Computer Translation of IUPAC Systematic Organic Chemical Nomenclature. 5. Steroid Nomenclature

D. I. COOKE-FOX, G. H. KIRBY,* M. R. LORD, and J. D. RAYNER

Department of Computer Science, University of Hull, Hull HU6 7RX, England

Received October 12, 1989

The grammar-based approach to the computer recognition and translation of IUPAC systematic organic chemical nomenclature, described in previous papers in this series, has been applied to semisystematic steroid names. Steroid nomenclature is based on trivially named steroid parents, which may be modified and substituted by following the systematic rules of IUPAC organic nomenclature, and a special numbering scheme. Grammar rules are presented for the stigmastane and lanostane series and for the stereochemical and structural modifications applied to steroids. Semantic processing of acceptable steroid names produces a concise connection table (CCT), and definitions of the extensions to this are given for some of the stereochemical and structural modifications allowed by the grammar. The production from steroid names of Standard Molecular Data (SMD) format and character graphics displays of structure diagrams in the preferred orientation are reported.

INTRODUCTION

Steroids constitute a large and very important class of compounds based on a common tetracyclic ring skeleton. A specialized nomenclature has developed for naming these compounds.[1] Rules developed specifically for naming steroids have been used in other nomenclatures to name diturpenes and triturpenes and for several of the alkaloids.

The principles followed in the naming of steroids are (1) clearly defined stem names with the stereochemistry implied within, (2) systematic application of the rules of organic nomenclature on these stem names, and (3) systematic application of the skeletal modification rules of steroid nomenclature. The use of trivial names for the stem, with systematic application of the rules of organic nomenclature, is particularly common in naming steroids. The skeleton of steroids may be named more systematically as a derivative of cyclopentanoperhydrophenanthrene but would then require the stereochemistry to be explicitly stated for each chiral center. Furthermore, as noted by Garfield,[2] "it is ridiculous to call phenanthrene systematic when it could properly be called benzonaphthalene".

The grammar-based technique for the recognition and translation of chemical nomenclature to structure diagrams, described in previous papers in this series,[3,4] has been applied to steroids. The purpose of this exercise was to demonstrate the application of the technique to a particular class of industrially important compounds where names are based on trivial nomenclature and a specialized numbering.

The systematic numbering of the stigmastane/lanostane steroid nucleus[1] is shown in Figure 1. The numbering is retained even if some of the lower numbered atoms, such as the methyl groups at positions 10 and 13, are missing. The concise connection table (CCT) entries produced in the semantic phase of nomenclature translation have to take into account this steroid specific numbering. The CCT was developed by Rayner[5] to describe structural fragments represented by the parts of a systematic name. It has since been extended to cover additional features including stereochemistry and the steroid nucleus.[6]

The only other study of nomenclature translation applied to steroids, reported by Stillwell,[7] used a small dictionary of steroid nucleus names and common substituent terms. By means of ad hoc processing methods, a tabular representation of molecular geometry was formed for each input name, and diagrammatic output was then produced. Stillwell made no use of formal syntactic methods and concentrated on the stereochemical aspects of steroid structures in his output.

SYNTAX ANALYSIS

To demonstrate the feasibility of applying the nomenclature translator to steroids, it was necessary to recognize a subset of the stem names. These are the stigmastane and lanostane series:

gonane	cholane	lanostane	cucurbitane
estrane	cholestane	tirucallane	protostane
androstane	ergostane	euphane	
pregnane	stigmastane	dammarane	

Only substitutive nomenclature is handled, as this is the system most often applied to steroids.

Derivatives. The above stems may be modified by the inclusion of double and triple bonds in the skeleton. Double bonds that are placed between numerically nonadjacent bridgeheads require both the start and end locants of the bond to be explicitly stated, for example, gon-5(10)-ene. All of the substituents and principal groups that are known to the full grammar, and reported in paper 2,[3] can be applied to steroids. Most of these substituents can be designated either suffixes or prefixes. A few substituents, such as the halogens, can only be named as prefixes, for example, 2-chloro-5α-stigmast-3-ol.

