# Systematic Synthesis Design. 6. Yield Analysis and Convergency†[1]

**James B. Hendrickson**

*Contribution from the Edison Laboratory, Brandeis University, Waltham, Massachusetts 02154, and the Department of Chemistry, University of Cape Coast, Ghana.* *Received June 23, 1976*

**Abstract:** Criteria of economy are developed from considerations of yields, to measure material and reagents consumed and time required in any synthesis plan. Application of these criteria to particular synthetic plans allows comparison of their relative economy, or efficiency, even in the early planning stages. Planning graphs of sequential synthetic steps are introduced to examine the convergency of different plans. From this analysis is derived a procedure to dissect a target structure into the relatively few bondsets that define the most economical convergent plan.

When planning an organic synthesis it is presently impossible to predict the yields of individual reactions, or indeed even whether they will succeed or fail (i.e., yield of 0%). Nevertheless, there are still valuable things to learn from a general consideration of the nature of synthetic sequences and the order of operations, in order to compare various alternative sequences with each other, and with previous syntheses of the same target, even early in the planning stage before details are fixed. These considerations can in turn yield valuable heuristics to apply to the initial target dissection, defining efficient general plans without dependence on individual yield prediction. The present analysis complements and expands the general protocol for synthesis design offered in ref 1c, both with respect to initial bondset selection and to evaluation of relative efficiencies of various proposed bondsets and their derived sequences.

**1. Nature of Synthetic Sequences.** A synthesis plan can be usefully broken down into the several functions served by its comprising reactions. These involved reactions are basically of two kinds: the *construction* reactions, creating C–C $\sigma$ bonds, which build the target skeleton; and the *refunctionalization* reactions, which alter the existent functional groups without changing the skeleton.[1b,c] The construction reactions in turn are further subdivided into the *affixations* (a), which unite separate synthons (intermolecular), and the *cyclizations* ($\Delta$r), which create skeletal rings (intramolecular).[1a] These construction steps will be interspersed with refunctionalizations in the total sequence. The sequence is further divided into three phases: (a) the initial phase of functional preparation (refunctionalization steps) of given starting materials for construction; (b) the central construction phase consisting of construction steps and the intervening refunctionalization steps necessary to prepare functionality for construction; (c) the final phase of refunctionalization of the constructed target skeleton to the correct functionality of the target structure itself.

The *bondset*[1c] of the target structure is the set of skeletal bonds or links (numbering $\lambda$) in it which are constructed in the synthesis plan; a bondset of a structure is the simplest description of a synthesis for that structure. The number of construction steps in the synthetic sequence is $\lambda = a + \Delta r$ and the number of component synthons or starting materials is $k$, to be joined by $a = k - 1$ affixations, i.e., $\lambda = \Delta r + k - 1$. If there are extra steps (e), devoted to refunctionalizations, then the total number of steps, $s = \lambda + e = \Delta r + k + e - 1$.

Since syntheses generally involve making larger molecules from smaller ones, only the construction steps are truly obligatory. Hence the shortest and most economical synthesis plan should be one with no refunctionalizations, in any of the three phases. Defined as an "ideal synthesis",[1b,c] such a plan begins with available starting materials requiring no initial functional preparation for construction, carries them through sequential constructions with no need for intervening functional alteration, and so arrives at a fully constructed target skeleton with correct functionality as well, no final refunctionalizations being needed. Such a sequence of constructions with no need for the intervention of functional repair steps is called a *self-consistent sequence*.[1c] This ideal synthesis is rarely possible owing to restrictions of starting materials and of chemistry, but represents an important conceptual goal and one that puts a premium on construction reactions.[1b,c]

Synthetic plans may profitably be abstracted as graphs, with the points representing involved compounds (starting materials, intermediates, and target) and connecting lines for their transformations. The $k$ starting materials are numbered ($i = 1 \rightarrow k$) and each passes through a synthetic path length of $l_i$ steps to the target. The *rank* of a starting material, or of any intermediate en route, is this path length or number of steps to target. The $k$ starting materials are ordered vertically by number ($i$) and horizontally by rank, $l_i$, then linked to intermediate points located horizontally by their rank and so to the target point of rank $l = 0$. The longest linear sequence of steps is the *main line*, i.e., the path ($l_1$) linking the starting material of $i = 1$ to the target. Two very efficient steroid syntheses are analysed as examples in Figure 1, which shows the target structures with the $\lambda$ skeletal bonds (the bondset) which are constructed shown with dotted lines and the separate synthons so isolated numbered ($i = 1 \rightarrow k$) so that the synthon with the largest rank is designated $i = 1$ and initiates the main line on the plan graph shown below the structure.

The points on the plan graphs represent compounds, those of degree 1 being starting materials and the target, those of degree 2 intermediates which are products of cyclization or refunctionalization, and those of degree 3 being affixation products. Horizontal lines are cyclizations and refunctionalizations, the former distinguished as double lines, while pairs of joining slanted lines are affixations; all lines are implicitly vectors to the right. The number of steps is the number of lines minus the number of affixations, $a = k - 1$. The refunctionalizations in the three phases are easily seen in Figure 1, the final refunctionalizing being the single horizontal lines after $l = 3$ in A, $l = 4$ in B (excluding double-line cyclizations), the central construction phase between first and final affixations (ten steps in A; five steps in B) on several lines, and no initial functional preparation in either one.

For examining only the component pairs of a synthesis plan, the complete plan graph may be reduced to a *construction plan* (CP), of construction steps only, by removing all single horizontal lines, the refunctionalization steps. Further reduction by removing all remaining horizontal lines (cyclizations) affords the *affixation plan* (AP) of affixation steps only. These

† This work is respectfully dedicated to Professor R. B. Woodward on the occasion of his 60th birthday this year.

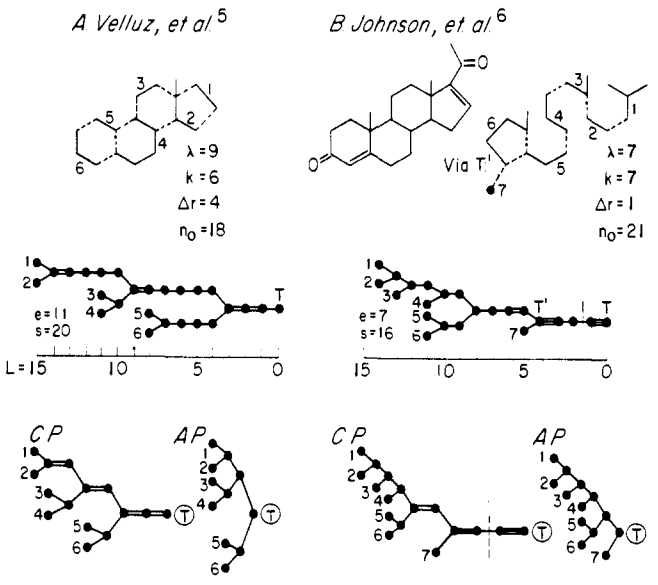



**Figure 1.** Examples of synthetic plans. Computed criteria for each plan in Table IV.

are shown at the bottom of the summary in Figure 1.

**2. Criteria of Economy.** The first requirement is the development of clear quantitative criteria with which to compare various synthesis plans to any given target. The primary basis for such comparisons is one of economy, economy both of time and of materials. The materials may be divided into the substance of the synthons used to construct the target molecule and the amounts of reagent required at each step. A single economy criterion of cost could be developed as the sum of the costs of the time, the starting materials, and the reagents required for a synthesis plan.[2] However, the three criteria are developed separately here, so as to allow the separate nature of each to remain visible and to avoid the approximations implicit in converting each to a common basis of cost.

At the outset we must eliminate the concept of "overall yield", which is necessarily (but implicitly) the yield of target from only one starting material. This is misleading since there are always a number of starting materials used, with different overall yields from each. The idea of overall yield is implicitly connected with a linear synthetic sequence and the first starting material employed on the line. With convergent sequences (see section 3) it is especially meaningless unless particularly defined. In particular the overall yield from one starting material is the inverse of the amount of that material required to yield a unit amount of final product, or target. Thus the nearest related criterion here is that of the total amount of all starting materials required, and it is this value which is most employed in the comparisons developed in the next sections.

The three criteria (starting materials, reagents, and time) are all functions of the nature of the plan, of the yields of the individual reactions, and of the molecular size of the synthons used. In order to compare plans at an early stage before chemical details are fixed, we shall use the number of skeletal carbons ($n_i$) as a measure of the size of synthon $i$ (to include skeletal nitrogen if desired). The actual molecular weight of the synthon is $M_i \simeq Fn_i$ using an average weight-per-carbon factor, $F$, to generalize the unformulated functionality on the synthon skeleton, either as the original starting material or the functionalized synthon as part of a larger intermediate en route to the target molecule of $n_0$ skeletal atoms ($n_0 = \Sigma_i n_i$). In general the number of heteroatoms associated with a synthon tends to decrease as it becomes incorporated in the target skeleton through constructions, but the overall average value

**Table I.** Values for Criteria Computation[a]

| Approx $x^l$ | $x^l$ | $l$ | $S_l$ | Approx $S_l$ | $Q_l$ | Approx $Q_l$ |
|---|---|---|---|---|---|---|
| 1.25 | 1.25 | 1 | 1.25 | 1 | 0 | 0 |
| 1.5 | 1.56 | 2 | 2.81 | 3 | 1.25 | 1 |
| 2 | 1.95 | 3 | 4.76 | 5 | 4.06 | 4 |
| 2.5 | 2.44 | 4 | 7.20 | 7 | 8.82 | 9 |
| 3 | 3.05 | 5 | 10.25 | 10 | 16.02 | 16 |
| 4 | 3.81 | 6 | 14.06 | 14 | 26.27 | 26 |
| 5 | 4.77 | 7 | 18.83 | 19 | 40.33 | 40 |
| 6 | 5.96 | 8 | 24.79 | 25 | 59.16 | 59 |
| 7.5 | 7.45 | 9 | 32.24 | 32 | 83.95 | 84 |
| 9.5 | 9.31 | 10 | 41.55 | 42 | 116.19 | 116 |
| 11.5 | 11.64 | 11 | 53.19 | 53 | 157.7 | 158 |
| 14.5 | 14.55 | 12 | 67.74 | 68 | 210.9 | 211 |
| 18 | 18.19 | 13 | 85.93 | 86 | 278.7 | 279 |
| 23 | 22.7 | 14 | 108.67 | 109 | 364.6 | 365 |
| 28 | 28.4 | 15 | 137.1 | 137 | 473.3 | 473 |
| 36 | 35.5 | 16 | 172.6 | 173 | 610.4 | 610 |
| 44 | 44.4 | 17 | 217.0 | 217 | 783.0 | 783 |
| 56 | 55.5 | 18 | 272.6 | 273 | 1000.0 | 1000 |
| 69 | 69.4 | 19 | 342.0 | 342 | 1272.6 | 1273 |
| 87 | 86.8 | 20 | 428.7 | 429 | 1614.6 | 1615 |

[a] Computed using 80% yield ($x = 1/y = 1.25$) with equations for $S_l$ and $Q_l$ shown in Table II: $x^0 = 1, S_0 = 0, Q_1 = 0, S_{l+1} = S_l + x^{l+1}$. $Q_{l+1} = Q_l + S_l$.

of $F$ is taken as a constant here since it is expected roughly to cancel in the comparison of two complete plans. As an example, acetyl chloride ($M_i = 78$) or methyl acetate ($M_i = 74$) used in an acylation construction yields an acetyl synthon ($M_i = 43$) on a growing skeleton; the value of $n_i = 2$ so that $F = 39, 37$, and 22, respectively. The factor $F$ generally varies from about 14 to 40 in simple starting materials but is less variant and nearer the lower figure as the intermediates grow toward the target.[4]

The variation in molecular weight of reagents is much more, so that the reagent criterion is left in molar rather than weight terms below. An examination of cost per mole for a number of common reagents also yielded no commonality, variations of more than a factor of 10 being common. When the true materials criteria are to be calculated for a specific known synthetic plan, the actual molecular weights of starting materials and reagents may be reincorporated into the formulas developed below.

