CH_2=C($SiMe_3$)$COCH_3$, 43209-86-5; (Z)-n-C_8H_{17}C(Br)=CH-$(CH_2)_2COCH_3$, 97071-73-3; (Z)-n-C_5H_{11}C(Br)=CH$(CH_2)_2COCH_3$, 97071-74-4; (Z)-PhC(Br)=CH$(CH_2)_2COCH_3$, 97071-75-5; (Z)-n-C_6H_{13}C(I)=CH$(CH_2)_2COCH_3$, 97071-76-6; (Z)-n-C_4H_9C(Br)=CHCHPhCH$_2COCH_3$, 97071-78-8; (Z)-n-C_4H_9C(Br)=CHCHPhCH$_2$COPh, 97071-79-9; (Z)-n-C_4H_9C(I)=CHCHPhCH$_2$COPh, 97071-80-2; (Z)-BrCH$_2$C(I)=CH$(CH_2)_2COCH_3$, 97071-81-3; CH_2=CHBr, 593-60-2; 1-decyne, 764-93-2; 1-heptyne, 628-71-7; phenylethyne, 536-74-3; 1-octyne, 629-05-0; 1-hexyne, 693-02-7; propargyl bromide, 106-96-7.

Supplementary Material Available: Syntheses and spectral data of compounds **7** (R = C_8H_{17}), **11**, **14**, and **15** (3 pages). Ordering information is given on any current masthead page.

A Logic-Based Program for Synthesis Design

James B. Hendrickson,* David L. Grier, and A. Glenn Toczko

Contribution from the Edison Chemical Laboratories, Brandeis University, Waltham, Massachusetts 02254. *Received January 16, 1985*

Abstract: A minicomputer program (SYNGEN) is described which generates all variable syntheses for any target structure within defined constraints and without user interaction. SYNGEN first dissects the skeleton to find all fully convergent bondsets which utilize available starting material skeletons found in (usually) only two successive levels of cuts. Then for each such bondset the necessary functionality is generated to initiate successive constructions of each designated bond, from real starting materials to the target functionality. The construction reactions are generated from mechanistic principles formulated in a generalized digital description and requiring no database library. The output is displayed by a second program (SYNOUT) which allows examination and selection of the resulting routes. Several examples are discussed showing actual syntheses reproduced as well as new routes equally short.

I. Introduction and Overview

We have discussed previously some new logical tenets for approaching a viable system for synthesis design.[1] This paper describes a working minicomputer program embodying those tenets. The program has two parts: first, a skeletal dissection of the target, second, a generation of functionality necessary to assemble the synthons so defined. In this paper we present an overview of goals, methods, and results as well as a discussion of skeletal dissection. The characteristics of the approach are summarized here and amplified below: (1) an executive program not interactive with the chemist; (2) assessment of all possible routes within clearly defined constraints; (3) digital expression of molecules and reactions; (4) initial skeletal dissection for efficient assembly; (5) orientation on available starting materials; (6) orientation on economy of steps; (7) limitation primarily to construction reactions; (8) generation of reactions from mechanistic logic; (9) no prediction of yields.

An executive program assures that the given set of rules and heuristics are consistently applied in all cases without operator bias so that all possible routes will be generated and assessed by the same criteria. The operator, of course, selects from the final output of a few optimal routes, but only after execution of the program is complete. The molecules and reactions are all expressed and manipulated as simple lists of digits expressing functionality. This serves to increase computer speed and lessen storage considerably, but more importantly it assures that all possible results are simply mathematical combinations and hence readily ascertained. Furthermore, the digital expression serves to abstract the essentials and coalesce trivial distinctions of functionality, making it possible to span the enormous potential search space of the problem.

The first operation of the program is to examine ways to break up the skeleton into the fewest pieces, or synthon skeletons, which actually exist as starting materials, thus defining optimal bondsets.[1,2] In the dichotomy of skeleton vs. functionality, we focus on the skeleton as first consideration, rather than an examination of target functionality to ascertain all possible last reactions to form it. We suspect this prior skeletal examination to be a very common mode of perception among synthetic chemists. It also has the advantage of formulating rational routes to saturated hydrocarbons and, even in functionalized targets, of discerning where to use dummy functional groups that are eliminated en route and leave no trace in the target functionality.

An ordered bondset[2] dictates the starting skeletons and is indeed the simplest overall description of any synthesis conceptually. Thus the search is focused to strike through the massive center of the synthesis tree, rooting the search on available starting material skeletons and so allowing it to converge rapidly. An ordered catalog of available starting materials is an important ensemble of data for the process of synthesis design and it can be used actively in this way to focus the selection of pathways. Our program interacts at present with a catalog of about 5000 basic starting materials, which represent 344 skeletons of connected C and N atoms.[3]

The central criterion of the program is economy, expressed as the fewest steps or operations in the most convergent order.[4] An ordered bondset shows both the number of constructions needed and the extent of convergency involved in assembling the skeleton. This is the key to assembling the target molecule in the most efficient way. Construction reactions are obligatory to the assembly but other reactions are not, and the intent of the program is to seek routes which are primarily composed of construction reactions. Other reactions are, however, included or implicit, especially when they are attendant on the constructions, as described in section IIIB below.

The number of synthetic reactions available is probably in the tens of thousands, depending on the detail of definition, and is constantly growing. Incorporation of a library of reactions in the

(1) Hendrickson, J. B.; Braun-Keller, E.; Toczko, A. G. *Tetrahedron, Suppl.* **1981**, *37*, 359.

(2) A *bondset* is the set of skeletal bonds actually constructed in any given synthesis.[1] An *ordered bondset* defines the order in which they are constructed as well.

(3) We obtained a computer listing of all Aldrich Chemical Company compounds from an EPA/NIH catalog of 170 000 compounds in connectivity-table form and accepted only those with connected C/N skeletons of $3 \leq n \leq 16$ atoms, e.g., compounds like RCOOR′ are catalogued as RCOOH as well as R′OH in the digital format ($z\pi$-lists) described below. This resulted in some 5000 starting materials, which appear in practice to contain all necessary ones. The catalog of course can be expanded. Suppliers could not unfortunately provide computer tapes of all their compounds in usable connectivity-table form.

(4) Hendrickson, J. B. *J. Am. Chem. Soc.* **1977**, *99*, 5439.

0002-7863/85/1507-5228$01.50/0

computer can at best provide only a portion of the whole and is a massive and continuing chore of uncertain consistency. It also occupies a huge computer storage difficult to search. Finally, it cannot create new reactions, being tied to past experience. We chose instead to define the range of all possible reactions in terms of the digital expression of functionality, abstracting the broad mechanistic possibilities to a relative small family of overall net structural changes in reaction. The net structural change in any reaction is defined by the digital changes in functionality at each involved carbon and so can be simply mathematically generated by adding a specified generator number to the digital list or number representing a product molecule in order to derive its substrates (or vice versa). This assembly of all possible net structural changes turns out to reflect exactly the mechanistic nature of the reactions, and so in any particular case a generated reaction can also be quickly assessed for mechanistic viability, as detailed later (section III).

Finally, we have elected not to use yield prediction to compare and select among generated reactions. Thousands of reactions must be compared in searching for optimal routes, and the precision of yield prediction at present is too poor to afford any confidence in the relative accuracy of these numerous comparisons. Therefore, the central criterion for selection of optimal routes remains that of fewest steps and most efficient (convergent) order of assembly, and these are the basis of the skeletal dissections which determine the best bondsets. Furthermore, the requirement of available starting materials and a sequence of construction reactions only (the fewest steps) puts stringent demands on the necessary functionality as well, and so yields a relatively tiny optimal set of actual routes.

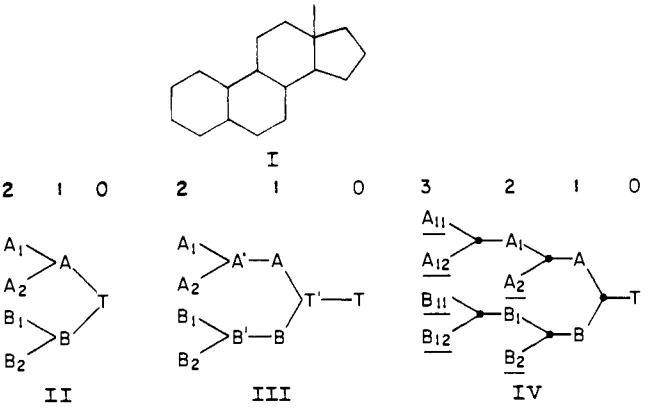
In summary, an overview of the procedure can now be seen as follows. The skeleton of the target (connected C- and or N-atoms) is first dissected all ways that provide convergent modes of assembly from the largest available starting material skeletons. This affords the optimal ordered bondsets. Then for each bondset the necessary functionality is generated to construct each bond in order, retrosynthetically from the target, with no intermediate refunctionalizing. This generates the intermediates back to and including the starting materials and accepts only routes in which the latter are available in the catalog.

II. Skeletal Dissection

A. General Perspective. The key to the first step in our analysis is to find the most efficient ways to assemble the skeleton of the target from available starting material skeletons. The number of possible ways to dissect the skeleton, without even considering reactions, is enormous.[4] Thus the number[5] of possible bondsets of λ cuts is given by $\binom{b}{\lambda}$. Since the number of possible ways to order these λ constructions for each such bondset is $\lambda!$, the total number of possible *ordered bondsets*[2] is then given by $b!/(b-\lambda)!$ We have found[6] that the average published synthesis constructs about one skeletal bond in four, i.e., $\lambda \sim b/4$, and that the average starting material has only three skeletal carbons. This implies that common targets of n = 15–30 carbons with up to four rings will require λ = 6–12 constructions.[5] Thus to assemble the skeleton (I) of estrone (n = 18) an average synthesis will require k = 6 pieces (starting material skeletons), hence λ = 9 constructions if all pieces are acyclic,[7] i.e., Δr = 4. As there are 21 bonds this means that there are $21!/(21-9)! = 1.07 \times 10^{11}$, or 107 *billion* possible ways to assemble the estrone skeleton with nine constructions.[1] The Torgov-Smith estrone synthesis (see examples in section V below) has only three constructions, but even three constructions can assemble the skeleton in nearly 8000 ways. Obviously the heuristics needed to decide on the best ordered bondsets here must be very selective.