Stereochemical Modifications. All of the principal stereochemical modifications for steroids are handled by the grammar. These include *ent*, *retro*, (±)-*rel*,*R*/*S*, $\alpha/\beta/\xi$ bonds, and *rac*.

This is the first class of compounds for which stereochemistry was introduced in the grammar. The asymmetric center can be represented by the $\alpha/\beta/\xi$ bond or the *R*/*S* system depending on whether it is attached to the tetracyclic ring skeleton or to a side chain, respectively. These notations have now been applied also to all the other classes of compounds currently handled by the full grammar.

Structural Modifications. Ring contractions and expansions are indicated by the prefixes nor and homo. Implementing this feature in the grammar resulted in lettered and ring locants being introduced for the first time, as in C(14a)-homo-5α-androstane. Dinor and dihomo are allowed when loss or insertion of two methylene groups is required.

Fission of a ring, with an additional hydrogen atom being created at each of the two new terminal nodes, is indicated by the prefix seco, for example, 2,3-seco-5β-cholestane. Additional ring formation, signified by cyclo, is allowed. Bond migration, indicated by *abeo*, has not been implemented.

Finally, the occurrence of hetero atoms in the steroid ring system is indicated by the "oxa-aza" replacement system of nomenclature.

0095-2338/90/1630-0128$02.50/0

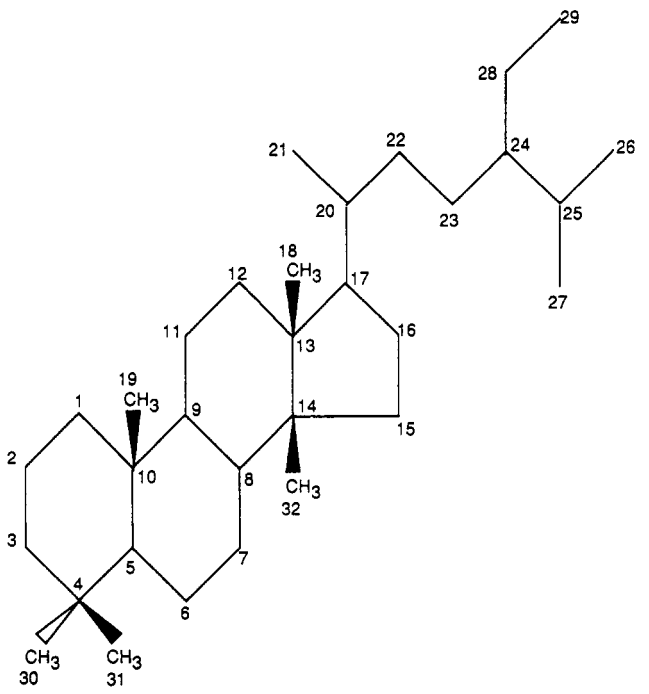

Figure 1. Systematic numbering within the steroid nucleus.

The newly created grammar rules for steroids are given in Appendix, in the format explained in paper 2,[3] together with restatement of earlier rules that needed modification. Steroids have been introduced into the grammar, at the highest level, by the new rule alternative

org-name = steroid-name,

Thus, the nonterminal symbol, "steroid-name", can be expanded, in a top-down fashion, to describe any steroid currently handled by the grammar. In addition, a new rule, defining a nonterminal symbol called "parent-part", has been introduced to allow steroid parents to have prefixed substituents. This modification has led to all those rules for certain classes of compounds, such as acids, alcohols, aldehydes, and ketones, which used to comprise an "aliph-part" and a unique class ending (e.g., ol for alcohol), now comprising a parent-part and the unique ending. This parent-part can either be aliph-part, as before, or the new "steroid-part". This methodology shows the ease with which a whole new class of compounds can be introduced, with only a few minor changes to the existing grammar.

The Greek symbols α, β, and ξ are input, via the keyboard, as alpha, beta, and xi. When this method is used, there needs to be an explicit hyphen between the locant and the modification. This is not the case when a Greek symbol is used, for example, 5α-stigmastane would be input as 5-alpha-stigmastane. Hence, the rules describing the locant and its stereo bond modification are

locant = num-mod number;

num-mod = config-mark hyphen, ..., $;

with the new terminal symbol "config-mark" representing alpha, beta, and xi.