The yields of individual steps cannot be predicted but they can be averaged in a similar way for overall comparisons of whole routes, in the expectation that variations will thus average and cancel in comparison. A standard value for average yield was taken as 80% so that $y = 0.8$ and its inverse, $x = 1/y = 1.25$ and these values were used to compute Table I. The assumption is based on the idea that application of mechanistic theory to trial synthetic reactions should generally allow development of detailed procedures to provide at least 80% yield in most cases. The table can easily be calculated for any average yield, of course. Reactions known to offer only lower yields, because of poor regio- or stereocontrol or because of unavoidable side reactions, can simply be incorporated as an appropriate number of extra steps. Thus, a reaction of 50% ($x = 1/y = 2.0$) is equivalent to three steps at 80% ($x = 1.25; x^3 = 1.95 \simeq 2.0$). Rounded values of $x^l$ and $S_l$ suitable for quick hand calculation are appended in Table I. Variations in yield from the average do not change the calculations much. Equal variations of ±20% over ten linear steps lower the overall yield, $Y_0 = y^{10}$, to $0.82Y_0$, i.e., lowered somewhat from that obtained with all step yields = 80%. To maintain an average yield of 80%, individual yields must vary equably from 64 to 100%.

The following paragraphs develop the separate criteria for

**Table II.** Criteria for Comparing Synthetic Plans

| | | |
|---|---|---|
| Actual synthesis | Starting materials weight | $W = \Sigma_i n_i x^{l_i}$ |
| | Reagents, mol | $R = \Sigma_i (S_{l_i} - S_{l_i'})$ with $l_i' = 0$ |
| | Total weight manipulated | $\mathrm{TW} = \Sigma_i n_i S_{l_i}$ |
| | Time | $T = s\left(\frac{\mathrm{TW}}{s}\right)^z$ |
| Development trials | Starting materials weight | $W_D = \mathrm{TW} - W + n_0$ |
| | Total weight manipulated | $\mathrm{TW}_D = \Sigma_i n_i Q_{l_i}$ |
| | Time (trials) | $T_{DT} = s$ |
| | Time (production) | $T_{D_p} = s\left(\frac{\mathrm{TW}_D}{s}\right)^z$ |
| Terms | $n_i$ = no. of carbons in synthon ($n_0 = \Sigma_i n_i$ = target size)<br>$x = 1/y$ where $y$ = average yield/reactions<br>$l_i$ = rank of starting materials $i$<br>$s$ = no. of steps in synthesis<br>$z$ = upscaling factor<br>$\sum_i = \sum_{i=1}^{k}$ where $k$ = no. of synthons<br>$S_l = \sum_{v=1}^{l} x^v$ $(S_0 = 0)$<br>$Q_l = \sum_{v=1}^{l-1} S_v$ $(Q_1 = 0)$ | |
| Extra steps added ($e$) | Increase in starting material weight for $k'$ involved synthons | $\Delta W = \sum_{i=1}^{k'} n_i x^{l_i}(x^e - 1)$ |
| | Increase in total weight for $k'$ involved synthons | $\Delta \mathrm{TW} = \sum_{i=1}^{k'} n_i (S_{l_i+e} - S_{l_i})$ |

comparing synthetic plans; some of the detailed derivations are provided in Appendix A and the final equations for each are collected in Table II in a form for the comparison of two plans by ratio.

**(a) Materials.** The weight of starting material of molecular weight $M_i$ required to carry synthon $i$ through a synthetic path length of $l_i$ steps to one mole of target is $W_i = M_i x^{l_i} = F n_i x^{l_i}$. The total weight ($W_0$) of required starting materials for a synthesis using $k$ components or synthons is given in eq 1, as well as the relative weight ($W = W_0/F$) for comparison purposes in which $F$ cancels as constant; the latter is the form used in the summary Table II.

$$W_0 = F \sum_{i=1}^{k} n_i x^{l_i} \quad \text{or} \quad W = \sum_{i=1}^{k} n_i x^{l_i} \tag{1}$$

**(b) Reagents.** In any sequence involving $s$ steps there are $s$ reaction products, numbered $j = 1 \rightarrow s$, each one at a linear path length of $l_j$ steps from the final product, or target. Assuming a 1:1 stoichiometry the total moles of required reagent is

$$R = \sum_{j=1}^{s} x^{l_j+1}.$$

It is more useful, however, to develop a sum based on the synthons $i$ rather than the steps or intermediate products $j$. To this end the intermediates $j$ are all arbitrarily assigned to individual synthons $i$ which pass through them on a direct line to the target. The first synthon ($i = 1$) initiates a linear sequence (the *main line*) incorporating all intermediates on its line to the target. In a simple linear synthetic plan, this includes all of the intermediates. In a more convergent plan other synthons initiate other linear subsequences, or sublines, including intermediates not incorporating the first synthon and ending at the juncture with the main line or even with other sublines already assigned to other starting materials. Thus each subline begins at rank $l_i$ and ends at rank $l_i'$. the rank of its last independent intermediate before its juncture with a prior line. The end rank of the mainline is $l_1' = 0$, i.e., at the target itself (rank = 0). Each intermediate $j$ is thus assigned to only one synthon $i$ which passes through it. This concept of separate linear sequences or lines is illustrated graphically in Figure 1.

For any simple linear sequence of length $l$ the sum of the (exponentially increasing) molar amounts used in each sequential step is conveniently summarized as

$$S_l = \sum_{v=1}^{l} x^v$$

and is calculable for any linear sequence length, $l$, and average yield, $y(x = 1/y)$. These sums refer either to molar amounts of synthon material required in all the steps or to those of reagents. Thus the total molar amount of reagents ($R$) required in the total synthetic plan is given by eq 2. In use the equation may be checked by the sum of the $S$ subscripts, which equals the number of steps, $s = \Sigma l_i - \Sigma l_i'$.

$$R = \sum_{i=1}^{k} (S_{l_i} - S_{l_i'}) \text{ where } l_1' = 0 \text{ and } S_0 = 0 \tag{2}$$

**(c) Time.** The effort or work involved in executing a synthetic plan is a function of the number of steps ($s$) and the net effective time for each. Even though it is clear that some reactions are harder or more time consuming than others, it will be assumed that there is a single standard net effective time, $T_0$, for all reactions and that in a comparison of two synthetic plans the time averaging involved will cancel. There is, however, an upscaling factor, $u$, for the several reactions which is dependent on the quantity of material which must be manipulated in each reaction, i.e., the time required to carry out a chemical operation increases with the weight of material used. Thus the total time becomes $T = sT_0u$, for $s$ steps. The concept was employed by Powers,[3] who found that the relation for nucleotide and nucleic acid synthesis was $T \propto W^z$ with an average exponential increase of $z = 0.3$, or a doubling of time for a tenfold increase in manipulated weight. In the absence of other published values for the net effective times of chemical operations, the exponential form of the upscaling factor and the exponent $z = 0.3$ were accepted for the calculations here.

The development of a formula for time then requires the weight of each synthon each time it is part of an intermediate used in a reaction, in order to compute this upscaling factor. This requires the total weight manipulated: each synthon starting weight times the number of steps it goes through, de-

Table III. Economy Comparisons of Linear and Convergent Affixation Plans[a]

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | $L_k = \frac{k(k+1)}{2} - 1$ | | | | | | | | | | | | $L_k = k(m+1) - 2^m$ [note: $I_{2k} = 2(L_k + k)$] | | | | | |
| Linear | | | | | | | | | | | | | $W = n(S_{k-1} + x^{k-1})$ | | | | | | Fully convergent | | | | $W = n(Ax^m + Bx^{m-1})$ | | | | | |
| | | | | | | | | | | | | | $TW = n(Q_k + S_{k-1})$ | | | | | | | | | | $TW = n(AS_m + BS_{m-1})$ | | | | Where | $2^{m-1} < k \leq 2^m$ |
| | | | | | | | | | | | | | $TW_D = n\sum_{\nu=1}^{k-1} Q_\nu + Q_{k-1}$ | | | | | | | | | | $TW_D = n(AQ_m + BQ_{m-1})$ | | | | | $A = 2k - 2^m$ |
| | | | | | | | | | | | | | $T = (k-1)\left(\frac{TW}{k-1}\right)^z$ | | | | | | | | | | $T = (k-1)\left(\frac{TW}{k-1}\right)^z$ | | | | | $B = 2^m - k$ |
| | | | | | | | | | | | | | $R = S_{k-1}$ | | | | | | | | | | $R = \sum_{\nu=1}^{m} 2^{\nu-1}x^\nu - Bx^m$ | | | | | $k = A + B$ |
| | | | $L_k$ | | | $W$ | | | $R$ | | | $TW$ | | | $T$ | | | $W_D$ | | | $TW_D$ | | | $T_{D_p}$ | | |
| $k$ | $m$ | (Con) $A$ | $B$ | Lin | Con | Ratio | Lin | Con | Ratio | Lin | Con | Ratio | Lin | Con | Ratio | Lin | Con | Ratio | Lin | Con | Ratio | Lin | Con | Ratio | Lin | Con | Ratio |
| 4 | 2 | 4 | 0 | 9 | 8 | 1.1 | 7 | 6 | 1.1 | 5 | 4 | 1.1 | 14 | 11 | 1.2 | 5 | 4 | 1.1 | 11 | 9 | 1.2 | 9 | 5 | 1.9 | 4 | 3 | 1.2 |
| 6 | 3 | 4 | 2 | 20 | 16 | 1.2 | 13 | 11 | 1.2 | 10 | 8 | 1.2 | 36 | 25 | 1.5 | 9 | 8 | 1.1 | 29 | 20 | 1.5 | 46 | 19 | 2.5 | 10 | 7 | 1.3 |
| 8 | 3 | 8 | 0 | 35 | 24 | 1.5 | 24 | 16 | 1.5 | 19 | 12 | 1.6 | 78 | 38 | 2.0 | 14 | 12 | 1.2 | 62 | 30 | 2.0 | 137 | 32 | 4.2 | 17 | 11 | 1.5 |
| 10 | 4 | 4 | 6 | 54 | 34 | 1.6 | 40 | 21 | 1.8 | 32 | 17 | 1.9 | 148 | 57 | 2.6 | 21 | 16 | 1.3 | 119 | 46 | 2.6 | 324 | 60 | 5.4 | 26 | 16 | 1.7 |
| 12 | 4 | 8 | 4 | 77 | 44 | 1.8 | 65 | 27 | 2.4 | 53 | 22 | 2.4 | 264 | 77 | 3.4 | 29 | 20 | 1.5 | 211 | 61 | 3.4 | 671 | 87 | 7.7 | 38 | 20 | 1.8 |
| 16 | 4 | 16 | 0 | 135 | 64 | 2.1 | 166 | 39 | 4.2 | 137 | 32 | 4.3 | 748 | 115 | 6.5 | 48 | 28 | 1.8 | 598 | 92 | 6.5 | 2314 | 141 | 16.4 | 68 | 29 | 2.3 |
| 20 | 5 | 8 | 12 | 209 | 88 | 2.4 | 411 | 54 | 7.7 | 342 | 44 | 7.8 | 1957 | 168 | 11.6 | 76 | 36 | 2.1 | 1566 | 135 | 11.6 | 6780 | 234 | 29.0 | 111 | 40 | 2.7 |