Several heuristic bases for selecting optimal bondsets have been described:[1] minimizing constructions by finding large starting skeletons; identification of equal (or near equal) halves; convergency; and multiple constructions. The first two are readily accomplished by the computer since each time it cuts a skeleton in two it will identify the two pieces formed both with the skeletons in the starting material catalog and also with each other. The process of comparison of skeletons is basic to the program and rapidly accomplished by a procedure of creating a unique, "maximal" adjacency matrix of any skeleton, in effect a unique canonical numbering, which allows any two skeletons to be compared or the catalog quickly searched for an identity match.[8]

B. Convergency. A simple procedure for establishing a fully convergent mode of target skeleton assembly has been described.[4] This consists essentially of cutting the skeleton into two pieces and then each piece again into two more, thus generating four pieces, i.e., starting skeletons. In order to separate a skeleton into two parts we allow the cutting of only one or two bonds, i.e., a maximum of only one ring opened. The first level of dissection affords two intermediate skeletons, A and B, and these at the second level, or cut, afford skeletons, A_1, A_2 and B_1, B_2, respectively. The process may be seen, either in the retrosynthetic direction (to the left) or in the forward direction (to the right), in a construction plan,[4] in which a forked pair of lines is an *affixation*, linking two pieces, and a horizontal line is a *cyclization*, forming a ring on one intermediate skeleton. The convergent construction plan here described for target T is shown in II with only affixations at each level (one bond cut each), for a total of λ = 3 (Δr = 0),[5] and in III with the maximum of three cyclizations, for a total of λ = 6 (Δr = 3); primed letters imply uncyclized intermediates.



The six pieces so generated by such cuts, i.e., A, B, A_1, A_2, B_1, B_2, are compared with the skeletons in the catalog. In our experience two levels of dissection are usually enough to find some convergent plans with all four starting skeletons available in the catalog. This would, for example, average C_5 synthons for a C_{20} target. For estrone with λ = 6 cuts there should be almost 40 million ordered bondsets (i.e., 21!/15!), but the *convergent* plans generated in this way number less than 2000. This is further reduced to 66 ordered bonsets, however, when we make the further demand that all four pieces produced at second level are actually skeletons in the starting catalog. Thus the two requirements, of convergency and found starting skeletons, massively reduce the total possible modes of skeletal assembly. When viable reactions are then sought to construct these bonds in order from real (functionalized) starting materials, the number of successful routes found will be much fewer (see section VA).

The number of options for first level cuts is reasonably small, but the implicit combinations swell this number by perhaps an order of magnitude for second level, and another for third level, making a full search at third level very costly. For targets up to

(5) For a target of n skeletal atoms and r rings the number of skeletal bonds is $b = n + r - 1$. If it is cut into k pieces, the number of bonds cut is $\lambda = k + \Delta r - 1$. If all pieces are acyclic and average three skeletal atoms, then $\lambda = n/3 + r - 1$.

(6) A survey of published syntheses for these conceptual correlations is in preparation; we thank Mr. Daniel Cohen for these preliminary results.

(7) One such synthesis of λ = 9, from Velluz, is analyzed in ref 4, Figure 1.

(8) The procedure for canonical numbering of skeletons and their comparison for identity is detailed in: Hendrickson, J. B.; Toczko, A. G. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 171.

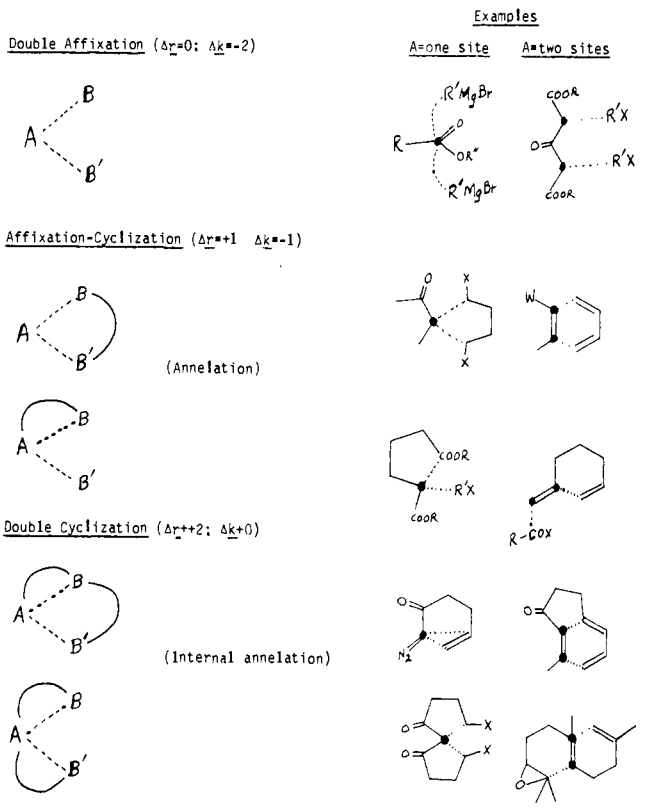


Figure 1. Modes of double construction.

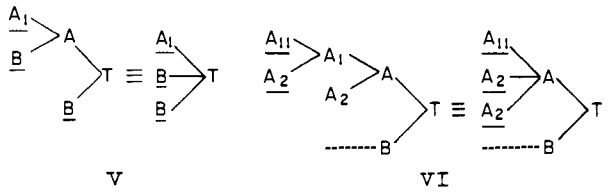
about C_{20}, fortunately, some solutions usually appear by second level. As these are necessarily the fewest constructions, they are the most economical solutions and the third level need not be generated. For larger targets the third level is often necessary because the second-level intermediates are often still too large to be found in the catalog. In such cases we allow third-level dissections to proceed only on one second-level intermediate of each half, i.e., construction plan (IV) in which the underlined starting skeletons are required to be found in the catalog. This generally serves to provide solutions up to about C_{30} targets.

C. Multiple Constructions. The greater efficiency of several constructions at once is a very powerful synthesis concept, one that has not apparently been explicitly described in general form before. If economy dictates the shortest sequence of construction steps, then any opportunity to do two or more constructions at once results in even shorter synthesis.

Double constructions are the simplest multiple constructions and may be systematically categorized as in Figure 1. The simplest form is a double affixation, joining one piece (A) to two others (B, B′) in one operation. The same site (e.g., carbon) on piece A is used in both constructions in examples like double alkylation of enolates or double Grignard attack on esters, but two sites may serve in examples like the double alkylation in the second example shown in Figure 1. In this case B and B′ must be the same. If any two of the components are already linked we generate two modes of affixation–cyclization, the first of which is an annelation (B–B′ linked), the second (A–B linked) a less common mode of assembly. Both join two pieces and create one ring. In the annelation both bonds formed are in the same ring and shared with no other ring, as in simple cycloadditions or Robinson annelation or, on one site, the (n + 1)-annelation (spiroannelation if A is already a ring) as illustrated. Finally, a double cyclization results if all three pieces are previously linked; the two cyclizations can be common to one ring as in the first instance, i.e., internal annelation, or form one bond in each of two rings as in the second. In either case the two bonds created on unit A may be at the same site or different sites as in the examples.

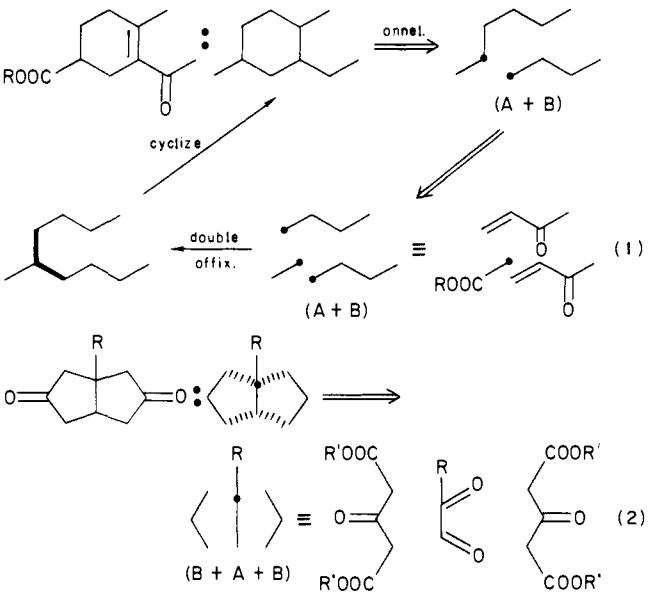
The program automatically locates annelations when the dissection into two pieces cuts two bonds in one ring. The functionality generator (section III) then creates only those pairs of constructions which can be done in one operation. This coalesces two constructions (one affixation and one cyclization).

However, the program requires a special skeletal search for the double affixations. The assumption of two identical components (B) both adding in one operation to component A also requires the same construction *site* on each molecule of B. The program locates the identity of two B components by comparing generated skeletons[8] at two subsequent levels rather than at the same level as in "equal halves" comparisons (A = B). The coalescence of constructions implied can occur at first level as shown in construction plan (V) or at second level as shown in (VI). Thus first-level double affixations are searched in the dissection for a duplicate $\underline{A_2}$ as in (VI), with found starting skeletons underlined.



