Synthetic Steroid Hormones. Part 1. Estrone and Related Estrapolyenes. J. Chem. Soc. 1963, 5072-5094.
(9) Hendrickson, J. B. Systematic Synthesis Design. 6. Yield Analysis and Convergency. J. Am. Chem. Soc. 1977, 99, 5439-5450.
(10) Each full bondset, at both levels, defines an ordered sequence of constructions, hence a family of routes differing only in the involved functional groups. The simplest overall description of any synthetic route is just its ordered bondset, i.e., simply the sequence of skeletal bonds constructed in the course of the synthesis.
(11) Hendrickson, J. B.; Braun-Keller, E. Systematic Synthesis Design. 8. Generation of Reaction Sequences. J. Comput. Chem. 1980, 1, 323-333.
(12) Wochner, M.; Brandt, J.; von Scholley, A.; Ugi, I. Chemical Similarity, Chemical Distance, and Its Exact Determination. Chimia 1988, 42, 217-225.
(13) Hendrickson, J. B. Descriptions of Reactions: Their Logic and Applications. Recl. Trav. Chim. Pays-bas 1992, in press.
(14) Allowing the first dissection of the target to cut off one- and two-carbon pieces generates many first-level bondsets but leaves the other intermediate piece so big that it is very unlikely that any further second-level dissection will then find real starting materials, and so such cuts are almost always wasteful.
(15) Estrone itself does not appear here since it is only successfully made by other bondsets.
(16) Billups, W. E.; Cross; Smith, C. V. A Synthesis of (±)-Grandisol. J. Am. Chem. Soc. 1973, 95, 3438-3439.
(17) For reasons involving the restrictions on generating these reactions[7] the exact dimerization does not appear.
(18) (a) Oppolzer, W.; Battig, K. A. Short and Efficient Synthesis of (±)-Modhephene by a Stereoelectrically-Controlled Ene-Reaction. Helv. Chim. Acta 1981, 64, 2489-2491. (b) Schostarez, H.; Paquette, L. A. Highly Stereocontrolled Synthesis of [3.3.3]Propellane Sesquiterpenes. (±)-Modhephene and (±)-Epimodhephene. Tetrahedron 1981, 37, 4431-4435.

Algorithm for Selecting the Parent Structural Unit of a Ring–Chain Assembly

SCOTT DAVIDSON*

Computer Data Systems, Inc., One Curie Court, Rockville, Maryland 20850

Received November 18, 1991

Selection of the parent structural unit in naming ring–chain assemblies by computer methods can be difficult. The IUPAC rules are sufficiently general to permit individual judgments in some cases. Nodal nomenclature employs a quite specific but very lengthy set of rules to give unique names. A recently reported improved method for uniquely identifying alkanes defines the main chain as the chain with the least complex side chains. This definition readily extends to ring–chain assemblies when rings are included, applies regardless of parent-unit size, and is consistent with the IUPAC guidelines and examples given, such as the preference for diphenylmethane over benzyl benzene in naming Ph–CH_2–Ph. An iterative procedure for selecting the parent unit and a simple method for linking units are described as part of a general skeleton-naming computer program.

INTRODUCTION

In organic chemical nomenclature, the selection of a parent structural unit is essential to naming an assembly of rings and chains. The selection process is complicated by the fact that in an assembly the individual units are independent with respect to simplification. In a single-unit carbon skeleton, selection of a main chain or ring usually removes most carbon atoms from the naming process, whereas with an assembly complete identification of one unit provides little information about the others, which can have arbitrary sizes and shapes. IUPAC nomenclature (Rule A-61.2)[1] provides two flexible, highly intuitive general guidelines for selecting the parent unit: "(a) the maximum number of substitutions into a single unit of structure; (b) treatment of a smaller unit of structure as a substituent into a larger". However, as reported by Goebels et al.[2] in reference to operation of the AUTONOM structure-naming program, such guidelines are not very well suited to computer implementation, but on the other hand the fixed rule of rings senior to chains employed by CAS can lead to overcomplicated names for some simple compounds.

Nodal nomenclature[3] provides a specific sequence of rules for selecting parent units (modules) with size as the highest level qualifier, followed by a set of 16 rules for numbering the remaining units. The basic concept of defining the entire skeleton first as single-bonded carbon and then overlaying bond and atom types is well suited to computerization. The use of the perimeter ring (largest ring) as the basic structural unit of multibridged systems in place of the bicyclic skeleton of the extended von Baeyer method simplifies structures and also avoids the proliferation of numbering schemes that became standard for common ring systems, bridged rings, and spiro structures before systematic nomenclature was developed.

Nomenclature rules commonly use linear sequences of tiebreaker tests to arrive at a unique name. The danger in this approach is that any one test that relies on incomplete information can cause the entire series to fail. For example, selecting the longest chain in an alkane requires examination of the complete structure. However, the first step in the IUPAC tiebreaker series (Rule 2.6a) for equal-length candidate main chains is to select the chain with the most side chains. This is a simple test—especially for a computer—because it relies on blind groping along a path. However, it ignores side-chain detail, allowing, for example, two methyl groups on one chain to override any single but highly complex side chain on the other. The first paper[4] shows how this can lead to nonunique names and describes a recursive algorithm for selecting an alkane main chain by side-chain complexity minimization. This alternative approach encodes and sorts structural details from the bottom up by depth-first search to give an ordered list (simple side chains + smaller lists encoding complex units). This can be compared with other lists representing alternate configurations from the top down as a tiebreaker series by symbolic substitution. The first value compared is the overall size of the skeleton, which of course is the same for all configurations, but permits a database of skeletons to be ordered in the same manner beginning with size. Recording details and then breaking ties at the lowest levels first ensure that all information is utilized in making decisions at the next higher level, etc. The main chain is

* Address all correspondence to this author at 240 Manor Circle 2, Takoma Park, MD 20912.