[a] Weight equations imply synthons of equal size, $n = n_0/k$; tabulated calculations for $n = 1$ ($k = n_0$); values rounded but ratios computed from full values.

creased by the yield loss each time, and these summed over all synthons $i = 1 \rightarrow k$. As with the reagents the derivation must convert from a sum over the number of steps ($j$) to a sum over the number of synthons ($i$). The formula for time so developed in Appendix A thus depends on the total weight of manipulated material (TW) in all steps, and the formulas for TW and for time ($T$) are shown in Table II.

**(d) Development Trials.** The criteria above apply to completed or projected complete syntheses as executed to provide 1 mol of target. In considering a synthesis plan in advance of execution it is important to provide some estimate of the time and material required to try out and assure success (i.e., 80% yield) for each synthetic step. These are the step trials and are taken to require $n$ trials, on $m$ moles and $t$ time for each such trial. Thus the total moles of material needed for $s$ steps is ($snm$) and will require ($snt$) time to try out. The second requirement will be the actual execution of the established steps to produce the ($nm$) molar quantities needed just to try out each of the subsequent steps, i.e., to bring up adequate quantities of the intermediates for the next step trials. In this development trial work, however, no significant amount of target is produced.

The material required of synthon $i$ for the last step is ($nmn_i$), for the penultimate step is ($nmn_i + nmn_ix$), i.e., step trials plus production of last step requirements. The next preceding step similarly requires $nmn_i$ ($1 + x + x^2$), and so on to the requirements on the starting material itself, at rank $l_i$, which are $nmn_i$ ($1 + S_{l_i-1}$). The weight of starting materials required for development is then $W_{dev} = Fnm(\Sigma_i n_i S_{l_i-1} + n_0)$ or eq 3.

$$W_{dev} = Fnm\left[\sum_{i=1}^{k} n_i(S_{l_i} - x^{l_i}) + n_0\right] = Fnm[TW - W + n_0] \quad (3)$$

The time required for step trials is ($snt$) while that for production of intermediates for the trials depends as before on the total manipulated weight for all these production steps. This is developed in a parallel way in Appendix A and the final form collected in Table II.

The formulas for the several criteria in Table II are easily applied by hand to any synthetic plans developed for a target molecule, using the rounded values in Table I for quick calculation. Values for different plans are then compared by ratio. In the following discussion the single criterion of required starting material weight ($W$) is the only one generally used. This most nearly approximates (though inversely) the "overall yield" idea, and it appears to be quite general that the other criteria roughly parallel $W$ when comparing different plans (see Table III).

These comparisons are intended to focus on aspects of synthetic plans which are independent of chemical detail, i.e., which are a function only of the sequence of events used, of synthon size and path length and convergency. This not only allows comparison of plans prior to execution but also leads to useful rules for target dissection and construction order in synthesis design.

The two examples of plans in Figure 1 were computed for the several criteria and the results shown in Table IV. The Johnson plan[6] (B) was conceived from the monocyclic subtarget shown, found at rank $l = 4$ on the plan, which then undergoes a triple cyclization to $l = 3$ and is cleaved and recyclized to dehydroprogesterone. The Velluz plan[5] (A) creates a target of $n_0 = 18$ in 20 steps from $k = 6$ synthons; the Johnson plan makes $n_0 = 21$ in only 16 steps from $k = 7$ synthons, but the criteria are nevertheless somewhat more favorable in the former, largely because in the latter there are more uninvolved synthons exposed to yield loss in the later reactions. If

**Table IV.** Computed Criteria for Figure 1

| Plan A | | | | | | | | | | Plan B | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | $n_i$ | $l_i$ | $x^{l_i}$ | $n_i x^{l_i}$ | $S_{l_i}$ | $n_i S_{l_i}$ | $l_i'$ | $S_{l_i'}$ | $\Delta S$ | $i$ | $n_i$ | $l_i$ | $x^{l_i}$ | $n_i x^{l_i}$ | $S_{l'}$ | $n_i S_{l_i}$ | $l'_i$ | $S_{l'_i}$ | $\Delta S$ |
| 1 | 2 | 15 | 28 | 56 | 137 | 274 | 0 | 0 | 137 | 1 | 4 | 14 | 23 | 92 | 109 | 436 | 0 | 0 | 109 |
| 2 | 4 | 15 | 28 | 112 | 137 | 548 | 15 | 137 | 0 | 2 | 2 | 14 | 23 | 46 | 109 | 218 | 14 | 109 | 0 |
| 3 | 2 | 11 | 11.5 | 22 | 53 | 106 | 10 | 42 | 11 | 3 | 3 | 13 | 18 | 54 | 86 | 258 | 13 | 86 | 0 |
| 4 | 5 | 11 | 11.5 | 58 | 53 | 265 | 11 | 53 | 0 | 4 | 3 | 11 | 11.5 | 35 | 53 | 159 | 11 | 53 | 0 |
| 5 | 2 | 8 | 6 | 12 | 25 | 50 | 4 | 7 | 18 | 5 | 3 | 11 | 11.5 | 35 | 53 | 159 | 9 | 32 | 21 |
| 6 | 3 | 8 | 6 | 18 | 25 | 75 | 8 | 25 | 0 | 6 | 5 | 11 | 11.5 | 58 | 53 | 265 | 11 | 53 | 0 |
| | | | | | | | | | | 7 | 1 | 5 | 3 | 3 | 10 | 10 | 5 | 10 | 0 |
| $\Sigma =$ | 18 | 68 | 90 | 278 | 430 | 1318 | 48 | 264 | 166 | $\Sigma =$ | 21 | 79 | 102 | 323 | 473 | 1505 | 63 | 343 | 130 |
| | ($n_0$) | ($L_k$) | | ($W$) | | (TW) | | | ($R$) | | ($n_0$) | ($L_k$) | | ($W$) | | (TW) | | | (R) |
| | | | | $T = 70.2$ | | | | | | | | | | $T = 62.5$ | | | | | |

the Velluz plan criteria are corrected by $7/6 = 21/18$, the ratio of synthons and target size between the two, then the values of $W$, $TW$, and $L_K$ (see below) become nearly identical, as do the reagent moles, $R$, if corrected by the 16/20 ratio of steps in the two.[7] The syntheses shown require 166 and 130 mol of reagent, respectively, to make 1 mol of target and 90–102 mol of starting materials ($\Sigma x^{l_i}$). The Johnson synthesis apparently actually requires 100 g of synthon 1 to create 10 g of target, which works out at 75% average yield instead of the 80% assumed in these calculations, but it is likely that more yield optimization is possible.

Finally the criteria can be graphically demonstrated in a weight chart with synthon size ($n_i$) plotted vertically and cumulatively and rank plotted horizontally but on a scale of $S_l$ as shown in Figure 2 for the Velluz steroid synthesis (A/Figure 1).[5] Each block then represents the weight used in the reaction at that rank. The initial blocks for each starting material show the required starting weight, $W_i$, while the entire area is TW, the total weight manipulated, exponentially related to the effective time required. A double line separates reactions in different lines in a convergent synthesis and allows a visual assessment of the relative manipulated weights in each line. The molar reagent requirement, $R$, is a linear length ($S_{l_i} - S_{l_i'}$) on the horizontal (molar) scale, one for each converging line, here being three lines, $15 \rightarrow 0$, $11 \rightarrow 10$, and $8 \rightarrow 4$, shown below the chart. The weight chart is another presentation of the plan graph of Figure 1.

**3. Convergency.** The concept of the convergent synthesis was introduced by Velluz et al.[8] In a convergent plan various parts of the target molecule are assembled separately and independently and then linked together afterwards near the end of the synthesis. Compared to a simple linear sequence, this has the effect of lowering the path length of the main line and raising the path length of some other synthons, but because of the exponential involvement of path length, $l$, in the criteria the overall effect is to make the convergent plan more economical than the simple linear sequence of the same number of synthons and steps. The qualitative basis for this economy lies in the idea that when a reaction is carried out on an intermediate, it usually involves only one or two of the synthons that make up the intermediate so that the other, uninvolved synthons comprising the intermediate are subjected to needless waste from yield loss in the reaction. Indeed the functionality present on the uninvolved synthons may contribute to yield loss through unwanted side reactions. Also the manipulated weights per step (TW), and hence the time required, are lower in a convergent plan in which fewer uninvolved synthons are carried along in each step (this is true even with 100% yields).

