In the remaining 20–30% the program refuses to produce a name, and prints a fatal diagnostic. The degree of acceptability of the names so produced is difficult to measure, since this is always a matter of taste and individual convention. In our eyes, naturally, the level of total unacceptability is extremely low (less than 0.1%). This can clearly not be expected to be a universal response, since it is heavily dependent on our particular implementation of the IUPAC recommendations. Hence the necessity of the form of the present paper in which we shall try to define the fundamentals behind the approach, in order to promote discussion on this important subject in which there is much current interest.[3]

The paper is obviously not intended to be a full definition of the relevant IUPAC rules. It is assumed that the heavily involved reader will read the appropriate sections of the Blue Book in parallel. On the other hand, for the reader who has no immediate access to this reference work, we have paraphrased the most important IUPAC recommendations and use many examples (some borrowed from the book) to make the overall approach more clear. We trust that the reader will bear with us as we try to walk the fine line between restating the obvious and losing ourselves in complex detail.

GENERAL REMARKS ON LIMITATIONS

(1) This paper is only concerned with fundamental questions in Version 1.0 concerning chains, carbocycles, and heterocycles (Sections A and B of the Blue Book). The numbering of the rules is dealt with in the order cited by IUPAC. Section C will be the subject of the next paper in this series.

(2) AUTONOM is built around the principles of **substitutive** nomenclature, with a limited application of **replacement** nomenclature (only for certain heterocycles, not for chains), **additive** nomenclature (hydro prefixes), and **subtractive** nomenclature (unsaturation in aliphatic compounds and cycloalkanes). The use of the alternative **radicofunctional** and **conjunctive** nomenclature is not supported, and (at present) this also applies to **multiplicative** nomenclature.

(3) In our eyes, the most important limitations to the power of AUTONOM lie (at present) in the complete neglect of stereochemical aspects and the inability to selectively deviate from the normal IUPAC guidelines when "common sense" dictates. The former problem will be solved in the course of further development. The latter problem finds its expression in the refusal of the program to name some surprisingly simple compounds, such as 1,2-dichloro-tetraphenylethane (see below). This point recurs frequently in the following text in the form "multivalent radicals are not applied in the case of assemblies".

We are also in the process of tackling the problem of "common sense", but the problem is not trivial, and certainly not solvable to universal satisfaction.

(4) The term "Ring Dictionary" is used extensively throughout the paper. We are referring to the nonalgorithmic part of AUTONOM, which consists of a collection of several thousand ring skeletons with their fusion names or trivial names. As described in our first paper, the determination of the precise description of hydrogenation states (including indicated -*H*) and symmetry elements of these skeletons is then carried out on a purely algorithmic basis, thus enabling a single entry to deal with tens or even hundreds of different cases based on the parent connectivity skeleton. This dictionary was built up on the basis of statistical occurrence in several million

Table I. Names for Saturated Chains of Length *n*

n	name	*n*	name
1	methane	21	henicosane
2	ethane	22	docosane
3	propane	23	tricosane
4	butane	24	tetracosane
5	pentane	25	pentacosane
6	hexane	26	hexacosane
7	heptane	27	heptacosane
8	octane	28	octacosane
9	nonane	29	nonacosane
10	decane	30	triacontane
11	undecane	31	hentriacontane
12	dodecane	32	dotriacontane
13	tridecane	33	tritriacontane
14	tetradecane	34	tetratriacontane
15	pentadecane	35	pentatriacontane
16	hexadecane	36	hexatriacontane
17	heptadecane	37	heptatriacontane
18	octadecane	38	octatriacontane
19	nonadecane	39	nonatriacontane
20	icosane	40	tetracontane

structures, and (although very broad based) is naturally not complete. As a general rule, an entry in the Ring Dictionary has precedence over purely algorithmic names for rings. Apart from this dictionary, all parts of the AUTONOM name are generated by the program code. The Ring Dictionary can be augmented at will in future versions. We will be grateful for any suggestions for additions or for information regarding errors or inconsistencies in use. For obvious reasons, the confines of the present paper do not allow a listing of the Ring Dictionary.

IMPLEMENTATION OF THE IUPAC RULES IN VERSION 1.0

A-1. Saturated Unbranched Acylic Hydrocarbons
A-1.1. Naming of Hydrocarbon Chains

AUTONOM uses the naming system for acyclic saturated alkanes as listed in the Blue Book. The names of Table I are used for the given chain length (*n*). Numbering starts with one terminal position and is consecutive.

Note: AUTONOM will handle all-carbon chains up to and including 44 atoms. Longer alkanes cause a fatal diagnostic.

A-1.2. Alkyl Radicals

Univalent radicals of saturated unbranched acyclic hydrocarbons are formed by replacing the ending ane with yl. The carbon of the free valence is numbered 1 and is not explicitly cited.

A-2. Branched Chains
A-2.1. Choice of the Principal Chain

The longest carbon chain present is chosen as the principal chain; e.g., structure **1** is a substituted pentane.


1
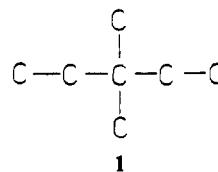
Note: Although the name isobutane is retained for the unsubstituted hydrocarbon, AUTONOM does not support the following names for unsubstituted branched hydrocarbons:[4] isopentane, neopentane, isohexane, etc. This is in accordance with a general policy of providing fully systematic usage in the aliphatic series for carbon chains of length 5 and higher.

A-2.2. Numbering of the Principal Chain in Branched Acyclics

The principal chain is numbered in such a way as to give the lowest number possible to the point of attachment of the first side chain. Should several choices then still be possible,

218 *J. Chem. Inf. Comput. Sci., Vol. 31, No. 2, 1991*

GOEBELS ET AL.

the second side chain is considered, and so on. Should a decision still not be reached, Rule A-2.4 (below) is applied. AUTONOM separates the chain terms by the use of a hyphen: e.g., 2,3,5-trimethyl-hexane (2).

2
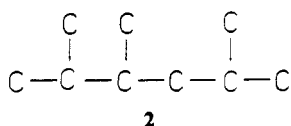
Note: AUTONOM can simultaneously maintain up to 24 paths of identical length in its search for the principal chain. This limit can be exceeded for extremely symmetrical large hydrocarbon structures with repeated branching ("explosion chains"). Such examples do not generally occur in practice. Should such a molecule be input, AUTONOM prints a warning along with the name.

A-2.25. Branched Radicals

The name of the substituent is formed by finding the main chain of the substituent in a manner analogous to that described above for the principal chain, and the procedure is applied recursively.