SEMANTIC PROCESSING AND CCT REPRESENTATION

The production of the concise connection table (CCT) in the semantic phase of the name translation has been described previously.[4] In the formation of a CCT representation for a steroid, the output from the semantic phase must include the special numbering scheme and stereochemical and structural modifications. Stereochemistry had not previously been handled by the CCT or by the chemical character graphics technique.[6]

LOCT	TIPE	SIZE	SUBS
1	5	17	0
6	6	6	5
13	1	1	2
10	1	1	2
17	1	2	0
20	1	3	0
24	1	2	0
25	1	1	0
24	1	1	0
28	1	1	0
4	1	0	1
4	1	0	1
14	2	1	1
20	2	1	2
17	2	1	1
5	2	1	0
8	2	1	2
9	2	1	1

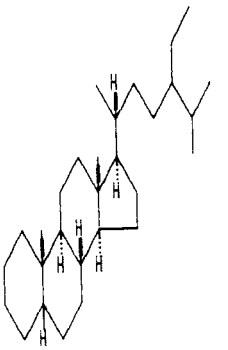

Figure 2. CCT for 5ξ-stigmastane and structure diagram drawn by character graphics.

Steroid CCT Definitions. A new TIPE, 5, was introduced for the steroid CCT main entry. The SIZE field contains a count of the number of CCT steroid segment entries (SSE), which follow the main entry to describe the particular parent. This count includes the steroid nucleus entry, describing the tetracyclic ring skeleton typically by the four ring-cue values 6 6 6 5. The SUBS field of the header entry contains the number of substituents or modifications on the parent, to give the desired overall structure.

The SSEs following the nucleus entry are arranged in the CCT so that, in the drawing phase, the entries in the atom table are placed in locant sequence. In all the segment entries, the LOCT field contains a positive integer that gives the position of attachment of the structure described by the CCT entry to the partially constructed steroid structure.

The TIPE field of an SSE can take one of two values. A TIPE value of 1 indicates that the entry is an aliphatic chain or dummy entry. If the SIZE field is 0 when the TIPE is 1, the entry is a dummy CCT entry and the SUBS field contains a count of the number of dummy entries that need to be added to the atom table to retain the nonsystematic numbering. This is necessary to find the position of attachment of, for example, a 30-chloro group to the lanostane series of parents which do not have the ethyl group (C-28 and C-29) at position 24.

If the SIZE field is not 0, the entry is an aliphatic chain whose length is given by the SIZE field. The SUBS field then has one of the values 0, 1, or 2 to indicate the stereochemistry of the bond connecting the first atom of the chain to the atom being substituted within the steroid nucleus. If the SUBS field is 1, the bond is an α bond and is below the plane of the steroid. If the SUBS field is 2, the bond is a β bond and is above the plane of the steroid, while a SUBS value of 0 indicates that the bond is a ξ bond and may be either above or below the plane.

A TIPE value of 2 indicates that the entry is an atom entry, in which the SIZE field contains the atomic number of the atom substituent and the SUBS field indicates the bond stereochemistry as described above.

The use of the SUBS field for describing bond stereochemistry is specific to the context of the steroid nucleus and contrasts with the use of ordinary bond modification entries in the case of substituents on the nucleus or elsewhere (see later under CCT extensions for steroid modifications, and Table II of ref 6).

The CCT for 5ξ-stigmastane constructed by the semantic processing and the corresponding structure diagram, drawn by character graphics, are shown in Figure 2. Following the

steroid header entry and steroid nucleus entry, the first two segment entries represent methyl groups (aliphatic chains of size 1) at positions 13 and 10 in the β orientation. The next six steroid segment entries, up to locant 28, describe the carbon chain attached at position 17 of the nucleus.

Not all steroid parents have the full complement of carbon atoms in this chain or all the methyl groups attached to the nucleus. Comparison of Figure 1 with 5ξ-stigmastane in Figure 2 shows that the latter parent lacks the two methyl groups at position 4, numbered 30 and 31. Hence, in the 5ξ-stigmastane CCT there are two dummy entries at locant 4, identified by being chains of size 0, with SUBS field value of 1 indicating the number of carbon atoms missing from the steroid numbering system. Although 5ξ-stigmastane lacks a methyl group at position 14, a dummy entry is not required because there is an explicit α hydrogen atom locant there.