The key reactions to convergency are just the affixations, which join the pieces together. These are brought out most clearly by reducing the plan to an affixation plan (AP), allowing us to focus on the essential nature and efficiency of convergent plans, with examples illustrated in Figure 3. First

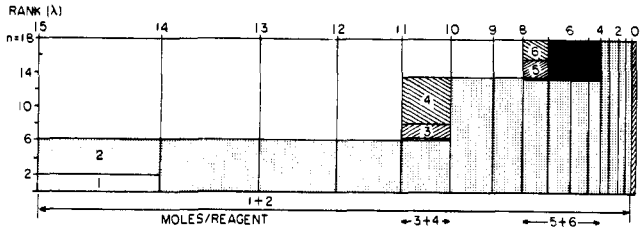



**Figure 2.** Weight chart for $C_{18}$-steroid synthesis.

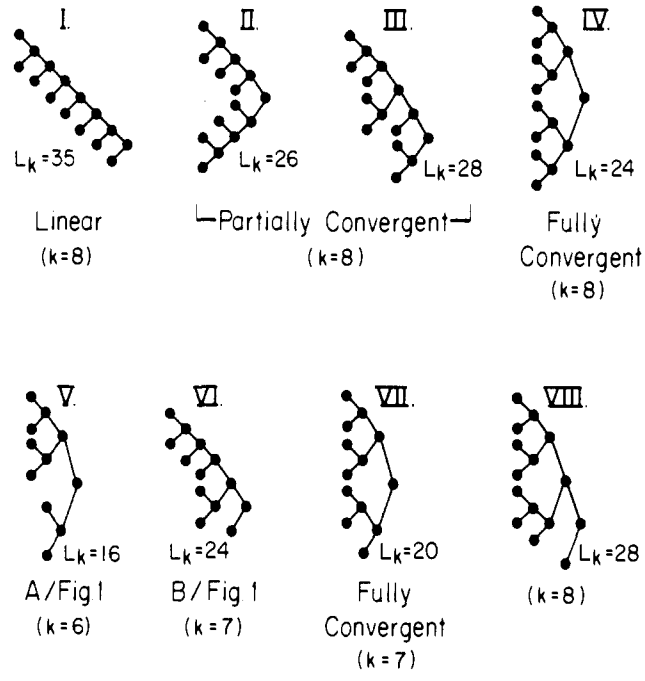



**Figure 3.** Affixation plans.

it is clear that (for $k > 4$) there are not only the extremes of a linear and a fully convergent plan but also a continuum of partially convergent plans between with intermediate efficiency. The Velluz plan (A in Figure 1) is one of two equivalent fully convergent affixation plans for $k = 6$ while the Johnson plan (B in Figure 1) is only partially convergent for $k = 7$. The number of possible affixation plans for $k$ synthons ($N_k$) is the number of possible trees with $k$ termini (points of degree 1), shown in Appendix B. Thus there are 23 possible plans for $k = 8$ synthons, of which four are illustrated in Figure 3.

Perfect convergency can be achieved only with $k = 2^m$ synthons, where $m$ = main line path length ($= l_1$; in the perfect convergent all path lengths $l_i = m$). In other fully convergent plans the main line, $m$, is defined by $2^{m-1} < k < 2^m$. A convenient index of the extent of convergency is the sum of all $k$

path lengths, i.e.,

$$L_k = \sum_{i=1}^{k} l_i$$

which records the sum of the reactions to which each synthon is subjected. (Equations for $L_k$ for affixation plans are given in Table III.) Since each affixation necessarily involves two synthons, then, for affixations only, the sum of uninvolved synthons in all affixations is $L_k - 2(k-1)$, and this is another measure of convergency. Thus the total of uninvolved synthons in all affixation reactions is readily found for comparing linear and fully convergent affixation plans for $k$ synthons. For $k = 6$ the ratio is 10/6 while for $k = 8$ it is 21/10, or more than twice as many synthons carried along uninvolved in a linear than in a fully convergent plan; for $k = 12$ the ratio is 2.5 times. In very large synthetic problems like that of long nucleic acids[3] the differences between linear and convergent plans are striking. For a 100-base nucleic acid chain, $k = 100$, $L_k$ is 5049 for a linear sequence but only 672 for a fully convergent one, and the ratio of uninvolved synthons for the two is just over ten times.

The path length sum $L_k$ is an index of convergency but is not a reliable measure of economy, which is exponentially related to path length (cf. $W$ in Table II). In order to perceive the range of economy variation we may calculate and compare the criteria for linear and fully convergent plans, taking only the affixation steps and assuming synthons of equal size, $n = n_0/k$. For a linear affixation plan the starting material weights become $W = n(x^{k-1} + x^{k-1} + x^{k-2} + x^{k-3} + \ldots + x) = n(S_{k-1} + x^{k-1})$ and the total weight manipulated is parallel, as $\mathrm{TW} = n(Q_k + S_{k-1})$. The required reagents are simply $R = S_{k-1}$ since there is only one synthetic line, the main line, of $l_1 = k - 1$.

The fully convergent synthesis always has only one value of $l_i = m$ for perfect convergents ($k = 2^m$) and only two values of $l_i = m$ and $(m-1)$ for the other fully convergent plans ($2^{m-1} < k < 2^m$). There are $A$ synthons of $l_i = m$ such that $A = 2k - 2^m$ and $B$ synthons of $l_i = m - 1$, numbering $B = 2^m - k$, the total synthons being $k = A + B$. Hence the weight of starting materials is $W = n(AX^m + BX^{m-1})$ and the total manipulated weight is parallel: $\mathrm{TW} = n(AS_m + BS_{m-1})$. The reagent moles are most easily derived from the number of intermediates (1, 2, 4, 8, . . .) at each rank in the next higher perfect convergent minus those missing from rank $m$ in the lines starting from the $B$ starting materials (of rank $m-1$). In this way

$$R = \sum_{v=1}^{m} 2^{v-1}x^v - Bx^m$$

Values for the three criteria (as well as for $L_k$) are tabulated in Table III, for both linear and fully convergent affixation plans with varying numbers of equal-size synthons. There is a continuously increasing difference between the linear and convergent mode as the number of synthons increases (at $k = 3$ they are the same) and this is reflected in increasing economy by each criterion (which increase also at similar rates), shown by the growing ratios of linear to convergent. The size of synthons used in existing syntheses[4] averages around $n_i = 4$, but if no aromatic rings are present to afford easy $n \geq 6$ synthons, the average $n_i < 3$. Thus the number of synthons, $k$, is 6–10 for traditional complex molecules ($n_0 = 18$–30). In this range the fully convergent mode is more economical than the linear by about 50%, with partially convergent plans in between. This economy refers only to the affixation steps and will be much greater when the other steps are entered (see below).

With synthons of disparate sizes the comparisons change. A very large synthon contributes heavily to the weight sums and so its rank is more critical, a low rank strongly favoring economy. Thus a large synthon can have a lower rank by being affixed last in a linear synthesis ($l = 1$) than it can have in a convergent synthesis ($l = m$ or $m-1$). However, even with one synthon alone equal to half the target size, for $k = 8$ synthons the convergent synthesis is still more economical than the linear plan with the large synthon affixed last, but the difference is now small ($W_{\mathrm{lin}}/W_{\mathrm{con}} = 1.12$, down from 1.51). The most economical plan in such a case is a partially convergent synthesis in which all the small synthons are joined first in a fully convergent mode and then finally affixed to the single large one (plan VIII, Figure 3). Although $L_k = 28$, larger than 24 for fully convergent (plan IV), the large synthon has rank $l = 1$ rather than 3 and $W_{\mathrm{VIII}}/W_{\mathrm{IV}} = 0.92$. In most cases synthons are not so disparate in size and the fully convergent mode is the most economical for affixations.

When cyclizations and refunctionalizations are added to the affixation plan, other changes will occur in the economy criteria depending on the placement of these added steps in the plan. For the addition of $e$ extra steps at a point that involves only $k'$ of the synthons, the additional starting material weight required will be a function both of the total size of the $k'$ synthons in the intermediate undergoing the $e$ extra steps as well as their starting ranks. The relation is given at the bottom of Table II, as well as the parallel one for $\Delta\mathrm{TW}$. This implies that cyclizations should be carried out as soon as the involved synthons are affixed (before extra uninvolved synthons have been added), and that the involved synthons should be themselves of the lowest possible starting rank.

For refunctionalizations the same equations for $e$ extra steps imply that a refunctionalizing preparation of a single starting material is much more efficient before it is initially affixed than later in the plan, and that functionality alteration at the end of the plan is the worst of all. Indeed for the latter, $\Delta W = W(x^e - 1)$, which implies a doubling of $W$ for three extra steps at the end (after the last construction). This may be compared with three extra steps executed on starting materials before construction in a perfect-convergent example of target size $n_0 = 24$ and $k = 8$ synthons of equal size ($n_i = 3$), for which $W = 8(3x^3) = 48$ and $\Delta W = W(x^3 - 1) = 48$ for final refunctionalization. Initial refunctionalization is only $\Delta W = 3x^3(x^3 - 1) = 6$. Similarly, $\Delta\mathrm{TW} = 27$ for initial and 216 for final refunctionalization of $e = 3$ extra steps.

A "total synthesis" of any target is a plan with one-carbon starting materials, i.e., the classical idea of synthesis from "coal, air, and water". In such plans, $n_i = 1$ and $k = n_0$, and all skeletal bonds of the target are constructed (the bondset is all skeletal bonds). There are many possible affixation plans for total syntheses, ranging from linear to fully convergent, and these plans define the order of constructions of the bonds. The calculations of Table III apply to total syntheses although the lin/con ratios shown are general for all cases of equal-size synthons.

The total synthesis, like any fully convergent plan of $k < 2^m$, is merely truncated from the next higher perfect convergent plan (for $k = 2^m$) by removing various pairs of starting points. This is illustrated in Figure 4 for the steroid of $n_0 = 21$, a total synthesis (solid lines) shown as part of the next higher perfect convergent plan of $m = 5$ or $2^m = 32$. The affixation plan is shown but circles are placed around the earliest possible points for cyclization in this example, for expanding AP to CP with minimum $W$. In expanding AP to CP each circle increases by one the rank of synthons leading into it.

In the simple affixation plan (without circles) there are $A$ starting points of rank $m = 5$ and the remaining $B = k - A$ synthons starting at rank $m - 1 = 4$, with the lines removed from the perfect plan shown as dotted. For the total synthesis $A = 2n_0 - 2^m$ ($A = 10$ here for $n_0 = 21$) and $B = 2^m - n_0 = 11$. The numbers $A$ and $B$ at each rank are fixed by $n_0$ but their relative location on the plan graph varies with different convergent total syntheses. However, all such fully convergent

total syntheses have the same criteria (Table III).

