COMPOUND/SELECTIONS	SM-1-	SM-2-	RXN-1-	RXN-2-
COST LIMIT 50$(1) 60$(2)	3	13	2	15
SM NOT REFUNCTIONALIZED	3	13	2	15
SM REFUNCTIONALIZED	0	0	0	0
INTERMEDIATES	4	0	4	0

REACTION/SELECTIONS	RXN-1-	RXN-2-	
UNCLEAR REGIO:KETONE	0	0	
UNCLEAR REGIO:PI-BOND	0	0	MOVE CURSOR TO
MIXED MODE CONSTRUCTION	0	0	LEVEL -1-
TWO-STEP ANNEL: ACID/BASE	4	3	LEVEL -2-
TWO-STEP ANNEL: RED.CYC.	0	5	DESC. FOR 1&2
BALDWIN RULES VIOLATION	0	0	ENTER
CHEM. EQUIVALENT REACTIONS	2	3	R FOR RETAIN
CHEM. EQUIV.: REVERSE-POLARITY	0	0	D FOR DELETE
CHEM. EQUIV.: REVERSE-ORDER	0	0	C FOR CLEAR
COMPATIBILITY: LEAVING GROUP	0	9	Clear All Flags
COMPATIBILITY: KETONE	4	12	HELP
COMPATIBILITY: ALDEHYDE	0	0	EXIT

Figure 8. Indirect selections: sample choices.

rapid, crude drawing of the target structure on a Tektronix mode graphics terminal and proceeds without operator input, as described above, to generate all possible sequential constructions of convergent bond sets of two levels only from real starting materials in its catalog. The program requires about 2 min for a steroid of the complexity in Figure 6. A number of sequential targets may be input with each going to batch mode for its output generation. These are stored when finished in a directory for operator examination later using the delete/retain modes for further selection from large outputs as described above. The reactions produced are generally sensible and constitute all possible combinations within the constraints applied. The results seen to date seem to mirror sensibly the expectations of those constraints. Further expansion of the program is currently under development.

ACKNOWLEDGMENT

We are grateful for support provided by the Eastman Kodak Co. and by a grant (CHE-86-20066) from the National Science Foundation.

REFERENCES

(1) Corey, E. J.; Long, A. K.; Rubinstein, S. D. *Science* **1985**, *228*, 408, and earlier references cited therein.
(2) Wipke, W. T.; Braun, H.; Smith, G.; Choplin, F.; Sieber, W. *ACS Symp. Ser.* **1977**, *61*.
(3) Bersohn, M. *Bull. Chem. Soc. Jpn.* **1972**, *45*, 1897. Bersohn, M.; Esack, M.; Luchini, J. *Comput. Chem.* **1978**, *2*, 105. Bersohn, M.; MacKay, K. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 137.
(4) Gelernter, H.; Sridharan, N. S.; Hart, H. J.; Yen, S. C.; Fowler, F. W.; Shue, H. J. *Top. Curr. Chem.* **1973**, *41*, 113. Gelernter, H.; Sanders, A. F.; Larsen, D. L.; Agarwal, K. K.; Bovie, R. H.; Spritzer, G. A.; Searlemen, J. E. *Science* **1977**, *197*, 1041. Agarwal, K. K.; Larsen, D. L.; Gelernter, H. *J. Comput. Chem.* **1978**, *2*, 75.
(5) Marsili, M.; Gasteiger, J.; Carter, R. E. *Chim. Oggi* **1984**, *9*, 11. Gasteiger, J. In *Computer in der Chemie*; Zeigler, E., Ed.; Springer: Berlin, 1984; p 207.
(6) Hendrickson, J. B. *Acc. Chem. Res.* **1986**, *19*, 274.
(7) Hendrickson, J. B. *J. Am. Chem. Soc.* **1977**, *99*, 5439.
(8) Hendrickson, J. B.; Grier, D. L.; Toczko, A. G. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 195.
(9) Hendrickson, J. B.; Toczko, A. G. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 171.
(10) Hendrickson, J. B.; Braun-Keller, E. *J. Comput. Chem.* **1980**, *1*, 323.

Multiple Constructions in Synthesis Design

JAMES B. HENDRICKSON* and PING HUANG

Department of Chemistry, Brandeis University, Waltham, Massachusetts 02254-9110

Received February 7, 1989

Multiple construction reactions that may take place in one laboratory operation are argued to be important keys to rapid creation of the target skeleton in synthesis design. The several paths for such constructions, forming two to four skeletal bonds, are logically formulated. Two of the major classes, double affixation and multiple cyclization, are then articulated as computer programs to find all skeletal variants for a given target and then to apply appropriate reactive functionality to the generated starting skeletons to produce viable routes to the target.

In the development of short, efficient synthesis routes to complex target molecules we have argued that the construction of the target skeleton should be the key consideration.[1] This arises from the basic observation that syntheses proceed from small starting molecules (average of only three carbons incorporated into the target) to form large target molecules. Hence, we concluded that only those reactions ("constructions") that link the small starting skeletons into the target skeleton are obligatory for synthesis. Therefore, the shortest syntheses will consist of only construction reactions, or at least will sharply minimize the refunctionalization reactions (those which alter functionality without changing skeletal bonds). To this end we developed the SYNGEN program to generate syntheses consisting of sequential construction reactions only, presumably the shortest routes.[1]

The ideal of the shortest synthesis then becomes one of minimizing refunctionalization reactions, but should also consider minimizing the number of construction steps as well. An average published synthesis constructs about one-third to one-fourth of the target skeletal bonds, i.e., for targets of 10–30 carbons, making about 3–9 skeletal bonds. In the SYNGEN program minimization of the number of constructions arises by dissecting the target skeleton into convergent assemblies of starting skeletons. These are found by cutting the target skeleton into two parts first and then, at the second level, cutting each of those two into two smaller skeletons, finally accepting only those ordered bondsets so derived that result in all four starting skeletons found in a catalog of available starting materials.[2] This convergent dissection procedure allows no more than two bonds cut to divide a skeleton into two pieces and so will result in no more than six bonds (and usually less) cut from the target, i.e., those bonds to construct in the synthesis.