V VI

Finally in cases of subsequent cyclization of the product, the order of constructions must be changed from that initially located. The example[9] in eq 1 shows target (T) dissected (double arrows) in normal mode (annelation) to A + B, then at the next level to A_1 + B + B. This must be reordered to show the (single-arrow) sequence via double affixation to A_1BB, then cyclized to T. The Weiss reaction,[10] shown in eq 2, is a particularly fine example of double affixation with two cyclizations.



III. Functionality Generation

The search problem is first simplified by examining only the target skeleton as described above, yielding optimal ordered bondsets, each of which then represents a subtree of the whole synthesis tree and designates which bonds are to be constructed and in what order. In the next operation we must define the functionality necessary to achieve construction reactions at those bonds in that order, and without resorting (as far as possible) to any refunctionalizing between constructions. This requires a second simplification, viz., of the description of the functionality.

The number of construction reactions, forming carbon–carbon bonds, is very large if the involved functionality is described in detail in the normal way. It is imperative for us to abstract this detail to its essentials and to simplify the description to a digital format which allows rapid manipulation but retains the central features of the net structural change characterizing any given

(9) Spencer, T. A.; Newton, M. D.; Baldwin, S. W. *J. Org. Chem.* **1964**, *29*, 787; improved by Mr. Ing-Lung Shih in our laboratory.
(10) Weiss, U.; Edwards, J. M. *Tetrahedron Lett.* **1968**, 4885.

construction. Then the problem in essence is to find all the possible sequences of construction reactions, for which the bonds to be made are pre-specified by an ordered bondset and the final functionality is that of the target. The digital format described here allows the functionality of the substrate for a particular reaction to be simply mathematically generated from that of any given product with a designated construction bond. In this way the procedure avoids a laborious library of known reactions. It can also generate new chemistry which is still rooted in viable mechanism. The net structural change in any construction reaction is closely linked to its mechanism since the abstract digital format used generalizes the mechanistic function of a functional group.

For structures, the digital system describes each involved carbon in terms of four fundamental types of attachments: R for σ-bond to another carbon, H for bond to hydrogen (or electropositive element), Π for π-bond to carbon, and Z for bond (σ or π) to electronegative atom.[1,11] For each carbon, then, the numbers of each attachment are σ, h, π, and z, respectively, and $\sigma + h + \pi + z = 4$. The functionality is $z + \pi$ and, since the skeletons are given, σ is known and h derives by subtraction. This means that any (connected) carbon functionality is described by two digits, z(**0–3**) and π(**0–2**), requiring four bits to designate in the computer. Any structure is then described by a $z\pi$-list of the carbons ordered by their skeletal numbering.[8] The major coalescence of detail lies in the designation of z for all heteroatom attachments. The mechanistic function these attachments serve in reactions is delineated further below as a subset of z (section IIIc).

For reactions, the system allows a clear generalization of all possible reactions in terms of the attachments gained or lost at each carbon. This has been used to create a basic "Beilstein system" for cataloguing all reactions into a logical, ordered system.[12] A *unit reaction* is defined as a unit exchange of attachments on each involved carbon.[11,12] This is designated by two letters, the first for the attachment made, the second for the attachment lost, hence $4 \times 4 = 16$ possible unit exchanges per carbon. Thus HZ represents a reduction of halide or carbonyl and ZH the oxidation of alcohols, aldehydes, etc. On two carbons HΠ·HΠ is hydrogenation of a double bond. At either carbon forming a carbon–carbon σ-bond in a construction reaction there can be only four exchanges: RH, RZ, RΠ, RR. The last implies a carbon–carbon bond broken as one is made (as in rearrangements) and is not used at present in our program, which does not accept carbon–carbon bond fragmentations.

A. The Basic Construction Reactions. A construction reaction may be seen as two linked *half-reactions* on each side of the bond formed. The skeletal format for a construction half-reaction is shown in eq 3; the strand of carbons is labeled α,β,γ out from the site of construction (the α-carbon). Functionality appears on these carbons and changes from substrate to product in a manner characteristic of a given half-reaction.[13] Further, unchanging functional attachments (π and z) may also appear on α-, β-, and γ-carbons as well as β' and, while these do not change in the reaction, they may affect its course.

(3)

Of the three construction half-reactions (RH, RZ, RΠ), the first two involve only one changing carbon, the α-carbon (RH may require an activating carbonyl at β but this does not change). The RΠ exchange implicates loss of a π-bond on construction, hence an exchange also of HΠ, ZΠ, or ΠΠ on the adjacent π-bonded carbon, the β-carbon. Thus simple π-addition half-reactions require two exchanging carbons (α,β) as RΠ·HΠ or RΠ·ZΠ. The

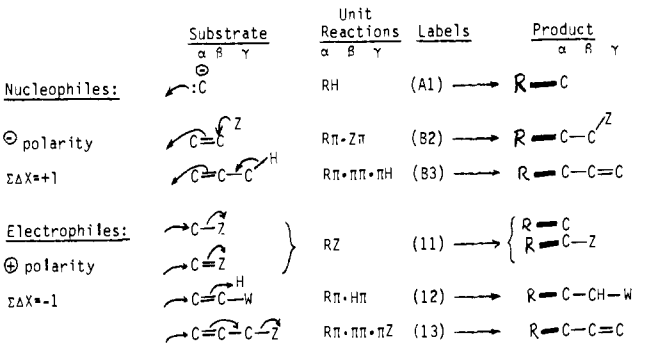


Figure 2. The six basic construction half-reactions.

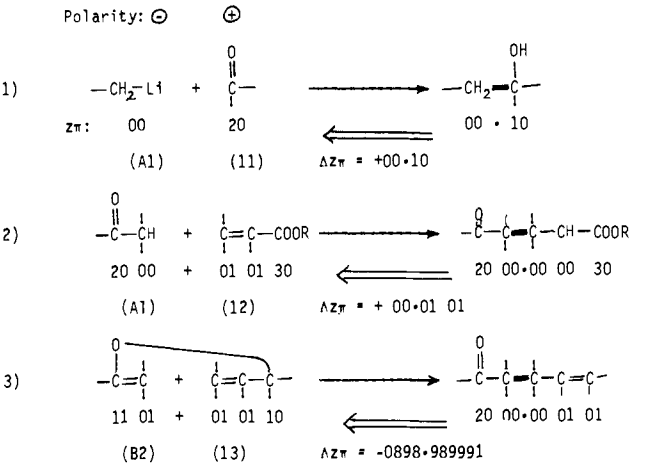


Figure 3. Illustrated simple constructions.

last group of RΠ half-reactions are allylic versions of RH and RZ requiring three changing carbons (α,β,γ): RΠ·ΠΠ·ΠH and RΠ·ΠΠ·ΠZ. The number of changing carbons involved (1–3) is the *half-span* (s′) of the half-reaction. These are all possible net changes of functional attachments on up to three carbons in a half-reaction.[13]

Since the oxidation state of each carbon is given by $X = z - h$, we find that three half-reactions are oxidative ($\Delta X = +1$) and three are reductive ($\Delta X = -1$), one of each half-span. The oxidative half-reactions are nucleophiles, designated as − polarity; the reductive ones are electrophiles (+ polarity). The six possible construction half-reactions are summarized in Figure 2. The labels given to the half-reactions (boldface in parentheses) have two characters, the second being the half-span. The first is a letter for nucleophiles (− polarity) or a number for electrophiles (+ polarity), representing the minimum necessary functionality level ($z + \pi$) of the α-carbon in the substrate. Two half-reactions of opposite polarity combine to create a full construction reaction with no overall oxidation and reduction, and with a full span equal to the sum of their two half-spans, i.e., 2–6 carbons.

The nine possible full constructions (from three − and three + polarities) may now be characterized by the minimum necessary $z\pi$-values for the functionality at each involved carbon in the substrate and the product. The list of changes in $z\pi$-values at each carbon is then characteristic of a particular construction. These changes may be expressed as a $\Delta z\pi$-list operator, or generator, which can generate the substrate by addition to the $z\pi$-list of the product retrosynthetically, or vice versa.[14] Examples are

(11) Hendrickson, J. B. *J. Am. Chem. Soc.* **1971**, *93*, 6847.
(12) Hendrickson, J. B. *J. Chem. Inf. Comput. Sci.* **1979**, *3*, 129.
(13) More than three carbons which do change in a half-reactions are virtually never found. The Diels–Alder diene has four changing carbons but two constructions, each of which can be simulated here with only 2–3 carbons changing in each half-reaction of the overall net change.

(14) The $z\pi$-lists in the examples are shown as normal decimal numbers, with two digits ($z\pi$) per carbon, as in Figure 3. In the program they are binary numbers with four binary bits per carbon instead of two numbers. Thus the $z\pi$-list of the two components in case (2), Figure 3, is written as 2000 + 010130 in decimal numbers but is 10000000 + 000100011100 in binary. The $\Delta z\pi$-list generator is obtained by subtracting the product $z\pi$-list from the substrate $z\pi$-list, oriented on the bond formed. In Figure 3 the $\Delta z\pi$-lists are shown as decimal numbers from that subtraction, although the program utilizes binary. Only the changing carbons are represented in the $\Delta z\pi$-list operator (for the others $\Delta z\pi = 0$).

