Complexity of synthetic routes: Linear, convergent and reflexive syntheses[1]†

Steven H. Bertz

Complexity Study Center, Mendham, NJ 07945, USA

Received (in New Haven, CT, USA) 2nd November 2002, Accepted 21st December 2002
First published as an Advance Article on the web 15th April 2003

Complexity and diversity indices can be used to evaluate the structural changes that take place during the course of a synthesis. From the generalized 'complexity *versus* step' plot and the definition of the 'ideal synthesis' we derive the concept of *excess complexity* and the heuristic, minimization of excess complexity. *Molecular complexity* and *synthetic complexity* are differentiated by identifying them with intrinsic and extrinsic complexity, respectively, and the effect of chirality on each is clarified. The role of symmetry in synthetic analysis is delineated by considering symmetry in molecules, their molecular graphs and the synthesis digraphs for their preparation, which leads to the concept of *reflexivity* and the heuristic, maximization of reflexivity. *Vulnerability* and *robustness* are introduced in connection with linear and convergent syntheses, which are also considered in relation to reflexivity.

Introduction

Complexity considerations are playing increasingly important roles in many areas of chemistry,[2,3] including synthetic chemistry.[4–16] Our approach to measuring the complexity of a system, such as a molecule or supramolecular complex, is to use topological invariants to quantify the relationships in it.[7–9,12–16] Definitions of useful mathematical terms and concepts were given in the preceding paper, which considered the complexity of reactions and the related transforms and disconnections.[1] In this paper we extend the discussion to series of reactions comprising synthetic routes.

The two main approaches to understanding the complexity of a system are (i) to study its operation, the way its parts interact with each other, which we call *intrinsic complexity*, and (ii) to examine its construction, the way the parts are assembled, which we call *extrinsic complexity*.[15] The main way that atoms in a molecule interact is by forming covalent bonds, but other, non-bonded interactions can also be considered.[1] These two kinds of complexity are intimately connected, that is, geometric relationships determine the way the atoms in a molecule interact, and they also influence how the atoms are assembled when the molecule is synthesized. The crucial difference between intrinsic and extrinsic complexity lies in their response to symmetry,[15] as the former is independent of it, whereas the latter is not. A complication is the inconvenient fact that the effect of symmetry is not always manifested (*vide infra*). Symmetry decreases the *number of kinds of interactions* in a system, but not the *total number of interactions*. The former is a measure of *diversity*, which is subsumed under extrinsic complexity, since it affects the difficulty of construction. The latter is a measure of intrinsic complexity.

When the system is a molecule, the intrinsic complexity is the *molecular complexity* and the extrinsic complexity is the *synthetic complexity*. Molecular complexity is intrinsic, since it depends solely on the structure of the molecule and not on the tools (reactions) used in its construction. A measure that is not affected by symmetry is needed, since a molecule with many identical subunits (*e.g.*, a polymer) has properties that are very different from a single one of them (*e.g.*, a monomer), owing to emergent phenomena.[2] In contrast, the synthetic complexity of a molecule is extrinsic, since it depends principally on the reactions used, that is on the state of the art of synthetic chemistry. Molecular diversity is extrinsic, since the kinds of interactions (*e.g.*, C–C, C–N, C–O, *etc.*) are not unique to any particular molecule. Moreover, it directly affects synthetic complexity, as a molecule with many different functional groups is generally more difficult to synthesize than one with the same number of identical ones. When symmetry is present, the synthetic complexity may be decreased by it, provided that synthetic methodology is available to take advantage of it.[17] Such synthetic methods have been called 'Odin reactions,'[18] and the development of additional ones would help approaches based on symmetry to reach their full potential. An example of the effect of symmetry on synthetic complexity is the reflexive synthesis of a molecule from two or more identical components (*vide infra*).[10,12,14] While the intrinsic complexity of a system at equilibrium is constant, the extrinsic complexity may vary. More than one way of constructing a system is usually possible, and each can have a different complexity. Nowhere is this more evident than in the case of molecular systems, which is one of the things that makes chemical synthesis a fascinating and challenging subject. Moreover, synthetic complexity changes with the state of the art, which is constantly evolving.

To measure the complexity of a molecule or supramolecular complex, we abstract it as a hydrogen-suppressed molecular graph M and then characterize it by using topological invariants $I(M)$.[1] The simplest (intrinsic) complexity index is the number of pairs of adjacent bonds η,[7,8] which has been very useful,[9,14] and it was combined with information theory to account for symmetry in the first general index of (extrinsic) complexity $C(\eta,\varepsilon)$,[7] which has also been applied to a number of interesting problems.[3–5,7,12]

More recently, we have introduced the 'all possible substructures' method, which features the number of kinds of substructures N_S and their total number N_T,[13] which can sometimes be approximated by the corresponding number of

† Electronic supplementary information (ESI) available: values of indices used to calculate analytical functions for routes 1–3. See http://www.rsc.org/suppdata/nj/b2/b210844p/

 DOI: 10.1039/b210844p

Table 1 All possible substructures of aldol

Entry (i)	Substructure	Chemical name	N_i	T_j	P_k
1	O=●–●–●(–O)–●	3-Hydroxybutanal (aldol)	1	–	–
2 ($j = 1$)	O–●–●–●(–O)–●	1,3-Dihydroxybutane	2	2	–
3	O=●–●–●–O	3-Hydroxypropanal	1	–	–
4 ($j = 2$, $k = 1$)	O–●–●–●–O	1,3-Dihydroxypropane	2	2	2
5	O=●–●–●–●	Butanal	1	–	–
6 ($j = 3$, $k = 2$)	O–●–●–●–●	1-Butanol	2	2	2
7 ($j = 4$)	●–●–●(–O)–●	2-Butanol	1	1	–
8	O=●–●–●	Propanal	1	–	–
9 ($j = 5$, $k = 3$)	O–●–●–●	1-Propanol	3	3	3
10 ($j = 6$)	●–●(–O)–●	2-Propanol	1	1	–
11 ($j = 7$, $k = 4$)	●–●–●–●	Butane	1	1	1
12	O=●–●	Acetaldehyde	1	–	–
13 ($j = 8$, $k = 5$)	O–●–●	Ethanol	4	4	4
14 ($j = 9$, $k = 6$)	●–●–●	Propane	2	2	2
15	O=●	Formaldehyde	1	–	–
16 ($j = 10$, $k = 7$)	O–●	Methanol	3	3	3
17 ($j = 11$, $k = 8$)	●–●	Ethane	3	3	3
18 ($j = 12$, $k = 9$)	O	Water	2	2	2
19 ($j = 13$, $k = 10$)	●	Methane	4	4	4
$N_S = 19$ ($T_S = 13$, $P_S = 10$)			$N_T = \Sigma_i N_i = 36$	$T_T = \Sigma_j T_j = 30$	$P_T = \Sigma_k P_k = 26$

kinds of heteroalkanes[19] T_S or linear heteroalkanes P_S and their total numbers T_T or P_T, respectively.[14] These topological invariants are among the few that fulfill the criteria for indices of complexity.[8,14,15,20] Furthermore, N_T, T_T and P_T are indices of intrinsic complexity, while N_S, T_S and P_S are indices of extrinsic complexity, specifically diversity. For our purposes conformational information is not included, as this convention has very useful consequences for synthesis planning (*vide infra*). The application of the 'all possible substructures' method to 3-hydroxybutanal (aldol) is illustrated in Table 1.