Note: Of the univalent branched radicals, AUTONOM supports the following unsystematic usage:[5] isopropyl, isobutyl, *sec*-butyl, and *tert*-butyl. These radical names may not be further substituted. In such cases, fully systematic usage is preferred.

A-2.3. Order of Citation of the Side Chains

The side chains are cited in alphabetical order, whereby IUPAC distinguishes between "simple" and "complex" radicals with respect to the alphabetization of multiplying terms (di, tri, etc.). Thus 2,3-dimethyl- is ordered according to "m" (for methyl), while 7-(1,2-dimethyl-pentyl)- is ordered according to "d": e.g., 5-ethyl-2,3-dimethyl-tridecane and 7-(1,2-dimethyl-pentyl)-5-ethyl-tridecane. The locants are not considered as part of the alphabetization, unless the side chains are word-identical. Then the lowest locant has precedence: e.g., 7-(1-methyl-butyl)-6-(2-methyl-butyl)-tridecane.

A-2.4. Numerically Equivalent Positions

Numerically equivalent positions are numbered according to the citation order: e.g., 4-ethyl-5-methyl-octane.

A-2.5. Multiplying Affixes

The multiplying affixes di, tri, tetra, penta, etc. are used for identical simple radicals. The affixes bis, tris, tetrakis, pentakis, etc. are used for identical complex radicals: e.g., 3,3-dimethyl-pentane and 5,7-bis-(1,2-dimethyl-propyl)-undecane.

Notes: (a) The complete expression denoting a side chain is set in parentheses as necessary; the use of primed numbers in this respect[6] is not supported.

(b) AUTONOM supports multiplying affixes to a limit of 44 (tetratetraconta and tetratetracontakis). Higher values cause a fatal diagnostic.
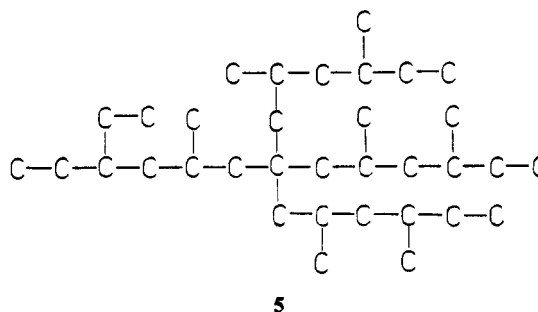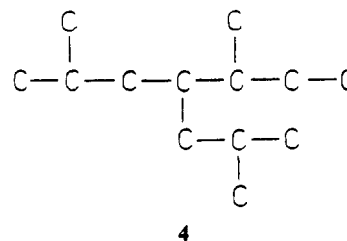
A-2.6. Seniority of Chains of Equal Length

The priority is decided in the following order:

(a) Greatest number of side chains: e.g., 2,3,5-tri-methyl-4-propyl-heptane (3)
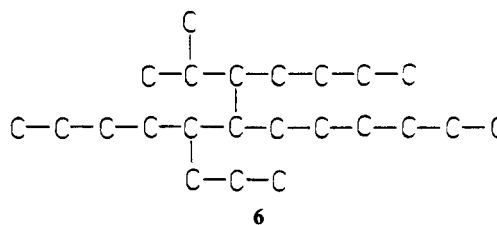

3
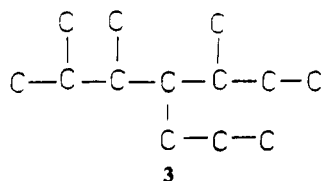
(b) Side chains with lowest numbered locants: e.g., 4-isobutyl-2,5-dimethyl-heptane (4)

(c) The greatest number of carbon atoms in the smaller side chains: e.g., 7,7-bis-(2,4-dimethyl-hexyl)-3-ethyl-5,9,11-trimethyl-tridecane (5)

4

5

(d) The least branched side chains: e.g., 6-(1-iso-propyl-pentyl)-5-propyl-dodecane (6)

6

Note: We are grateful to a referee for pointing out that rule A-2.6 in the IUPAC recommendations does not lead to unique names for some highly branched hydrocarbons, and there is indeed a clear contradiction between A-2.6 c,d and C-13.11 j,k. This problem is apparently under review in the Commission on Nomenclature of Organic Chemistry and we will follow the resulting recommendation with great interest. For the meantime, AUTONOM supports the implementation as described here.

A-3. Unsaturated Acylic Hydrocarbons

A-3.1. Unbranched with Double Bonds

The ending ane is replaced by ene. The chain is numbered in such a way as to give the lowest possible number to the position of unsaturation. Only the first locant of each multiple bond is cited.

Notes: (a) AUTONOM consistently places the locants for multiple bonds immediately before the corresponding syllable ene etc. and not simply at the start of the name of the chain, as is often used in the CAS dialect: e.g., hex-2-ene, not 2-hexene; hexa-2,4-diene, not 2,4-hexadiene. We consider this usage to be more logical, since it places the descriptor in a direct relationship to the designated object of its description. It has also been used for many years as the preferred form in the publications of the Chemical Society. In the Blue Book, this usage is also illustrated in the examples of the von Baeyer System for bridged hydrocarbons (bicyclo[2.2.2]oct-2-ene, not 2-bicyclo[2.2.2]octene), but is less frequently used for hydrocarbon chains, although IUPAC makes it quite clear that the use of the CAS convention in the examples of the Blue Book is not to be considered as part of its recommendations.[7] The present AUTONOM usage is in any case in full accord with the latest IUPAC draft recommendations.[8] (b) The use of ethylene and allene for ethene and propadiene, respectively, is not supported.

A-3.2. Unbranched with Triple Bonds

The ending ane is replaced by yne. Otherwise the comments of A-3.1 apply.

Note: The use of acetylene for ethyne is not supported.