The remaining six steroid segment entries in the example CCT are atom entries for hydrogen (SIZE = 1) that need to be explicit to indicate the stereochemistry at the bridgehead of the nucleus (positions 5, 8, 9, and 14) and at the asymmetric centers at positions 17 and 20. The latter two entries result in the structure diagrams for those steroid parents that have asymmetric centers at locants 17 and/or 20 (i.e., all but androstane, estrane, and gonane) displaying explicit hydrogen atoms at these positions. In fact, the presence of these explicit hydrogen atoms is not necessary for the display of those parents in the stigmastane series, because for all these parents the carbon chain attached at position 17 is assumed to be β oriented and the C21 methyl group at position 20 is assumed to be α oriented.

However, with the lanostane series this is not the case, and the above modification can indicate clearly that the carbon chain attached at position 17 is α oriented for tirucallane, euphane, and protostane and, second, that the C21 methyl group at position 20 is β oriented for tirucallane. Without this distinction, the parent structures for tirucallane and euphane would appear the same.

Although the syntactic recognition of the *R*/*S* terminology is handled, semantic processing is extremely difficult and has not yet been implemented. The rules for determining whether a substructure should be labeled *R* or *S* are very complex, involving detailed attention to the structure of the parent molecule in each individual case. Beyond that, there are potential problems in structure display since two-dimensional representations are ambiguous and existing display packages typically do not provide a sufficient variety of bond symbols to make the stereochemistry explicit.

CCT Extensions for Steroid Modifications. As all the CCT extensions for steroids are modification entries, they all have a LOCT field value of zero and a nonzero value in the TIPE field.

Additional ring formation, cyclo, has a TIPE field value of 3 to indicate a bond modification:

LOCT	TIPE	SIZE	SUBS
0	3	3	0
locant 1	config 1	config 2	locant 2

The 3 in the SIZE field represents the cyclo bond, with the start and end locants and their corresponding stereochemical configurations being given in the next CCT entry, as indicated above by locant 1, locant 2, config 1, and config 2.

The ring fission entry, seco, is also a bond modification entry depicted by a SIZE field value of 4 with the start and end locants given in the following CCT entry.

LOCT	TIPE	SIZE	SUBS
0	3	4	0
locant 1	0	0	locant 2

The elimination of methyl groups, the shortening of side chains (nor), and ring contraction and expansion modifications (nor, homo) are all depicted as hetero atom modifications (TIPE = 2):

LOCT	TIPE	SIZE	SUBS	
0	2	0	locant	nor for chains
0	2	0	ring	nor for ring contraction
0	2	6	ring	homo for simple ring expansion
0	2	6	locant	homo for ring expansion in
ring 1	0	0	ring 2	bridgehead

A zero in the SIZE field indicates that the carbon atom will not be replaced by anything. Hence, in the case of elimination of a methyl group and shortening of side chains the SUBS field contains the locant of the carbon atom to be removed. With ring contraction, the SUBS field holds the ring concerned. Rings are represented by 100 for A, 200 for B, 300 for C, and 400 for D, to distinguish them from locants. A 6 in the SIZE field (atomic number of carbon) implies an expansion. The SUBS field gives the ring concerned. However, describing a ring expansion, created by formal insertion of a methylene group between directly linked bridgeheads, requires an additional following CCT entry. This gives the ring(s) affected by the insertion, as shown above, with ring 2 in the SUBS field being zero if only one ring is affected.

Any hetero atoms occurring in the steroid ring system are represented by a hetero atom modification (TIPE = 2) with the SIZE field containing either the atomic number of oxygen or the atomic number of nitrogen, depending on whether an oxa or aza modification is present, respectively.