These equivalent, optimal syntheses differ in the pairs of points truncated from rank $m = 5$ so that the two "halves" ($\alpha/\beta$) at $l = 1$, finally joined to create the target, may vary from $n = 8$–16. Similarly the four intermediates at $l = 2$ may vary from $n = 4$–8 without any change in overall economy, and each of these intermediates itself always exhibits a fully convergent affixation plan as its own subsynthesis. The particular dissection shown is one of $\alpha/\beta = 12/9$, so that of $A = 10$ carbons starting at $m = 5$ there are eight in the $\alpha$ half and only two in the $\beta$ half.

Any actual synthesis plan is truncated from some total synthesis plan by accepting as given, or preconstructed, available starting materials of $n$ carbons which are seen as intermediates in the corresponding total synthesis plan. These actual syntheses are then partial synthesis plans, the one illustrated in Figure 4 starting from starting materials of $n$ = 2–4, all in rank $l = 3$. Because some starting materials are given, there is less construction to do, and so the partial synthesis plan is of course more economical than the corresponding total synthesis plan.

**4. Consequences for Synthesis Design.** However various synthetic plans for a certain target may be derived, the economy criteria can always be profitably used to compare their relative potential. Beyond this straightforward use in evaluation, however, general conclusions can be drawn which lead to heuristics useful in the design process itself. The traditional approach to synthesis design is one of functionality dissection,[9] by which the target functionality guides the choice of successive synthetic reactions in a retrosynthetic direction. This "stepwise-backwards" approach suffers from not revealing the required starting materials until the end of the process, and from the many individual yield-predictive decisions made en route to prune the rapidly growing tree of choices.

An alternative to the stepwise approach has recently been advanced.[1c] This fundamentally different approach seeks to start with both target and starting materials together, to collect rough clusters of whole routes linking them, only loosely defined at first, then to refine them in stages, selecting without yield prediction. The protocol starts with consideration of *bondsets* (sets of λ target bonds to be constructed, as in Figure 1) which in turn define synthon skeletons, then considers the *order* of construction in the bondsets. This is largely a skeletal dissection rather than a functionality one and yields a plan of successive constructions without any functionality detail. Then a search is made mechanically for the requisite functionality (still only partially or broadly defined) to place on the synthon skeletons so as to achieve sequences of constructions not requiring any intervening refunctionalization (self-consistent sequences).

These sequences define a "natural" functionality resulting on the intermediates and the target skeleton (and only sequences resulting in or near the actual target functionality are accepted), and at the same time they define the actual corresponding starting materials required. In this approach then final plans are reached through successive refinement of whole routes rather than stepwise accretion of single steps back from the target. The conception is based on a search for ideal syntheses for a given bondset. This is carried out by examining the central construction phase of the synthesis first in order to define all self-consistent sequences, and their requisite functionality, to do all constructions among the given synthons without intervening refunctionalization. This then defines the functionality of the necessary starting materials for each of these self-consistent sequences and allows choices based on available starting materials not requiring preliminary functional preparation. At the same time the choice of sequences is also further limited to those minimizing final refunctionalization, i.e., those arriving directly (or nearly directly) at the exact target structure.

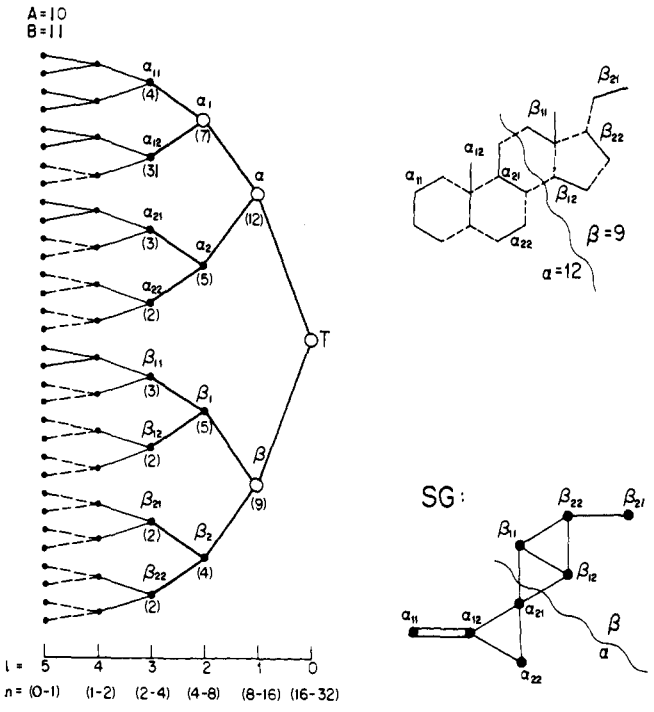



**Figure 4.** Partial and total syntheses from perfect convergent plan. Solid lines = total synthesis; heavy lines = partial synthesis; circle points = earliest place for cyclizations; numbers in parentheses = sizes ($n$).

What remains is the need to extract optimal bondsets from the target skeleton itself in order to set this selection process in motion for each one. It is here that the present criteria of economy can be early applied not only to evaluate various bondsets but also to direct the actual selection of particular optimally convergent bondsets.

**(a) Bondset Selection by Convergency.** The most powerful dissection tool to drive from these economy criteria is this: *full convergency defines whole bondsets and their construction order.*[10] Furthermore, only a relatively small number of bondsets satisfy full convergency and these are readily derived.

The derivation is designed to yield a fully convergent total synthesis if carried to completion, i.e., to synthons of $n = 1$. However, by accepting larger available synthons, the derivation leads to partial, but still fully convergent, syntheses which are contained in and truncated from the plan of the corresponding total synthesis.[11]

In the general convergent plan (IV, Figure 3) for common complex targets of $n_0 = 16$–32 there is a primary division into two "halves" ($\alpha$ and $\beta$) at $l = 1$, of $n = 8$–16 each and these in turn are each divided to leave four units of $n = 4$–8 at $l = 2$. These intermediate units are then further subdivided to leave eight starting materials at $l = 3$ of $n = 2$–4. Starting materials of this size are considered to be always available, primary sources; indeed in present practice the average starting material for nonaromatic targets[4] is $n \simeq 3$. Thus no target of $n_0 \leq 32$ needs any affixation path of more than $l = 3$, or more than $k = 8$ starting materials. Since the primary sources are $n = 2$–4, they will virtually always be acyclic and all target rings must be formed.

The procedure for defining all possible fully convergent bond sets is simple.

1. Divide the target skeleton ($n_0 = 16$–32) all ways such that neither half ($\alpha$ or $\beta$) is less than $n = 8$ or more than $n = 16$ and cutting as few bonds as possible; $\Delta r_1$ = rings cut = bonds cut − 1, and this is minimized.

2. Divide each half again the same way to yield four skeletal units, all of $n = 4$–8; $\Delta r_2$ = total rings cut, minimized as before.

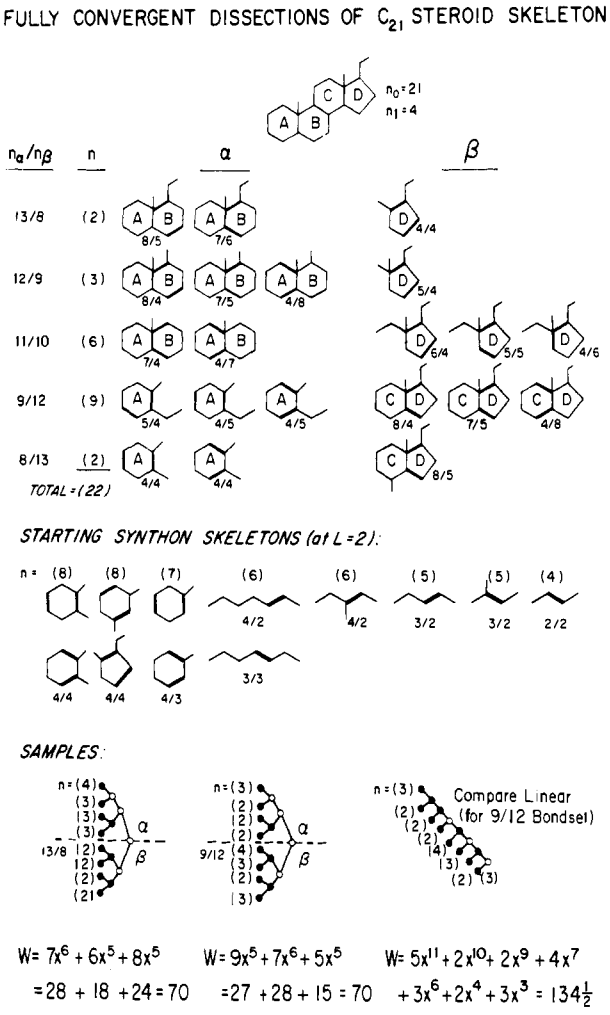



**Figure 5.** Fully convergent dissections of $C_{21}$ steroid skeleton.

3. Divide each of the four units again into two of $n = 2$–4; $\Delta r_3$ = remaining rings cut. $r_0 = \Delta r_1 + \Delta r_2 + \Delta r_3$.

4. Accept only final bondsets from (1)–(3) for which all three operations are successful, $\Delta r_1$ and $\Delta r_2$ are minimized, and $\Delta r_3$ is maximized.

5. These steps define directly all possible construction plans (CP), all bearing the same full convergency and minimum economy criteria.[12]

The number of bondsets and corresponding construction plans (CP) so derived is remarkably few, as may be illustrated by the case of the $C_{21}$-steroid skeleton (Figure 5). The first dissection must yield two intermediates of sizes $n_\alpha/n_\beta = 13/8$, 12/9, and 11/10. There are only five such dissections with minimum $\Delta r_1 = 1$ (cutting only one ring). The next dissection in every case needs to cut only one ring in each half ($\alpha$ and $\beta$), so for $\Delta r_2 = 2$ there result a total of only 22 bond sets to four units each, i.e., 22 four-unit prestructs.[1a] The allowed second dissections of $\alpha$ and $\beta$ are $n = 13 \rightarrow 8/5, 7/6$; $n = 12 \rightarrow 8/4$, $7/5, 6/6$; $n = 11 \rightarrow 7/4, 6/5$; $n = 10 \rightarrow 6/4, 5/5$; $n = 9 \rightarrow 5/4$; $n = 8 \rightarrow 4/4$; and all such cuts are represented in the 22 derived prestructs.

These 22 prestructs from two primary dissections contain only nine different synthon skeletons among them, constructable 12 ways, as shown in Figure 5, to fit rules (3) and (4). The total combinations for the three dissections then result in 45 bondsets: 13/8 (6); 12/9 (6); 11/10 (16); 9/12 (15); 8/13 (2). These are all bondsets of $k = 8$, $\lambda = 11$, and all have the same fully convergent plans with the same $W$ and TW criteria.[12] The stringency of this bondset selection becomes apparent on noting that these 45 prestructs are all in the set $O_8$ on the construction grid of ref 1a (Figure 9),[13] for which there are calculated to be a total of 2 324 084 possible eight-component acyclic prestructs! Many of these possible prestructs, of course, are synthetically valueless combinations with a few very large components and/or many of $n = 1$ (such prestructs cannot have fully convergent plans). The bond set selection here gives only synthons of $n = 2$–4 and of those only the collection which may then be constructed on a fully convergent plan.