Another way to minimize the number of construction steps is to employ reactions that make more than one construction per step, i.e., multiple construction steps, and this approach

0095-2338/89/1629-0145$01.50/0

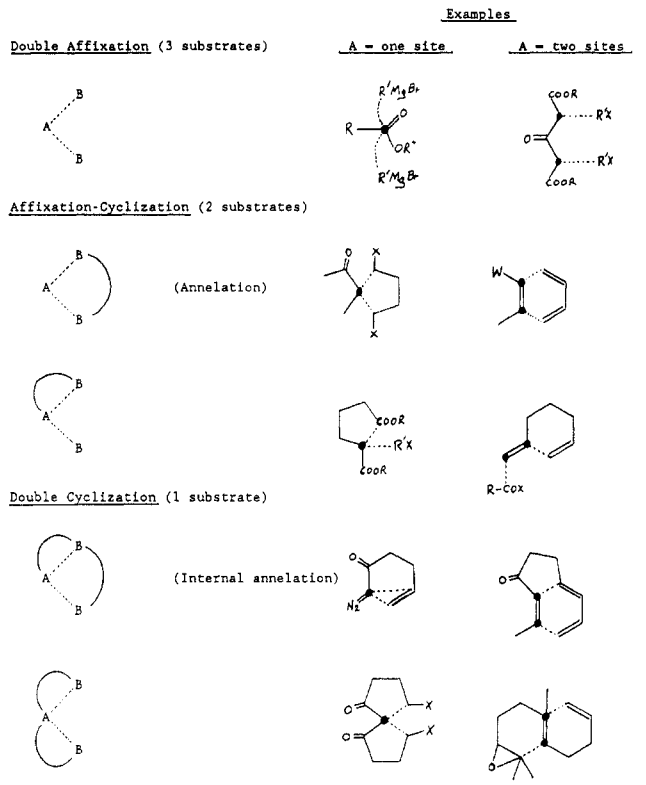


Figure 1. Modes of double construction.

is considered here. It is prefigured in classical chemistry by reactions such as Grignard reactions with esters that add two alkyls at once, forming two skeletal bonds in one laboratory step, and cycloadditions, which also form two bonds at once. It is convenient to distinguish intermolecular constructions as *affixations* (of two pieces) from intramolecular constructions as *cyclizations* (of one piece). Thus, the simplest multiple construction will be a double construction, which can be any of the three possible combinations of affixations and cyclizations, summarized in Figure 1. These three combinations are also distinguished by the number of substrates joined together in the double-construction operations.

The simplest case is the double affixation in which two substrate units, B, are joined to a third, A, by two constructions. For this to occur in one step requires that both units B are the same and use the same construction site on each. This operation then joins three units, i.e., three substrates. The unit A may undergo its two constructions at the same site, as shown in the left example, or at two different (but similarly reactive) sites, as in the right example. The construction sites on unit A are indicated with a heavy dot.

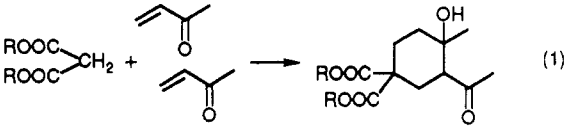
We can systematically generate all the other possible double constructions by formally creating prior links between pairs of units all ways. With one such link, joining either A and B or B and B, the double construction now joins only two molecules and cyclizes spontaneously, i.e., an affixation–cyclization. The latter case is the familiar annelation mode, affixation followed by cyclization, shown again with both constructions to one carbon site on the A unit or with the two constructions to different sites. The A–B-linked case is less common, as the two examples imply, i.e., cyclization usually (but not necessarily) is the first construction, followed by affixation.

If we designate two prior links between the three units, the result is a double cyclization on one substrate, and this can also be conceived formally in two ways: either linking unit A to each unit B or linking A–B and B–B. In the former case the construction bonds are in two rings; in the latter they are both in one ring, i.e., an internal annelation. As before, there are examples with both constructions to the same site on unit A or to two different sites.

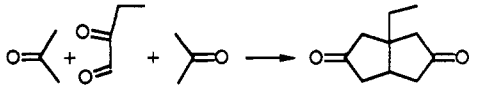
More extensive multiple constructions may now follow by adding on to the modes in Figure 1 successive further constructions. In practical terms these are likely only to be cyclizations since further affixations in succession are unlikely to stop and so are suitable only for polymerizations.

The several formal modes of Figure 1 can be systematically searched on a given target. As in the SYNGEN program this is most efficiently done by simplifying the problem to dissect the target skeleton first and then in skeletally acceptable cases to apply the needed functionality. The SYNGEN program operates by dissecting the target skeleton first all ways into two pieces, thus retrosynthetically cutting either one acyclic bond or two bonds in one ring. The latter instance reveals all the annelation modes of Figure 1, but none of the others, which are the subject of the new programs described below.

Double Affixations. The constraints here are that the target skeleton must be dissected into three pieces such that two are identical, each attached to the third and attached from the same site on each one. For an acyclic molecule this involves cutting only two bonds, the two created in the forward direction by the double construction. In cyclic molecules we must cut more bonds to result in three starting pieces. In the forward direction this corresponds to a double affixation followed by one or two cyclizations. In the example of eq 1[3] the retro-



(1)



(2)

synthetic dissection to produce three pieces (with two alike) cuts three bonds. In the very efficient Weiss reaction,[4] eq 2, four bond cuts are required. The order of the constructions in the forward direction is always double affixation first, followed by the cyclization(s). We limit the number of cuts to four, as in eq 2.