0095-2338/92/1632-0215$03.00/0

selected on the basis of how it affects the entire skeleton, as measured by the complexity of the resulting side chains. No candidate main chain, whether for the alkane or one of its side chains, is compared with the current best until all of its potential side chains have been completely characterized. The resulting main-chain definition, "the chain with the least complex side chains", extends readily to assemblies by including cyclic paths. Because structural units are independent, it is not necessary to constrain the definition to the set of longest paths as in an alkane. The meaning of the word "chain" in the definition then broadens to include parent units of arbitrary sizes and shapes in assemblies. The seniority order for alkanes (size, length, locant) is modified to become (size, length, type, locant), where type is either ring or chain, with rings senior. Two more extensions are required: (1) all cyclic side chains are complex and (2) simple side chains (C_1–C_4) cannot be parent units but do contribute to the overall size of a candidate parent unit. For example, if an assembly consists of two units that have the same total size and longest path, but one is a ring unit and the other a chain unit, then the ring unit is more complex as a side chain and is thus preferred as the parent unit.

FROM GUIDELINES TO ALGORITHM

An assembly may be visualized as a string of fishing line sinker weights varying in size and shape. If the assembly were suspended by each weight in turn to give two strings of weights, then the parent unit would be the weight that most nearly balances the assembly, as detailed under IUPAC (Rule 61.2) guideline a below. Guideline b is a special case where an end unit is the parent (one null string), while a single-unit skeleton has two null strings. With this picture in mind, the two general guidelines can be made more specific in terms of side-chain complexity (beginning with the simpler case) as follows:

(b) "...smaller unit of structure as a substituent into a larger". All four of the IUPAC examples (61.4) show two unsubstituted units; however, this guideline can be extended somewhat as follows: Given A – B, where it is assumed that A ≥ B (size), then parent unit A is a single unit that can include small side chains and B is either a single unit or represents a contraction of units. If B contains several units, then as parent, end-unit A must of course be larger than all other units combined.

Examples: 1-cyclobutylpentane; cyclobutylmethylcyclohexane (Figure 11); hexylcyclohexane; butylcyclopropane (an exception because simple side chains cannot be parent units)

(a) "the maximum number of substitutions into a single unit..." Three or more units are contracted to the form A – P – B, where P is a candidate parent unit and A and B are the two largest side chains of P.

Case 1. If A ≥ B and A < P + B (less than half the total carbons), then this configuration is accepted (as a unique solution because of the two above constraints).

Example: Ph–CH_2–Ph, diphenylmethane

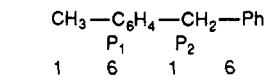
The most complex side chain has six carbons, benzyl benzene has seven. By the IUPAC guidelines, b is impossible, unless bent to include contractions of units that have trivial names (α-phenyltoluene). Therefore guideline a controls by default because there are no other choices. On the other hand, where there are choices, guideline a does not appear to be a very effective test (as shown below), but instead serves mainly to confirm the observation that the more substituents a candidate parent unit has, the less likely that one of them will be larger than all the others combined. If guideline b is further bent to permit the named unit to be the one to which B is connected, then this guideline also applies to A < P + B cases, giving phenyl(phenylmethane) in the above example, which then simplifies. However, it seems much simpler to turn the test around and examine the side chains of candidate parent units.

Case 2: If A = P + B (exactly half) and A is a single unbranched unit, then P is the parent unit, because PB is a more complex side chain having B branch carbons compared with none for A.

Example: cyclopentylcyclohexylmethane (*not* cyclopentylmethylcyclohexane)

More generally, parent-unit ties may be represented in the form A – B – C – D where A + B = C + D; A and D are the number of side-chain carbons of the adjacent candidate parent units B and C. If B > C then A < D, and C is the parent unit because side chain AB has fewer branch carbons.

Example:

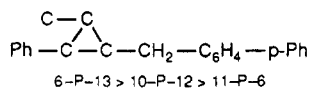


a phenyltolylmethane (*not* benzyltoluene)

Since the largest side chain of each candidate parent unit has seven carbons, the tie is broken by comparing the number of side-chain branch carbons. Tolyl with one is less complex than benzyl (phenylmethyl) with six. Note that neither IUPAC guideline is very helpful in this case.

Case 3: If A > P + B then P is disqualified and another configuration tested.