Results and discussion

Synthesis and retrosynthesis

An effective method for planning a synthesis is retrosynthetic analysis,[21] in which the chemist begins with the target molecule and works backwards, breaking bonds until relatively simple starting materials are obtained. A retrosynthetic *transform* τ is conceptually the exact reverse of a synthetic reaction σ, and *vice versa*. Mathematically, $\tau = \mathscr{R}(\sigma)$ and $\sigma = \mathscr{R}(\tau)$, where $\mathscr{R}$ is the 'retro' operator, which converts any process into its microscopic reverse. For example, the Diels–Alder transform is identical to the retro-Diels–Alder reaction, τ(Diels–Alder) = $\mathscr{R}$[σ(Diels–Alder)] = σ(retro-Diels–Alder). The retro-Diels–Alder reaction can also be used synthetically,[22] in which case the transform is the Diels–Alder reaction: τ(retro-Diels–Alder) = $\mathscr{R}$[σ(retro-Diels–Alder)] = σ(Diels–Alder).

The *retrosynthetic scheme* (or *retrosynthesis*) $T = \{\tau_1, \tau_2, \tau_3, \ldots, \tau_n\}$ is a series of transforms τ_j ($j = 1, \ldots, n$) from the target to the starting materials. The *synthetic scheme* or *synthesis plan* $\Sigma = \{\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n\}$ is the microscopic reverse of the retrosynthetic scheme, $\Sigma = \mathscr{R}(T)$, that is a series of synthetic reactions σ_i ($i = 1, \ldots, n$) from the starting materials to the target. [*N.B.*, this operator is symmetrical, so that $T = \mathscr{R}(\Sigma)$.]

A *strategic transform* τ_s is a transform that provides major simplification, and the first one usually sets the course of a synthesis. Often, it does not correspond to the first transform (τ_1) in a retrosynthetic scheme, as adjustments to the skeleton or functional groups in the target must be made before a powerful transform can be used. The corresponding synthetic step is the *strategic reaction* $\sigma_s = \mathscr{R}(\tau_s)$. Strategic reactions are characteristic of convergent synthesis plans (*vide infra*) and in a multiply convergent plan there are two or more of them, as brilliantly illustrated by Kishi *et al.*'s quadruply convergent synthesis of halichondrin B.[23] In contrast, linear routes do not have strategic reactions, since no transform provides major simplification.

A *disconnection* δ is the minimal subset of transform bonds that contains the skeletal (sigma) bonds that are broken or formed and the shortest paths (if any) joining their termini. To maintain continuity from structure to structure, each disconnection in a multistep analysis contains the bonds from all disconnections. In mathematical terms $\delta = \mathscr{C}(\tau)$, where $\mathscr{C}$ is the 'cut' operator, which removes any bonds that do not conform to the definition. A *topological retrosynthesis* $\Delta = \{\delta_1, \delta_2, \delta_3, \ldots, \delta_n\}$ is a series of disconnections δ_j ($j = 1, \ldots, n$) from the target to the starting materials, $\Delta = \mathscr{C}(T)$, and the corresponding *topological synthesis plan* $\Gamma = \{\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_n\}$, where $\gamma_i = \mathscr{R}(\delta_i) = \mathscr{C}(\sigma_i)$, is the microscopic reverse, $\Gamma = \mathscr{R}(\Delta) = \mathscr{C}(\Sigma)$. They allow us to focus on the topological changes that take place over the course of a synthesis or retrosynthesis and clearly see any inefficiencies present (*vide infra*, Scheme 2).

The disconnection δ_s that corresponds to a strategic transform τ_s is a *strategic disconnection*, $\delta_s = \mathscr{C}(\tau_s)$, and it has been discussed from both heuristic and topological perspectives.[12,14] Since the bonds broken (or formed) in δ_s are the *strategic bonds*, it is also called the *strategic bond disconnection*.[21] Many ways of disconnecting a complex target are usually possible and for each topological retrosynthesis there are a number of retrosynthetic schemes, since every disconnection can give rise to several transforms. Thus, many synthetic routes are possible for a given target molecule of even moderate complexity and deciding which one to reduce to practice can be a difficult problem. As will be shown below, molecular complexity considerations can give valuable insights for synthetic analysis, which has been based on literature precedents and heuristic principles, but not rigorous mathematics.

Complexity *versus* step plots

The *total complexity* C(total) of a synthesis plan is defined as the area under its 'complexity *versus* step' plot (C–s plot). (*N.B.*, for integration each step has width $\Delta s = 1$.) The portion of it above the absolute minimum required for the corresponding 'ideal synthesis' is the *excess complexity* C_x, which can be calculated as the sum of the complexities C_i of the

intermediates I_i between the starting materials and the target [eqn. (1)].[9] The *ideal synthesis* converts the starting materials quantitatively into the target in a single step. (Other authors have their own criteria for this important concept.[24,25]) In analogy to Whitlock's ΔS_m (*vide infra*), we define ΔC_m as the average magnitude of the change in complexity per step in eqn. (2), where ΔC_j ($j = 1, \ldots, k$) is the change in complexity for the jth step and k is the total number of steps. As a general principle, we have proposed the heuristic of *minimization of excess complexity*,[9] which is justified below. Wender and Miller have discussed the related 'maximization of target-relevant complexity,'[26] and Trost has highlighted 'atom economy,'[27] which also tends to minimize C_x.

$$C_x = \Sigma_i\ C_i \qquad (1)$$

$$\Delta C_m = \Sigma_j |\Delta C_j|/k \qquad (2)$$

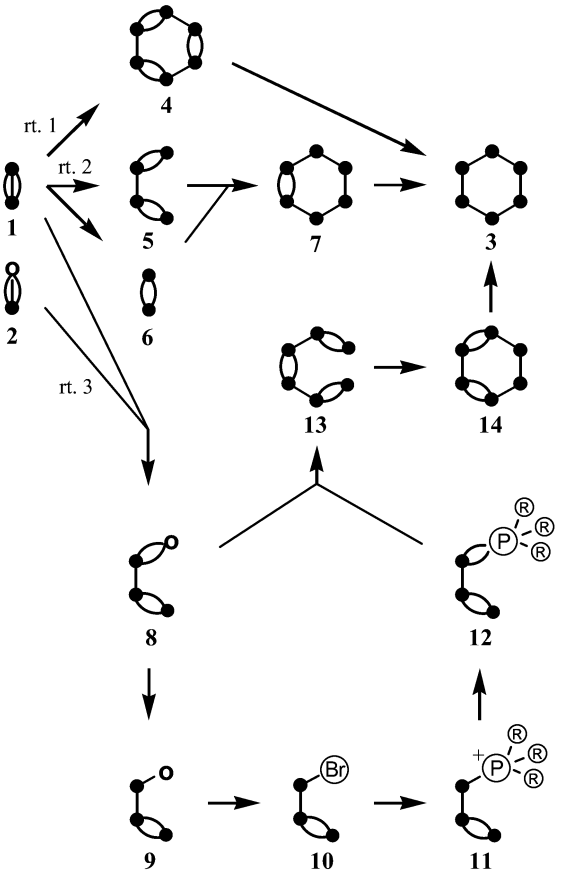
The derivation of eqn. (1) is based on the difference between the total complexity C(total) of a synthetic route and the complexity C(ideal) of its ideal counterpart. If we assign a complexity contribution of $\frac{1}{2}C(A) + \frac{1}{2}C(B)$ to a reaction that converts reactant A into product B,[28] then the total complexity of a route from starting material S to target T with n intermediates I_i is given by $C(\text{total}) = [\frac{1}{2}C(S) + \frac{1}{2}C(I_1)] + \ldots + [\frac{1}{2}C(I_{i-1}) + \frac{1}{2}C(I_i)] + [\frac{1}{2}C(I_i) + \frac{1}{2}C(I_{i+1})] + \ldots + [\frac{1}{2}C(I_n) + \frac{1}{2}C(T)] = \frac{1}{2}C(S) + \Sigma_i\ C(I_i) + \frac{1}{2}C(T)$. Thus, each intermediate I_i ($i = 1, \ldots, n$) has a contribution of $\frac{1}{2}C(I_i)$ from the reaction that makes it and another $\frac{1}{2}C(I_i)$ from the reaction that takes it on to the next intermediate (or the target) for a total contribution of $C(I_i)$. The ideal synthesis that converts S into T in one step has complexity $C(\text{ideal}) = \frac{1}{2}C(S) + \frac{1}{2}C(T)$.[28] Then, the excess complexity C_x over this unavoidable minimum is given by $C_x = C(\text{total}) - C(\text{ideal}) = \Sigma_i\ C(I_i)$. Defining $C_i = C(I_i)$ gives eqn. (1).