The dissection of the target skeleton proceeds by cutting all possible combinations of two to four bonds such that the target skeleton is broken into three pieces. Each time the dissection produces three pieces, the bondset so generated is tested with three successive requirements: (1) two of the pieces must be skeletally identical; (2) they must each have been linked to the third piece by a cut bond; and (3) those cut bonds must have been attached at the same skeletal carbon on each of the identical pieces. In this way we derive a group of bondsets defining two identical pieces, i.e., the "doubling pieces" (B), each to be affixed to a third piece, the "common piece" (A). For synthetic efficiency this double affixation has special value only if the doubling piece is large enough so that a major increase in skeleton results from linking two of them. Accordingly, we placed a further constraint that the doubling piece must have at least three skeletal carbons. The bondsets so validated (of two to four bonds) are then ordered so that, in the forward direction, the two affixation bonds are made first and the cyclization(s) last. If there are two cyclizations, both orders are set for them.

With these ordered bond sets the program now proceeds as in SYNGEN, working stepwise backward, from the target through the bond-set bonds in order, to generate the functionality. Thus, first all valid construction reactions are retrosynthetically generated from the target functional groups for the last bond of the bondset. Then the intermediate functionality so generated is used in turn to generate valid constructions for the next bond of the bondset and so forth

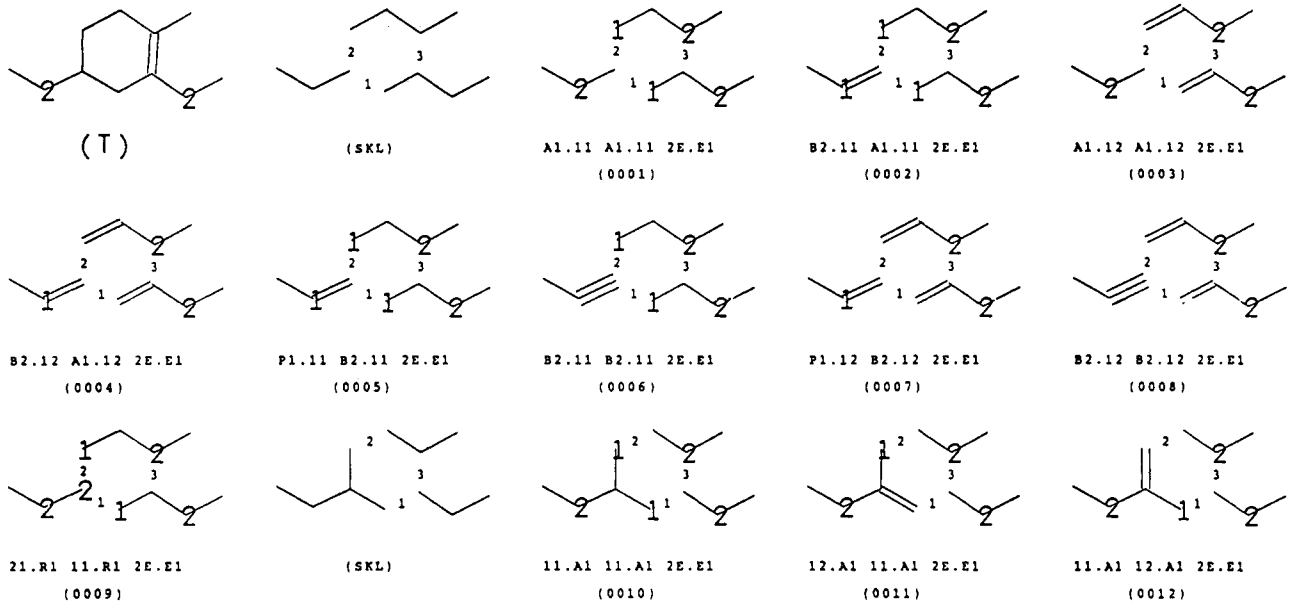


Figure 2. Double affixation output.

until the bondset (two to four bonds) is complete, generating finally the functionality on the three starting pieces. As in SYNGEN, the construction reactions are generated from all valid combinations of an electrophilic and a nucleophilic half-reaction and subsequently validated also by rough mechanism tests.[1]

In the double-affixation program further constraints on the generated construction sequences are also applied. First, the two doubling pieces were defined as identical, and so the program rejects any sequences generated in which the two doubling pieces are not exactly the same in functionality as well as skeleton. The program also requires that both doubling pieces react by the same kind of half-reaction, i.e., electrophile or nucleophile. The other constraints involve the generation of carbanion nucleophiles by prior reduction (as in Grignard half-reaction preparation, RX + Mg → RMgX). On the grounds that such reduction cannot be selectively available during the sequence in a one-pot operation, these half-reactions are disallowed for the cyclization step(s) and allowed for the two affixations only if applied to the doubling pieces, i.e., disallowing only one reductive half-reaction or any successive pair of reductions on the common piece. The implication is that the reductive carbanion preparation may occur on the doubling piece before the common piece is added, but not on the common piece, which must require a dianion or successive reductions and constructions.

This procedure then locates more than just simple double affixations, but also finds multiple constructions of up to four constructed bonds in one step by locating also those cases in which cyclizations occur subsequent to the initial double affixation. This is shown by the output for a simple cyclic case in Figure 2, closely analogous to the example[3] in eq 1. The output is recorded with numbers for the functional groups, i.e., z values, the number of heteroatom bonds to the numbered carbon: thus, 1 is a singly attached N, O, or X and 2 is a carbonyl (Figure 1, ref 1). The three constructions are shown below each case, in order, as pairs of half-reactions identified in Figure 3 of ref 1.

The multiple constructions of Figure 2 occur from two bondsets, labeled (SKL), one for the first 9 entries, the other for 12 more, only 3 of which are shown. Basically the reactions represent sets of simple variants on combinations of modified reactants. Thus, examples 1, 2, 5, and 6 [shown as (0001), (0002), etc. in Figure 2] imply double displacement of halides on the doubling pieces by ketone (1), enol/enamine type (2, 5), or electrophilic substitution twice on an acetylene (6).