Example:



4-phenyl-1-(3-methyl-2-phenylcyclopropylmethyl)-benzene

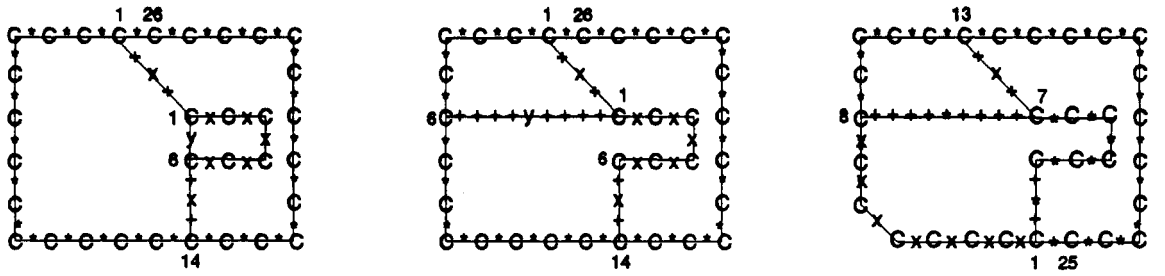
From left to right, the first two candidate parent units are rejected because the largest (most complex) side chain contains more than half of the 23 total carbons. The third parent unit is accepted without further testing because 11 is less than half. Note that guideline a selects the cyclopropane ring as the parent unit while guideline b and the allowed use of trivial names selects biphenyl as a unit. (As in nodal nomenclature, complexity naming considers biphenyl to be two ring units to avoid trivial names.) If A is a single unit (including simple side chains) then a further contraction A – PB reduces to type b with A as parent unit.

Example: $C_6H_3(CH_3)_2$–CH_2–Ph, a benzylxylene

The hard rule in nodal nomenclature and the guideline in IUPAC nomenclature that the parent unit be as large as possible also makes sense for complexity minimization, but only as a visual aid or computer strategy. While the least complex configuration is more likely to have the largest unit as parent than as part of a side chain, this strategy fails badly in the simple case of diphenylmethane and similar structures. For computer evaluation, we could simply test each unit as a candidate parent unit. However, the usual approach is not to test units but to look for the largest ring or longest chain. With either method, processing can stop when the largest side chain has less than half the total carbons in the skeleton or after two ties have been evaluated as indicated above—and in the discussion of programming details that follows.

SKELNAME PROGRAM

The present skeleton-naming program (SKELNAME) began as a simple alkane-naming program written and distributed via a journal note in 1973.[5] The initial program was limited to C_1–C_4 side chains and later extended to handle one complex side chain (there was no facility for ordering two or more). In a way, this limitation was better for instructional use in beginning organic chemistry because the student was asked to design, name, and then input the 2-D structural diagram



(a) (b) (c)

[1,14]-([1,6]hexano)cyclohexacosane [1,14]-([1,6']-hexano)cyclohexacosane [7,13]-[1,8]-hexanocyclopentacosane

1c (Nodal): tricyclo[025.6^{1,8}0^{7,13}]hentriacontane

1c (IUPAC): tricyclo[12.11.6.0^{19,31}]hentriacontane (outside ring, reverse #)

Figure 1. Computer-named bridge skeletons: (a) nested, (b) side, (c) main-ring tie [(b) – one C simplifies to two independent bridges].

(on cards) for checking. The restriction to one complex side chain acted to discourage haphazard constructions. The precise detail required to "teach" a computer the IUPAC rules and extensive testing were what eventually led to the detection of discrepancies in the main-chain tiebreaker rules (A-2.6a...d) previously reported.[4] In recent years the program has been redesigned to name general alkanes and then extended to name general carbon skeletons. Currently, the Beilstein Institute's AUTONOM program[2] is by far the most comprehensive development in this area, naming general organic structures algorithmically by the IUPAC rules and employing a dictionary of common ring names where applicable. SKELNAME is limited to carbon skeletons, but will name—within reason—arbitrary assemblies of arbitrary skeletons. Its primary value has been in providing insights and experience in the development of a systematic naming method based upon complexity minimization. This method also generates numeric rankings that can be compressed to yield reversible, compact codes.[6] Features of both IUPAC and nodal nomenclature systems are used as follows:

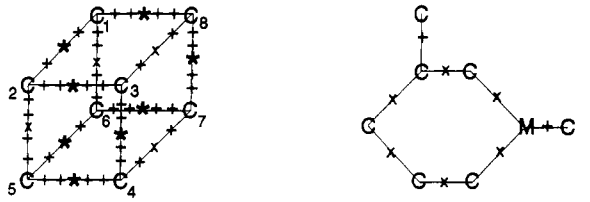
IUPAC: Chain, ring, and bridge names (methano, etc.); local numbering of complex side chains; bridges part of main-ring framework (thus more complex than side chains).

Nodal: Perimeter ring (largest ring) is the main structural unit of polycyclic skeletons (not a bicyclic skeleton); spiro system = ring + bridge; nested bridges (not cycloalkano—the bridge comes first); poly rings (terphenyl, etc.) are treated as assemblies of individual rings.

Both of the above methods minimize low locants; complexity naming minimizes high locants in reverse order (pairs for bridges—see Figure 2A). For rings, numbering begins at the most complex bridge (by size first, then length) or side chain if there are no bridges. The direction of numbering is by the shortest path to the other bridgehead (high locant) or, if equal, determined by the second most-complex substituting unit. When all bridges are equal, this method minimizes the length of the path that includes all bridgehead carbons, consistent with the numbering of, for example, a trimethyl benzene. (The high locant of the main bridge is not always paired with locant 1.) IUPAC Rule 32.31(c) "The main ring shall be divided as symmetrically as possible by the main bridge" is also implemented automatically for rings with two equal bridges. (This rule sometimes creates conflicts between symmetry and simplicity in main-ring tie situations, as shown below.)