Each step in a multistep synthesis exposes the evolving structure to reagents or energy that can induce side-reactions. For prediction of the overall yield, the probability p_s of desired product from any step s in a t-step synthesis of target T is the fraction of molecules of (limiting) reactant that is converted into the desired product. It is usually called the yield Y_s (expressed as a mole fraction) and it can be written in terms of the probability p_s' or yield Y_s' of side-products for step s by $p_s = 1 - p_s'$ or $Y_s = 1 - Y_s'$. If the probability of side-reactions during step s is proportional to the complexity C_s of this step, then $p_s' = k_s C_s$ or $Y_s' = k_s C_s$, where k_s is the constant of proportionality for step s. Then, the probable overall yield of target T is $Y(T) = \Pi_1^t Y_s = \Pi_1^t (1 - Y_s') = \Pi_1^t (1 - k_s C_s)$. Assuming that to a first approximation C_s is determined by the complexities $C_s(A_i)$ of the reactants A_i and $C_s(B_j)$ of the products B_j of step s, then $C_s \approx [\Sigma_i\ C_s(A_i) + \Sigma_j\ C_s(B_j)]$ and Y(T) is given by eqn. (3).

$$Y(T) \approx \Pi_1^t\ \{1 - k_s[\Sigma_i\ C_s(A_i) + \Sigma_j\ C_s(B_j)]\} \qquad (3)$$

The polynomial form of eqn. (3) is complex and difficult to evaluate; nevertheless, we can say that minimizing $C_s(A_i)$ and $C_s(B_j)$ will maximize Y(T). Furthermore, since $C_x = \Sigma_i\ C_i$ includes most of the terms $C_s(A_i)$ and $C_s(B_j)$, we can say that minimizing the excess complexity of a route tends to maximize its overall yield. Minimizing the number of steps t also tends to maximize Y(T). Like all such generalizations, these are based on 'all other things being equal.'[29] Nevertheless, this 2-dimensional heuristic is an improvement over the classical 1-dimensional minimization of the number of steps.

Three synthetic routes from acetylene **1** and (in one case) carbon monoxide **2** to cyclohexane **3**, based on the reactions calculated in the previous paper,[1] are illustrated in Scheme 1, and their C–s plots are drawn in Fig. 1. [The data used to construct them are given in the Electronic supplementary information (ESI).] Although it is not affected by heteroatoms,



Scheme 1 Three pedagogical routes to **3**.

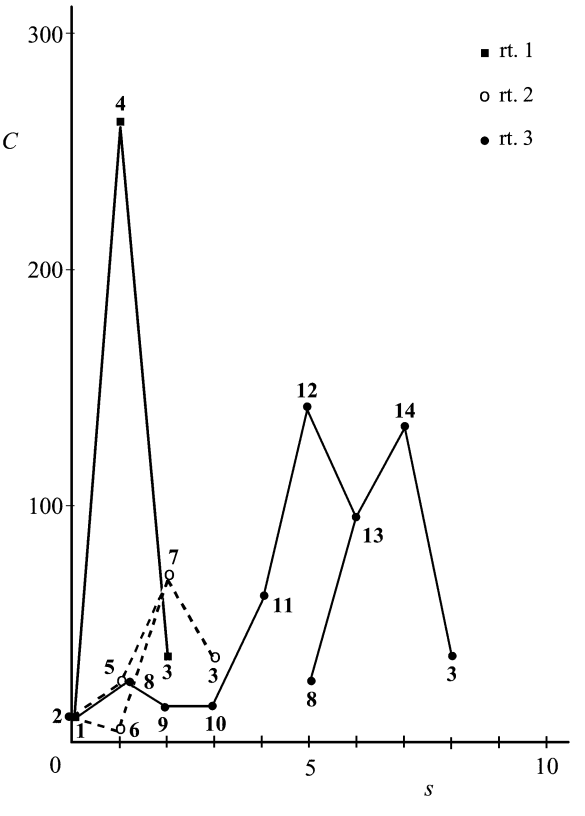


Fig. 1 Complexity *versus* step plots for Scheme 1.

we use N_T as the invariant for the *C–s* plots discussed here, since all things considered, it is the most robust topological complexity index.[1] Heteroatoms increase the probability of side-reactions, but also decrease the synthetic difficulty,[21] so that their effects tend to cancel. All reactions are taken from the literature, except for the specific Wittig reaction used in route 3. While laboratory syntheses of cyclohexane have only pedagogical value, the preparation of isotopomers, such as labeled with ^{11}C (β^+ emitter), ^{13}C (stable) or ^{14}C (β^- emitter), have practical value.[30]

In route 1 acetylene **1** is cyclotrimerized with a Pd or Pt catalyst to benzene **4**,[31] which is reduced by catalytic hydrogenation (Ru complex) to cyclohexane **3**.[32] Dihydrogen and catalysts are considered to be reagents and not intermediates, and consequently their complexities are not part of the calculations.[1] Therefore, the excess complexity in this case is the molecular complexity of the lone intermediate, benzene, C_x(route 1) = N_T(**4**) = 261.

Route 2 also proceeds from **1** as the sole starting material. The first stage of the synthesis has two parallel reactions, the preparations of butadiene **5** and ethylene **6**.[33] These intermediates are then joined in the convergent step by σ_s(Diels–Alder) to afford cyclohexene **7**,[34] which is reduced by catalytic hydrogenation to **3**.[32] Thus, we have C_x(route 2) = N_T(**5**) + N_T(**6**) + N_T(**7**) = 100.

Route 3 has two starting materials, **1** and **2**, which give acrolein **8** in a Pd- or Rh-catalyzed reaction.[35] Half of the product is reduced to allyl alcohol **9**,[36] which is converted to allyl bromide **10** with HBr.[37] Bromide **10** is treated with triphenylphosphine to give phosphonium salt **11**, and this precursor is deprotonated with MeLi to afford phosphorane **12**,[38] which is allowed to react with the second half of the **8**. The portion of *Z*-1,3,5-hexatriene **13** from σ_s(Wittig) is cyclized *via* a [3,3]-sigmatropic rearrangement to 1,3-cyclohexadiene **14**,[39] which is reduced by catalytic hydrogenation to **3**.[32] When calculating C_x, we add the complexity of **8** once, since the half that is converted into **12** is considered to be an intermediate, while the half that is used as a reactant in σ_s(Wittig) is considered to be a starting material. Then, we have C_x(route 3) = N_T(**8**) + N_T(**9**) + N_T(**10**) + N_T(**11**) + N_T(**12**) + N_T(**13**) + N_T(**14**) = 489.