A-3.3. Combinations of Double and Triple Bonds

Numbers as low as possible are given to double and triple bonds, irrespective of which. However, should an equivalent choice arise, the double bonds are given the lowest numbers: e.g., pent-3-en-1-yne (**7**).
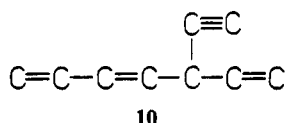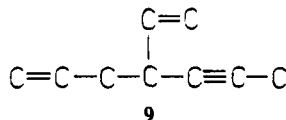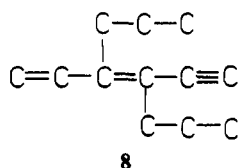
$$C{\equiv}C{-}C{=}C{-}C$$

7
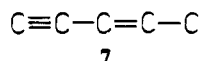
A-3.4. Unsaturated Branched Chains

In accordance with the IUPAC recommendations, unsaturated branched acyclic hydrocarbons are named as derivatives of the chain containing the **maximum number** of unsaturated bonds (not necessarily the **longest** chain, which is a variation frequently found in the literature, unfortunately). AUTONOM distinguishes between double bonds and triple bonds in the following manner:

Precedence is given to

(1) the chain with the maximum number of multiple bonds

(2) the longest chain

(3) the chain with the maximum number of double bonds

in that order: e.g., 3,4-dipropyl-hexa-1,3-dien-5-yne (**8**); 4-vinyl-hept-1-en-5-yne (**9**); and 5-ethynyl-hepta-1,3,6-triene (**10**).

$$
\begin{array}{c}
C{-}C{-}C \\
| \\
C{=}C{-}C{=}C{-}C{\equiv}C \\
| \\
C{-}C{-}C
\end{array}
$$

8

$$
\begin{array}{c}
C{=}C \\
| \\
C{=}C{-}C{-}C{-}C{\equiv}C{-}C
\end{array}
$$

9

$$
\begin{array}{c}
C{\equiv}C \\
| \\
C{=}C{-}C{=}C{-}C{-}C{=}C
\end{array}
$$

10

Note: The application of this rule is one of the major differences between pure IUPAC and the dialect practiced by CAS. IUPAC recommends the order 1-2-3, while CAS gives precedence to the longest chain.

A-3.5. Unsaturated Substituents

The names of unsaturated substituents are derived by using the endings enyl, ynyl, etc. Combinations are expressed in the order -en-ynyl with the appropriate multiplying affixes and locants where necessary. The locant of attachment is always 1 and is not explicitly cited.

Note: The following IUPAC retained names are supported:[9] vinyl (for ethenyl), allyl (for prop-2-enyl), and isopropenyl (for 1-methyl-ethenyl). The radical name propenyl is used unequivocally by AUTONOM to mean prop-1-enyl, since allyl replaces prop-2-enyl.

A-4. Bivalent and Multivalent Radicals

A-4.1. Bivalent and Trivalent Radicals at Single Carbon Atoms

Bivalent and trivalent radicals at single carbon atoms are formed by replacing yl with ylidene or ylidyne, respectively. The IUPAC recommendations A-4.2 to A-4.5[10] (**multivalent bridges**) are not supported when applied to assemblies.

Notes: (a) The IUPAC prefix methylene is retained for the radical $CH_2{=}$.

(b) A referee has kindly pointed out that the use of ylidene and ylidyne is under review in the Commision on Nomen-

clature of Organic Chemistry.

A-11. Monocyclic Hydrocarbons

Numbering of Monocyclic Hydrocarbons. The problem of assigning locants to the ring reduces to two questions:

Where does one start?

In which direction around the ring is the numbering continued?

The following general precedence is supported by AUTONOM:

(i) Minimize the locant of the free valence (in a radical). In a monocyclic substituent the free valence always has the number 1 but is not explicitly cited by AUTONOM.

(ii) Minimize the locants of the multiple bonds. Should there be an equivalent choice, double bonds have precedence over triple.

(iii) Minimize the locant of the substituent first encountered on proceeding around the ring. Should there be an equivalent choice, consider the locant of the next substituent, and so on.

(iv) Assign locants alphabetically according to Rule A-2.4 (above).

This procedure will be extended later in the discussion of more complex rings (multicyclic, heterocyclic) and can be referred to in the following shorthand form:

yl position

cited unsaturation

substituent locants

substituent order

A-11.1. Saturated Parent Rings

Saturated monocyclic hydrocarbons are named by attaching the prefix cyclo to the name of the corresponding acyclic hydrocarbon. The prefix cyclo is not separated by a hyphen. Numbering is consecutive around the ring. In monosubstituted hydrocarbons the locant 1 is redundant and is not explicitly cited in AUTONOM: e.g., ethyl-cyclodecane.

Note: AUTONOM sets a limit of 44 for the size of the largest ring. The presence of a larger ring than 44 atoms causes a fatal diagnostic.

A-11.2. Cycloalkanes as Radicals

Univalent radicals derived from cycloalkanes are named by replacing ane by yl.

A-11.3. Unsaturated Monocylic Hydrocarbons

The names are formed by substituting ene, yne, etc. for ane in the corresponding cycloalkane. The name benzene is retained.

A-11.4. Unsaturated Monocyclic Hydrocarbons as Radicals

The endings enyl and ynyl are used. The position of unsaturation is indicated by locants analogously to A-3.3. The name phenyl is retained: e.g., cyclohexa-2,4-dienyl-benzene.

Note: In such cases, the yl position, when not cited, is implicitly 1. In the Blue Book the locant position is explicitly cited.

A-11.5. Bivalent Radicals at Single Carbon Atoms

The terms ylidene, x-enylidene, x-ynylidene, and x-en-x'-ynylidene are used with the corresponding locants (x and x') and multiplying affixes as necessary: e.g., 7-cyclopenta-2,4-dienylidene-cyclohepta-1,3,5-triene (**11**).
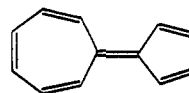
11

A-11.6. Multivalent Bridges

This IUPAC recommendation[11] is not supported in AUTONOM when applied to assemblies.

A-12. Substituted Aromatic Compounds

A-12.1 to A-12.4. Alkylbenzenes

Table II. IUPAC Retained Carbocyclic Trivial Names

(1) pentalene	(19) chrysene
(2) indene	(20) naphthacene
(3) naphthalene	(21) pleiadene
(4) azulene	(22) picene
(5) heptalene	(23) perylene
(6) biphenylene	(24) pentaphene
(7) as-indacene	(25) pentacene
(8) s-indacene	(26) tetraphenylene
(9) acenaphthylene	(27) hexaphene
(10) fluorene	(28) hexacene
(11) phenalene	(29) rubicene
(12) phenanthrene	(30) coronene
(13) anthracene	(31) trinaphthylene
(14) fluoranthene	(32) heptaphene
(15) acephenanthrylene	(33) heptacene
(16) aceanthrylene	(34) pyranthrene
(17) triphenylene	(35) ovalene
(18) pyrene	

AUTONOM names all substituted benzenes in a strictly systematic fashion, i.e., as benzene derivatives, with the following exceptions:[12] cumene, styrene, cymene, toluene, mesitylene, and xylene.

Notes: (a) The above trivial names may neither be further substituted nor may they be used in the radical form (exceptions: styryl, tolyl).

(b) AUTONOM does not support the use of ortho, meta, and para for systematically named 1,2-, 1,3-, and 1,4-disubstituted benzenes: e.g., 1-isopropyl-3-pentyl-benzene.