The convergent dissection procedure above generates perfect convergent affixation plans all possible ways, with $k = 8$, $m = 3$, and, for affixation only, $W = n_0X^3 = 2n_0$. The rule about cyclizations at highest rank minimizes the size of intermediates involved in cyclizations and so minimizes $W$ for the full CP as well. Points in the AP which are circles indicate intermediates to be cyclized when the AP is expanded to CP. The calculated weights of the CP for these steroid constructions is $W = 70$ (Figure 5) compared to a likely linear variant from the same pieces of nearly twice as much ($W = 134.5$); similarly, TW = 238 (con) and 560 (lin).

The overall affixation plan from the derivation occupies only three ranks (heavy lines in Figure 4) and has synthons there of $n = 2$–4. If larger synthons are known to acceptable starting materials they in turn occupy points on rank $l = 2$ (for $n = 4$–8). When they are taken as given, the synthesis has an advantage of fewer constructions. In particular this occurs with synthons of $n = 4$ which may appear either at ranks $l = 2$ or $l = 3$. Since these are accepted as starting materials at $l = 3$, they must also be acceptable at $l = 2$ (as in $\beta_2$/Figure 4) and so may avoid the further dissection into $n = 2$ (at rank $l = 3$) and save steps. This puts an immediate premium on plans with the maximum number of $n = 4$ at rank $l = 2$, i.e., in 16/22 cases from Figure 5.

**(b) Bondset Evaluation.** Bondsets selected by other criteria, or developed by stepwise dissection, have still a number of possible construction plans, reflecting the various possible orders of construction steps among the synthons represented by any bondset. We can analyze any given bondset to determine the best plan(s) by which it may be constructed.

The target skeleton with an imposed bondset is itself a structure graph showing the synthon skeletons and the links ($\lambda$) between them that must be constructed. It may be reduced to a *synthon graph* (SG) by coalescing each synthon to a single point, leaving the bondset as the lines. Thus the synthon graph has $k$ points and $\lambda$ lines, and $\Delta r$ cycles, where $\Delta r$ = number of cyclizations implied by the bondset.[14] As illustrated in Figure 6, the two-membered cycles in SG represent simple annelations, whereas cycles of more than two represent target rings constructed of more than two synthons. The number of constructions required of each synthon is the degree ($d_i$) of its point in SG, and this in turn is the minimum number of construction steps each synthon must undergo in the construction plan, i.e., the rank of the synthon, $l_i \geq d_i$. Any bondset then implies a minimum value (for CP) of $L_k = \Sigma_i d_i = 2\lambda$, and also of $W = \Sigma_i n_i X^{d_i}$ and $TW = \Sigma_i n_i S_{d_i}$ with which final derived plans may be compared.

The synthon graph is now a reduced skeleton composed of synthons as points, and it may be dissected according to the same procedure above for deriving optimal convergent construction plans. Thus it is successively cut into units of (2–4), (1–2), (0–1) and only those of minimum $\Delta r$, and this must now proceed to total synthesis, i.e., until each synthon point is separated from all others. This will provide optimal convergent plans with earliest cyclization for best economy. These are illustrated in Figure 6 below the synthon graphs of the three $C_{18}$-steroid skeleton cases. The plans are shown as AP with circles to indicate points of cyclization for expanding to CP. The weight values are calculated for the full resultant CP (ranks = lines + circles to target). The first case is the simple Torgov synthesis (A), a perfect convergent on $k = 4$ synthons,

derived from SG by dividing first into two halves of two synthons each. The use of the larger synthon ($n_o$ = 10, naphthalene skeleton) here is a major factor in reducing the bond set and improving economy. However, the same use of $n$ = 10 synthon for the B/C ring (case B) cannot be convergent and leads only to linear plans of poorer economy. This case illustrates that some SG, like some full structures, are incapable of convergent plans;[11] such SG are further discussed in Appendix B.

The same skeleton is also shown as case C utilizing the same bondset as illustrated in Figure 1A. Both affixation plans exhibit $W$ = 33.5 but the expansion of AP to CP in Figure 6, derived by the convergency dissection procedure above, is more economical than the CP shown in Figure 1A because of more efficient placement of cyclizations. The latter shows $W$ = 70 while the Figure 6 dissection is $W$ = 57. For comparison a linear sequence of the same synthons (taken in numerical order) with earliest possible cyclization gives $W$ = 85, typically about 1.5 times the weight of the best convergent plan.

When comparing different plans it is useful to have available some standards of comparison to indicate the range of possible criteria. The best available fully convergent plans for both total and partial synthesis ($l$ = 3; $n$=2–4; $k$=8) are readily obtained from the dissection procedure above, with the best placement of cyclizations. The comparable linear options are usually at least 50% higher but hard to generalize except for equalsize synthons as in Table III (see a particular comparable linear plan calculated for the $n_0$ = 21 steroid at the bottom of Figure 5). The worst placement of cyclizations is at the end and easily computed for any plan by multiplying the affixation value of $W$ by $x^{\Delta r}$. There is also special economy in using the largest synthons possible so as to reduce both $k$ and $\lambda$, and these are best placed at lowest rank, i.e., added last to a convergent assembly of the others if possible. In serious consideration of possible plans, of course, it will be important to examine not only the criterion of starting weight ($W$) but also those of reagent quantities and time, as well as the time and material required for development of the synthesis, as summarized in Table II.

The dissection outlined above allows derivation of the convergent construction plans of optimal economy before any details of actual necessary functionality are considered. This derivation yields a sequence of construction events on the synthon skeletons, onto which may now be placed such functionality as will yield a self-consistent sequence, or ideal synthesis.[1c] Ideal syntheses are rare in practice and so there usually arises the need for refunctionalization, adding $e$ extra steps and swelling $W$. It is the addition of refunctionalization steps, however, that rapidly destroys the economy of any construction plan. In general this ballooning of $W$ and TW with $e$ extra steps is least with initial preparation of separate synthon functionality before affixation, and worst with refunctionalization carried out after the full skeleton is assembled, as discussed above. Both of these refunctionalizations are independent of the degree of convergency in the plan.

The refunctionalizations in the central construction phase of the plan are caused by the need to adapt functionality after one construction to make it suitable for the next. Here the convergent synthesis has an important advantage, however, since the number of successive constructions in any one line is rarely more than 3 (for $n_0$ = 16–32) whereas in a linear plan there are $\lambda$ constructions all in one line, i.e., commonly more than 10. As an example, if one refunctionalization is required after each construction in the steroid target of Figure 5, there are $\lambda = k + \Delta r - 1 = 8 + 4 - 1 = 11$ constructions and so $e$ = 11 refunctionalizations. In the fully convergent case (the 9/12 dissection sample was calculated) $W$ increases from 70 to 234 but in the linear plan for the same bond set $W$ increases from 134 to 1118, since the main line now has $l_1$ = 22 steps! Thus the economy ratio (lin/con) in $W$ rises from 1.9 to 4.8

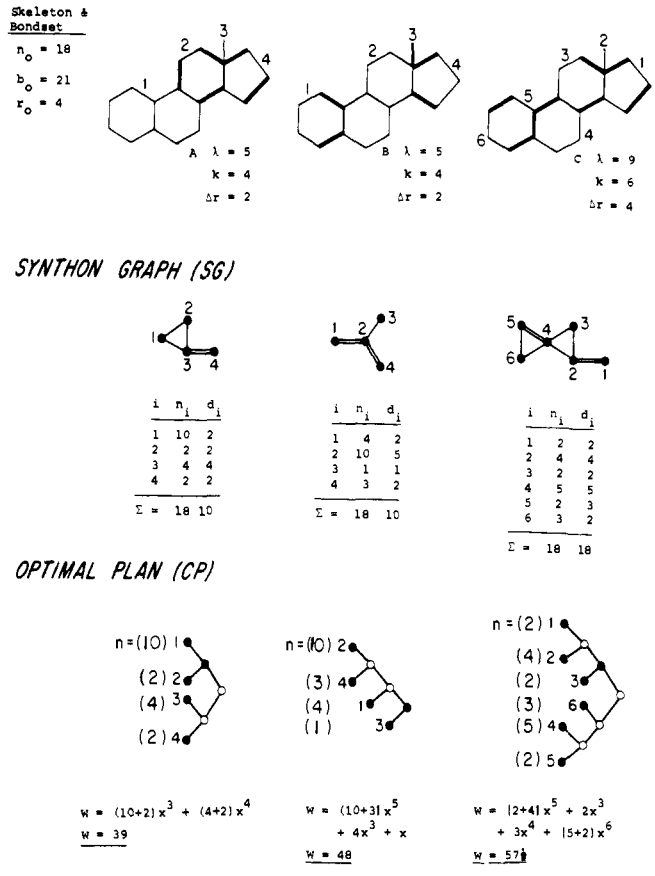



Figure 6. Synthon graphs.

when the refunctionalizations are added, and the TW ratio, as well as those for development ($W_D$ and $TW_D$), also rise to over 5 times in the linear over the convergent plans.

Therefore, although the advantage of the fully convergent plan may seem only modest at the CP stage, it becomes very much greater as the chemical demands for functionality adjustment are laid on. This advantage of convergency also works to favor the search for ideal, or self-consistent, sequences in the design protocol of ref 1c, since the chances of finding many sequences of several self-consistent constructions rapidly diminish above about 3–4 sequential constructions in a row.