From the half-reaction descriptions (Figure 3 in the preceding paper[1]) example 1 operates from the anion of the simple ketone at left, i.e., half-reaction A1 twice. In example 2 an enol ether or enamine equivalent is the nucleophile, reacting via B2 in the first construction, then as the resultant ketone (anion), i.e., A1, in the second. Example 5 implies both constructions occur as B2 from enol ether/enamine, leaving the product ketone only after the second construction. The acetylene example (6) represents two B2 half-reactions in succession, the first going to an enol ether intermediate and then on to the final ketone. Then there are the parallel four (examples 3, 4, 7, 8) that employ methyl vinyl ketone as the electrophile. All cyclize last with an aldol condensation (2E·E1). The last entry (9) entails double addition of an organometallic derivative of 4-halo-2-butanone to pyruvaldehyde. The second bondset has more entries because the symmetrical identity of some entries is not perceived, cf. examples 11 and 12. The analogy to eq 1 is found as entry 3.

The Weiss reaction (eq 2) involves cyclizing two rings and so there are more generated combinations, sampled in Figure 3. There are eight possible orders for the bondset shown (SKL) consistent with double affixation, i.e., the first two bonds formed successively join the two doubling pieces to the common piece. Hence, there are a number of variations in reaction sequence from the same set of starting materials. In fact, the program produces 137 sequences from only 18 combinations of starting materials.

Samples are shown in Figure 3, selected from three most likely bondset orders. The first two show an aldehyde common piece alkylated and cyclized, while the next (example 8) is an enamine/enol ether variant of the common piece. A presumed sequence for the Weiss reaction is example 83, then a reductive (Reformatsky) variant in example 84 and one enamine variant in example 85. One of the methyl ketone versions is illustrated in example 97, while Wittig and Grignard initiation are seen in examples 31 and 35, respectively. The other two selected utilize an acetylene aldehyde, cyclizing twice onto the triple bond.

The value of this analysis in synthesis design becomes clear when we examine some target molecules for double affixations. We examine both the target itself and a set of intermediates formed, as in SYNGEN, by cutting off portions of the target. Taking estrone as an example, we applied the SYNGEN program to cut it into two skeletal pieces and develop the chemistry to

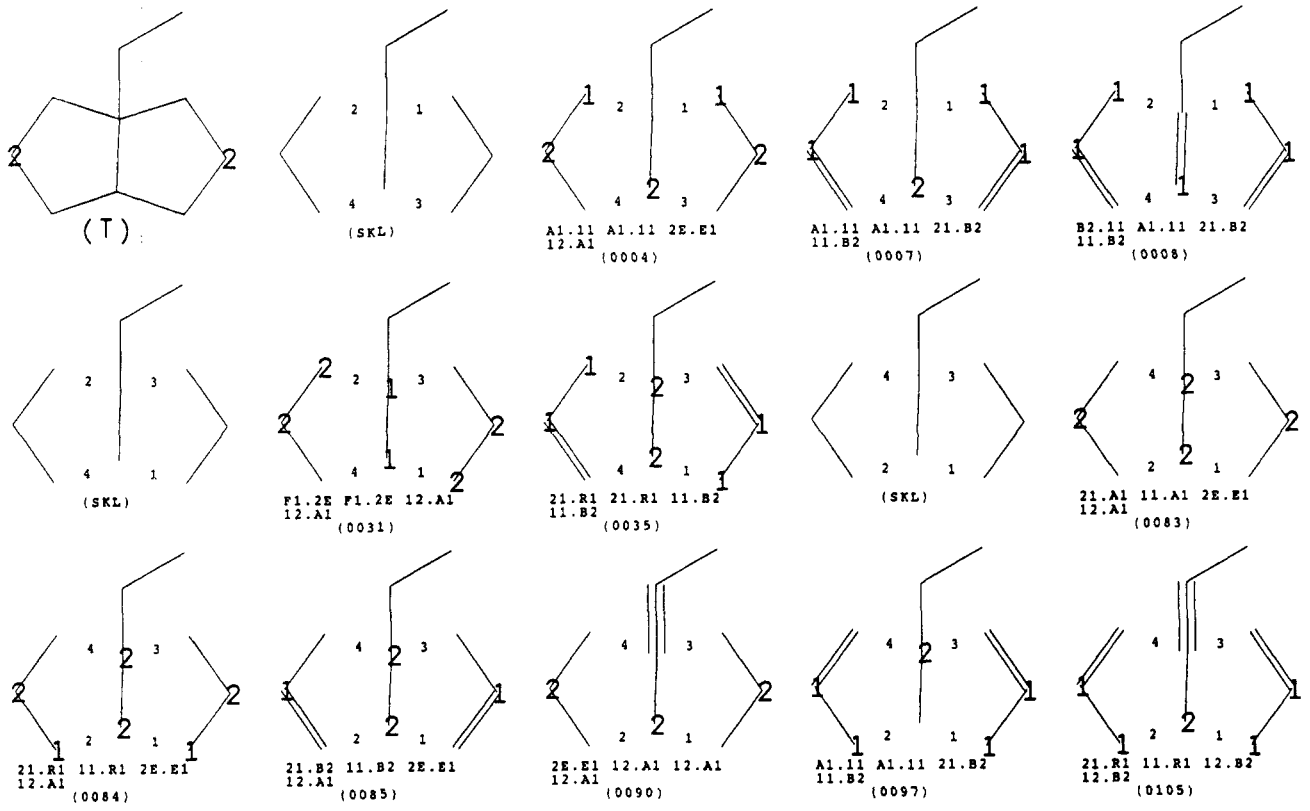


Figure 3. Weiss reaction: selected output.