IUPAC local numbering of complex side chains has been extended to bridges for consistency. Therefore, it becomes necessary to define one additional construction—the side bridge. This is usually a bridge between the main bridge and the main ring, with the phenalene skeleton being the most common example. To distinguish from a nested bridge, a prime (') is added to the main-ring locant. More generally, a prime indicates a connection between atoms that are two nesting levels apart. A difference of three requires two primes (") etc. Side-bridge structures are relatively rare, and casual attempts to construct examples often result in simplification by lengthening the intended main ring or main bridge, or else the main bridge becomes part of an alternate main ring, reducing the side bridge to a normal secondary bridge. Primed locants are considered senior (more complex). Therefore main rings are usually numbered from the main bridge so as to minimize a primed locant (see Appendix). In most cases, names would still be specific if the prime were omitted. However, with some effort, ambiguous cases can be constructed, as shown in Figure 1a and 1b. Figure 1c shows what happens to the side bridge if just one carbon atom is removed from the C_{26} ring. The IUPAC rules do not distinguish between secondary bridges that cross the main ring and those that cross a smaller ring (side bridges) except by numbering. [Side bridges are easily identified by one locant that is larger than the main ring, but this information is not used in the rules until Rule 32.31(d).] Thus in the main-ring tie situation in Figure 1c, Rule 32.31(c) (ring symmetry) would retain the main bridge of Figure 1b. It seems more logical that a structure with two 1-path bridges is simpler than one with a 3-path bridge, especially considering the difficulty of constructing examples of the latter that do not simplify. Analogously, 3-ethyl-2-methylpentane is simpler than 3-isopropylpentane—despite the loss of symmetry—and is the correct configuration by both Rule 2.6(a) and by minimum side-chain complexity. Nodal nomenclature also avoids the side bridge in similar situations by minimizing locants—including the main bridge pair. (IUPAC and nodal names are also given for Figure 1c.)

Input to the current program is the C-skeleton itself, using 'C' and any odd number of '+' characters in a straight line for bonds, including diagonal directions. Combinations of the eight bond directions can be employed in future versions to let the user specify stereochemistry. Crude 3-D constructions are sometimes possible (cubane in Figure 2). Output consists of the input skeleton overlaid by ring and chain paths by nesting level (main = *; 1–3 = x,y,z), followed by the complexity-ordered name. The ring count prefix (tricyclo, etc.) is omitted because it can be obtained from the name (unit main rings + bridges). Bridge locant pairs appear inside brackets



(A) (B)

Complexity: [2,5]-[1,6]-[4,7]-[3,8]cyclooctane 1-methyl-[1,1]-(2-methylpentano)

Nodal: pentacyclo[08.0^{1,4} 0^{2,7} 0^{3,6} 0^{5,8}]octane [1,1]-(4-methylpentano)methane

Figure 2. SKELNAME features: (A) 3-D input, (B) modified to find asymmetric C atoms as substituted methanes ('M' name given).

([1,4] etc.). Bis or tris repeat factors for complex side chains have been replaced by di, tri, etc.

RING SEARCH

In naming an alkane, exploring all paths from each chain end guarantees a correct (if not efficient) solution, including the correct numbering of the main chain. Equivalently, this solution is the least complex side chain of the set of side chains created by attaching each chain end one at a time to a much longer ("master") chain. In exploring rings, all ring carbons become start atoms analogous to chain ends. Downs et al.[7] have recently reviewed ring perception methods, and the method used by SKELNAME according to their terminology would be classified as a "walk around the connection table", except that the program explores the skeleton itself and uses labeling to store (and easily display for debugging) much of the state of the search. A reversible depth-first search marks each atom visited on a forward move and unmarks it during backtracking, thus restoring the skeleton to its initial state upon return to the start atom. The first time this happens, the path trace circles the ring in the opposite direction. Initially, all carbon atoms are in the starting list, resulting in a path trace from each ring carbon. Therefore, one of these paths will contain the correct ring numbering. The necessity for $2n$ ring traces of an unbridged n-carbon ring makes a renumbering routine imperative for large rings to avoid excessive run times. (The number of candidate main chains or rings for a given structural unit is not memory limited because all candidates are evaluated *during* the search, and therefore only the current best needs to be stored.) Following initial numbering of the first-found ring, each most-complex side chain or bridge locant is evaluated as C-1 to minimize the locant series in descending order. After renumbering a given ring, all of its atoms can be removed from the list. Both-way path tracing is retained for cyclic side chains (since C-1 is fixed) and also to provide a fallback position for special situations. For example, in the case of a substituted bicyclo[2.2.2]octane there are three candidate main rings, analogous to the three chains of a trialkylmethane skeleton. After two of these had been evaluated by the renumber-and-remove procedure, there would be no start atoms from which to explore the third left in the list.