LOCT	TIPE	SIZE	SUBS
0	2	atomic number	locant

The stereochemistry of substituents, rather than that within the steroid nucleus, is represented by modification entries using the bond codes 0, 1, or 2 (Table II of ref 6):

LOCT	TIPE	SIZE	SUBS	
0	3	0	0	xi bond
0	3	1	0	alpha bond
0	3	2	0	beta bond

Thus, the CCT for 3α-chloro-5β-gonane is represented by

1	5	17	1
	⟨the 17 SSEs⟩		
3	2	17	1
0	3	1	0

The *ent* stereochemical modification is dealt with by replacing all the α bond CCT entries by β bond entries, and vice versa, in the SSEs and any appropriate substituent entries.

CCT Generation. The CCT is constructed from appropriate entries as determined by reference to a dictionary of name fragments or morphemes.[4] To conserve space in this fragment dictionary, not all the structure of each steroid stem is stored, merely the modifications that need to be made to a predefined default steroid. The CCT for this default steroid is created by the semantic rules and then modified by the CCT entries in the fragment dictionary for the stem being processed. The structure chosen to be the default steroid is 5ξ-stigmastane, the only parent that has the complete side chain on carbon 17 (C-20–C-29). The use of ξ at position 5 is introduced as it is the only ring junction that is not implied implicitly by the stem used. Hence, there is no semantic information in the fragment dictionary for stigmastane other than the steroid nucleus entry 6 6 6 5. The CCT for 5ξ-stigmastane contains entries for all those positions on the ring system which may be different in any of the other steroid parents. This method requires the presence of two dummy entries both at locant 4, because all the lanostane series have two methyl groups at that position. Thus, to translate the default parent CCT to any other parent CCT only requires modification of the existing entries, rather than addition of new ones or deletion of old ones. In the fragment dictionary, only the entries that differ from

the default one for a particular parent are stored following the steroid nucleus entry. During the semantic processing, the steroid nucleus entry is expanded to the full CCT for the 5ξ-stigmastane parent structure (Figure 2), modified first by those entries for the particular steroid parent and, second, by any additional substituents or modifications on that parent, such as unsaturations, appearing explicitly in the name. The latter stage always involves the addition of new CCT entries, but may result in the deletion of a steroid segment entry if, for example, a substituent is replacing a parent group or atom.

CCT to SMD Conversion for Steroids. Steroid CCTs are converted to Standard Molecular Data (SMD) format[8] by the following method.

First, a procedure is executed, which constructs the partial atom and bond linked lists for the first 17 locant positions of the 5ξ-stigmastane skeleton (the default structure). At this stage the carbon atom at locant 17 only has two bond entries making up the five-membered ring and no implied hydrogens.

The complete parent structure is then built by investigating in turn each of the CCT steroid segment entries (SSE) following the nucleus segment and using the information to add to or modify the existing data structures. For each SSE the TIPE and SIZE fields are used to determine the entry type, which can be chain, dummy, or atom. The atom and bond lists are then constructed by using a combination of the entry type and locant for the appropriate SSE.

DISPLAY OF STEROID MOLECULAR STRUCTURES

The character graphics technique[9] has proved adequate for simple derivatives of those steroid parents that can be handled (see Figure 2). To allow the stereochemistry at the bridgeheads of a steroid skeleton to be explicitly shown, a number of new chemical characters have been implemented.[6] α bonds are represented by a dashed line, β bonds by a solid line, and ξ bonds by the normal bond character.

Many steroid features that can be handled by the grammar cannot be drawn at present by character graphics. For example, 3α,5-cyclo-5α-cholestan-6β-ol has a bond connecting atoms 3 and 5 in the steroid ring system. The graphic characters required for this bond are not present in the font since it is not possible to allow for all the variations of such characters in a fond of 128.

The software to display steroids by means of the character graphics technique has been written so that the preferred orientation of the steroid nucleus is displayed with the A ring to the lower left and the D ring on the upper right. Use of molecular graphics packages that generate their own coordinates from our SMD CT block[6] does not necessarily ensure the correct orientation. This problem has been recognized by Shelley.[10]

DISCUSSION AND CONCLUSIONS

It has been shown possible to recognize some trivally named steroid parents and to modify the steroid nucleus with a grammar derived from the systematic rules of IUPAC organic nomenclature. While changes to the IUPAC recommendations on the nomenclature of steroids are imminent, these are not expected to result in the need for major changes to the grammar presented here or to our current software.