A-13. Substituted Aromatic Radicals

With the exception of tolyl, all substituted benzene radicals are described as substituted phenyl radicals. The terms mesityl[13] etc. are not supported. The use of phenylene as a bivalent radical is also not supported. The following trivial names for radicals having a single free valence in the side chain are retained:[14] benzyl, benzhydryl, phenethyl, styryl, and trityl.

Notes: (a) No further substitution in the side chain is allowed for these trivial names.

(b) The usage of Rule A-4 is supported for the following: benzylidene, benzhydrylidene, and phenethylidene.

(c) Ring substitution is allowed for benzyl and benzylidene only.

(d) The name cinnamyl is not supported.

### A-21. Fused Polycyclic Hydrocarbons

A-21.1. Retained Trivial Names

The examples listed in the Blue Book[15] are supported (see Table II). The numbering in these compounds is predefined. However, the presence of symmetry often allows possible alternatives in substituted examples, or radicals, which must be separated where necessary.

In such cases the following precedence order is supported by AUTONOM. Minimize locants for

    yl position
    indicated H
    hydro prefixes
    substituent locants
    substituent order

in that order: e.g., 9-ethyl-8-methyl-2,6-dihydro-1*H*-phenalene (**12**).
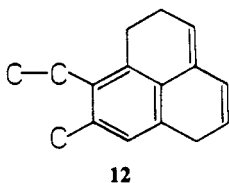
12

Note: (a) One can detect a fundamental difference between two otherwise similar means of assigning locants to cyclic systems. There is the method by which the position of un-

saturation is **directly cited** (e.g., monocyclic hydrocarbons, see above in section A-11.4) and the method by which the positions of unsaturation are **indirectly cited** (as here). Obviously, the difference comes from the choice of either a fully saturated or fully unsaturated name as the parent. However, it is convenient to relate the difference in this manner, since it is also relevant for heterocycles (see below).

(b) The hydro prefixes in AUTONOM are **nondetachable** in nature (i.e., these always appear immediately before the name of the corresponding unreduced hydrocarbon). This conforms perfectly to the IUPAC recommendations,[16] although these also allow the detachable form. On the other hand, it is not quite clear in the IUPAC rules what is exactly meant by detachable and nondetachable, and how far the implications must logically be carried. Here we really mean "excluded from the normal alphabetization procedure applying to the other prefixes and cited after all substituent prefixes".

A-21.2. Straight Linear Arrangements of Fused Benzenes

Apart from the examples contained explicitly in the above list (pentacene, hexacene, heptacene), the use of -acene for linear fused benzene rings[17] is not supported for eight or more rings. A fatal diagnostic is given.

A-21.3 to A-28.1. Other Arrangements of Fused Rings

These rules[18] are fully supported by numerous examples in the Ring Dictionary. However, not all possible combinations are listed. Should the appropriate ring system not be found, AUTONOM gives a fatal diagnostic for ring systems tetracyclic and higher. Hydrogenated compounds are dealt with in strict accordance to Rule A-23.1, in which the prefixes dihydro, tetrahydro, etc. are used as nondetachable prefixes. The use of indicated H to eliminate ambiguity is fully supported where necessary. All bicyclic and tricyclic fused systems are dealt with on a fully algorithmic basis should no entry in the Ring Dictionary be found: e.g., 4,5,6,7,8,9-hexahydro-1*H*-cyclopentacyclooctene (**13**).
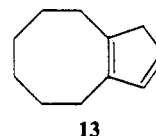
13

Notes: (a) The term perhydro is not supported. When a polycyclic hydrocarbon is fully saturated, AUTONOM does not list the locants: e.g., 1,4-dihydro-naphthalene and decahydro-naphthalene (fully saturated).

(b) Of the trivially named exceptions of partially hydrogenated cyclic hydrocarbons listed in the Blue Book,[19] only the name indan is supported.

(c) The method of A-23.5 (as an alternative to A-23.1) is not supported.

(d) The exceptions naphthyl, anthryl, and phenanthryl[20] are not supported. The preferred usage in AUTONOM is naphthalen-*x*-yl, anthracen-*x*-yl, and phenanthren-*x*-yl, where *x* is the appropriate locant of the free valence.

(e) Rule A-24.4 (multivalent radicals) is not supported in AUTONOM for assemblies.

A-31 to A-34. Bridged Hydrocarbons

The IUPAC recommendations use an extension of the von Baeyer system[21] for some (but not all) bridged hydrocarbons. To clarify the situation with respect to AUTONOM, it is convenient to distinguish between the cases of true bridges (non-zero bridges) and ortho-fused systems (zero bridges), where the von Baeyer system is also used in some cases (see above).

(1) **Non-Zero Carbon Bridges.** AUTONOM uses this system exactly as described[22] in the Blue Book for all **bicyclic and tricyclic** non-spiro bridged systems, provided that an alternative ring name is not first of all found in the Ring Dictionary. Ring systems **tetracyclic and higher** are not dealt with
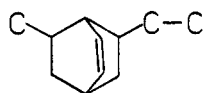
AUTONOM. 2

J. Chem. Inf. Comput. Sci., Vol. 31, No. 2, 1991 **221**

Table III. Key to Nomenclature for Bicyclic Ortho Fused Systems

ring 1		ring 2			nomenclature used by AUTONOM
hetero	size	hetero	size		
yes/no	≤4	yes/no	≥3		von Baeyer
yes	5–10	yes	5–10		1 dictionary
yes	5–10	no	5–10		or (less favored)
no	5–10	yes	5–10		2 replacement nom.
no	5 or >6	no	5 or >6	R1 = R2	..alene
				R1 <> R2:	cyclo..acyclo..ene
no	6	no	≥7		benzo..
yes	5–10	no	≥11		1 dictionary or (less favored)
					2 replacement nom.
no	5–10	yes	≥11		replacement nom.
yes	5–10	yes	≥11		(standard)
		Exceptions to These Rules			
no	5	no	6		indene (and indan)
no	5	no	7		azulene
no	6	no	6		naphthalene

using the von Baeyer system. They are either found in the Ring Dictionary or a fatal diagnostic is given.

Notes: (a) The use of the von Baeyer system for tricyclic compounds demands the use of superscripts for secondary bridges (e.g., tricyclo[4.4.1.1^{1,5}]undecane). AUTONOM delivers its name as an ASCII string, and there are no ASCII characters for all the superscripted numbers. Nevertheless, the superscripted numbers may be easily recognized by the fact that they are punctuated by commas (,) not dots (.).