The SKELNAME program is designed to find the largest ring or longest chain of the entire assembly, with rings preferred in the event of ties. This initial solution is printed before testing for a possible rerun, because it will correspond to the one obtained from nodal nomenclature in many cases. A structure specification page (Appendix) codes the linked units, with the parent unit on the top line. Thus the most efficient way to evaluate other candidate units would be to relink the page by editing rather than to continue searching the structural diagram. However, due to programming staff shortages at this installation (my PC), the easier route was taken. Since all evaluated main-chain or ring carbons (ties included) are removed from the starting list to reduce repeat searches, this FORTRAN program is simply rerun from near the top using the G-word. For the diphenylmethane skeleton, both rings are disqualified (as ties), and only one rerun is required. For phenyltolylmethane, the methyl side chain (if not explicitly removed) is also evaluated as a tie for longest chain during the rerun. This property can be utilized because the page gives an isomorphism check on any substructure. A minor modification that loops through the start-atom list to identify a skeleton as a set of substituted methanes can label asymmetric carbon atoms and meso combinations (see 1,3-dimethylcyclohexane example in Figure 2). Pseudocode for the rerun modification is shown in Figure 3. The naming of the 1,2-dichlorotetraphenylethane skeleton—described by Goebels et al.[2] as indicative of a general, nontrivial problem in search of a common-sense solution—is shown in Figure 4 as an example. Figure 5 shows three assemblies containing the tercyclopropyl skeleton.

Page editing is utilized for renumbering main rings and also for selecting the correct numbering direction of spiro paths. During the search of a spiro-fused ring for side chains, the

```
BEST-SPEC = <huge>; NTIES=0; NRUN=1
Find CTOTAL (of skeleton)
BIGSIDE = CTOTAL (largest side chain of current main path)
Do while BIGSIDE ≥ %CTOTAL and NTIES < 2
      Find the set of longest remaining paths (rings + chains)
      If set is empty then STOP (error?)
      Do type = rings, chains (rings first)
            Do while more rings (chains) of this type, size (new run)
                  (Encode, test complexity of assembly linked through new parent unit)
                  Complete TEST-SPEC page (includes new BIGSIDE value)
                  If TEST-SPEC < BEST-SPEC then BEST-SPEC = TEST-SPEC
                  If BIGSIDE = %CTOTAL then NTIES=NTIES+1
                  Disqualify this ring (chain)
            End(run)
      If NRUN=1 then print longest-path name
      NRUN=NRUN+1
      End(type)
End(BIGSIDE)
Print minimum-complexity name
DONE
```

Figure 3. Complexity minimization algorithm for assemblies.

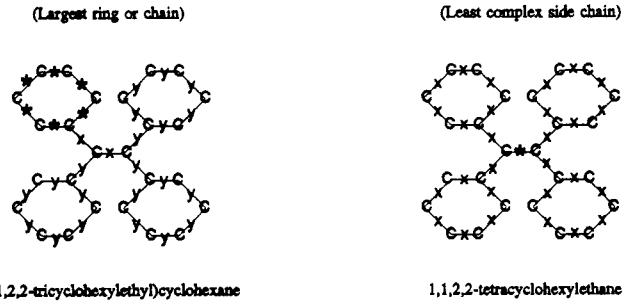
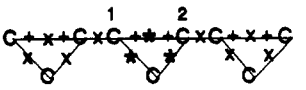
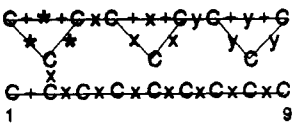


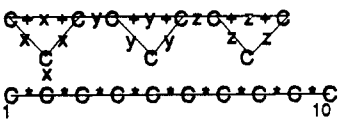
Figure 4. Initial and final solutions to naming an assembly.



(A) 1,2-dicyclopropylcyclopropane



(B) 2-(2-cyclopropylcyclopropyl)-1-(1-methyloctyl)cyclopropane



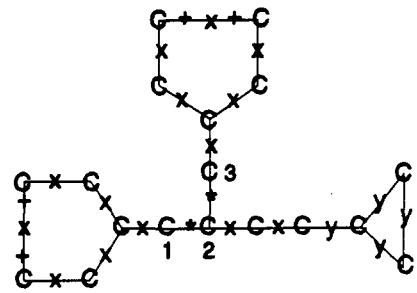
(C) 2-(2-(2-cyclopropylcyclopropyl)cyclopropyl)decane

Figure 5. Tercyclopropyl assemblies showing shift of parent unit (*).

Nodal: Tricyclo[06.3^{1,1}.4^{8,8}]tridecane
[1,1]-([2,2]butanopropano)cyclohexane

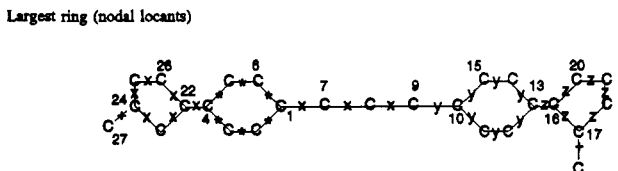
[3,3]butano-[1,1]pentanocyclobutane

Figure 6. Nodal name requires numbering shorter bridge first.

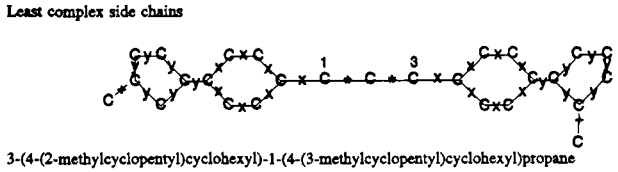


1,3-dicyclopentyl-2-(2-cyclopropylethyl)propane

Figure 7. C_3 chain of acyclic unit reduces largest side chain to C_5.