Complexity *versus* step plots have characteristic profiles; for example, some have one or more peaks, which substantially overshoot the complexity of the target. Such plots are typical of convergent syntheses (*vide infra*), which employ powerful construction reactions in the key steps (σ_s). Corrective steps are then needed to bring the complexity back down to that of the target, or optimally the peak is the target. Precisely where along the route a peak occurs depends on how long it takes to set the stage for σ_s. All three *C–s* plots in Fig. 1 have at least one peak significantly higher than the complexity of the target. In route 1 no functional group modification is necessary for σ_s(Reppe–Vollhardt) and the peak occurs immediately. In route 2 the reactants for σ_s(Diels–Alder) must be prepared, but only one step is required for each, and the peak is more or less in the middle of the route. In route 3 several steps are needed to prepare **12**, the key intermediate for σ_s(Wittig), and the peak occurs late in the synthesis.

Other syntheses have *C–s* plots in which complexity increases incrementally until the target is reached, and they tend to be linear (*vide infra*). Generally speaking, linear syntheses maximize target relevant complexity,[26] but do not necessarily minimize excess complexity, since they tend to have more steps. Of course, these examples are archetypes and everything in between is possible. Typical situations were illustrated by seven syntheses of modhephene in the communication that introduced *C–s* plots.[9] An extensive article on polyquinane synthesis by Chanon *et al.* graphically illustrated numerous complexity *versus* step plots and also introduced similarity *versus* step plots.[4] Serratosa *et al.* used *C–s* plots to analyze several syntheses of the unnatural polyquinane, dodecahedrane, projected or accomplished.[40] Whitlock used complexity *versus* step plots in conjunction with his *S* index to analyze some historically important syntheses.[41] Complexity (and similarity) *versus* step plots for a number of classic syntheses have been presented by Barone *et al.*[5]

The three routes described above can be evaluated by using the classic heuristics and the new ones introduced here. Route 3 is significantly longer than the other two and it has considerably higher excess complexity, most of which is associated with the two peaks at the end. It contains functional group interchanges that are not part of construction reactions, for example, a C–O bond is replaced by a C–Br bond, which is replaced by a C–P bond in order to activate the molecule for carbon–carbon bond formation. Such reactions contribute to excess complexity without building up the molecular skeleton. Fortunately, they occur in the first half of the synthesis, where complexity is relatively low and consequently side reactions have low probabilities. The (unoptimized) yield of **8** from the first reaction is poor[35] and purification is problematical, but the starting materials are not expensive. Of more concern, the reaction of **8** with phosphorane **12** has not been reported, and literature precedent implies a more or less equal mixture of *E* and *Z* products.[38] Route 3 is not atom economical[27] and its by-products (CH_4, LiBr, Ph_3PO) present substantial waste disposal problems. These factors eliminate route 3 from consideration, even though it has elements of convergence and reflexivity (*vide infra*).

Routes 1 and 2 are very competitive. Whereas route 1 has higher excess complexity, it has fewer steps. The relatively high value of C_x(route 1) is not as bad as it may seem at first glance, since the intermediate (benzene) is aromatic and has only one reactive functional group (the π-sextet), which limits side reactions. Thus, one should not look at a complexity value without also looking at the molecule to see what contributes to it. The two extra steps before σ_s in route 2 are not as inefficient as they would be in a linear synthesis, since they can be run simultaneously. Both routes are atom economical and use powerful construction reactions for σ_s. There is literature precedent for both and the decision as to which one to investigate first would be based largely on practical considerations, such as the costs of chemicals, equipment and labor.

At the current state of the art, very few targets can be prepared by using only construction reactions, much less in a single step (*cf.* ideal synthesis[9,24]). Excess complexity is unavoidable, owing to the need for functional groups to activate bond-forming processes. Most such functional groups are not present in the target and must be modified or removed in extra, yield-sapping steps. The reactions discussed above were chosen to reflect the invaluable role played by transition metals in modern synthetic chemistry.[42] They can dramatically decrease excess complexity, since auxiliary functional groups are usually not needed. Furthermore, they can serve as templates to enable the incorporation of three or more reactants in a single step,[43] as in the Reppe–Vollhardt,[44] Pauson–Khand,[45] and Dötz–Wulff[46] reactions. These construction reactions correspond to extremely powerful disconnections,[1] which can empower highly convergent and reflexive syntheses.

Synthetic complexity

Chemists have not always drawn a sharp distinction between molecular complexity and synthetic complexity,[6,7,21,41] which has been clarified as our series of papers on the subject has evolved.[12,14,15] For example, molecules with six-membered rings have been relatively easy to prepare because of well-developed synthetic methods such as the Diels–Alder cycloaddition,[47] Robinson annulation,[47] Birch reduction,[47] Posner MMRC process,[48] and the Reppe–Vollhardt reaction.[44] For this reason their synthetic complexities are relatively low.

The synthetic complexities of compounds with five- and seven-membered rings have been higher, since they have not been as accessible. However, this situation is changing, as advances in the Pauson–Khand,[45] Weiss–Cook,[49] and methylenecyclopropane–olefin [3 + 2]-cycloaddition[50] reactions are lowering the synthetic complexity of the former, while vinylcyclopropane–olefin [5 + 2],[51] oxyallyl–diene [4 + 3],[52] and π-allyl–alkyne–alkyne [3 + 2 + 2][53] cycloadditions are doing the same for the latter. Another potentially useful 7-ring synthesis is Hendrickson's pyrylium–olefin cycloaddition,[54] which generates useful functionality in the adduct that is ready for further construction reactions. Since their structures remain the same, the molecular complexities of these ring systems are not changing. What is evolving is the state of the art of synthetic chemistry, which largely determines synthetic complexity.

Another issue that has caused some confusion is the relationship between chirality and complexity. Since chirality affects the difficulty of construction and is a function of symmetry, it is an aspect of extrinsic (synthetic) complexity. As far as intrinsic (molecular) complexity is concerned, the *R* and *S* enantiomers of a chiral compound are equal. Nevertheless, one of them may have a lower synthetic complexity, for example, when it or a precursor is available from the 'chiral pool.'[55] The starting materials in Nicolaou *et al.*'s synthesis of the oligosaccharide portion of calicheamicin γ_1 (*L*-serine, *D*-galactose, *N*-hydroxyphthalimide, *D*-glucose, 3,4,5-trimethoxytoluene and *L*-rhamnose) were largely drawn from the chiral pool.[56] The synthetic complexity of chiral compounds is decreasing over time, as progress is made in developing new methods of asymmetric induction[57] and double asymmetric induction.[58] It can be argued that a mixture of *R* and *S* enantiomers (*e.g.*, the racemate) has a higher molecular complexity, or in this case supramolecular complexity, than either pure one. (*N.B.*, the racemic mixture has different physical properties from the constituent enantiomers.[47]) On the other hand, except for the chiral pool the synthetic complexity of the racemate is generally lower, which further illustrates the fundamental dichotomy between molecular complexity and synthetic complexity.