Difficulties do exist, however. The steroid parents cardanolide and bufanolide behave differently to those currently handled, and the necessary additional rules for these are expected to be rather more complex. Particularly, their behavior with functional group derivatives is problematic, and these names will therefore not be included in the foreseeable future. The same applies to spirostan and furostan, although the ending "-an" is similar to the present "-ane" ending, making their implementation more straightforward.

Problems arise not in the construction of a grammar to recognize the nomenclatural forms that are given in the IUPAC steroid rules but rather in generating a numeric representation of the steroid structure, and the many amended forms, and in checking that the CCT produced is structurally valid. In the semantic tree,[4] the steroid segment entries are held under the parent entry of the semantic tree node in the same way as aromatic ring segment entries. They are therefore not particularly easy to access and check for the validity of the parent. The CCT has been shown to be extendable in handling some of these many forms that are implemented in the grammar. However, some of these extensions do not strictly conform with the original definitions of the LOCT, TIPE, SIZE, and SUBS fields.[5]

The exercise reported here shows that it is reasonable to provide translators for subsets of IUPAC nomenclature for specialist classes of compounds, which may involve semisystematic nomenclature. However, such translators rely on a system for general nomenclature to handle substituents, principal groups, and bond modification. Where, as here, the class is built around one structural skeleton with its own numbering scheme, the semantic processing can use this by providing the necessary modifications to the basic skeleton. The display of stereochemistry has been accomplished with the character graphics technique, but general coverage of ring formation, fission, contraction, and expansion is beyond its capabilities.

ACKNOWLEDGMENT

This work was supported by the U.K. Laboratory of the Government Chemist, and we are particularly grateful to E. W. Godly and I. Cohen of its Chemical Nomenclature Advisory Service for their help and advice.

APPENDIX: CURRENT STEROID GRAMMAR FOR THE IUPAC SYSTEMATIC ORGANIC CHEMICAL NOMENCLATURE

TERMINALS

```
...
steroid-suffix      = e ;
...
steroid-mark        = gon , estr , androst , pregn , chol ,
                      cholest , ergost , stigmast , lanost ,
                      tirucall , euph , dammar ,
                      cucurbit , protost ;
...
config-mark         = alpha , beta , xi ;
seco-mark           = seco ;
nor-mark            = nor ;
ring-locant         = A , B , C , D ;
let-locant          = a , b , c , d , e , f , g , h , i , j , k ,
                      l , m , n , o , p , q , r , s , t , u , v ,
                      w , x , y , z ;
homo-mark           = homo ;
h-mark              = H ;
h-ket               = ' ) ' ;
oxa-aza-mark        = oxa , aza ;
stereo-hyphen       = '-' ;
ent-mark            = ent ;
retro-mark          = retro ;
rac-mark            = rac ;
rs-mark             = R , S ;
rel-mark            = rel ;
sign-mark           = '+' , '-' ;
...
```