(b) In the von Baeyer system, numbering starts with one of the bridgeheads, proceeds by the longest possible path to the second bridgehead, and then by the longest path back to the first bridgehead. Numbering is then continued from the unnumbered atom next to the first bridgehead. This numbering is predefined. However, due to symmetry, choices may still arise. These are settled by the method of **direct citation**, in analogy to section A-11: e.g., 5-ethyl-8-methyl-bicyclo-[2.2.2]oct-2-ene (**14**).

14

(2) Zero Carbon Bridges (Fused Systems). The common case of bicyclic structures with a zero bridge deserves some comment. AUTONOM applies the various IUPAC methods according to the size of the two rings (Ring 1 and Ring 2) of Table III (here for carbocyclics, i.e., hetero = no).

In the case of **tricyclic** systems with a zero bridge, it sometimes occurs that a von Baeyer name is given by AUTONOM, although the program determines (algorithmically) that the ring system should belong in the Ring Dictionary, but is not present. In this case a warning diagnostic is given.

Rule A-34[23] (hydrocarbon bridges using methano etc.) is supported in AUTONOM to some extent directly by the Ring Dictionary, but also for all bridged systems with a single non-zero bridge across a ring system already present in the Ring Dictionary.

A-41 to A-43. Spiro Hydrocarbons

AUTONOM supports the Method 1 outlined in Rules A-41.1 to A-41.3[24] for all monospiro and dispiro alicyclic ring systems, which do not include a fused ring component. Should a fused ring component be present, or a trispiro system (or higher), a fatal diagnostic is given unless the system is found in the Ring Dictionary.

Simple spiro compounds are named after the alkane corresponding to the total number of ring atoms in the spiro system. The nature of the rings is indicated in square parentheses by the number of carbon atoms attached between individual spiro atoms, in the order of numbering of the system.

The numbering starts at an atom immediately next to the spiro atom of the smaller terminal ring and proceeds along the periphery through the next spiro attachment by the shortest path. After this predefined numbering, the method of **direct citation** is used to indicate unsaturation as necessary: e.g., 4-ethyl-spiro[4.4]non-1-ene (**15**).
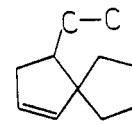
15

A-51 to A-56. Hydrocarbon Ring Assemblies

Assemblies are formed when two or more identical ring systems are linked directly together with a single or double bond. IUPAC proposes two alternative methods[25] for the case of two identical ring systems, of which only one is supported by AUTONOM, the biradical method (e.g., bicyclopropyl, not bicyclopropane). Three to seven identical ring systems are dealt with by using the underlying hydrocarbon (e.g., tercyclopropane, not tercyclopropyl). As an exception to the latter, one should note the linear unbranched assemblies of benzene rings (quaterphenyl, not quaterbenzene).

AUTONOM only supports linear assemblies. A branched assembly gives rise to a fatal diagnostic.

AUTONOM uses the following multipliers for assemblies: 2, bi; 3, ter; 4, quater; 5, quinque; 6, sexi; and 7, septi.

Numbering of the individual ring systems follows the general rules for the individual cyclic system concerned. The rings are distinguished by the use of primes (′) starting from one unprimed terminal ring system. The decision as to which terminal ring system of the chain is unprimed is decided according to the following order of precedence (carried out from both ends of the assembly, if necessary, into the center of the chain until a decision is reached):

> lowest locant for points of assembly (from before to)
> yl position
> highest degree of hydrogenation
> lowest locants for indicated H (where applicable)
> lowest locants for hydro prefixes
> highest degree of substitution
> lowest locants for substituents
> alphabetization of substituents

Once the unprimed ring system has been determined, the chain of ring systems is linearly attributed primes (second ring system in chain one prime, third ring system two primes, etc.).

Notes: (a) The classification of a structural subunit as an assembly (or not) is a recurring problem in practice, which is sometimes at odds with any set of rules, no matter how carefully worded. In principle, an assembly must consist of

identical rings. However, the question of when rings are identical is not so clearly defined. Individual preference might have chosen to include stereochemical aspects, for instance, although this is unclear from the Blue Book. Even more fundamental, however, is the question of location of saturation or unsaturation, since this is particularly sensitive to the uncertainties of mesomeric representation and tautomeric equilibria. The problem also touches the use of nondetachable or detachable hydro prefixes (see above), as a referee has kindly pointed out (see also example 17, below). Instinctively, we feel that cyclohexyl-benzene is not an assembly (i.e., a hexahydro-biphenyl), while 1,2,3,3',4,4'-hexahydro-1,1'-binaphthyl clearly is, irrespective of the state of hydrogenation of the individual rings. The Blue Book explicitly quotes these examples in Rule A-53, which supports our instincts. However, we acknowledge certain difficulties in this approach, but recognize distinct advantages from the common sense point of view.

AUTONOM accordingly distinguishes between two cases:

(i) Assemblies where hydro prefixes are not used, i.e., where double bonds are **explicitly cited**. These include assemblies of monocyclic hydrocarbons, von Baeyer systems, and spiro compounds. In these cases the rings of the assembly must all be identical with respect to **degree and position of unsaturation**, otherwise they are treated as nonidentical ring systems according to Rule A-23 (see also Rule 53.3): e.g., 5-bicyclo-[2.2.1]hept-2-ylidene-bicyclo[2.2.1]hept-2-ene (**16**).

16

(ii) Assemblies where hydro prefixes are used in the corresponding single ring systems (**indirect citation**). These include all ortho fused systems containing at least two five-membered (or larger) rings (see Table III). In these cases the individual systems of the assembly do not have to be at the same level of unsaturation to be considered an assembly. The saturated positions are indicated by the locants of the hydro prefix, using primes as necessary: e.g., 2',3'-dihydro-1H,1'H-[1,2']biindenyl (**17**). An expert referee has suggested
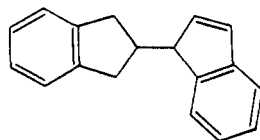
17

for this structure 2',3'-dihydro-1,2'-bi-1H-indenyl, thus implicitly drawing attention to the problems mentioned above. In this form, a structure with indicated hydrogens at the 1 and 3 positions would **not** be considered an assembly, in contrast to the current conventions of AUTONOM. We respect this position, but await future developments.

(b) AUTONOM supports the description of assemblies to a limit of 44 carbon atoms in the assembly, or seven systems, whichever limit is first reached. Larger assemblies cause a fatal diagnostic.

(c) Multivalent bridging radicals of assemblies are not supported.