Table I. Construction Half-Reactions

	substrate α β γ	product α β γ	minimum zπ-list change (retro) product α	β	γ	⇒ substrate α	β	γ
			I. Nucleophiles					
A1	HC(=O)—C—	R—C(=O)—C—	00	20		00	20	
	HC(E)—	R—C(E)—	10			10		
	HC≡C—	R—C≡C—	01	01		01	01	
E1	H_2C(=O)—C—	R—C(=O)—C—	01	20		00	20	
	H_2C(E)—	R—C(E)—	11			10		
F1	HC(E)—	R—C—	01			10		
b2	C≡C—	R—C≡CZ—	00	10		01	01	
	C=CZ—	R—C(=O)—C—	00	20		01	11	
C2	C≡C—	R—CH—C(=O)—	00	20		02	02	
B3	C≡C—CH—	R—C—C≡C—	00	01	01	01	01	00
A3	H_2C—C=C—	R—C=C—CH—	01	01	00	00	01	01
			II. Reductive Carbanions					
R1	ZC—	R—C—	00			10		
	ZC≡C—	R—C≡C—	01	01		11	01	
R2	CO(E) C≡C—	CO(E) R—C—CH—	00	00		01	01	
R3	C=C—CZ—	R—C—C≡C—	00	01	01	01	01	10
RT	HC=C—CZ—	R—C=C—CH—	01	01	00	01	01	10
			III. Electrophiles					
I1	ZC—	R—C—	00			10		
	C(=O)—	R—CZ—	10			20		
	ZC=C—C(=O)(E)	R—C=C—C(=O)(E)	01	01		11	01	
12	C=C—C(=O)(E)	R—C—CH—C(=O)(E)	00	00		01	01	
	C≡C—	R—C=CH—	01	01		02	02	
13	C=C—CZ—	R—C—C≡C	00	01	01	01	01	10
21	C(=O)—	R—C—	01			20		

illustrated in Figure 3. There may be extra functionality present, but the net structural change is the same. The $\Delta z\pi$-list operators are characteristic for each of the (nine) possible full constructions, and the labels for the pairs of corresponding half-reactions comprising the full constructions are shown in parentheses as well in Figure 3, in each case nucleophile (−) at left and electrophile (+) at right. An examination of these constructions shows that, while they derive from all possible net structural changes on the involved carbons, they also faithfully represent reaction mechanisms, and so further examination of their viability in particular settings can be mechanistically addressed (section IIIC,D).

B. **Expanded Half-Reactions.** While in principle we do not accept in our program any refunctionalization reactions (those which do not alter skeleton), there are several broad groups of constructions which implicate attendant refunctionalizations, either spontaneously or in the same laboratory operation. We recognize three types: prior reduction to create carbanions; spontaneous elimination across the constructed bond; and concomitant tautomerization. A survey of all refunctionalization families[12] showed no other comparably general combinations with construction. The three types allow for an expansion of the 9 basic construction half-reactions to a full catalog of 15, presented in Table I.

The prior reduction is characteristic of Grignard and similar organometallics (RX + M → R:−) as well as metal reduction of unsaturated carbonyls and similar systems. In terms of attachment-exchange these are respectively (HZ + RH) or (HΠ·HΠ + RH), as well as allylic reductions leading to allylic carbanions. These amount to reduction followed by **A1** or **B3** half-reactions and afford in nucleophiles the same net structural change as the electrophile reactions of the same half-span, i.e., **R1** ≡ **11**, **R2** ≡ **12**, and **R3** ≡ **13**. All are nucleophiles and labeled with a letter (**R**), listed as section II in Table I.

Elimination to form a Π-bond across the newly constructed bond is typical of aldol and Wittig condensations and occurs after construction by ΠZ exchange at the electrophilic α-carbon and either ΠH or ΠZ at the nucleophilic α-carbon. The former is designated as a **21** half-reaction, i.e., (RZ + ΠZ), and the latter as **E1** (RH + ΠH) or **F1** (RH + ΠZ), respectively. The net structural changes are listed in Table I. The tautomeric half-reactions reflect the uncertain final position of a double bond in allylic carbanions and are seen in Table I as reaction **A3** as well as the reductive combination **RT**.[15] Keto-enol tautomerism is

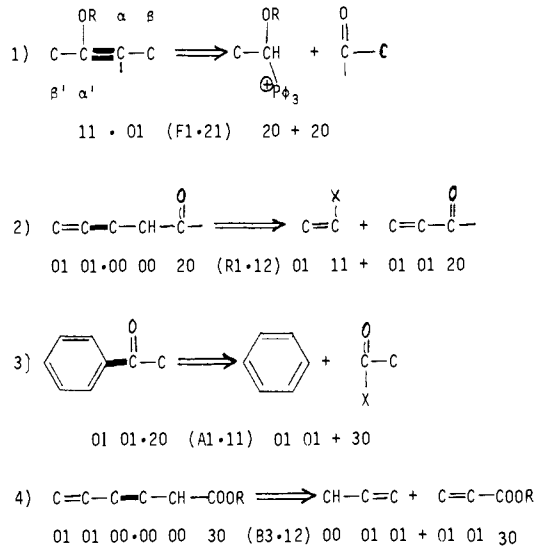


Figure 4. Sample constructions retrosynthetically.

reflected in the **C2** half-reaction, i.e., electrophilic addition to a triple bond followed by tautomeric reversion to ketone; construction on a triple bond without subsequent tautomerism is already reflected in **B2** at triple bond level, i.e., $z\pi$-lists 02.02 → 01.11.

The 15 half-reactions are collected in Table I to show the simplest characteristic functional change (minimum necessary functionality) for each one and the retrosynthetic $z\pi$-list changes for each at the α-, β-, and γ-carbons. Some are illustrated at common higher functionality levels as well. In any case further functionality (z or π) may be placed on the unused valency locations. Several further examples of the expanded half-reactions are illustrated in Figure 4, this time in the retrosynthetic direction as used by the program; the $\alpha\beta\gamma$-strand is, of course, ordered backwards in the left synthon. The Wittig reaction in (1) bears an extra Z substituent (-OR). The reductive conversion of vinyl halide to organometallic is implied in (2) by half-reaction **R1** and used as a construction with conjugate addition (**12**). Example 3 is a Friedel-Crafts acylation: reaction **11** designates alkylation at its lowest functionality level (**10** → **00**), carbonyl addition at the next (**20** → **10**), and acylation at the level of **30** → **20**. The net structural change on the benzene ring is simply RH, hence reaction **A1**. The ene reaction in (4) is shown with an ester which activates but is unchanged on the γ-carbon of the electrophilic half-reaction **12**, i.e., conjugate addition.

Table I shows 15 half-reactions, of which 11 are nucleophilic and only 4 are electrophilic. Combining two half-reactions into a full construction requires one each. Furthermore, two of the nucleophiles (**E1**, **F1**) create double bonds at the construction site and so may only be matched with electrophile **21**, creating two full constructions with double bonds. The remaining nine nucleophiles and three electrophiles combine to create 27 full constructions with single bonds. Hence the total number of full constructions is not (15 × 15) but only 29.

C. Restrictions on Reactions. The half-reactions summarized in Table I are defined by their net structural change but the definition is closely correlated with mechanism, initially in seeing the half-reactions as nucleophiles or electrophiles and then in the details of functionality change. However, we cannot accept many reactions which would be generated this way owing to the ancillary effects of proximal functionality which affects the reaction but does not itself change. Thus, the simple RH half-reaction, **A1**, is perceived as requiring activation, as an electron-withdrawing group, for removal of H. Without this restriction the reaction RH + R′X → R–R′ would appear for any bond. Similarly, not all heteroatoms are displaceable in the **11** half-reaction. However, the effects of proximal functionality on any reaction can be examined by quick numerical tests on the values of z, π, σ, and h for the involved atoms, α, β, β', γ, and γ' (eq 3). The aim has been to restrict half-reactions, in accord with mechanism, enough to eliminate most nonviable constructions without losing much of the possibility of new chemistry. The problem is to do this selection tightly enough to minimize an otherwise excessive output of results without missing good sequences. Hence the restrictions are programmed in a flexible way as modules for easy modification if desired. The balance struck at present is still far more lenient than a detailed reaction library would be, and of course is vastly more compact in both computer time and storage owing to the simple numerical generation and testing of intermediates. Nevertheless, the price paid for allowing new chemistry in this way is still a large output and much of it chemically nonviable, at least with present technology. Other modes of pruning or sorting this output are under study.

In order for the necessary mechanistic functions of attached heteroatoms to be distinguished, we must more closely define z, so far defined for a carbon only as the number of its bonds to heteratom(s). The necessary abstraction is retained if these are defined by *mechanistic function* rather than by heteroatom type (O, N, X, P, etc.). Thus for one heteroatom bond (z = 1) its function may be L = leaving group, E = electron-withdrawing group, or O = electron-donating or neutral (as –OH, –OR) group. For z = 2 or 3, if one bond is E or both are O, the definitions are the same but a carbonyl serves two functions: analogous to L in its additions or electron-withdrawing for an adjacent carbon; and so a carbonyl is designated as W.

Functional groups are of two kinds: those required for activation of a half-reaction and those which reject a reaction because of impeding activation and/or improper regioselectivity, or because of a preferred side reaction. Required activation takes several forms. Electron-withdrawing groups are required for carbanions as E_α, W_β, or W_β' and for Π-electrophiles as E_β or W_γ, where E = heteroatom electron-withdrawing ($-SO_2R$, $-POR_2$, $-PR_3^+$, etc.) and W is used for carbonyl or cyano in the skeleton. The carbanion half-reactions so activated are **A1**, **E1**, **R2**, **R3**, **A3**, and **RT** and the Π-electrophiles are **12** and **11** with a π-bond (third **11** in Table I). Markovnikov activation of simple π-bonds is tested for regioselectivity by comparing σ-values ($\sigma_\alpha > \sigma_\beta$ in product invalidates) or by testing electron-donating or withdrawing groups on the double bond. Also some reactions have lower requirements if the reaction (usually a π-bond reaction) is a cyclization.