**5. Relation to Other Synthesis Concepts.** (a) Any procedure for synthesis design must at some stage draw on a known catalog of available starting materials. These should be organized in such a way as to make all candidates for a given operation immediately accessible, but current commercial catalogs are nearly useless in this respect since they are arranged alphabetically. The present approach focuses on synthon skeletons and the number of possible skeletons is few. The 32 possible acyclic and monocyclic possibilities for $n \leq 6$ are assembled in Figure 7, but most of the cyclic ones are not available starting materials. A very useful organization of starting materials would order them by size and skeleton as in Figure 7, and within these by functionality (f-lists).[1b,c]

(b) The design protocol in ref 1c suggests a number of general ways to create optimal bondsets, including seven skeletally based heuristics, which may be examined in connection with the convergency dissection described here. Among these seven heuristics are convergency itself, available starting material skeletons, and the need to construct quaternary and tertiary centers. The dissection here produces a small set of all convergent bond sets, and their orders of operation in the form of construction plans, utilizing starting materials of size $n$ = 2–4. With such synthons all quarternary and most tertiary centers are automatically constructed. The recognition of larger

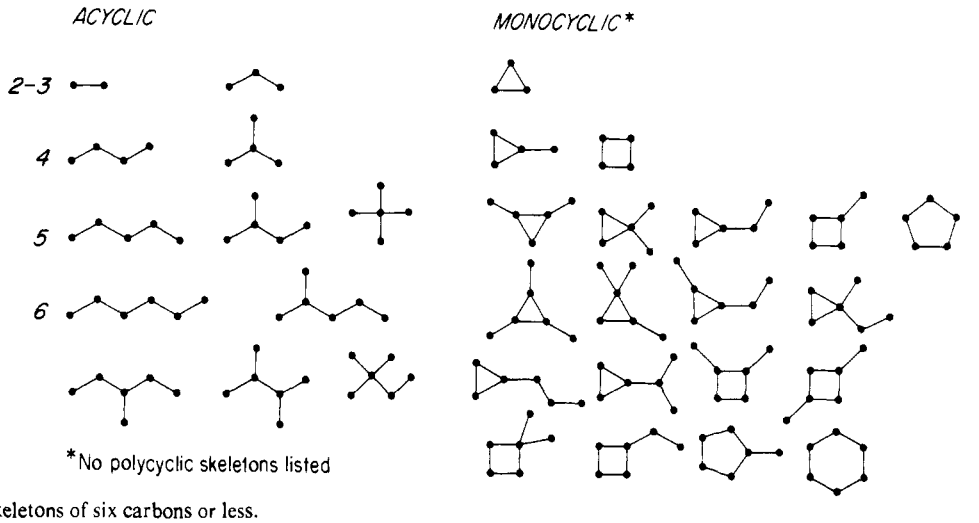



**Figure 7.** Synthon skeletons of six carbons or less.

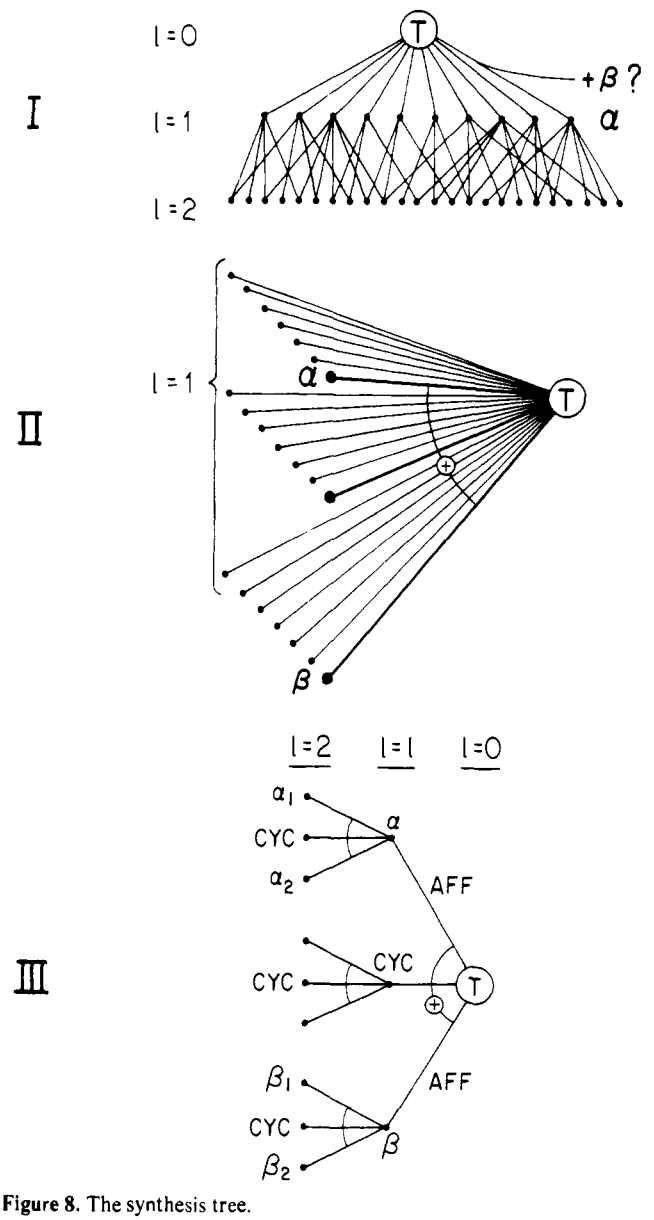



**Figure 8.** The synthesis tree.

available synthons has been discussed and may indeed provide a more economical plan if that synthon is affixed last after convergent construction of the rest.[15] The other heuristics are discussed briefly in the next two paragraphs.

(c) As to stereochemistry, since the synthons derived in this approach are all $n = 2$–4, then all skeletal chiral centers are to be constructed in the plan. Since the timing of their construction on the CP is also set, general principles of stereocontrol can be considered in the early planning stage. Poor stereocontrol and resultant product separations must be counted as extra steps equivalent to their lower yields. The timing of construction of chiral centers on the CP also allows an immediate assessment of the earliest time to undertake resolution of racemates. Resolution (yield < 50%) must also be accounted as extra steps, $e \geq 4$ (equivalent to a yield of $y = (0.80)^4 = 41\%$). From the analysis of extra steps above resolution must be done as early as possible for economy.

(d) In the category of extra steps must also be considered protecting and activating groups which require later removal. Unless they exist in the starting material they require $e = 2$ extra steps and a serious consequence of increasing the various economy criteria. Clearly, their initial attachment should be carried out on the starting material before any affixation, but their removal must always come in the central or final phase after some constructions. Minimization of such groups is certainly desirable.

They have another less obvious bad effect on economy. In the formulas for the criteria is the factor $F$, the amount of total molecular weight carried by each skeletal carbon. These groups (ketal for protection, carboalkoxy for activation, etc.) add excessive weight to the synthons and this weight is carried even though not accounted for in the TW calculation (where $F$ is an average unlikely to take extra groups like these into account). This puts a premium on the development of methods to allow selective reaction without added facilitating groups, but also especially focuses on the low molecular weight olefin as a potential protected form for other functions and so indicates the need for development of more (and more selective) refunctionalizations and constructions at simple double bond sites, presumably by organometallic and pericyclic reactions.

(e) The plan graphs (Figure 1) used here constitute selected single syntheses which may be found intact in the general synthesis tree (cf. Figure 2 in ref 1c) for which each level of the tree is the rank of all intermediates found there. However, in the standard two-dimensional representation of the synthesis tree, with points as intermediates and lines as reactions (I, Figure 8), any second synthon or intermediate affixed to a given intermediate on a line of the tree is not readily indicated. If intermediate $\alpha$ in Figure 8 is affixed to another intermediate, $\beta$, as in a convergent synthesis, it is only implicit in the ordinary drawing. However, if the planar tree is seen from the side as in II (Figure 8) and expanded into three tiers at $l = 1$, the top

and bottom sets of intermediates may be seen as linked (+ as shown) synthons for affixation, while the middle set consists of intermediates with the same number of carbons undergoing a last reaction of cyclization or refunctionalization. Further expansion of the tree, still seen from the side, is shown as III for constructions only, the horizontal lines being cyclizations, the linked lines being affixations, $\alpha + \beta$ at $l = 1$, $\alpha_1 + \alpha_2$ or $\beta_1 + \beta_2$ at $l = 2$, etc. In this way the construction plans of Figure 1 are seen as individual syntheses taken from the total tree represented this way in three dimensions.

The value of the present procedure is to find the less obvious split paths which are shorter because they are convergent. By contrast most of the syntheses reported to date are linear, as a quick inspection of ref 4 will show. Some traditional cast of mind has favored linear syntheses. The traditional approach of working stepwise backwards from the target in synthesis planning is implicitly linear in concept and usually leads to a linear plan. The approach does not specifically seek convergency, and if key construction links for convergency are not dictated by target functionality they are not found. Indeed such key links may often be constructed via "dummy" functionality which leaves no residual functional group in the target. The present approach differs in specifically seeking a convergent skeletal dissection and afterwards applying the requisite functionality to the synthons generated.[1c] It has the advantage of pointing to synthons not readily perceived in the traditional mode, and often also of demanding the development of new chemistry. In any case the definition of criteria of economy here serves to set more stringent standards on the process of synthesis planning.

**Acknowledgment.** The author expresses gratitude to the University of Cape Coast, Ghana, and to the U.S. State Department for a Fulbright Professorship and a peaceful setting that allowed these ideas to develop.

## Appendix A. Derivation of Time Criteria

$\mathrm{TW}_i = Fn_iS_{l_i}$ — TW = total weight manipulated, for synthon $i$ ($= 1 \rightarrow k$) or for all material through all steps ($\mathrm{TW}_0$)

$\mathrm{TW}_0 = F\sum_{i=1}^{k} n_iS_{l_i}$ — $S_{l_i} = \sum_{v=1}^{l_i} X^v$

$\mathrm{TW}_0 = \sum_{j=1}^{s} W_j$ — $W_j$ = weight manipulated in step $j$ ($= 1 \rightarrow s$)

$W_{av} = \mathrm{TW}_0/s$ — $W_0$ = unit weight requiring unit time $T_0$

$T = T_0\sum_{j=1}^{s}\left(\frac{W_j}{W_0}\right)^z$ — $T$ = overall required time for plan

It was found on a number of trials for particular plans that the average weight per step ($W_{av}$) could be used in place of $W_j$ within 3% correlation. This allows conversion to a sum over $i$ instead of $j$. Hence

$$t, sT_0\left(\frac{W_{av}}{W_0}\right)^z = sT_0\left(\frac{TW_0}{sW_0}\right)^z = sT_0\left[\frac{F\Sigma_i n_iS_{l_i}}{sW_0}\right]^z$$

Removing constants,

$$\mathrm{TW} = \sum_{i=1}^{k} n_iS_{l_i}$$

$$T = s^{1-z}\mathrm{TW}^z$$

For trial production, weight manipulated/last step = $nmn_ix$ for each synthon $i$, and in step previous = $nmn_ix^2$ for last step trials + $nmn_ix$ for penultimate step trials, etc. For all $(l_i - 1)$ steps

$$\mathrm{TW}_i = nmn_i[(l_i - 1)x + (l_i - 2)x^2 + (l_i - 3)x^3 + \ldots + x^{l_i-1}]$$