(d) Hydro prefixes are treated with a higher priority than substituents. This is consistent with the nondetachable use of hydro in AUTONOM, and this policy is also used in the choice of locants, in a slight variation to Rule C-15.11 d.[26]

A-61 to A-75. **Cyclic Hydrocarbons with Side Chains**

This section deals with hydrocarbons more complex than those envisioned in Rule A-12. The IUPAC rules at this point are not very computer-friendly(!). The choice "is made so as

Table IV. Heteroatom Prefix Priority Table (Hantzsch-Widman)

n	atom	descr	n	atom	descr
(1)	O	oxa	(9)	Bi	bisma
(2)	S	thia	(10)	Si	sila
(3)	Se	selena	(11)	Ge	germa
(4)	Te	tellura	(12)	Sn	stanna
(5)	N	aza	(13)	Pb	plumba
(6)	P	phospha	(14)	B	bora
(7)	As	arsa	(15)	Hg	mercura
(8)	Sb	stiba			

Table V. Hantzsch-Widman Stems for Rings of 3-10 Atoms

| ring size | rings without nitrogen | | rings with nitrogen | |
	unsatd	satd	unsatd	sat
3	irene	irane	irine	iridine
4	ete	etane	ete	etidine
5	ole	olane	ole	olidine
6	ine/inine[a]	ane[b]/inane	ine	ane
7	epine	epane	epine	epane
8	ocine	ocane	ocine	ocane
9	onine	onane	onine	onane
10	cine	ecane	ecine	ecane

[a]Rings containing P, As, Sb, or B as element of last priority. [b]Rings containing O, S, Se, or Te as element of last priority.

to provide the name which is the simplest permissible or the most appropriate for the chemical intent".[27] The problem here is that the rules do not indicate a general seniority of rings before chains or vice versa. Factors such as "treatment of a smaller unit of structure as a substituent into a larger" and "the maximum substitutions into a single unit of structure" are weighed off against each other, and the "most appropriate" choice is made. Although this is an excellent basis for readily understandable and concise nomenclature, it is not the sort of seniority procedure which lends itself well to computerized nomenclature. Nevertheless, it is our intention in future versions of AUTONOM to attempt to implement a user-friendly nomenclature of this sort.

For the meantime, let's consider the very simple and commonplace compound 1,2-dichloro-tetraphenylethane cited in recent correspondence[28] concerning the user-friendliness of systematic nomenclature. AUTONOM possesses a very keen sense of seniority rings and chains and usually applies (in contrast to the above IUPAC recommendations) the rule that rings are senior to chains in the absence of characteristic groups.

Now the use of AUTONOM on the above structure results in a clear refusal to name the compound (since the program refuses to choose arbitrarily among four identical rings). We consider this to be a better response than to give the name based on a preference for rings: 1,1',1'',1'''-(1,2-dichloroethane-1,2-diyl)tetrakis[benzene]. However, we acknowledge that this is a matter of taste.

Note: The terpene exceptions listed in Rule A-71 to Rule A-75[29] are not supported in AUTONOM Version 1.0. Later versions will support stereochemistry.

B. Fundamental Heterocyclic Systems

Monocyclic Heterocyclic Rings with 3-10 Ring Atoms.

B-1.1. Extension of the Hantzsch-Widman System

AUTONOM does not directly support the system described in Rule B-1[30] for monocyclic rings containing one or more hetero atoms, but uses the later extended IUPAC version,[31] which is simpler for rings of six to ten members. Hetero atoms are given descriptors and priority according to Table IV. The size and state of saturation is given by the stem according to Table V.

B-1.2. Partially Saturated Hantztsch-Widman Names

AUTONOM treats all partially saturated Hantzsch-Widman names according to the indirect citation method, using

Table VI. IUPAC Retained Heterocyclic Trivial Names

(1) thiophene	(23) indazole	phenomercurin
(4) thianthrene	(24) purine	isoarsinoline
(5) furan	(25) quinolizine	arsinoline
(6) pyran	(26) isoquinoline	arsanthridine
(7) isobenzofuran	(27) quinoline	acridarsine
(8) chromene	(28) phthalazine	arsanthrene
(9) xanthene	(29) [x,y]naphthyridine	isophosphinoline
(10) phenoxathiine	(30) quinoxaline	phosphinoline
(12) pyrrole	(31) quinazoline	phosphanthrene
(13) imidazole	(32) cinnoline	tellurophene
(14) pyrazole	(33) pteridine	selenophene
(14a) isothiazole	(35) carbazole	selanthrene
(14b) isoxazole	(36) β-carboline	phenothiarsine
(15) pyridine	(37) phenanthridine	phenoxantimonine
(16) pyrazine	(38) acridine	phenoxarsine
(17) pyrimidine	(39) perimidine	phenoxaphosphine
(18) pyridazine	(40) [x,y]phenanthroline	phenoxatellurine
(19) indolizine	(41) phenazine	phenoxaselenine
(20) isoindole	(42) phenarsazine	pyrrolizine
(22) indole	(44) phenothiazine	phenomercazine
isophosphindole	(46) furazan	phenarsazine
phosphindole	(47) phenoxazine	phenophosphazine
isoarsindole	phenotellurazine	phenoselenazine
arsindole		

explicit locants and (where necessary) indicated *H* in order to specify partial saturation: e.g., 2,5-dihydro-1*H*-silole (**18**).

18

B-1.3 to B-1.4. Multiplicity of Heteroatoms

Multiplicity of identical heteroatoms is indicated by use of the affixes di, tri, etc. Different types of heteroatom are cited in the priority order of Table IV.

B-1.5. Numbering of Heteroatoms

As noted for monocyclic hydrocarbons, the question of numbering reduces to the selection of a starting point and the direction of counting.

The locant 1 is always given to that hetero atom which
    is as high as possible in Table IV
    and
    from which the lowest locants for heteroatoms (as a
        complete set) can be derived by further numbering
in that order.

Should there still be a choice, this is then settled by the following order of precedence: Minimize locants for
    individual heteroatoms in the order of Table IV
    yl position
    indicated *H*
    hydro prefixes
    substituent locants
    substituent order
in that order: e.g., 4-methyl-3-(6-methyl-[1,3]dithian-4-yl)-4*H*-[1,4]oxazine (**19**).

19
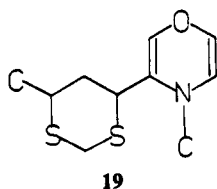
Notes: (a) Monocyclic heterocycles with more than 10 ring atoms are named by using replacement nomenclature with the corresponding cycloalkane name (see Table II). Numbering is carried out in full analogy to the above, with the exception that the direct citation method is used for unsaturation.