Functional groups which reject a reaction include several kinds of restrictions such as incorrect activation giving rise to improper regioselectivity, preference for an alternative reaction such as β-elimination of a leaving group or for a different construction, or simply that the structural demands of the strand (from Table I) do not afford a place for some attachment found in the reaction product.

These restrictions applied for each of the 15 half-reactions are summarized in Table II. The presence of heteroatom(s) ($z \neq 0$) on each carbon (α, β, β', γ) of the product strand is further designated as the functional types L, E, O, and W. Then for each half-reaction those that are required for activation are denoted with an R, and those that invalidate the half-reaction are denoted with an X. When a given product strand $\alpha\beta\beta'\gamma$ (out from a bondset-designated construction bond) is to be tested for the success of each half-reaction, its attached heteroatoms are first checked for R to be sure any required activation exists, then checked again for X to see if any present functionality invalidates it. If both checks are successful,[16] the $\Delta z\pi$-list generator for that half-reaction is used to generate the substrate, as a new $z\pi$-list on $\alpha\beta\gamma$, including the z-variants (L, E, O, W) as a subset definition of each z-value. The list of half-reactions is divided into groups by the various possible presences of π-bonds on the strand as

(15) The other logical tautomeric allylic anion is HC≡C—CH → R—C≡CH—CH, in which the γ-H is removed but the double bond tautomerizes back to is original position after construction. This is not listed separately since the net structural change is the same as in the **A1** reaction.

(16) Further checks are also made for Markovnikov regioselectivity and restriction of simple **11** alkylations to invalidate tertiary carbon sites. These are not shown in Table II since they involve tests of σ instead of z.

described below in section IV; hence some of the 15 half-reactions appear more than once, for a total of 27 entries, including some eased restrictions for cyclizations.

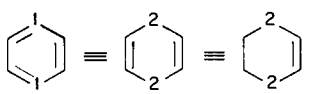
Detailed examination of the entries in Table II allows the scope of the restrictions applied to be surveyed. Thus the requirement of electron-withdrawing groups on α and/or $\beta(\beta')$ is seen as R-entries for **A1** in a saturated product, with leaving groups at the β-carbons invalidating (X-entries), as they are for the reductively created carbanions (**R1, R2, R3, RT**). A carbonyl (W) is disallowed at the β-carbon and at the β'-carbon also (improper regioselectivity) in the conjugate addition half-reaction (**12**), but electron-withdrawing groups are required as activation, either E_β or W_γ.

D. Nitrogen and Phenyl. The skeleton of the target can be input as connected carbon atoms only, or it may be entered including nitrogens in the skeleton as well, actually only when they contain more than one bond to carbon. These nitrogens behave as mechanistically analogous to carbon and are treated simply as special carbons. Hence their presence in the reactive strand ($\alpha\beta\gamma$) is a basis for special restrictions on the allowance or rejection of each half-reaction. These are summarized in the same way as the variance in attached heteroatoms in Table II. For example, the simple **A1** reaction (no π-bond) requires (R-entry) either an electron-withdrawing group (E_α or W_β) or a nitrogen (N) at α, but is indifferent to N at β, or γ. Similarly, in **A1** for a $\pi_{\alpha\beta}$ target, nitrogen at α is acceptable (imine alkylation) but not at β (imine as π-nucleophile for reaction at carbon by an electrophile) or β' (enamine as π-nucleophile at N—C*=C), while nitrogen at γ is acceptable (normal enamine alkylation or acylation). In this fashion all the restrictions the program currently adopts for constructions on strands containing nitrogen atoms can be read from Table II.

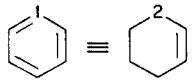
The same approach has not been applied to oxygen and sulfur atoms, which are simply regarded as "z" functionality even if found in a ring. The rationale is that their bond formation is typically refunctionalization in kind and subject to less difficulty than construction reactions. Indeed nitrogen attached to the skeleton by only one bond is also simply treated as a functional group.

Any reaction which attempts to use the "double bonds" of an aromatic ring (phenyl, pyridine, etc.) is disallowed as is attack by a nucleophile (i.e., **11** half-reaction when α is in the ring). When used solely for activation, however, an aromatic ring is considered as an olefin, thus providing sufficient activation to allow the **A1** and **R1** reactions (α-carbon on the aromatic ring) and to activate **B2** when only the γ-carbon is phenyl, i.e., a styrene. Also, when α is in the ring, a β-leaving group (in the ring) does not disallow the **A1** and **R1** reactions (cf., Table II).

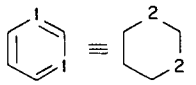
In order to broaden the usefulness of aromatic rings we have also incorporated certain refunctionalizations, i.e., Birch reductions and quinone–hydroquinone interconversions. Thus, whenever an α-atom is encountered in a six-membered ring meeting certain requirements as to placement of heteroatoms and Π-bonds, one or more refunctionalizations are performed. Thus any form in eq 4, 5, or 6 which is encountered in a product automatically also generates for consideration the equivalent one(s) by refunctionalization. The rings are shown with the z-values of functionalized carbons drawn in, e.g., the first two in eq 4 are hydroquinone and quinone, respectively, also equivalent to the quinone cycloadduct form. Equation 5 creates the common Birch reduction conversion and eq 6 the reduction of resorcinols to β-diketones, or vice versa.



(4)



(5)



(6)

IV. The Program

Two programs are involved: SYNGEN accepts the target structure and proceeds to create syntheses, sorting, and storing the results, while SYNOUT is used separately at any later time to display and examine the stored output. On our DEC 11-23 minicomputer the times for complete synthesis generation by SYNGEN range from 2.5 min for jasmone to 27 min for testrone among the examples below, a range common for all cases we have run. The size of the two programs, written in FORTRAN and some assembly language, totals about 5850 lines for SYNGEN and 5600 lines for SYNOUT.

In practice the user first draws the target structure on the graphics terminal. The program then regularizes the drawing (equal bond lengths and angles) and proceeds, leaving the user free to input further target structures if he wishes. The program reorganizes the skeleton as all connected C and N atoms, creates an adjacency matrix of the skeleton with an ordered list of functionality attached (as $z\pi$-values of the numbered skeletal atoms), and locates all rings, identifying aromatic rings separately. The skeleton is dissected all possible ways into two pieces, cutting no more than two skeletal bonds and no aromatic rings; this is the level 1 cut. The user was asked to specify the minimum size for allowed skeletons at this level; otherwise the program accepts no cuts yielding skeletal units smaller than one-fourth of the target skeleton. The intermediate skeletons so created are matched against the catalog of starting material skeletons[17] and flagged, if found, as are identical skeletons, i.e., A = B in plans II or III.

The second level cut is now made all ways on each intermediate (A and B, in plans II or III) of more than four skeletal atoms, creating second level bondsets. Each starting skeleton so found is checked against the catalog, and cuts without all matches are rejected (with one caveat, below). Double affixation possibilities for first level are located here, i.e., A_2 = B in plan V, with identical construction sites on each doubling fragment, $B(A_2)$, attached to A. All level-2 fragments are then further dissected to search for double affixation of the second level kind (plan VI). In either found case the order of constructions is rearranged to afford two successive affixations first, followed by any cyclizations, as in eq 1.

Third-level cuts create too many possibilities so that normally starting skeletons must be found in the catalog by second level. However, the third-level cuts are made solely for the purpose of locating possible double affixation opportunities, as described above. Furthermore, we have found that virtually any product of 5–9 carbons which is not found in the catalog can be made in one construction. Hence such second-level skeletons which are not found in the catalog are kept anyway as intermediates but marked as constructable in principle at the third level. This follows the discussion of plan IV above for large targets, and the program allows only one such intermediate in each convergent half (as in plan IV). If adequate syntheses are generated without using these third-level constructions, they may then be discarded.

In the second phase of SYNGEN the required functionality for the skeletal constructions is generated. For the first level the target functionality (and designated construction bonds) is taken as product and will generate, for each successful construction, the functionality on the first-level intermediate skeletons. These in turn are taken as products to generate the functionality on the second-level intermediates. These functionalized intermediates so generated are now looked up in the catalog[17] to identify real starting materials, i.e., now with functionality as well as skeleton. Only syntheses generating actual available starting materials are retained (except for the C_5–C_9 second-level intermediates marked

(17) Matching with the catalog is achieved by maximizing the adjacency matrix of the skeleton and comparing it as a maximal binary list with the numerically ordered maximal binary list of skeletons in the catalog.[8] Within each maximally numbered skeleton, the several functional variations are numerically ordered by their $z\pi$-lists, so that the functionalized skeleton (full structure) can subsequently be searched as well as the skeleton itself. Thus the acyclic 3-carbon skeleton is maximally numbered 213 and acetone, acrylic acid (esters, etc.), propionic acid (esters, etc.), and 1,3-propanediol then appear in that order, with $z\pi$-lists of 200000, 013001, 003000, and 001010, respectively.