Nodal: Tetracyclo[06.1:7(3)9:10(06)13:16(05)17:21(1)4:22(05)24:27(1)]heptacosane
4-(3-methylcyclopentyl)-1-(3-(4-(2-methylcyclopentyl)cyclohexyl)propyl)cyclohexane



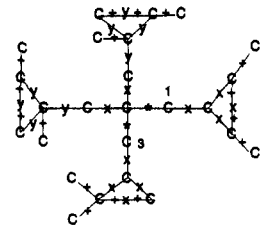
3-(4-(2-methylcyclopentyl)cyclohexyl)-1-(4-(3-methylcyclopentyl)cyclohexyl)propane

Figure 8. Main tie example[3] resolved by methyl locants.

program will find and store two consecutive "bridges" from the spiro atom (Figure 2B), differing of course only in the numbering of any side chains on the spiro path. Comparison of the two lines added to the page permits elimination of the more complex line and its parent code (See Appendix).

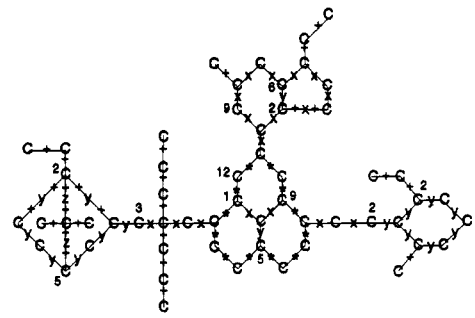
SPECIAL CASES

Figures 6–10 show special or more difficult cases. Figures 6, 8, and 10 are nodal nomenclature examples taken from ref 3. Figure 6 is a special nodal-naming case where the shorter [*n*,*n*] bridge of a dispiro skeleton must be numbered before a longer one. In minimum complexity naming, linear polyspiro skeletons evaluate similarly to the tercyclopropyl skeletons of Figure 5, with the main ring shifting toward the center as new rings are added. Figure 7 illustrates that the longest chain of any acyclic unit is not necessarily the chain that minimizes side-chain complexity. To use the early exit of the basic algorithm, it is necessary to first evaluate all chains of an acyclic unit whose longest chain has four or more atoms (Figure 10 has a similar case). Figure 8 has a main ring tie. Methyl locants break the tie in both nodal and complexity naming. The IUPAC name obtained by choosing the same disubstituted parent unit would be the same, except the main-chain numbering is reversed to give the lower locant series (1,4,2). Figure 9 has four candidate main rings (three different, one duplicate) and six candidate main chains. Ties can only be broken by comparing dimethyl locants. The decreasing complexity order is (3,2) > (2,2) > (2,1). Complexity is minimized by placing the most complex units at the highest levels. Thus the main ring of the largest-ring name (not shown) is 2,3-dimethylcyclopropane. The minimum complexity name is obtained by placing the two most complex side chains at the first level, i.e., connected directly to the main chain (R-, not RCH_2-), and giving locant 1 to the more complex of these. Figure 10 is a complex example whose nodal name is given on the last page of ref 3. The derivation of the name from the structure graph is thoroughly documented. The parent unit is evident at first glance. Likewise, the C_3 chain connected to locant 2 of the parent unit would appear to be the main chain of an acyclic unit. In fact this is the case, because the C_5 chain created by removal of a propyl side chain would add one atom to the attached C_{11} cyclic unit, increasing complexity as much as is possible with one atom by pushing this unit to a lower level.



3-(2,2-dimethylcyclopropyl)-1-(2,3-dimethylcyclopropyl)-2,2-di(1,2-dimethylcyclopropyl)methylpropane

Figure 9. Six-way main chain tie resolved by dimethyl locants.



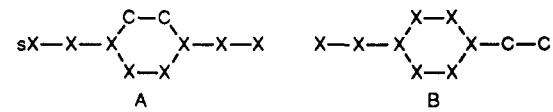
8-(2-(6-methyl-2-ethylcyclohexyl)ethyl)-11-(8-methyl-5-ethyl-[2,6]cyclononyl)-2-(2,2-dipropyl-3-(2-ethyl-[2,5]-1,1-dimethylmethanocyclohexyl)propyl)-[1,9]-([1,5']-methano)cyclododecane

Figure 10. Complex example with documented nodal name derivation.[3]

COMPUTER LINKING OF UNITS

Computer procedures for linking structural units of assemblies are utilized extensively in organic synthesis and generic structure handling software but are seldom described in detail. This section describes a simple computer method for linking units, including ring–ring and spiro–ring "chains". While most likely not novel, the method is easily illustrated, as shown below.

In attempting to extend the chain search routine to include rings, it soon became evident that it would be necessary to store several candidate paths because chain paths can include part of a ring, and if found first must be disqualified, e.g., 1,4-diethylcyclohexane:



The search for bonds to carbon atoms ('C') proceeds in a clockwise direction beginning with a look to the right. Thus, a search starting at the left ethyl chain end (s) would mark the X-path shown in A and find an 8-carbon main chain. The next forward search after backtracking (B) will hit an 'X' to

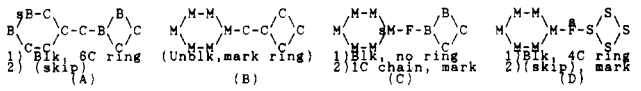


Figure 11. Linking structural units of a ring–chain assembly.

close the ring. To disqualify the octane chain, it would be necessary to find a match between chain and ring carbons, since the naming method does not require smaller rings to be senior to chains. Because large structures could store numerous false paths, the single search-routine approach was abandoned in favor of two routines—a new one for rings (RINGER) and the original one for chains (CHAIN).