As an empirical measure of complexity for synthetic analysis, Whitlock introduced *S* [eqn. (4)], where *R* is the number of rings, *U* is the number of unsaturations, *A* is the number of asymmetric centers and *H* is the number of heteroatoms.[41] As far as *U* is concerned, a triple bond is counted as two double bonds and aromatic unsaturations are not included in his scheme. Whitlock used *S* to construct complexity *versus* step plots, which were analyzed in terms of ΔS_m, the average magnitude of the change in *S* per step, calculated from eqn. (5), where ΔS_j ($j = 1, \ldots, k$) is the change in *S* for the *j*th step and *k* is the total number of steps. In analogy to C_x, we define S_x as the excess complexity based on *S* (or excess *S*) of a synthesis with *z* intermediates, calculated by using eqn. (6), where S_y is the *S* value for intermediate I_y ($y = 1, \ldots, z$). Since the coefficients depend on the perceived synthetic difficulty (*i.e.*, the state of the art), we believe that *S* is best viewed as an index of synthetic complexity. Thus, indices of molecular complexity or synthetic complexity can be used to calculate excess complexity and the interpretation is similar, with the *caveat* that empirical indices of the latter type depend upon the state of the art, while the former do not. Barone and Chanon have introduced another linear combination approach with more terms,[5] and they find it closely parallels $C(\eta,\varepsilon)$. Fuchs has introduced the concept of intricacy,[6] which is closely related to complexity. (One of the dictionary definitions of complex is intricate.)

$$S = 4R + 2U + 2A + H \quad (4)$$

$$\Delta S_m = \Sigma_j |\Delta S_j| / k \quad (5)$$

$$S_x = \Sigma_y \, S_y \quad (6)$$

Table 2 Analytical functions for routes 1–3

Function[a]	N_T	N_S	T_T	η	$C(\eta,\varepsilon)$	S
Route 1						
C_x (S_x)	261	25	105	15	69.4	10
ΔC_m (ΔS_m)	238.0	19.5	84.5	10.5	59.2	6.0
Route 2						
C_x (S_x)	100	28	72	16	66.1	12
ΔC_m (ΔS_m)	22.7	6.0	14.5	2.5	18.7	1.0
Route 3						
C_x (S_x)	489	145	299	62	357.0	34
ΔC_m (ΔS_m)	36.7	9.5	19.5	1.7	21.7	1.7

[a] When C_x (ΔC_m) is calculated from *S* (ΔS), it is called S_x (ΔS_m).

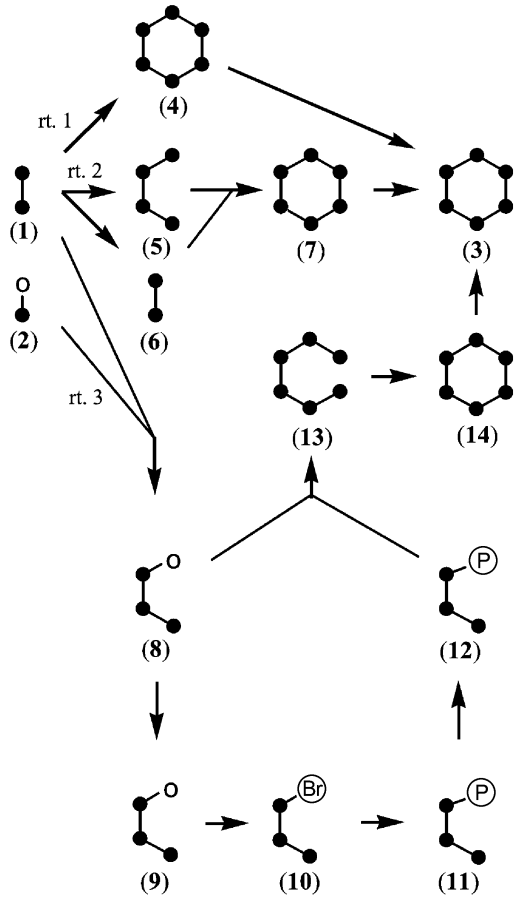
The analytical functions C_x, ΔC_m, S_x, and ΔS_m for routes 1–3, derived from selected complexity indices, are summarized in Table 2. The values of complexity *C* and changes ΔC used in the calculations are summarized in the ESI. While they are not usually included,[41] we count aromatic unsaturations when calculating *S* values for route 1, since benzene is the only intermediate, and $S_x = 0$ without it. Using $C = N_T$, T_T or $C(\eta,\varepsilon)$ to calculate excess complexity, the order of routes is C_x(route 2) < C_x(route 1) < C_x(route 3). Excess *S* gives a different order, with S_x(route 1) < S_x(route 2) < S_x(route 3), which is the same as that for C_x based on $C = N_S$ or η. It is interesting to note that for all indices except N_T and T_T, C_x(route 1) ≈ C_x(route 2).

With one exception the trends in ΔC_m parallel those in ΔS_m, *viz.* ΔS_m(route 1) > ΔS_m(route 3) > ΔS_m(route 2); η gives ΔC_m(route 1) > ΔC_m(route 2) > ΔC_m(route 3). (*N.B.*, whereas it is desirable to minimize C_x, it is desirable to maximize ΔC_m.) Excess complexity (C_x or S_x) identifies route 3 as the worst of those discussed here; however, in all cases but one (ΔC_m based on η), ΔS_m and ΔC_m fail to red-flag route 3, since they do not differentiate positive from negative changes. All of the functions identify the best synthesis as either route 1 or route 2, which was also the conclusion based on heuristics (previous section). This congruence suggests that the heuristics may have a mathematical (topological) basis.

The repeated appearance of identical structures is the salient feature of the topological synthesis plans Γ for routes 1–3 (Scheme 2). One can clearly see the inefficiencies in these routes, caused by functional group interchanges that do not increase target-relevant complexity and reduction reactions that do not control stereochemistry. Even this simple example puts the lie to statements that chemistry is a 'mature' science that does not need further development itself (and that consequently it has lost its 'identity').[59] Many additional construction reactions are needed to give synthetic chemists dramatically more diversity in initial and final functionality, so that adjustments will not be needed, and the ideal of using only construction reactions will become a reality.[24] (As a practical matter, we will approach this state asymptotically.) Finally, increased diversity will allow industrial chemists to find more 'green' conditions for chemical manufacturing,[60] for example, the importance of 'alternative synthetic design' for pollution prevention has been emphasized.[61]

Linear and convergent syntheses

In a *linear synthesis* relatively small molecules are added sequentially to the gradually growing chain of intermediates until the target is reached. For example, a 27-step linear synthesis of phorboxazole B *via* 'complex Aldol reactions' has been reported by Evans *et al.*[62] A classic example of linear synthesis is the preparation of polypeptides by the Merrifield solid phase method,[63] which involves the iterative addition of (protected) amino acids to the growing chain, anchored to a polymer bead.



Scheme 2 Topological synthesis plans for routes 1–3

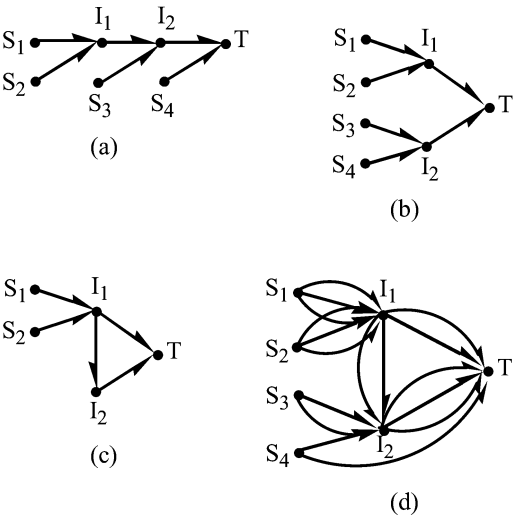


Fig. 2 (a) Linear, (b) convergent, (c) reflexive and (d) composite synthesis digraphs.

High molecular weight polymers such as polyethylene, polystyrene and polybutadiene are prepared *via* reactions that repeatedly add monomers and are, in fact, extremely efficient linear syntheses.

In a *convergent synthesis* two molecules of approximately the same size,[64] or in our scheme the same complexity,[9] are joined in the key step by the strategic reaction (σ_s). Heuristically, a convergent synthesis is more efficient than a linear one of the same target.[64–67] As explained by Hendrickson:[65] "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." This argument is similar to our rationale for the minimization of excess complexity (*vide supra*).