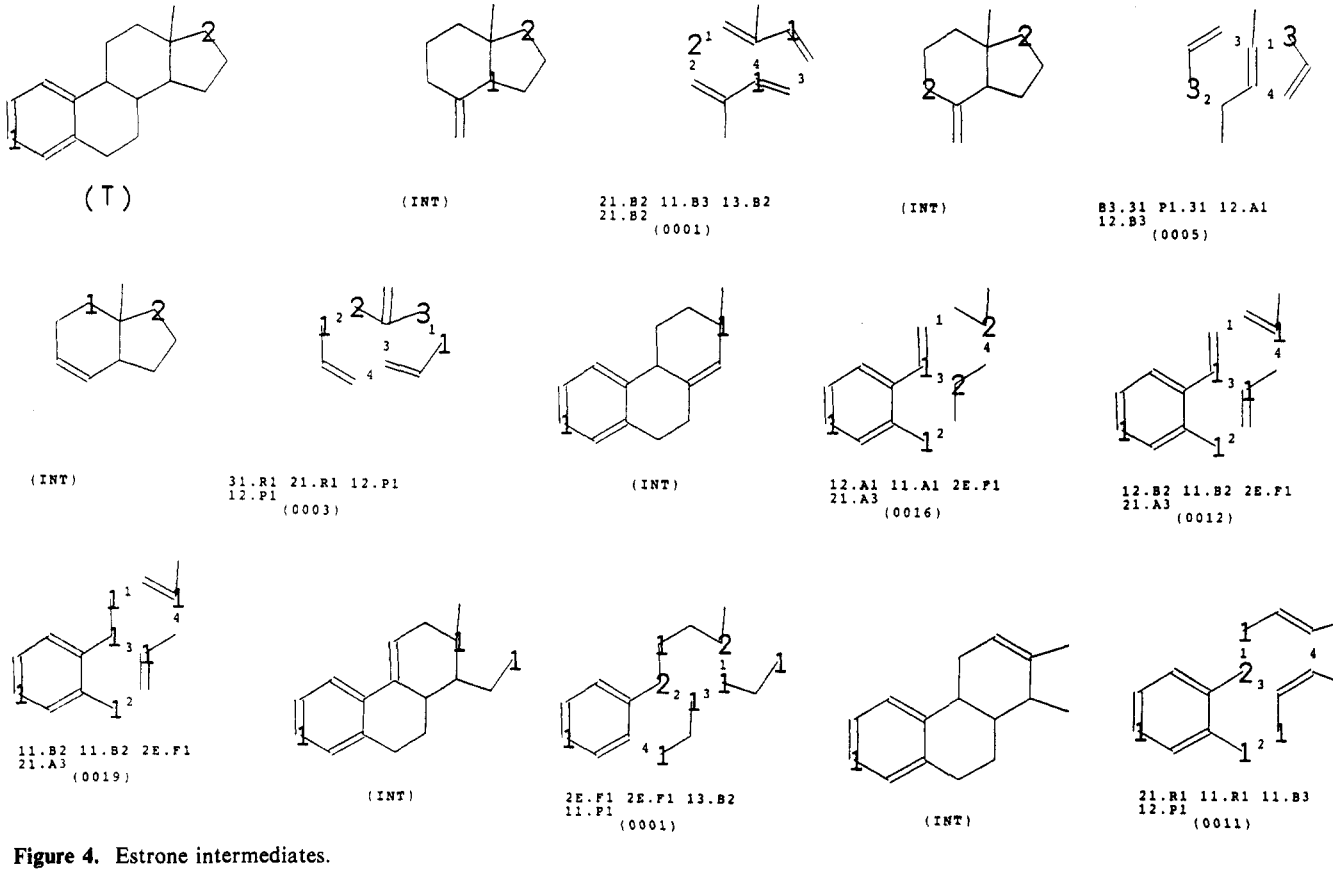


Figure 4. Estrone intermediates.

define these as functionalized intermediates. Each of these was then scrutinized by the double-affixation program. Analysis of estrone afforded no double affixations, but four of its SYNGEN-generated intermediates yielded six double-affixation sequences, generally of little chemical credibility. Estrone itself is usually synthesized via derivatives bearing extra functionality,[5] and so we examined several dehydroestrones with double bonds in ring C (four ways). These yielded a total of 74 sequences to 28 intermediates. Chemically interesting selections from these sequences are shown in Figure 4, with estrone itself shown as T and the intermediates labeled INT.

In a similar vein the hydrindans with potential steroid functionality have been frequent synthetic targets.[5] The sim-

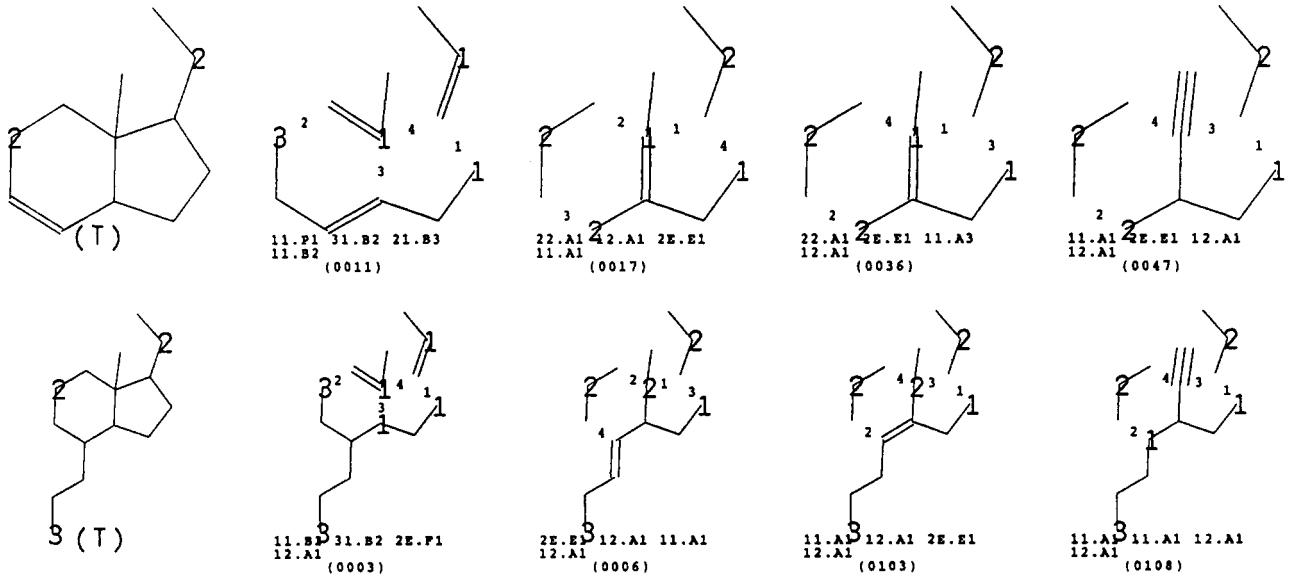
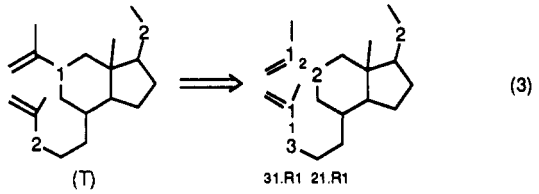