Chain ends are easily detected by counting bonds at each C visited. However, this test fails in assemblies where chain ends can be capped by rings (1,6-dicyclopropylhexane) and also duplicates effort because the depth-first search eventually explores (and could count) all paths from each C anyway. A faster method is to test whether the last move was forward whenever all forward moves are blocked. Fortunately this method was chosen, because it led to a simple solution for linking assemblies: CHAIN is retained as a "dumb" routine that can be fooled into reporting a chain end. RINGER is called first to count carbons and find rings (+ bridges). When a ring is found, the ring-closure carbons are blocked from visits. This also prevents the reverse-direction ring from being explored and the carbons from being counted twice. A ring is stored only when the start atom is on the ring. In such case the call to CHAIN is skipped; otherwise the blocks allow CHAIN to isolate a connecting-chain segment by blocking entrance to any rings. After each set of searches the main path is marked, the blocks are removed, and the sequence is repeated.

Figure 11 shows the action of RINGER (1) and CHAIN (2) in linking units ('s' marks the start carbon). Panel A shows blocks of the ring-closure atoms ('B') required to count the 11 carbons in the structure. The call to (2) is skipped because a 6-C ring was found. The stored main ring is marked 'M' in panel B. In C(1), the search for cyclic side chains from s finds and reblocks the 4-C ring. A 5-C side chain is reported; however, no ring is stored because locant 1 of the side chain is not on the ring. Therefore, (2) is called and finds a 1-C chain ('F') that becomes the current main chain because the forward move is blocked. In D, the 4-C ring is found and marked ('F' and 'S' denote first and second-level side chains).

CONCLUSION

A relatively simple method for selecting the parent unit of a ring–chain assembly has been presented in the form of an algorithm that can serve to replace imprecise guidelines and complex rules. With either visual or computer naming, the majority of known assemblies, such as those that occur in natural products, can be named by finding the unit whose largest side chain has fewer than half the total number of atoms. It is anticipated that extending the complexity minimization technique to include atom and bond types will be straightforward, leading to a natural seniority sequence such as size, shape, what (atom), how (bonded), and where (locant).

APPENDIX: EXTENDED SPECIFICATION PAGE FOR BRIDGED RINGS

The specification page (see Chart I) is translated by reading from right to left in a depth-first fashion to give the complexity-ordered name, in close agreement with the former IUPAC rules[8] that gave this method as an alternative to alphabetical ordering. The page format was designed for 16-bit computers, emphasizing readability to facilitate debugging. The results of structure searches are coded as follows to compare configurations for relative complexity:

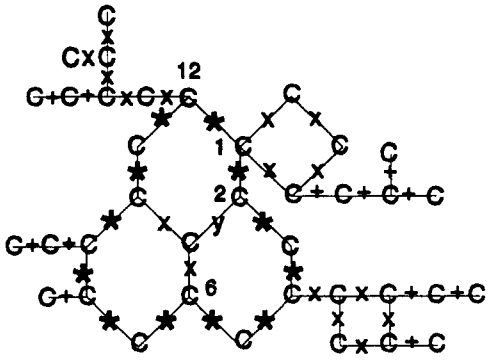
Simple Side Chains: 3 digits, SLL; S = simple side-chain code 1..8 (Me, Et, Pr, i-Pr, Bu, s-Bu, i-Bu, t-Bu), LL = locant.

Complex Side Chains: 5 digits, 1NNLL; NN = line where structure is further simplified, LL = locant.

Bridges: 5 digits, 2NNBB; NN as above, BB = the number of bridge locants stored as their actual value on line NN (always 02 in current version, but adaptable to a naming

Chart I. Specification Page

line	C(total)	C(side)	side chain codes					
1	37	2501	20802	20202	10604	11012	209	108
2	1	0	20302	10	6	0	0	0
3	0	0	34	1	0	0	0	0
**	1	0	20502	5	1	0	0	0
**	0	0	42	1	0	0	0	0
6	7	301	202	103	0	0	0	0
**	7	301	204	103	0	0	0	0
8	7	400	701	1	1	0	0	0
**	7	400	703	6	6	0	0	0
10	7	300	202	103	0	0	0	0
**	7	300	202	103	0	0	0	0
**	7	300	402	0	0	0	0	0



8-methyl-9-ethyl-12-(3-methyl-2-ethylbutyl)-4-(3-methyl-2-ethylcyclobutyl)-[6,10]-([1,2']-methano)-[1,1]-(1-isobutylpropano)cyclododecane

method that permits a complex bridge framework to be anchored at more than two sites). Side-bridge locants (primed) are augmented by 32 (max ring size) to indicate the prime.