The synthesis of *N*-glycopeptides provides the perfect illustration, since the linear approach suffers from the fact that "some of the *O*-glycosidic bonds present in complex oligosaccharides are not completely stable to the acidolytic deprotection conditions normally used in peptide synthesis."[68] Convergent plans also enable significant time savings, provided that the personnel and equipment are available to run parallel reactions. Perhaps the most famous example is vitamin B_{12}, which was divided between the Woodward and Eschenmoser groups.[69] Each half was further subdivided so that, overall, the target was quartered, and the synthesis was triply convergent. Other important compounds prepared by convergent syntheses include steroids,[44,46,64] prostaglandins,[70] and antibiotics.[71] Both convergent solution-phase and linear solid-phase syntheses of combinatorial libraries have been demonstrated by Boger *et al.*[72] Convergent syntheses have also been performed on solid supports.[73]

A synthesis plan for target T can be abstracted as a *synthesis digraph* (directed graph) $D_{syn}(T)$, in which the points represent compounds and the arcs (directed lines) represent reactions.[1] Synthesis digraphs for linear and convergent routes to a target are shown in Fig. 2(a) and (b), respectively. Since they do not have multiple arcs or cycles, these synthesis digraphs are trees; therefore, each arc is a *bridge*, meaning its removal results in a disconnected (multi-component) graph. Whether linear or convergent, such synthesis plans are *vulnerable*:[11] if a single reaction fails and a replacement cannot be found, then the synthesis cannot be completed. Corey and Cheng refer to this situation as the 'Damocles sword of synthesis.'[21] A heuristic such as convergence is not a law of nature and exceptions are expected; thus, Warren cautioned, "Convergent syntheses are better than linear ones only if all other things are equal. There is no magic about convergence and a bad step can be just as disastrous here as elsewhere."[66] Sometimes, all other things are *not* equal:[29] Nicolaou *et al.* reported a linear synthesis "after several abortive attempts to construct the hemibrevetoxin B polycyclic skeleton by a convergent approach ...."[74] The convergent and reflexive approach to quassin failed because the key intramolecular Diels–Alder reaction did not work as planned (*cf.* next section).[75]

The *composite synthesis plan* is the superposition of all individual synthesis plans to a target T and analogously the *composite synthesis digraph* $\breve{D}_{syn}(T)$ is the union of all individual synthesis digraphs $D_{syn}(T)$.[1] Conversely, each $D_{syn}(T)$ [Fig. 2(a), (b) and (c)] is a subgraph of $\breve{D}_{syn}(T)$ [Fig. 2(d)].[1] To minimize the risk of catastrophic failure, the complexity of the composite plan should be maximized with the goal of eliminating bridges by providing alternative routes to the target. In addition, seeing all the routes together can lead to cross-pollination between them.

One way to estimate a plan's resistance to failure is to calculate its *robustness numbers* R_i [eqn. (7)], which are the maximum numbers of (directed) line-disjoint paths $p_{ld}(S_i, T)$ from starting material S_i to target T. (*N.B.*, line-disjoint paths can share points, but not lines.[1]) According to Menger's theorem, R_i is also equal to the minimum number of lines that must be removed to make it impossible to transit from S_i to T.[11] For $\breve{D}_{syn}(T)$ in Fig. 2(d), $R_1 = p_{ld}(S_1, T) = 3$, $R_2 = p_{ld}(S_2, T) = 3$, $R_3 = p_{ld}(S_3, T) = 2$ and $R_4 = p_{ld}(S_4, T) = 2$, which is a reasonably robust plan, since there are no bridges (*i.e.*, $R_i > 1$). The goal is to make R_{min}, the smallest value of R_i,

as large as possible for $\check{D}_{syn}$(T) (in this case $R_{min} = 2$.) Conceptually, the simplest way to accomplish this is to have several back-ups for every reaction in a synthesis plan, so that all the arcs are multiple ones. In practice, this is not so simple, as linear syntheses are often based on one kind of chemistry and convergent syntheses are usually based on one strategic reaction. When the key step in the synthesis plan for quassin was changed from σ_s(Diels–Alder) to σ_s(Aldol), the rest of the reactions had to be changed as well.[75] A good strategy is to screen δ_s for those with more than one τ_s, and then to choose the τ_s with the most highly developed σ_s.

$$R_i = p_{ld}(S_i, T) \quad (7)$$

$$R_t = \Sigma_i\ R_i \quad (8)$$

Another measure of resistance to failure is the *total robustness* R_t, the sum of the robustness numbers R_i [eqn. (8)], which should also be maximized: $R_t = 10$ for Fig. 2(d). If one abstracts the structures in Scheme 1 as points, then it becomes a composite synthesis digraph $\check{D}_{syn}$(**3**) with $R_1 = p_{ld}$(**1**, **3**) = 3, $R_2 = p_{ld}$(**2**, **3**) = 1, $R_{min} = 1$ and $R_t = 4$. Therefore, the failure of one reaction can scuttle a synthesis of **3**, when the unique path between **2** and **3** is part of it. The composite plan could be substantially improved by simply adding another path from **2** to **3**, for example, the Liebeskind quinone synthesis applied to the preparation of *p*-benzoquinone from **1** and **2**,[76] followed by exhaustive reduction.[77] Finally, the conversion of **8** to **13** *via* electrodimerization and diol elimination has been reported.[78]

Reflexive syntheses

A *reflexive synthesis* is one for which there is symmetry in the synthesis digraph and *reflexivity* is defined as economy in a synthesis that results from symmetry in its synthesis digraph.[10] Reflexive syntheses can be linear or convergent; consequently, four cases are possible: convergent and reflexive, convergent (non-reflexive), linear and reflexive, and linear (non-reflexive). Generally speaking, the first is the most efficient and the last is the least. The two middle cases do not have a universal order; there have been convergent (non-reflexive) syntheses of peptides that are not as efficient as the Merrifield synthesis,[63] which is linear and reflexive. It is the reflexive nature of the latter that is responsible for its great efficiency (*vide infra*).

A reflexive synthesis plan results in greater economies of time and effort than the corresponding non-reflexive one, whether it is linear or convergent.[10,12,14] The efficiency of the reflexive plan can be demonstrated by comparing it to the non-reflexive plan that is as similar as possible, that is the reactions must be essentially the same in both cases. Our approach is to compare the synthesis of a compound with the same synthesis of an isotopically labeled analog. (Kinetic isotope effects can be minimized by judicious selection of the labeled atom.) Thus, water **15** or oxygen-labeled (^{17}O or ^{18}O) water **15*** reacts with acetylene **1** to afford acetaldehyde **16** or oxygen-labeled acetaldehyde **16***, respectively.[79] The (directed) aldol reaction of **16** with itself or with **16*** yields aldol **17** or labeled aldol **17***, respectively.[1] The reflexive synthesis digraph D_{syn}(**17**) can be collapsed into a simpler one (Fig. 3), whereas D_{syn}(**17***), which is not reflexive, cannot be simplified. It is easy to see how the amounts of time and effort required for one person to prepare **17** are approximately two-thirds of that for **17***, as two reactions need to be run instead of three. As a limit, the excess complexity of the reflexive route approaches $(1/c)$ times that of the non-reflexive one, where c is the number of components joined. This leads to the heuristic principle of maximization of reflexivity.