This sum rearranges to a sum of the previous sum terms $S_l$, which is

$$Q_l = \sum_{v=1}^{l-1} S_v$$

$$\mathrm{TW}_{dev} = Fnm\sum_{i=1}^{k} n_iQ_{l_i}$$

$$T_{dev} = snt + sT_0\left[\frac{Fnm\Sigma_i n_iQ_{l_i}}{sW_0}\right]^z$$

For ratio comparisons, separate and remove constants:

$$T_{(\text{trials})} = s;\ T_{(\text{production})} = s^{1-z}\mathrm{TW}_D{}^z$$

where

$$\mathrm{TW}_D = \sum_{i=1}^{k} n_iQ_{l_i}$$

## Appendix B. Ancillary Relations of Structures and Plan Graphs

(1) The number of AP for $k$ synthons is

$$N_k = \tfrac{1}{2}\sum_{i=1}^{k} N_{k-1}N_i - Z$$

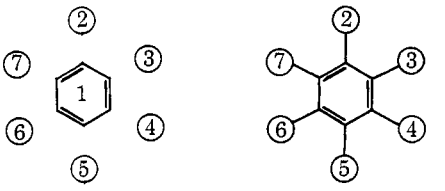
where $Z = 0$ for odd $k$

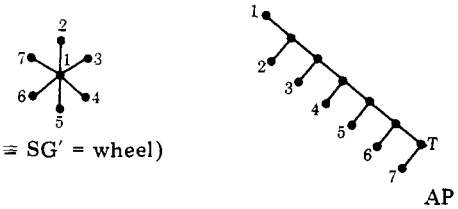
$Z = \tfrac{1}{2}N_{k/2}(N_{k/2} - 1)$ for even $k$

$N_1 = N_2 = N_3 = 1$

| $k$ | $N_k$ |
|---|---|
| 4 | 2 |
| 5 | 3 |
| 6 | 6 |
| 7 | 11 |
| 8 | 23 |

(2) Whether a fully convergent synthesis is possible for any SG may be explored by reducing it to any acyclic SG′ by removal of $\Delta r$ cyclization bonds and so leaving only bonds for affixation in SG′. The number of SG′ for $k$ synthons is the same as that for the number of possible AP to assemble them (Appendix B, section 1), i.e., the number of acyclic connected graphs on $k$ points. There is not a simple correlation, however, since linear AP are available for all SG′ but fully convergent AP only for SG′ with no more than two points of degree = $2^m$ (and none of higher degree), and those two must be linked at the center dividing bond of the SG′. Of the SG′ for $k = 8$ synthons only 3/23 can be fully convergent; for $k = 7$ only 4/11, but any linear SG′ can be fully convergent. On the other hand, the most branched SG′ is the "wheel" graph with a central point (synthon) of degree $k - 1$, and this can have only a linear AP. An example is the polysubstitution of benzene,[16] illustrated below.



## References and Notes


(1) Previous papers in this series: (a) part 3. J. B. Hendrickson. *J. Am. Chem. Soc.*, **97**, 5763 (1975); (b) part 4. *ibid.*, **97**, 5784 (1975); (c) part 5. *Top. Curr. Chem.*, **62**, 49 (1976).

(2) In his treatment of nucleic acid synthesis design,[3] Powers used time as a unitary criterion, converting material manipulation into units of time.

(3) G. J. Powers, R. L. Jones, M. Carruthers, H. van de Fande, and H. G. Khorana. *J. Am. Chem. Soc.*, **97**, 875 (1975).

(4) A survey of about 90 natural product syntheses from N. Anand, J. S. Bindra, and S. Ranganathan, "Art in Organic Synthesis", Holden-Day, San Francisco, Calif., 1970, was taken as a basis for comparison.

(5) L. Velluz, G. Nomine, J. Mathieu, E. Toromanoff, D. Bertin, J. Tessier, and A. Pierdet, *C. R. Acad. Sci.*, **250**, 1084, 1511 (1960).

(6) W. S. Johnson, T. T. Lee, C. A. Harbert, W. R. Bartlett, T. R. Herrin, B. Stastiun, and D. H. Rich, *J. Am. Chem. Soc.*, **92**, 4461 (1970).

(7) The Velluz plan[5] has been adapted[6] to create $C_{21}$ steroids by an increase of ten refunctionalization steps at the end, making it less economical. Thus $L_K$ goes to 128 and $W$ increases by almost tenfold to 2622.

(8) L. Velluz, G. Valls, and J. Mathieu, *Angew. Chem., Int. Ed. Engl.*, **6**, 778 (1967).

(9) R. E. Ireland, "Organic Synthesis", Prentice-Hall, Englewood Cliffs, N.J., 1969; E. J. Corey and W. T. Wipke, *Science*, **166**, 178 (1969); E. J. Corey, *Q. Rev., Chem. Soc.*, **25**, 455 (1971).

(10) The *order* of operations in a synthetic plan is the reverse of the rank on the plan graph, so that the order of all steps is clearly sequential in a linear plan but arbitrary for steps of equal rank on separate lines in a convergent synthesis. Thus it is unimportant in what order the four initial affixations at $l = 3$ are performed in the convergent plan IV, Figure 3, or whether all the steps on one line are performed before or after those on another before the rank at which the lines join.

(11) A very few highly branched target structures cannot have a fully convergent AP for total synthesis from $n = 1$ synthons, the simplest examples being neopentane and hexamethylethane skeletons. However, even such targets may often have perfect convergent dissections down to synthons of $n = 2$–4 when the branched $n = 4$ (isobutane) skeleton is allowed as a starting material.

(12) The affixation plans (AP) for all these derived fully convergent dissections have identical $W$ but the full $W$ for the CP (including cyclizations) will vary slightly from one dissection to another depending on the detailed size ($n$) of the intermediate being cyclized in each case.

(13) The plan graph for any bondset indicates its route through the construction grid of all possible construction combinations in ref 1a (Figure 9). The bondset itself defines $\lambda$ and $\Delta r$ and so the grid position $R_k$ of the prestruct of $k$ synthons. The order of operations on the plan graph is now a sequence of affixations and cyclizations, i.e., successive lines on the grid leading either down or to the right, respectively. With convergent syntheses several routes through the grid are equivalent,[10] one for the example in Figure 4 being $O_8 \rightarrow O_4 \rightarrow A_4 \rightarrow A_2 \rightarrow B_2 \rightarrow C_2 \rightarrow C_1 \rightarrow D_1$.

(14) The difference $\Delta r$ is between the number of rings in the target skeleton ($r_0$) and the total rings of the starting materials,[1a] and is easily counted as the number of target skeleton rings which include any bonds of the bond set.

(15) Since normal synthons are $n = 2$–4 this places a special premium on available aromatic synthons of $n \geq 6$; not only for aromatic targets, but also for nonaromatic synthons in a target. This in turn implies the need for development of more effective methods of dearomatization.

(16) J. B. Hendrickson, *J. Am. Chem. Soc.*, **93**, 6854 (1971).


# Detection of the Furanose Anomers of D-Mannose in Aqueous Solution. Application of Carbon-13 Nuclear Magnetic Resonance Spectroscopy at 68 MHz


**David J. Wilbur, Carol Williams, and Adam Allerhand***

*Contribution No. 3018 from the Department of Chemistry, Indiana University, Bloomington, Indiana 47401.*

*Received March 28, 1977*


**Abstract:** Natural-abundance $^{13}C$ NMR spectra (at 67.9 MHz) of aqueous D-mannose (4 M in $H_2O$, 36 °C) yield identifiable resonances of five carbons of $\alpha$-D-mannofuranose and three carbons of $\beta$-D-mannofuranose. Integrated intensities indicate the presence of 0.6 ± 0.1% $\alpha$-D-mannofuranose and 0.3 ± 0.1% $\beta$-D-mannofuranose.


Proton NMR spectroscopy has been used to determine the proportions of the predominant (pyranose) anomers of various common aldohexoses in aqueous solution, by taking advantage of the relatively resolved resonances of the anomeric hydrogens.[1,2] Some aldohexoses have also yielded observable resonances for the anomeric hydrogens of the furanose anomers.[2] Notable exceptions are glucose and mannose. Angyal and Pickles[2] estimated that the proportion of the furanose forms in the anomeric equilibrium of each of these sugars in water is considerably less than 1%. Recently, $^{13}C$ NMR spectra (at 15 MHz) of aqueous D-glucose yielded resolved resonances of carbons 1, 2, and 4 of $\beta$-D-glucofuranose.[3] Integrated intensities yielded 0.14 ± 0.02% for the proportion of this anomer at 43 °C.[3] Carbon-13 NMR spectra of D-mannose (**1**), obtained at 15 MHz, exhibited some weak signals attributable to one or both furanose anomers, but interference from sidebands of the strong resonances of $\alpha$-D-mannopyranose (**1a**, Figure 1) and $\beta$-D-mannopyranose (**1b**, Figure 1) prevented an unambiguous interpretation.[4] In this report we show that $^{13}C$ NMR spectra at 67.9 MHz (63.4 kG) yield identifiable resonances of five carbons of $\alpha$-D-mannofuranose (**1c**, Figure 1) and three carbons of $\beta$-D-mannofuranose (**1d**, Figure 1). The intensities of these resonances yield fairly accurate values for the proportions of the furanose anomers. This study strongly suggests that the resolution and sensitivity available at high magnetic field strengths should greatly facilitate the use of natural-abundance $^{13}C$ NMR spectroscopy in studies of trace components of complex carbohydrate mixtures.

## Experimental Section

**Materials.** We used three samples of **1**: Sample 1 was obtained by once recrystallizing commercial D-mannose ("Ultrex" grade from J. T. Baker Chemical Co., Phillipsburg, N.J.) into **1a**;[5] sample 2 was obtained by once recrystallizing the "Ultrex" D-mannose into **1b**;[6] sample 3 was D-mannose from Sigma Chemical Co., St. Louis, Mo., used as received.

**Methods.** Proton-decoupled natural-abundance $^{13}C$ NMR spectra were obtained at 67.9 MHz (63.4 kG) on a spectrometer consisting of a Bruker high-resolution superconducting magnet, Bruker 10-mm probe, home-built radiofrequency electronics, and a Nicolet 1085 computer. The spectrometer was not equipped with a field-frequency lock. For $^{13}C$ excitation, 90° radiofrequency pulses of 16 $\mu$s duration were used, and the frequency was set 118 ppm downfield from $Me_4Si$. Time-domain data were accumulated in 8192 addresses, with a spectral width of 73.6 ppm, 16 384 scans, and a recycle time of 1.03 s (5 h accumulation time per spectrum). Fourier transformation was done on 16 384 addresses (by adding 8192 addresses with a zero value at the end of the accumulated time-domain data points), with 0.4 or 0.8 Hz digital broadening. The $^1H$ irradiation (at 270 MHz) had a peak field strength of about 0.8 G (3.4 kHz). Other proton-decoupling conditions are given in the caption of Figure 2 and in footnote *d* of Table I. Chemical shifts are reported in parts per million downfield