(b) For certain monocyclic heterocycles with 5 or 6 ring atoms, the trivial names are preferred (see following section).

B-2. Trivial and Semi-Trivial Names

The examples of trivial names listed[32] in the Blue Book are retained (Table VI). (Numbers, when present, are as listed in the original.)

Note: AUTONOM also supports the following:

pyrindine	isothiochromene	phenazasiline
thiochromene	isoselenochromene	phenoxasiline
selenochromene	isotellurochromene	thiopyran
tellurochromene	telluranthrene	selenopyran
isochromene	telluropyran	

B-2.12. Saturated and Partially Saturated Trivial and Semi-Trivial Heterocyclic Compounds

The following IUPAC recommended names[33] are retained in AUTONOM, but are not used in fusion names or assemblies:

isochroman	pyrrolidine	piperidine
chroman	imidazolidine	piperazine
pyrazolidine	morpholine	

Not retained by AUTONOM are the following:

pyrroline	pyrazoline	indoline
imidazoline	quinuclidine	isoindoline

Note: The names piperidino and morpholino[34] are not retained for substituents connected via the nitrogen atoms.

B-3. Other Fused Heterocyclic Systems

Ortho fused heterocyclic systems are named in accordance with Rule B-3, and all examples are taken from the Ring Dictionary: e.g., 8-ethyl-9,10-dimethyl-9,10-dihydro-8*H*-naphtho[1′,8′:4,5,6]cyclohepta[1,2-*c*]pyrrole (**20**). The choice

20
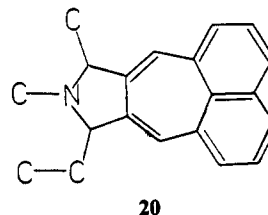
of the base component of the fused system is made according to the rules of Ring Seniority, discussed below.

B-4 to B-15. Replacement Nomenclature for Heterocycles

AUTONOM supports the use of replacement nomenclature only in the following cases (see Table III):

Monocyclic rings: only for ring size 11 and higher (otherwise the Hantzsch-Widman system or trivial names are used).

Bicyclic rings and higher: all carbocyclic systems which can be named by AUTONOM in the absence of heteroatoms. These include on a systematic basis:
    all bicyclic fused systems (but see note a)
    all bridged ring systems named according to the von
        Baeyer system
    all bicyclic and tricyclic non-fused mono and dispiro
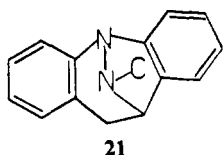   systems

Notes: (a) The Ring Dictionary has first priority. The general order of priority is as follows:
    Ring Dictionary (trivial and fusion names)
    Hantzsch-Widman names
    replacement names (carbocycle from Ring Dictionary)
   replacement names (von Baeyer, spiro)

(b) Bridges such as epiazano are supported in individual cases in the Ring Dictionary. Also, all bridged systems with a single non-zero bridge across a ring system in the Ring Dictionary are named algorithmically. Otherwise the von Baeyer name is used.

(c) Heterocyclic assemblies are treated naturally, depending on the type of name for the corresponding single system. E.g.
    12-methyl-10,11-dihydro-5,10-epiazano-dibenzo[*b*,*f*]-
        azepine (**21**)

1',4',5',6'-tetrahydro-[5,7']biindolylidene (**22**)

21
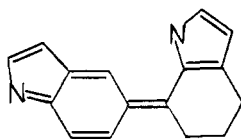
22

SENIORITY OF SYSTEMS

The question of seniority is fundamental to any hierarchical concept. In IUPAC nomenclature, such choices must often be made, for example, to determine which structural component becomes the parent in the absence of decisions dictated by characteristic groups. The seniority of chains has been mentioned in Rules A-2 and A-3. The problems of ring versus chain seniority has been mentioned above. Ring system seniority is dealt with in detail by Rules C-0-14[35] and B-3.[36] These can be summarized here in the following short form:

(i) All heterocycles before all carbocycles: e.g., furan senior to chrysene.

For heterocycles:

(ii) Nitrogen-containing rings before all rings with no nitrogen: e.g., pyrrole senior to benzofuran.

(iii) Rings containing a heteroatom as high as possible in Table IV: e.g., furan senior to benzo[c]thiophene.

(iv) Greatest number of rings: e.g., benzofuran senior to [1,4]dioxane.

(v) Largest expression for the set of smallest rings: e.g., quinoline senior to indole.

(vi) Largest total number of heteroatoms: e.g., imidazole senior to pyrrole.

(vii) Largest variety of heteroatoms: e.g., [1,3]azaphosphole senior to imidazole.

(viii) Multiplicity of most senior heteroatom at first position of difference using Table IV: e.g., oxazole senior to thiazole senior to [1,3]azaphosphole.

(ix) Lowest locants for heteroatoms, taken first as a complete set, then successively in the order of Table IV: e.g., quinoline senior to isoquinoline; [1,2]azaphosphole senior to [1,3]azaphosphole; and [1,2,5]-oxathiazinane senior to [1,5,2]oxathiazinane.
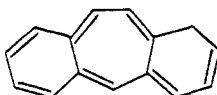
Continue as for carbocyclics from (xii) below.

For carbocycles:

(x) Greatest number of rings: e.g., anthracene senior to azulene.

(xi) Largest Expression for the set of smallest rings: e.g., benzo[1,2:3,4]dicycloheptene (**23**) (i.e., 7,7,6) is senior to 1H-dibenzo[a,d]cycloheptene (**24**) (i.e., 7,6,6).
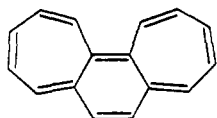
23

24

Note: The order of fusion of the rings is not considered at this stage, simply the size of the rings. The pre-

cedence according to fusion is considered in (xii) to (xvi), below.

(xii) Largest number of atoms in common among rings (length of bridge): e.g., bicyclo[2.2.2]octane, bicyclo[3.3.1]nonane, decahydro-naphthalene, spiro-[5.5]undecane, and bicyclohexyl, in that order.

(xiii) Lowest letters in the expression for ring junctions: e.g., dibenzo[a,h]anthracene is senior to dibenzo[a,-j]anthracene.

(xiv) Lowest numbers at the first point of difference in the numerical expression for ring junctions: e.g., tricyclo[5.3.1.10^{2,4}]undecane is senior to tricylo-[5.3.1.10^{3,5}]undecane.

(xv) Lowest state of hydrogenation: e.g., benzene is senior to cyclohexane.

(xvi) For trivial names, alphabetic precedence: e.g., pentacene senior to pentaphene senior to perylene senior to picene.

Should no decision occur at this point, AUTONOM limits its recursive treatment (for reasons of response time) and prints a fatal diagnostic.