A convergent synthesis becomes reflexive when the molecules joined in σ_s are identical (or are prepared from identical intermediates). When there is symmetry in the target molecule

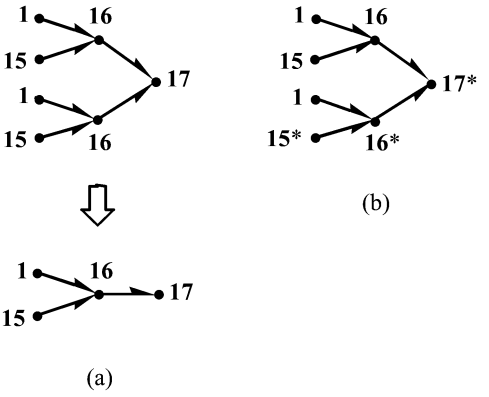


Fig. 3 (a) Reflexive and (b) non-reflexive routes to **17** and **17***, respectively.

T or its molecular graph M(T), a disconnection corresponding to a reflexive synthesis is usually obvious, as in the syntheses of a growing number of natural products with a rotational axis.[80] Alvarez and Serratosa have given symmetry rules for the retrosynthetic analysis of symmetrical and unsymmetrical targets.[81] It is possible that the symmetry in M(T) is not present in the corresponding three-dimensional structure of T, owing to conformational differences between substructures; therefore, the use of (unlabeled) molecular graphs in retrosynthetic analysis is the most general approach (*vide infra*).

There is a second kind of reflexivity in which no element of symmetry is present in T or M(T), but it is nevertheless possible to find a τ_s that results in two or more identical components. Barton *et al.*'s classic synthesis of usnic acid[82] was used to illustrate this situation in the communication that introduced reflexivity.[10] Subsequent discussions focused on the synthesis of carpanone,[12,21] which was prepared *via* an intramolecular hetero-Diels–Alder reaction.[83] Since then, some excellent examples of the difference between the two kinds of reflexive targets have appeared: papuamine has a C_2 axis, but haliclonadiamine, which is identical except for one epimeric C–N bond, does not; nevertheless, a reflexive synthesis is possible in both cases.[84] In petrosin two homochiral moieties are tethered together by two five-carbon chains and a C_2 axis is present; in petrosin A the two moieties are enantiomers and an S_2 axis (inversion center) is present; in petrosin B they are enantiomers but no symmetry element is present, owing to a different relative configuration of the chains.[85] Syntheses with varying degrees of reflexivity are possible for all three. Reflexive syntheses of the second kind can be more difficult to discern; nevertheless, their number is growing, too.[86] Whether or not there is symmetry in T or M(T), all reflexive synthesis plans have symmetry in D_{syn}(T).

An effective general approach for finding reflexive syntheses is to look for isomorphic subgraphs in M(T) and $M(I_i)$, the molecular graphs of target T and intermediates I_i generated from it by retrosynthetic analysis. Isomorphic graphs have the same adjacency matrix for some labeling (*cf.* Prolegomenon[1]). The 'all possible substructures' approach to the complexity of molecules does not explicitly include conformations, which may be a drawback for some applications, but is advantageous for problems such as this one. If conformational information were to be included, many opportunities for synthetic efficiency would be missed, for example, chair and boat cyclohexane are distinct in three dimensions, but they become isomorphic when abstracted as graphs.[1]

To demonstrate the application of this method, we look for disjoint isomorphic (DI) subgraphs S_i^{DI} in M(aldol), the molecular graph of aldol (*cf.* Table 1). (Disjoint subgraphs do not share points or lines.[1]) The largest possible DI subgraphs have $n/2$ points, when M has an even number of points n, and

$(n-1)/2$ points when n is odd. In this case $n/2 = 3$, and there are three kinds of subgraphs on 3 points (entries 12–14). The only DI ones are two pairs of ethanol subgraphs, out of a total of four pairs. (Either of the two ethanol subgraphs in the acetaldehyde subgraph can be taken with the distal ethanol subgraph in the 2-propanol subgraph.)

This prompts us to look for an *isomorphic transform* τ_{iso} that preserves at least one DI subgraph in each of the resulting components and $\tau_{iso} = \tau$(Aldol) fulfills this requirement. In this case the (two acetaldehyde) molecules resulting from τ_{iso} are identical; nevertheless, the analysis can be repeated on them to see whether the synthesis can be made multiply reflexive. In those cases where there are no DI subgraphs in M(T), the same analysis can be performed on $M(I_i)$, the intermediates generated by retrosynthetic analysis.

In general, the components from τ_{iso} need not be the same. The goal is then to find further transforms that will give identical molecules, for example, one of the components resulting from τ_{iso} is transformed to match the other one, or they are both transformed until identical molecules are obtained. Since there is no symmetry in aldol, this example illustrates the most general case. The extension to three or more molecules from τ_{iso} is straightforward; benzene has three DI ethylene subgraphs, which are preserved in the three acetylene molecules from $\tau_{iso} = \tau$(Reppe–Vollhardt) (*cf.* route 1, Scheme 1). Benzene also has three pairs of DI propene subgraphs, any one of which forms the basis of route 3, which is reflexive, since intermediate **8** is used twice. This method is formally introduced in a separate communication.[87]

Dramatic examples of reflexivity are found in polymer synthesis, where 'living' reactions can be considered linear reflexive syntheses, and condensation polymerizations can be considered convergent reflexive syntheses. In the former method, monomers are added sequentially to the 'live' end of the growing chain until the reaction is terminated.[47] If it is assumed that $(\text{monomer})_n \approx (\text{monomer})_{n-1}$, then each propagation step has the same reactants (monomer and polymer chain), and the synthesis is highly reflexive. This approximation gets better as the chain grows longer. (A polymer has translational symmetry, rather than the point group symmetry of small molecules.) Likewise, the Merrifield polypeptide synthesis is reflexive because to a first approximation the amino acids are all treated in the same way and the n-mer of the growing chain is essentially the same as the $(n-1)$-mer. Reflexivity allowed the Merrifield synthesis to be automated before sophisticated laboratory robots were available and it has revolutionized the synthesis of polypeptides,[63] oligonucleotides,[47,63] and oligosaccharides.[88]

All of the linear syntheses discussed above are 'one-directional,' as the substrate is modified at one site per step. A related, reflexive strategy is 'two-directional synthesis,' in which a molecule is modified at two sites in the same step.[89] This concept can be extended to the n-directional synthesis of star molecules,[90] and it can be further elaborated to the limiting case of dendrimers,[91] where each stage in propagation creates a new star. A spectacularly efficient dendrimer synthesis has been reported by Hecht and Fréchet.[92] The first stage involves the two-directional synthesis of 'convergent' dendrons at both ends of 2-butyne-1,4-diol, and the second stage involves trimerization of this monodisperse intermediate *via* σ(Reppe–Vollhardt).

The archetypes 'linear' and 'convergent' are limiting cases and different degrees of convergence are possible. Thus, most convergent syntheses have linear segments and many linear syntheses add at least one molecule that is significantly more complex than the rest. A hybrid 'segment synthesis–condensation' approach has been described by Kaiser *et al.*,[93] in which peptide segments are prepared linearly on a solid support and then condensed in solution in the convergent step. This approach can be multiply reflexive, since one segment can be used in two or more of the condensation steps of a multiply convergent route. (Often, the repeat segment is lengthened before the subsequent condensation.) The hybrid approach has been applied to the synthesis of the amyloid protein of Alzheimer's disease, which is difficult to obtain in pure form when prepared by the strictly linear route.[94] Another interesting hybrid synthesis is the preparation of glycopeptides, in which linear syntheses of oligosaccharides and polypeptides converge in the final condensation step and the only additional reaction needed to reach the target is deprotection.[68] These syntheses approach the optimal situation in which σ_s gives the target directly without additional modification of the skeleton or functionality. Considering the above examples, it appears that reflexivity is higher in the hierarchy of heuristics than convergence, since it makes both convergent and linear syntheses more efficient.