Figure 5. Hydrindan intermediates.

pler target at the top in Figure 5 showed 50 sequences in only two basic bond sets; typically the computer develops all possible orders (here there are eight) in a number of chemically equivalent modes, as exemplified by the variations in Figure 2. As a result, the number of real chemical ideas developed is actually rather small, well represented by the samples shown in Figure 5. Examples 17 and 36 differ only in the order in which the four bonds are formed, resulting in two different sequences of the four half-reaction pairs shown below. The lower target in Figure 5 is the intermediate for a recent steroid synthesis by Stork,[7] which was distinguished in the next step by one of the few double affixations in a published synthesis, i.e., eq 3, which is also reproduced by the program.



(3)

Modes for synthesis of Stork's intermediate (here, target) are shown in the lower half of Figure 5. There were 145 sequences generated in two bond sets, one of which developed all eight bond orders in 141 of the sequences, but as before the samples shown are representative of the whole set. The chemistry implied in these entries is quite complex and, while the reactions as written may not be viable as four-in-sequence, still the implicit ideas for rapid assembly should stimulate new directions in considering fast syntheses of these key intermediates.

Multiple Cyclizations. The double affixations, although they proceed in one laboratory "step", commonly consist of a sequence of mechanistic steps, such as deprotonation preceding each construction (cf. Figure 2). If we consider how several constructions may proceed in one formally concerted mechanistic step, we can most simply ask what unit reaction[8] can accommodate several constructions. A linear string of unit reactions at each carbon necessarily terminates in ±H or Z changes at the two terminal carbons. Hence, the intermediate carbons involve only ±R or Π; cyclic unit reactions[8] involve only ±R or Π. This is illustrated in the linear and cyclic strings in Figure 6, which show as well the conventional part-structure mechanisms. These are exactly reflected by the unit reaction letter strings, as shown. The linear form is simply a vinylog of the standard construction (ref 1, Figure 2) with one or two interpolated π-bonds between the standard nucleophilic and

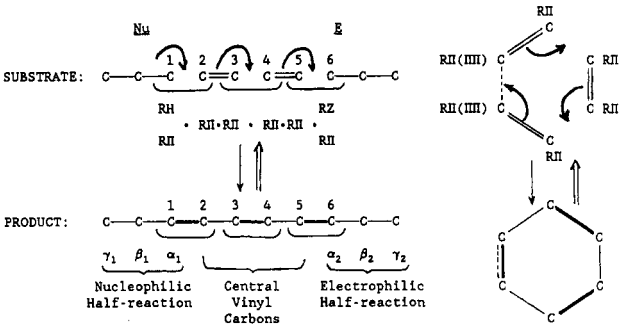


Figure 6. Single-step multiple constructions for cyclizations.

electrophilic half-reactions, constructing two or three new C–C bonds, respectively. The standard half-reactions then will be those of Figure 3 in ref 1.

If the pieces in Figure 6 are all linked, the reactions become multiple cyclizations and serve as the model for another program, discussed here, which aims to discern all such opportunities for multiple cyclization to a polycyclic product, taken here as the target molecule. Except for some organometallics, most of the examples in Posner's excellent review of multiple cyclizations[9] fit the mechanistic pattern of Figure 6.

As before, the skeleton is examined first, and then functionality appropriate to the cyclization is derived afterward to fit both the bondset bonds and the final product functional groups, i.e., those of the target. In some cases a final refunctionalization is required to alter the functional groups of the immediate product of multiple cyclization to create the actual given target. The bondset bonds dissected are always alternating bonds on a string of carbons such that each dissected bond cuts one ring and all cuts leave a single starting skeleton, as required in Figure 1. The program seeks separately the set of cases constructing two bonds and the set constructing three, and it seeks both the linear strings of Figure 6 and the internal cyclic ones, the [2 + 4] and [2 + 2 + 2] cycloadditions. The model for only two constructions lacks the second vinyl component at atoms 4–5 in Figure 6.

The following constraints are placed on the selection of skeletal bonds from a polycyclic target:

(1) No aromatic rings are allowed on atoms 2–5 nor are double bonds to off-strand atoms, but double bonds between 2 and 3 and between 4 and 5 are allowed.

(2) Any double bond at cut sites in the target (atoms 1–2, 3–4, 5–6) is replaced by an attached (unspecified) heteroatom,