RULES

```
name                = org-name ;
```

```
org-name        = aliph-name ,
                  arom-name ,
                  acid-name ,
                  conj-name ,
                  alcohol-name ,
                  ketone-name ,
                  aldehyde-name ,
                  radico-name ,
                  steroid-name ,
                  special-case ;
...
parent-part     = aliph-part ,
                  steroid-part ;
...
acid-name       = acid-mark acid-space acid-parent ;
acid-parent     = oic-part parent-part ,
                  dioic-part parent-part ,
                  carb-suffix carb-root ,
                  unsub-aliphacid ,
                  sub-aliphacid ;
...
carb-root       = aliph-suffix parent-part ;
...
alcohol-name    = ol-part alcohol-parent ;
alcohol-parent  = parent-part ,
                  alc-en-mark arom-parent sub-part ,
                  trivalcohol-part sub-part ;
...
ketone-name     = one-part ketone-parent ;
ketone-parent   = parent-part ;
...
aldehyde-name   = al-part parent-part ,
                  dial-part parent-part ,
                  carbald-suffix carb-root ,
                  glyoxal ,
                  aldehyde-mark triv-alds ;
...
steroid-name    = steroid-suffix steroid-part;
steroid-part    = steroid-parent sub-part
                  stereo-mod ;
steroid-parent  = sat-seq steroid-stem ndprefix-part ;
steroid-stem    = steroic-root loc-part ;
steroid-root    = steroid-mark ;
ndprefix-part   = nondet-prefix ndprefix-part ,
                  $ ;
nondet-prefix   = ring-form ,
                  ring-break ,
                  size-change ,
                  hetero-mod ;
ring-form       = cyclo-mark loc-part ;
ring-break      = seco-mark loc-part ;
size-change     = rg-contractn ,
                  rg-expansion ;
rg-contractn    = nor-mark gm-part hyphen
                  rg-loc-part ;
rg-loc-part     = num-loc-part ,
                  rg-loc-seq ;
num-loc-part    = number num-loc-seq hy-part ;
num-loc-seq     = comma number num-loc-seq ,
                  $ ;
rg-loc-seq      = ring-locant hy-part ,
                  ket let-locant number bra
                  ring-locant secnd-rg-loc hy-part ;
secnd-rg-loc    = ring-locant ,
                  $ ;
rg-expansion    = homo-mark gm-part hyphen
                  rg-loc-part ;
hetero-mod      = oxa-aza-mark loc-part ;
stereo-mod      = stereo-hyphen stereo-desct,
                  $ ;
stereo-desct    = rs-mod ,
                  ent-mod ,
                  retro-mod ,
                  rac-mod ,
                  rel-mod ;
rs-mod          = rs-loc-part ;
rs-loc-part     = ket rs-locant rs-loc-seq bra ;
rs-locant       = rs-mark number ;
rs-loc-seq      = comma rs-locant rs-loc-seq ,
                  $ ;
ent-mod         = ent-mark ;
retro-mod       = retro-mark ;
rac-mod         = rac-mark ;
rel-mod         = rel-mark sign-part ;
sign-part       = hyphen ket sign-mark bra ,
                  $ ;
...
locant          = num-mod number ;
num-mod         = ket number bra ,
                  config-mark hyphen ,
                  h-ket h-mark bra config-mark
                  hyphen ,
                  $ ;
...
ROOTSYMBOL      name
```

REFERENCES AND NOTES

(1) IUPAC/IUB. Definitive Rules for Nomenclature of Steroids (1971). *Pure Appl. Chem.* **1972**, *31*, 283–322.
(2) Garfield, E. An Algorithm for Translating Chemical Names to Molecular Formulas. In "*The Awards of Science and Other Essays*; ISI Press: Philadelphia, 1985; pp 463–464.
(3) Cooke-Fox, D. I.; Kirby, G. H.; Rayner, J. D. Computer Translation of IUPAC Systematic Organic Nomenclature. 2. Development of a Formal Grammar. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 106–112.
(4) Cooke-Fox, D. I.; Kirby, G. H.; Rayner, J. D. Computer Translation of IUPAC Systematic Organic Chemical Nomenclature. 3. Syntax Analysis and Semantic Processing. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 112–118.
(5) Rayner, J. D. A Concise Connection Table Based on Systematic Nomenclatural Terms. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 108–111.
(6) Cooke-Fox, D. I.; Kirby, G. H.; Lord, M. R.; Rayner, J. D. Computer Translation of IUPAC Systematic Organic Chemical Nomenclature. 4. Concise Connection Tables to Structure Diagrams. *J. Chem. Inf. Comput. Sci.* (preceding paper in this issue).
(7) Stillwell, R. N. Computer Translation of Systematic Chemical Nomenclature to Structural Formulas—Steroids. *J. Chem. Doc.* **1973**, *13*, 107–109.
(8) Bebak, H., et al. The Standard Molecular Data Format (SMD Format) as an Integration Tool in Computer Chemistry. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 1–5.
(9) Rayner, J. D.; Milward, S.; Kirby, G. H. A Character Set for Molecular Structure Display. *J. Mol. Graphics* **1983**, *1*, 107–110.
(10) Shelley, C. A. Heuristic Approach for Displaying Chemical Structures. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 61–65.