The IUPAC rules continue in the following fashion:

(xvii) Lowest locant for indicated H.

(xviii) Lowest locant for yl position.

(xix) Lowest locant for suffix functions.

(xx) Maximum number of substituents as prefixes.

(xxi) Lowest locants for substituents taken as a set.

(xxii) Lowest locant for substituents in the citation order.

Notes: (a) The application of (ix) is always decided on the basis of the numbering of the particular ring system under consideration. This may be predetermined (as in the case of all entries in Tables II and VI, or algorithmically generated (as in the case of von Baeyer, spiro, Hantzsch-Widman and assemblies nomenclature). Since (ix) has precedence over all criteria (xii) and upwards, the following examples are worth noting:

2-aza-bicyclo[4.3.1]decane is senior to 3-aza-bicyclo-[3.2.2]nonane [compare example to (xii)].

Naphtho[1,2-f]quinoline (**25**) (nitrogen is in 4-position) is senior to Benzo[a]phenanthridine (**26**) (nitrogen is in 6-position) [compare example to (xiii)].
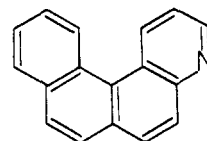
25
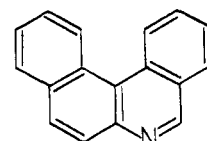
26

However,

1H-pyrrolo[3,2-c]quinoline is senior to 1H-pyrrolo-[2,3-g]quinoline [application of rule (xii) before rule (xiii), since in both cases the nitrogens are in positions 1 and 5, and thus rule (ix) offers no decision].

(b) The temporary application of (xvi) is necessary in AUTONOM since the IUPAC rules give no direct recommendation on this matter. We await a ruling on this point with interest.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Part 1: Wisniewski, J. L. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 324–332.
(2) International Union of Pure and Applied Chemistry, Organic Chemistry Division, Commission on Nomenclature of Organic Chemistry. *Nomenclature of Organic Chemistry, Sections A, B, C, D, E, F, and H*, 1979 ed.; Rigaudy, J., Klesney, S. P., Eds.; Pergamon Press: Oxford, England, 1979; 559 pp.
(3) Cooke-Fox, D. I.; Kirby, G. H.; Rayner, J. D. Computer Translation of IUPAC Systematic Organic Chemical Nomenclature. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 101 and references therein.
(4) Ref 2, p 6.
(5) Ref 2, p 7.
(6) Ref 2, p 9.
(7) Ref 2, p 80.
(8) Weissbach, O. Private communication.
(9) Ref 2, p 13.
(10) Ref 2, p 14.
(11) Ref 2, p 17.
(12) Ref 2, p 18.
(13) Ref 2, p 19.
(14) Ref 2, p 20.
(15) Ref 2, p 21.
(16) Ref 2, p 108.
(17) Ref 2, p 20.
(18) Ref 2, pp 22–27.
(19) Ref 2, p 28.
(20) Ref 2, p 30.
(21) Rücker, G.; Rücker, C. Nomenclature of Organic Polycycles out of the Computer—How to Escape the Jungle of the Secondary Bridges. *Chimia* **1990**, *44*, 116.
(22) Ref 2, p 31.
(23) Ref 2, p 35.
(24) Ref 2, p 38.
(25) Ref 2, p 42.
(26) Ref 2, p 106.
(27) Ref 2, p 46.
(28) Haberfield, P. *Chem. Eng. News* **1990**, *68* (34), 4.
(29) Ref 2, p 47.
(30) Ref 2, p 53.
(31) International Union of Pure and Applied Chemistry, Organic Chemistry Division, Commission on Nomenclature of Organic Chemistry, Revision of the Extended Hantzsch-Widman System of Nomenclature for Heteromonocycles (Recommendations 1982). *Pure Appl. Chem.* **1983**, *55*, 409–416 and references therein.
(32) Ref 2, p 56.
(33) Ref 2, p 62.
(34) Ref 2, p 64.
(35) Ref 2, p 101.
(36) Ref 2, p 64.

Atom-by-Atom Searching Using Massive Parallelism. Implementation of the Ullmann Subgraph Isomorphism Algorithm on the Distributed Array Processor[†]

PETER WILLETT* and TERENCE WILSON

Department of Information Studies, University of Sheffield, Western Bank, Sheffield S10 2TN, U.K.

STEWART F. REDDAWAY

Active Memory Technology, 65 Suttons Park Avenue, Reading RG6 1AZ, U.K.

The AMT Distributed Array Processor (DAP) is a massively parallel SIMD processor array that contains thousands of processing elements. This paper describes the implementation of atom-by-atom searching on the DAP using Ullmann's subgraph isomorphism algorithm. Two alternative algorithms are discussed. The first of these allows rapid processing of a single structure, the adjacency matrix of which is distributed across the array of processing elements. The second is much slower in execution for a single molecule but allows very large numbers of structures to be searched in parallel. With current codes, the first algorithm is faster and out-performs a large mainframe; however, developments of the second algorithm are expected to make this the faster. Combined algorithms are also described that utilize both approaches.

INTRODUCTION

Parallel processing involves the use of some, or many, processors operating in parallel to reduce the elapsed time that is required to carry out some computational task.[1-3] Parallel processing techniques have been used for scientific and engineering applications for many years; more recently, there has been increasing interest in the use of such techniques for applications that involve extensive nonnumeric or symbolic computation. One such application is that of database searching, which involves the matching of a query record against a large database of analogous records to identify those that satisfy some retrieval criterion. While each individual matching operation may be quite rapid in execution, searching can be highly demanding of computational resources when a large database containing many records needs to be processed:

* To whom all correspondence should be addressed.
† Dedicated to Michael Lynch to celebrate his 25 years at the University of Sheffield.

this suggests the use of a *data parallel*[4] system architecture that allows the query record to be compared with as many database records as possible at the same time. The last few years have seen many reports of the use of parallel computer hardware for a range of chemical and biochemical database applications. Examples of this include studies of searching protein sequence databases,[5,6] of clustering chemical structure databases,[7,8] and of both atom-by-atom[9-11] and maximal common substructure searching.[12]

In this paper, we report the initial results of an investigation of the implementation of atom-by-atom searching using a massively parallel computer, the Active Memory Technology (AMT) Distributed Array Processor (DAP). The second section describes the particular atom-by-atom searching procedure that we have investigated, this being based on Ullmann's subgraph isomorphism algorithm.[13] The third section then describes the DAP. The fourth section presents two, very different, ways of implementing the Ullmann algorithm on the