Table II. Restrictions on Half-Reactions

	α					β′					β					γ				
	L	E	O	W	N	L	E	O	W	N	L	E	O	W	N	L	E	O	W	N
										R—C≡C (s' = 1)										
A1	X	X	X	X	X	X	X	X	X	X	–	–	–	X	–	–	–	–	X	–
R1	X	X	X	X	X	X	X	X	X	X	–	–	–	X	–	–	–	–	X	–
11	X	X	X	X	X	X	X	X	X	X	–	–	–	X	–	–	–	–	–	–
										R—C=C—C										
A1	X	–	X	X	–	–	–	X	–	X	–	X	–	X	X	–	X	–	X	–
R1	–	–	–	X	X	X	–	–	–	–	X	X	–	X	X	X	X	–	X	–
11	–	–	–	X	X	–	–	–	–	–	–	R	–	X	R	–	–	–	R	–
B2	X	–	X	X	–	–	–	–	–	X	R	X	R	X	X	–	–	–	X	–
R2	X	R	X	X	R	X	X	X	R	X	X	X	X	X	X	X	X	–	X	–
12	X	X	X	X	X	–	–	–	X	–	X	R	X	X	R	–	–	–	R	X
(cyc)[a]	X	X	X	X	X	–	–	–	X	–	X	–	X	X	–	–	–	–	–	X
A3	X	R	X	X	X	X	X	X	R	X	X	X	–	X	X	X	X	–	X	R
(cyc)[a]	X	–	X	X	X	–	–	–	–	X	X	X	X	X	X	X	X	X	X	–
RT	X	R	X	X	X	X	X	X	R	X	X	X	–	X	X	X	X	–	X	X
										R—C (s' = 1)										
A1	–	R	–	X	R	X	–	–	R	–	X	–	–	R	–	X	–	–	–	–
R1	–	–	–	X	X	X	–	–	–	X	X	–	–	–	X	X	–	–	–	–
11	–	–	–	–	X	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
										R≡C (s' = 1)										
E1	–	R	–	X	R	X	–	–	R	X	X	–	–	R	X	X	–	–	–	–
F1	–	–	–	X	X	X	–	–	–	–	X	–	–	–	–	X	–	–	–	–
21	X	X	X	X	X	–	–	–	–	X	–	–	–	–	X	–	–	–	–	–
										R—C—C (s' = 2)										
B2	X	–	X	X	–	–	–	–	–	X	R	X	R	R	X	–	–	–	X	–
C2	X	–	X	X	–	–	–	–	–	X	X	X	X	R	X	–	–	–	X	–
R2	X	R	X	X	R	X	–	–	R	X	X	X	–	X	X	X	–	–	X	–
12	X	X	X	X	X	–	–	–	X	–	–	R	–	X	R	–	–	–	R	–
(cyc)[a]	X	X	X	X	X	–	–	–	X	–	X	–	X	X	–	–	–	–	–	X
										R—C—C=C (s' = 3)										
B3	X	–	X	X	–	–	–	–	–	X	–	X	–	X	X	–	–	–	X	–
R3	–	R	–	X	–	X	–	–	R	X	X	X	–	X	X	–	X	–	X	X
13	–	X	–	X	X	–	–	–	–	X	–	–	–	X	X	–	–	–	X	X

[a] Same reaction on cyclization.

as above for third-level construction). Routes with the earliest found starting materials are given priority.

Large starting materials[4,18] are especially favored: if the skeleton has been found in the catalog but without the exact functionality required by the generator, then a prior refunctionalizing to repair it is allowed. It is a simple matter to calculate the "chemical distance" between any two compounds,[19] i.e., the number of unit reactions required to convert one to the other. For large starting materials[18] we accept prior refunctionalization of one or two steps from available compounds.

In order to examine any product with a designated construction bond, the program must first select the active strands ($\alpha\beta\gamma$). Because of branching there may be several possible strands out from each end of a designated bond. The reactive strands are selected on the basis of the presence of π-bonds on the α-, β-, γ-carbons, since the half-reaction products group naturally into families defined by the positions of π-bonds in the product, as seen in the groupings of Table II. The positions of π-bonds in the product serve both to determine which branch is to be designated as the $\alpha\beta\gamma$-strand and to limit the possible half-reactions which may be used. In some cases more than one strand is used. For each appropriate half-reaction, the required functionality on all involved carbons is first checked all at once by matching a single list of the LEOWN characteristics of $\alpha\beta\beta'\gamma$ of the given product with the R-list for the reaction in Table II. If this succeeds, the X-list of reactions is then similarly matched, and other requirements checked.[16] If all succeed, the half-reaction generator may be applied to create the substrate functionality on those carbons. Full constructions are then selected from all half-reaction pairs across the designated bond with opposite polarity. When the generated substrate has a skeleton in the catalog, then the functionality on it is also matched against the catalog to ascertain if it is a real starting material, and if so the generated route is retained and stored.

The synthesis generation protocol above is one program (SYNGEN) which proceeds to completion without user intervention and stores its output on a disk for examination later. A second program (SYNOUT) then affords a detailed examination of the output in a variety of ways and is directed by the user to make further selection. The SYNOUT program provides a user menu allowing one to view all bondsets, intermediates, starting materials, or reactions at either level and to select some or all of these for final hardcopy output from an electronic hardcopy unit, a digital plotter or a printer. Thus one can delete unwanted bondsets, accept as given any intermediates (not to see their various syntheses if trivial), or reject any undesired starting materials. The chemist can in this way scan the output of the program to make a final selection, but he makes this from a set of bondsets and routes which must contain all possible shortest paths of sequential constructions from real starting materials.

Certain areas of the output are sometimes unnecessarily large and SYNOUT affords pruning options. Thus many intermediates are relatively trivial linear molecules of C_5–C_{10} with many ways to construct and these may be waived, i.e., accepted as given compounds. Also, much output consists of minor variations on a single reaction theme because of the mechanical generation procedure. Here SYNOUT identifies these "chemical equivalents" and sets them aside in order to examine initially only the salient primary chemistry. If a particular primary reaction is deemed interesting, its equivalents can then be brought up for examination.

A frequent objection to the protocol employed here is that refunctionalizations are disallowed, with only sequential con-

(18) The size definition of a large starting material is left to user discretion but is normally taken as 8 skeletal atoms.

(19) Hendrickson, J. B.; Braun-Keller, E. *J. Comput. Chem.* **1980**, *1*, 323.

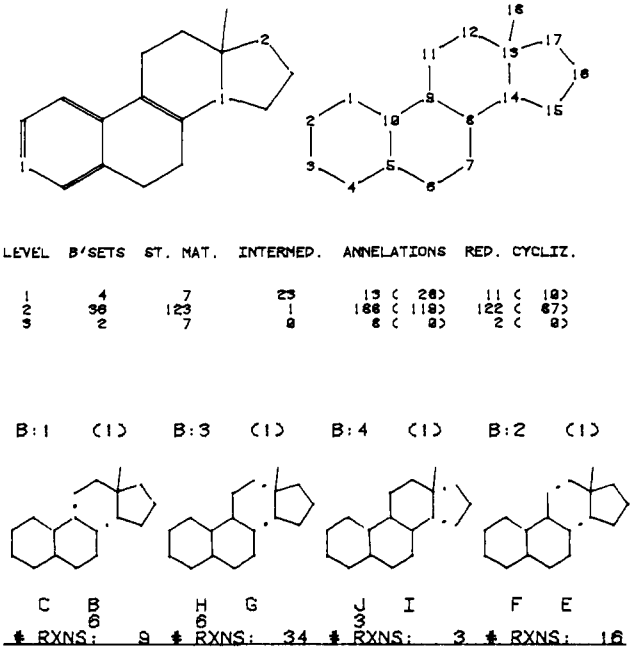


Figure 5. Overview of testrone synthesis results.

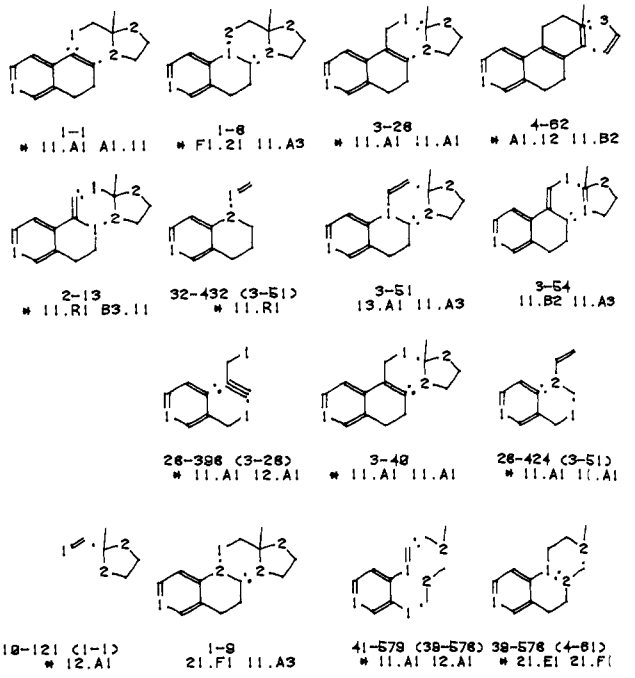


Figure 6. Selected reactions from testrone syntheses.

structions sought, whereas real syntheses usually employ refunctionalizations; in fact our survey of existing syntheses shows a nearly 2:1 predominance over construction reactions. However, as we seek the shortest syntheses, it is clearly more efficient to find routes of construction only. As long as acceptable chemistry is turned up by this procedure, it is hardly reasonable to look beyond for longer routes which incorporate refunctionalization. In fact, however, refunctionalizations are implicit in several ways in our program.

First, at starting material level, we accept larger available starting materials (presently $\geq C_8$)[18] which require 1-2 steps (i.e., unit exchange reactions) of prior refunctionalization in order to be incorporated in the construction sequence. Secondly, there are refunctionalizations involved in six of the fifteen basic half-reactions, as well as the special treatment of aromatics for Birch and quinone refunctionalization. Thirdly, the correct variants (L, E, O, W) of z-functionality are required during the full course of sequential constructions but implicitly may be altered either before or after this central sequence of constructions, i.e., altering the nature of a functional group at any site (as R-X → R-OH, etc.) either on a starting material before construction or on the target skeleton at the end to afford the correctly functionalized target. However, the program may be loosened so as to operate without the z-variant distinctions (L, E, O, W) and so will turn up routes implicitly requiring functional change at any site between constructions. Finally, the compatibility of distant functional groups is not tested here, and so the use of protecting groups (refunctionalization) may actually be required in execution even though this is not specified in the program.