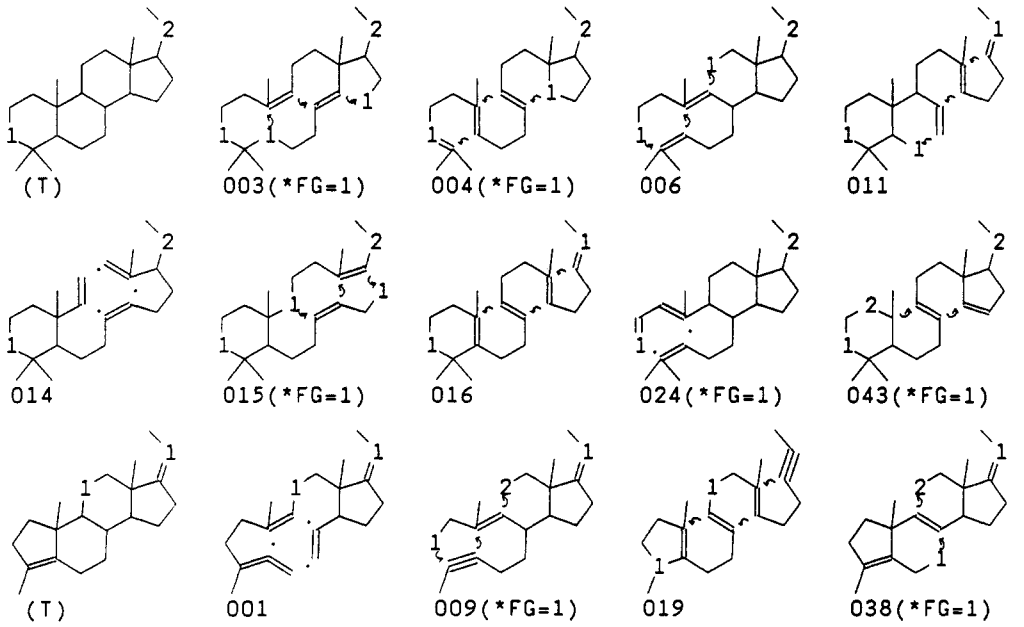


Figure 7. Polycyclization of steroid intermediates (*FG = number of refunctionalization steps).

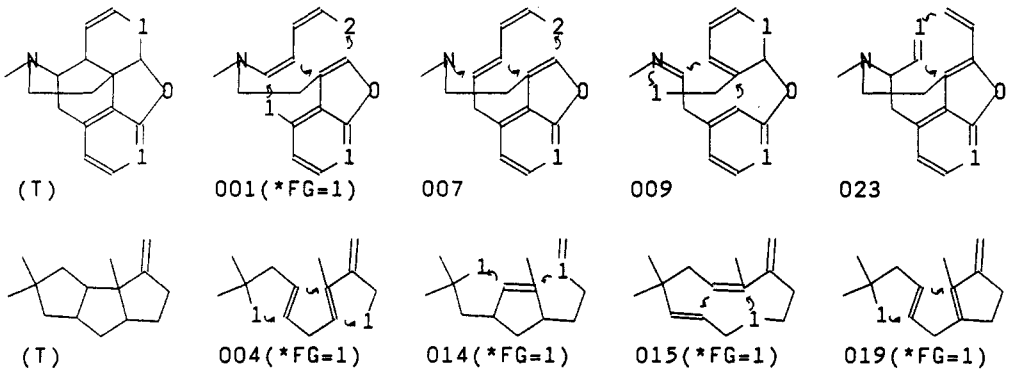


Figure 8. Polycyclizations for morphine and hirsutene (*FG = number of refunctionalization steps).

i.e., $z = 1$, and any $z = 2$ functionality (i.e., carbonyl, etc.) is changed to $z = 1$. These are recorded as refunctionalization steps.

(3) Markovnikov's rule is followed to determine the direction of electron flow (atoms 1 → 6) in any linear string, but ignored in any unit (atoms 2–3 or 4–5) with a functional group. (a) Units containing skeletal heteroatoms (N, O, S) must flow from C to Z; i.e., the heteroatom will have the higher number, 3 or 5 in Figure 6. (b) Units must have $\sigma_a \geq \sigma_b$, where $a < b$ in number and σ is the number of attached carbons (Figure 1, ref 1).

(4) The terminal atoms (1 and 6) are evaluated as in SYNGEN with a further strand of up to three atoms ($\alpha\beta\gamma$) out from the dissected bond to accommodate product functionality for a nucleophilic half-reaction from atom 1 and an electrophilic one from atom 6. (a) Nucleophilic half-reactions: A1, B1, R1, B2, R2, B3, R3 (Figure 3, ref 1). (b) Electrophilic half-reactions: 11, 21, 31, 12, 13. (c) If no appropriate functional groups for activation of these half-reactions are present on the target, then appropriate functionality is introduced, i.e., $z_\alpha = 1$ for B1, $z_\beta = 2$ for A1, $\pi_{\alpha\beta}$ for P1, $z_\alpha = 1$ for 21. These are also recorded as refunctionalizations. (d) Markovnikov tests of $\sigma_\beta \geq \sigma_\alpha$ are applied for P1 and B2 at atom 1 and $\sigma_\alpha \geq \sigma_\beta$ for 12 at atom 6.

For each successful skeletal bondset so derived the substrate functionality is created by introducing π-bonds between atoms 2 and 3 and between atoms 4 and 5 and the appropriate functional group changes for the two terminal half-reaction strands ($\alpha\beta\gamma$) as above and in ref 1. These substrates are then recorded and displayed, with arrows indicating the direction of electron flow and an indication of the number of refunctionalization changes required.

The well-known multiple cyclizations of Johnson to synthesize steroids[10] were modeled after the biocyclization of squalene. Two of Johnson's targets are illustrated in Figure 7. The first produced 18 reactions making three rings and 35 more making only two. The 9 examples in Figure 7 were selected in part to exclude precursors with large rings; one such example is shown (3) from the 12 produced in the output. Otherwise, cases were selected to include cycloadditions as well as linear cyclizations and then on the relative apparent ease of synthesis of the precursors.

As before, only z values (of 1 or 2) are shown for functionalized carbons, but the actual chemistry is generally easy to elaborate from these. Since the program does not seek 4-bond cyclizations, a true "squalane cyclization" forming all four rings does not appear, but is mimicked by the two examples 4 and 16. The latter is very close to one used by Johnson.[10] The program does not show all functional variants at the two termini, so that the enamine/enol ether nucleophile in examples 11 and 16 could as well be a triple bond, as Johnson actually used. The second Johnson target, requiring ring A ozonolysis and recyclization to a 6-membered ring, is shown as the lower example in Figure 7, which produced 23 three-bond cyclizations and 30 making only two bonds. Similar selections afford the four examples shown, of which Johnson's was example 19.

The morphine molecule is the target at the top of Figure 8, yielding 9 and 29 conversions for three- and two-bond cyclizations, respectively. Four selections from this output are

shown, the last being similar to a plan of Dalton,[11] but no other morphine syntheses have used this concept. The cyclizations of the benzofurans in examples 1 and 7, however, seem to represent an especially efficient approach. The lower example in Figure 8 is the tricyclic terpene hirsutene, which yielded 9 and 18 conversions of three- and two-bond cyclizations, respectively. No examples of the use of multiple cyclizations exist in laboratory syntheses,[12] but the biosynthesis[13] does involve polycyclization of an 11-membered ring humulene precursor. The nearest such case produced here is example 15.