A phenalene skeleton, substituted by a spiro bridge, complex, and simple side-chains is coded and displayed below as an example that contains all structural types. The main path length is stored as its complement—the number of side-chain carbons. Thus Code 2501 indicates 25 side-chain carbons on C_{12} main ring. Codes in this column that end in 00 denote acyclic side chains (less complex than cyclic). Code 20802 specifies the most complex side chain (where numbering begins)—the spiro bridge detailed on line 8. Code 20202 points to the [6,10] methano bridge on line 2 that points to the side bridge ('y') on line 3. Code 34 here (32+2) translates to the 2' locant. Code 10604 is for the cyclic side chain (simplified on line 6) at locant 4, etc. Lines with asterisks (**) show more complex or identical configurations (alternate main chains or numberings) that were rejected. Notice the reverse-numbered cyclobutyl group in line 7 and the isopropyl code in column 3 of the last line—indicating a main-chain tie in the substituted butyl side chain. In addition to the reverse-numbered isobutyl group (removed by page editing), line 9 shows the [6,6] initial numbering of the spiro junction, indicating that the winning search began at C-6. The C-atom positions are read into the starting list line by line. Thus, the first-found ring began at C-12. The program retained C-6 in the starting list as a special case when the others were removed.

REFERENCES AND NOTES

(1) IUPAC. *Nomenclature of Organic Compounds*; Pergamon Press: New York, 1979; p 46.
(2) Goebels, L.; Lawson, A. J.; Wisniewski, J. L. AUTONOM: System For Computer Translation of Structural Diagrams into IUPAC-Compatible Names. 2. Nomenclature of Chains and Rings. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 216-225.
(3) Lozac'h, N.; Goodson, A. L.; Powell, W. H. Nodal Nomenclature—General Principles. *Angew. Chem., Int. Ed. Engl.* **1979**, *18*, 887-899.
(4) Davidson, S. An Improved IUPAC-Based Method for Identifying Alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 151-155.
(5) Davidson, S. Computer Program for Identifying Alkane Structures. *J. Chem. Educ.* **1973**, *50*, 707.
(6) Davidson, S. Compact Numeric Alkane Codes Derived from IUPAC Nomenclature. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 417-421.
(7) Downs, G. M.; Gillet, V. J.; Holliday, J. D.; Lynch, M. F. Review of Ring Perception Algorithms for Chemical Graphs. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 172-187.
(8) IUPAC. *Nomenclature of Organic Compounds*; Butterworths: London, 1958; pp 8-9.

Thermodynamic Implications of Substituent and Solvent Effects on Reactivity

WOLFGANG LINERT

Institute of Inorganic Chemistry, The Technical University of Vienna, Getreidemarkt 9, A-1060 Vienna, Austria

Received November 25, 1991

The theory of the isokinetic relationship and its connection to generalized linear free-energy relationships are outlined, and its practical implications are discussed on the basis of experimentally observed substituent and solvent effects. It is further shown in what way the isokinetic relationship may be applied for detecting changes in reaction pathways within a given series of reaction.

1. INTRODUCTION

Information concerning substituent and solvent effects on chemical reactivity can be correlated in a variety of ways, for example, by means of the Hammett relationship,[1] Brønsted relationships,[2] or correlations of equilibrium and rate constants with solvent parameters such as donor or acceptor numbers,[3] or Reichardt's E_T values.[4] Such relationships fall into the large class of linear free-energy relationships (LFERs) which is, in a generalized form, the subject of this present paper. Broadly speaking, linear free-energy relationships or LFERs are produced when rate constants or equilibrium constants (or related functions describing kinetic or thermodynamic chemical behavior) *of a series of reactions* are plotted versus a characteristic quantity measured by means of another reaction series.[5-7] Whenever a LFER is found, some quantity must be present that appears in *the same functional form* in both the "test reaction series" and the "reference reaction series". In what follows this parameter will be given the symbol ξ. As long as only one parameter appears, it may always be expressed in the form of a *linear* relationship. Furthermore, since thermodynamic functions will be concerned to some extent, the abbreviation x for $1/RT$ will be used, partly to draw attention to the similar meaning of the two parameters. Wherever it is not necessary to distinguish between rate constants and equilibrium constants, both will be referred to as K.

Isokinetic behavior is commonly found[7] for reaction series in which only one reaction mechanism is followed, and, in fact, this is a necessary condition for the occurrence of an LFER described by not more than one parameter ξ. This produces in turn a common point of intersection of the Arrhenius lines or of the van't Hoff lines ($\ln K$ versus x). This is called the isokinetic relationship (IKR). This effect has also been often referred to as the "compensation effect". However, the latter usually means a proportionality between enthalpic and entropic contributions to K which may occur as an artifact.[7,8]

Such common points of intersection are also found when the same set of data is depicted for different temperatures in LFER plots. Whereas in the case of the IKR the characteristic intersection abscissa has the dimension of a reciprocal temperature ($x_{iso} = 1/RT_{iso}$, where T_{iso} is referred to as the isokinetic temperature), in the latter case a characteristic LFER parameter (ξ_{iso}) is found. In the following, some general properties of these relationships will be explored and then applied to some actual experimental reaction series (involving both substituent and solvent effects) in order to illustrate the practical value of these concepts.

2. THEORY

2.1. Interrelationships between LFERs and IKRs. There are several possibilities open to us for defining the IKR, both in integrated as well as in differential form.[7] For the present

0095-2338/92/1632-0221$03.00/0