One final observation about the procedure is that it ignores stereochemistry. Certainly the mode of abstraction applied in order to examine all constructions does not allow for stereochemical differentiation. However, once the operator sees all generated routes, he can easily assess which ones are amenable to stereocontrol, or which intermediates can be stereochemically altered as needed en route. An implicit thesis of the approach taken is that the dissection into the most efficient synthons and order of assembly is more critical than an initial consideration of stereochemistry in many cases.

V. Examples

A. Estrone. The first main test of the program is that it should produce known syntheses that fit the short, sequential-constructions format. The shortest synthesis of estrone[20,21] is that of Torgov and Smith,[22] which uses only sequential constructions from simple starting materials to an estrone-skeleton derivative ($\Delta^{8,9}$; 14-OH) which was labeled "testrone" and applied to SYNGEN as a target. The summary of results is shown (as displayed by SYNOUT) in Figure 5 with the target structure and its skeletal numbering as input, and the numbers of routes found at each level, e.g., 4 bondsets at the first level using 7 starting materials and 23 intermediates (created at second level) in 13 "true" (one-step) annelations, with 28 chemical equivalents and 11 two-step annelations[23] (with 10 chemical equivalents). The four first-level bondsets are displayed below, e.g., for bondset number 3 the joining skeletons shown are labeled H and G and are joined in 34 reactions, 6 of which use skeleton H as a found starting material. The actual (functionalized) starting materials and intermediates so labeled may all be displayed as well, and selections made for further display of the actual reactions.

Since bondset 2 shows neither piece available as a starting material, it may be deleted, leaving only four primary annelations at the first level, shown in the first row of Figure 6. In these displayed reactions carbons bearing heteroatom functionality are just labeled with their z-values; the bondset and reaction numbers are shown below the structures and above the half-reaction labels for the two constructions in the annelation. The single and double dots at the sites of bond formation imply the first (affixation) and second (cyclization) constructions, respectively. Thus in reactions 1-8 the first is a Wittig (**F1.21**) followed by an allylic double-bond shift and cyclization onto the ketone of the D-ring (**11.A3**). A chemical equivalent is 1-9 below with the Wittig components

(20) Some 16 syntheses of estrone are recorded,[21] but most of the longer ones were obviously conceived primarily for ring-A nonaromatic steroids.

(21) Syntheses of estrone (and homoestrone) are reviewed: ApSimon, J. In "The Total Synthesis of Natural Products"; Wiley: New York, 1973; Vol. 2. Later syntheses: (a) Danishefsky, S.; Nagel, A. *Chem. Commun.* **1972**, 373. (b) Cohen, N.; Banner, B. L.; Blount, J. F.; Tsai, M.; Saucy, G. *J. Org. Chem.* **1973**, *38*, 3229. (c) Bartlett, P. A.; Johnson, W. S. *J. Am. Chem. Soc.* **1973**, *95*, 7501. (d) Danishefsky, S.; Cain, P. *Ibid.* **1975**, *97*, 5282; **1976**, *98*, 4975. (e) Kametani, T.; Nemoto, H.; Ishikawa, H.; Shiroyama, K.; Fukumoto, K. *Ibid.* **1976**, *98*, 3378. (f) Oppolzer, W.; Battig, K.; Petrzilka, M. *Helv. Chim. Acta* **1978**, *61*, 1945. (g) Byron, T. A.; Reichel, C. J. *Tetrahedron Lett.* **1980**, 2381. (h) Grieco, P. A.; Takigawa, T.; Schillinger, W. J. *J. Org. Chem.* **1980**, *5*, 2247. (i) Quinkert, G.; Weber, W.-D.; Schwartz, U.; Durner, G. *Angew. Chem., Int. Ed. Engl.* **1980**, *19*, 1027. (j) Stork, G.; Sherman, D. H. *J. Am. Chem. Soc.* **1982**, *104*, 3758.

(22) Ananchenko, S. N.; Torgov, I. V. *Tetrahedron Lett.* **1963**, 1553. Smith, H. et al. *Experientia* **1963**, *19*, 394; *J. Chem. Soc.* **1963**, 5072.

(23) Those annelations with reaction conditions compatible with creating both constructions in one operation are separated from those in which the second (cyclization) step is a reductive one and so cannot be done in the same operation.

reversed. Three of the chemical equivalents of 3-28 (bondset 3) are also shown in Figure 6, all with reverse construction order but otherwise illustrating allylic variants in ring C (3-40, 3-51) or enolic derivatives in ring D (3-54). One of the deleted annelations from bondset 2 is added for illustration (2-13), an allylic reductive carbanion on ring B adding to the ring-D diketone, followed by allylic cyclization to testrone.

The D-ring diketone in 3-28 and 3-51 is a found starting material; the others are intermediates constructed at second level. Thus on the second line (Figure 6) reaction 32-432 creates the intermediate for 3-51 by an organometallic from halo-ethene attacking 6-methoxytetralone. This is the Torgov-Smith synthesis as found by SYNGEN and displayed in Figure 6 as the following reaction sequence: 32-432 → 3-51 → testrone. Alternate and novel syntheses also appear, as in 28-398, a cyclization of the acetylene formed by alkylation of a 4-substituted 1-butyne, which creates the precursor for annelation 3-40, which in turn yields testrone. An alternate second-level construction of the 3-51 intermediate is seen in the same bondset (28) via alkylating and cyclizing in methyl vinyl ketone in 28-424. A different but equally short synthesis is indicated in 1-9, and one of the second-level constructions which derives from it is shown as 10-121, read as conjugate addition to a vinyl phosphonium to prepare the Wittig intermediate for 1-9. Finally, bondset 4 is rare and only arises because of a dramatic assembly of the large tricyclic intermediate in 4-62 by way of a double affixation followed by double cyclization at second level as illustrated in 41-579 → 39-578 → 4-62, a sequence of dubious practicality, however.

B. **Jasmone.** It is instructive to examine the program output for a target that has been often synthesized. Here we selected jasmone (and dihydrojasmone), for which some 45 syntheses are recorded,[24] varying from 4 to 13 steps (up to 9 for dihydrojasmone). In order to compare these synthesis plans, and because actual yields were not often given, we used a standard weight ratio of starting materials to target[4] based on 70% average yield for all reactions.[25] For economy, the lowest weights are the best routes.

SYNGEN afforded seven bondsets at first level for jasmone, with 36 primary reactions (and 41 chemical equivalents) from 19 found starting materials and 42 intermediates made at second level. The result for dihydrojasmone was similar except that more first-level pieces were found to be starting materials. The overlap of this program output with the published syntheses can be roughly divided into four groups for comparison. In group A six of the first-level reactions from SYNGEN duplicated 13 of the 45 published syntheses, the variations in the latter arising from different uses of activating or protecting groups.[26] The six routes in group A use three of the bondsets and are shown in Figure 7. The shortest is 2-17, the notation implying γ-enolization, α-alkylation, and reconjugation of the cyclic piece (chemically equivalent routes from each β,γ-double bond isomer were also generated). This synthesis (2-17) was executed by Yoshida[27] and showed the best weight[25] of the 45 examples. The next lowest weight was (1-1), done by Frank,[28] who used a Grignard and dehydration on cyclization instead of the equivalent Wittig (**F1.21**) generated by SYNGEN. The others in group A, which correspond to the other

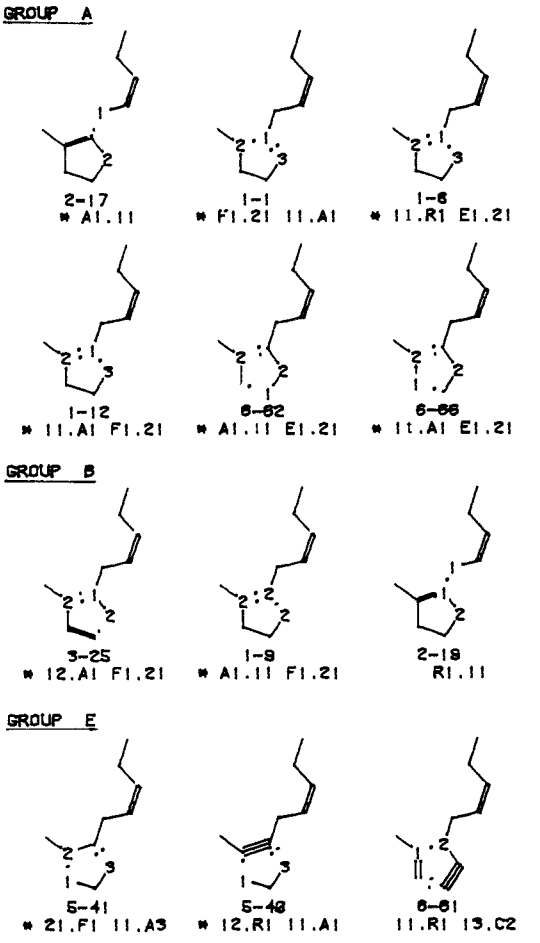


Figure 7. Jasmone: sample first level reactions.

four SYNGEN-derived syntheses shown, are somewhat longer owing to use of protection or activation (e.g., an ester later decarboxylated) which is not shown in the generated plans. The 13 actual routes which match this generated set are the best (lowest weight) of the published syntheses.