In summary, these two multiple construction programs offer new planning ideas for the efficient synthesis of a number of complex targets. Even when the detailed chemistry presented or implied here is not fully viable, variations incorporating the central theme may lead to successful syntheses that are often not likely to be conceived through common intuitive planning. Example 5 for hirsutene in Figure 8 illustrates this since the right-hand double bond has an added heteroatom placed on its lower carbon, whereas -OR on the upper carbon instead would provide a much better initiation of cyclization. To expand the potential of the SYNGEN program these two programs will now be added to it as modules offering a different aspect of the search for efficiency in synthesis design.

ACKNOWLEDGMENT

We are grateful for support provided by the Eastman Kodak Co. and by a grant (CHE-86-20066) from the National Science Foundation.

REFERENCES AND NOTES

(1) Hendrickson, J. B.; Toczko, A. G. *J. Chem. Inf. Comput. Sci.* (preceding paper in this issue and references cited therein).
(2) This procedure creates four starting skeletons, which for a C_{20} target will average C_5 in size. Our experience with our starting materials catalog (some 6000 compounds) implies that the available functional group variety on the various skeletons falls off sharply above C_5. Therefore, it appears reasonable to expect of this procedure a fairly productive generation of routes from real catalog starting materials. A survey of published syntheses shows the average starting material to have only three skeletal carbons. Since this procedure will find routes from average C_5 compounds, it further serves to minimize the construction reactions required.
(3) Spencer, T. A.; Newton, M. D.; Baldwin, S. W. *J. Org. Chem.* **1964**, *29*, 787; improved by Ing-lung Shih in our laboratory.
(4) Desphande, M. N.; Jawdosiuk, M.; Kubiak, G.; Venkatachalam, M.; Weiss, V.; Cook, J. M. *J. Am. Chem. Soc.* **1985**, *107*, 4786, and references cited therein.
(5) Taub, D. In *Total Synthesis of Natural Products*; ApSimon, J., Ed.; Wiley: New York, 1973, Vol. 2, p 641; 1984, Vol. 6, p 1.
(6) Jung, M. E.; Halweg, K. M. *Tetrahedron Lett.* **1981**, 3929. Bal, S. A.; Helquist, P. *Ibid.* **1981**, 3939. Denmark, S. E.; Germanas, J. P. *Ibid.* **1984**, 1231.
(7) Stork, G., et al. *J. Am. Chem. Soc.* **1981**, *103*, 4948; **1982**, *104*, 3758.
(8) Hendrickson, J. B.; Braun-Keller, E. *J. Comput. Chem.* **1980**, *1*, 323.
(9) Posner, G. H. *Chem. Rev.* **1986**, *86*, 831.
(10) Johnson, W. S., et al. *J. Am. Chem. Soc.* **1976**, *98*, 1038, 1039; **1977**, *99*, 8341; **1982**, *104*, 3508. *Bioorg. Chem.* **1976**, *5*, 51.
(11) Dalton, D. R., private communication.
(12) Heathcock, C. H.; Graham, S. L.; Pirrung, M. C.; Plavac, F.; White, C. T. In *Total Synthesis of Natural Products*; ApSimon, J., Ed.; Wiley: New York, 1983; Vol. 5, p 1.
(13) Cane, D. E. In *Biosynthesis of Isoprenoid Compounds*; Porter, J. W., Spurgeon, S. L., Eds.; Wiley: New York, 1981; Vol. 1, p 336.

An Improved IUPAC-Based Method for Identifying Alkanes

SCOTT DAVIDSON*

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

Received December 19, 1988

The IUPAC rules for naming alkanes with more than one main-chain candidate do not always reduce the largest side chains and in some cases fail to break ties, leaving a compound without a unique name. Recognizing that alkane and side-chain structures differ only by level (side chains have main chains that have side chains, etc.) leads to an improved algorithm based upon a simple, recursive definition of a main chain: *the chain with the least complex side chains*. The described computer algorithm breaks down the carbon skeleton into simple and complex side chains, which are coded onto a tree-structured, complexity-ordered specification page suitable for comparison with alternate configurations to minimize side-chain complexity.

INTRODUCTION

Alkane nomeclature in organic chemistry provides the foundation for building sets of rules to describe compounds of increasing complexity. Traditionally, study of alkane nomenclature in college chemistry courses has been limited to finding the longest chain and selecting the correct chain end to label as C-1. This generally suffices to move on to substituted structures. However, there is more to the topic. Consider the alkane shown in Figure 1. This compound has a symmetric, easily visualized structure reflected in the name 3-isopropylpentane. However, according to IUPAC Rule A-2.6(a) (Appendix I), the correct name is 3-ethyl-2-methylpentane. This rule simplifies by breaking the isopropyl group into a methyl group and a segment of an alternate main chain. Similarly, in the Rule 2.6(a) example, a potential 5-carbon side chain is simplified. IUPAC Rules 2.6(a-d) are a sequence of four tiebreaker tests to be employed when alkanes having two or more longest chains are encountered. (The reader is urged to study the examples carefully.)

Main-chain ties are actually a rather common feature of alkanes, for example, every 2-methyl, 3-ethyl, etc. In fact, all alkanes except methane have at least one main chain tie—the same main chain numbered from the other end. This will be shown later to have relevance to the naming of alkanes beyond numbering.

Now consider the substituted decane shown in Figure 2, which is certainly no more complicated than the examples in Appendix I. The C-skeleton shows two candidate main-chain segments (1-4 and 1′-4′) branching from C-5, each with methyl and ethyl side chains. The only difference is that the positions of methyl and ethyl are reversed. Now, applying IUPAC Rules 2.6(a-d) yields the following: (a) both have the same number of side chains; (b) both have the same

* Address correspondence to this author at 240 Manor Circle, No. 2, Takoma Park, MD 20912.

0095-2338/89/1629-0151$01.50/0