Group B represents 14 more of the published syntheses, somewhat higher in weight, which correspond in bondset to five of the seven generated bondsets from the program but differ in carrying masked functionality requiring refunctionalization which is not used in the program, usually variants of the 1,4-diketone precursor of jasmone. Thus the computer does not allow $z = 2$ to be a ketone (z-variant W) in one construction and then an acyl anion equivalent (z-variant E) in the next, implying a refunctionalization step between the constructions. The SYNGEN solution in 3-25 is to add the acyl anion equivalent in conjugate addition (**12.A1**) followed by an independent Wittig (**F1.21**) so that the product of the sequential constructions may be refunctionalized only after the skeleton is fully constructed. Bondset 1 (cf., 1-9) was employed several times in practice with a furan for the C_5 synthon as a masked 1,4-diketone. Reaction 2-19 implies a nucleophilic vinyl halide on the cyclic synthon; in practice (two cases) the polarity was reversed in these syntheses. Bondset 5 was used only once in a synthesis[29] closest to 5-41 in group E below with an activating ester instead of withdrawing heteroatom at the $z = 1$ carbon, later decarboxylated.

Of the remaining 18 syntheses, eight (group C) utilized C-C bond cleavages and are not found by the program. The basis of our generation is one of direct, sequential constructions only and so cleavages are regarded as retrograde steps at present. We plan to incorporate those which have an efficient synthetic rationale in our next stage of development. Two of these jasmone syntheses[30,31] are quite efficient here in using 3- or 4-ring cyclo-

(24) Most of the jasmone syntheses are reviewed by Ho: Ho, T.-L. *Synth. Commun.* **1974**, *4*, 265. Subsequent syntheses: (a) Torii, S.; Tanaka, H.; Tomotaki, Y. *Bull. Chem. Soc. Jpn.* **1977**, *50*, 537. (b) Stetter, H.; Kuhlmann, H. *Synthesis* **1975**, 379. (c) Pattenden, G.; Storer, R. *J. Chem. Soc., Perkin Trans. 1* **1974**, 1603. (d) Murata, S.; Matsuda, I. *Synthesis* **1978**, 221. (e) Wakamatsu, T.; Akasaka, K.; Ban, Y. *Tetrahedron Lett.* **1974**, 3883. (f) Smith, A. B., III; Branca, S. J.; Toder, B. T. *Ibid.* **1975**, 4225. (g) Clark, R. D.; Kozar, L. G.; Heathcock, C. *Synth. Commun.* **1975**, *5*, 1.

(25) Total weight of starting materials[4] is $W = \Sigma_i \mathcal{M}_i x^{l_i}/\mathcal{M}_T$, where $\mathcal{M}$ = molecular weight, l_i = number of steps for synthon i at average yield y, and $x = 1/y$ or $x = 1.43$ for 70% yields.

(26) The program at present takes no account of the incompatibility of nonproximal function groups. Hence an *implicit* need for functional protection or activation may arise in a synthetic route generated by SYNGEN.

(27) Yoshida, T.; Yamaguchi, A.; Komatsu, A. *Agri. Biol. Chem.* **1966**, *30*, 370.

(28) Frank, R. L.; Arvan, P. G.; Richter, J. W.; Vanneman, C. R. J. *J. Am. Chem. Soc.* **1944**, *66*, 4.

(29) Elliott, J. *J. Chem. Soc.* **1956**, 2231.

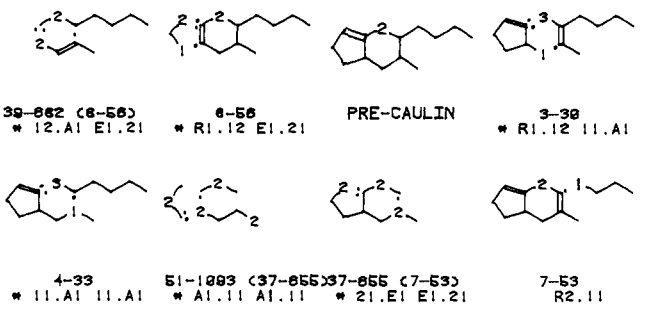


Figure 8. Selected reactions from ptilocaulin syntheses.

additions followed by cleavage of one of the two bonds so created, but the average weight ratio for the group C cleavage routes in general is larger than in A and B and so tends to justify our basis of not generating such routes.

The final group (D) of ten syntheses is a miscellaneous collection of routes not produced by SYNGEN for clear reasons, the main one being that they are not nominally convergent, i.e., the last one or two constructions do not join two pieces (A and B) of three or more carbons. This group has the worst average weight ratio. The simplest overall conclusion is that the best of the published syntheses (in terms of the criterion of economy) are directly generated by the program, and that, specifically, the 13 group A syntheses which are all generated, average better in calculated weight ratios than those not produced by SYNGEN. The program of course also offers some interesting synthetic ideas not mirrored in the published group, several of which are collected as group E in Figure 7.

C. **Ptilocaulin.** The Snider synthesis[32] of ptilocautin also exemplifies a sequential construction route to a precursor labeled "pre-caulin" in Figure 8. Using this as target SYNGEN produced nine first-level bondsets with 26 primary annelations (32 equivalents), among them the Snider synthesis, which is 39–682 → 8–56 → pre-caulin. Other bondsets are represented in the first-level annelations 3–30, 4–33, and 7–53, which also yield pre-caulin via the reactions labeled, the last (7–53) being a reductive alkylation of an intermediate which was in turn created by the only double affixation found for the target. This sequence, an aldol-Michael combination, is shown as 51–1093 → 37–655 → 7–53 → pre-caulin.

VI. Summary

The SYNGEN program was written to fit the requirements and protocol outlined at the outset, and it proves itself by producing known and reasonable syntheses. It operates within specific constraints without user intervention and assesses all possible paths within these constraints. Basically, the constraints are the following: (1) skeletal dissection into ordered bondset families which exhibit convergent assembly of synthons; (2) generation of consecutive constructions from real starting materials for each bondset. In order to avoid the necessity of a library of reactions, the reactions here are generated from broad mechanistic guidelines. To encompass all possibilities, the functionality is abstracted to a digital format ($z\pi$ at each carbon) and all reactions generated by adding $\Delta z\pi$-list operators all possible ways to product $z\pi$-lists. When extensive output appears it may be sorted and selected by the user in numerous ways to facilitate examination, using the SYNOUT program.

The intent of the program is to provide an optimal set of all the shortest, convergent syntheses. These can then serve as standards against which syntheses invented by practicing chemists may be compared.

Acknowledgment. We gratefully acknowledge the financial support of the National Science Foundation through Grant CHE-8102972.

Registry No. Estrone, 53-16-7; jasmone, 488-10-8; ptilocaulin, 78777-02-3.

(30) Wenkert, E.; Mueller, A.; Reardon, E. J.; Sathe, S. S.; Scharf, D. J.; Tosi, G. *J. Am. Chem. Soc.* **1970**, *92*, 7428.
(31) Weinreb, S. M.; Cvetovich, R. J. *Tetrahedron Lett.* **1983**, 861; *J. Am. Chem. Soc.* **1984**, *106*, 1443.
(32) Snider, B.; B.; Faith, W. C. *Tetrahedron Lett.* **1983**, 861; *J. Am. Chem. Soc.* **1984**, *106*, 1443.

Vinyl Cations. 42. Synthesis and Solvolysis of Substituted 1-Cyclobutenyl Nonaflates[1]

Michael Hanack* and Gerhard Auchter

Contribution from the Institut für Organische Chemie, Lehrstuhl für Organische Chemie II der Universität Tübingen, Auf der Morgenstelle 18, D-7400 Tübingen 1, West Germany.
Received February 15, 1985

Abstract: The 2-, 3- and 4-substituted as well as bicyclic nonaflates **19–27** were prepared by treating the corresponding cyclobutanones with nonafluorobutanesulfonic acid anhydride in the presence of 2,6-di-*tert*-butyl-4-methylpyridine as the buffer. Some of the substituted cyclobutanones were prepared by slight modification of the reported procedures. The kinetic and product studies of the 1-cyclobutenyl nonaflates were carried out in TFE–water mixtures. The solvolysis of all the 1-cyclobutenyl nonaflates is shown to proceed by an S_N1 mechanism involving nonclassical 1-cyclobutenyl cations **58**, which can rearrange to the cyclopropylidenemethyl (**59**) and homopropargyl ions **61** and **62**. The solvolysis products are derived by nucleophilic substitution of the solvent with one or more of the cations **58–62**. The kinetics of the nonaflates **19–27** indicate that the rate of solvolysis is strongly dependent on the substituent pattern of the cyclobutenyl system. The substituent effects are interpreted with the formation of the nonclassical structure **58** for the derived cation, with positive charges at C-2 and C-3.

1-Cyclobutenyl nonaflate (**1**)[2] solvolyzes in the highly ionizing and slightly nucleophilic solvent 2,2,2-trifluoroethanol (TFE) with a surprisingly high rate constant[3,4] via an S_N1 mechanism with the formation of four-membered ring products **2** and **3** only.[2,5] According to ab initio and MINDO/3 calculations, the reactive intermediate formed during the solvolysis reaction, the 1-cyclobutenyl cation **4**, has a bridged nonclassical structure in which

(1) Vinyl cations part 41: Collins, C. J.; Hanack, M.; Stutz, H.; Auchter, G.; Schoberth, W. *J. Org. Chem.* **1983**, *48*, 5263.
(2) (a) Subramanian, L. R.; Hanack, M. *Angew. Chem.* **1972**, *84*, 714; *Angew. Chem., Int. Ed. Engl.* **1972**, *11*, 714. (b) Subramanian, L. R.; Hanack, M. *J. Org. Chem.* **1977**, *42*, 174.
(3) Subramanian, L. R.; Hanack, M. *Chem. Ber.* **1972**, *105*, 1465.
(4) Stang, P. J.; Rappoport, Z.; Hanack, M.; Subramanian, L. R. "Vinyl Cations"; Academic Press: New York, 1979.
(5) Hanack, M.; Carnahan, E. J.; Krowczynski, A.; Schobert, W.; Subramanian, L. R.; Subramanian, K. *J. Am. Chem. Soc.* **1979**, *101*, 100.

0002-7863/85/1507-5238$01.50/0