Conclusion

R.B. Woodward exhorted,[95] "Synthesis must always be carried out by plan, and the synthetic frontier can be defined only in terms of the degree to which realistic planning is possible, utilizing all of the intellectual and physical tools available." We have introduced a number of new intellectual tools: complexity indices, complexity *versus* step plots, excess complexity, composite synthesis digraphs, robustness and reflexivity. Synthetic complexity has been differentiated from molecular complexity, as the former changes with the state of the art, whereas the latter does not. Symmetry can be a simplifying factor for synthetic complexity, for example symmetry in the synthesis digraph can lead to efficient reflexive syntheses. Maximizing the complexity of the composite synthesis plan helps to maximize the probability of reaching the target, while minimizing the excess complexity of an individual synthesis plan helps to maximize its overall yield.

Acknowledgements

The author thanks T. Sommer for helpful discussions and W.F. Wright for valuable literature searches.

References

1 Contribution 20 in a series on applications of discrete mathematics to chemistry. Number 19 is the previous paper in this issue: S. H. Bertz, *New J. Chem.*, 2003, DOI: 10.1039/b210843g.
2 (*a*) L. B. Kier and B. Testa, *Adv. Drug Res.*, 1995, **26**, 1–43; (*b*) D. Bonchev and W. A. Seitz, in *Concepts in Chemistry: A Contemporary Challenge*, ed. D. H. Rouvray, Wiley, New York, 1997, pp. 353–381; (*c*) *Complexity: Metaphors, Models, and Reality*, eds. G. A. Cowan, D. Pines and D. Meltzer, Addison-Wesley, Reading, MA, USA, 1994; (*d*) W. Y. Tam, in *1990 Lectures in Complex Systems*, eds. L. Nadel and D. L. Stein, Addison-Wesley, Redwood City, CA, USA, 1991, pp. 99–123; (*e*) I. R. Epstein, in *1989 Lectures in Complex Systems*, ed. E. Jen, Addison-Wesley, Redwood City, CA. USA, 1990, pp. 213–269; (*f*) P. G. Wolynes, in *Lectures in the Sciences of Complexity*, ed. D. L. Stein, Addison-Wesley, Redwood City, CA, USA, 1989, pp. 355–388.
3 (*a*) K. Ośmialowski and R. Piekośe, *J. Solution Chem.*, 1991, **20**, 241–248; (*b*) M. Thompson and U. J. Krull, *Anal. Chem.*, 1991, **63**, A393–405; (*c*) B. Lang and T. Grochowski, *Meteoritics*, 1989, **24**, 290–291; (*d*) M. D'Amboise and M. J. Bertrand, *J. Chromatogr.*, 1986, **361**, 13–24.
4 M. Chanon, R. Barone, C. Baralotto, M. Juillard and J. B. Hendrickson, *Synthesis*, 1998, 1559–1583.
5 (*a*) R. Barone, M. Petitjean, C. Baralotto, P. Piras and M. Chanon, *J. Phys. Org. Chem.*, 2003, **16**, 9–15; (*b*) R. Barone and M. Chanon, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 269–272.
6 P. L. Fuchs, *Tetrahedron*, 2001, **57**, 6855–6875.
7 S. H. Bertz, *J. Am. Chem. Soc.*, 1981, **103**, 3599–3601.

8 (*a*) S. H. Bertz, *J. Chem. Soc., Chem. Commun.*, 1981, 818–820; (*b*) S. H. Bertz, *Bull. Math. Biol.*, 1983, **45**, 849–855; (*c*) S. H. Bertz, *Discrete Appl. Math.*, 1988, **19**, 65–83.
9 (*a*) S. H. Bertz, *J. Am. Chem. Soc.*, 1982, **104**, 5801–5803; (*b*) S. H. Bertz, in *Chemical Applications of Topology and Graph Theory*, ed. R. B. King, Elsevier, Amsterdam, 1983, pp. 206–221.
10 S. H. Bertz, *J. Chem. Soc., Chem. Commun.*, 1984, 218–219.
11 S. H. Bertz, *J. Chem. Soc., Chem. Commun.*, 1986, 1627–1628.
12 S. H. Bertz and T. J. Sommer, in *Organic Synthesis: Theory and Applications*, ed. T. Hudlicky, JAI Press, Greenwich, CT, USA, 1993, vol. 2, pp. 67–92.
13 (*a*) S. H. Bertz and T. J. Sommer, *Chem. Commun.*, 1997, 2409–2410; (*b*) S. H. Bertz and W. C. Herndon, in *Artificial Intelligence Applications in Chemistry*, ACS Symposium Series 306, eds. T. H. Pierce and B. A. Hohne, American Chemical Society, Washington, DC, 1986, pp. 169–175.
14 S. H. Bertz and W. F. Wright, *Graph Theory Notes of New York (NY Acad. Sci.)*, 1998, **XXXV**, 32–48.
15 S. H. Bertz and C. M. Zamfirescu, *Commun. Math. Comput. Chem. (MATCH)*, 2000, **42**, 39–70.
16 S. H. Bertz, *Chem. Commun.*, 2001, 2516–2517.
17 (*a*) S. H. Bertz, *Tetrahedron Lett*, 1983, **24**, 5577–5580; (*b*) see also: T.-L. Ho, *Symmetry: A Basis for Synthesis Design*, Wiley, New York, 1995.
18 D. J. Milner, *J. Chem. Technol. Biotechnol.*, 1995, **63**, 301–312.
19 A *heteroalkane* is an alkane where h carbon atoms have been replaced by heteroatoms, which are represented by coloring the points of the molecular graph with a unique color label for each element. Alkanes are subsumed as the trivial case for which $h = 0$.
20 J.-M. Lehn, *Supramolecular Chemistry: Concepts and Perspectives*, VCH, Weinheim, 1995.
21 (*a*) E. J. Corey and X.-M. Cheng, *The Logic of Chemical Synthesis*, Wiley, New York, 1989; (*b*) see also: E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak and G. Petersson, *J. Am. Chem. Soc.*, 1975, **97**, 6116–6124.
22 (*a*) G. Bernáth, G. Stájer, F. Fülöp and P. Sohár, *J. Heteocycl. Chem.*, 2000, **37**, 439–449; (*b*) A. J. H. Klunder, J. Zhu and B. Zwanenburg, *Chem. Rev.*, 1999, **99**, 1163–1190; (*c*) N. E. Schore, *Synth. Commun.*, 1979, **9**, 41–47.
23 T. D. Aicher, K. R. Buszek, F. G. Fang, C. J. Forsyth, S. H. Jung, Y. Kishi, M. C. Matelich, P. M. Scola, D. M. Spero and S. K. Yoon, *J. Am. Chem. Soc.*, 1992, **114**, 3162–3164.
24 (*a*) J. B. Hendrickson, *Chemtech*, 1998, September, pp. 35–40; (*b*) J. B. Hendrickson and P. Huang, *J. Chem. Inf. Comput. Sci.*, 1989, **29**, 145–151.
25 P. A. Wender, S. T. Handy and D. L. Wright, *Chem. Ind.*, 1997, 765–769.
26 P. A. Wender and B. L. Miller, in *Organic Synthesis: Theory and Applications*, ed. T. Hudlicky, JAI Press, Greenwich, CT, 1993, vol. 2, pp. 27–66.
27 (*a*) B. M. Trost, *Science*, 1991, **254**, 1471–1477; (*b*) B. M. Trost, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 259–281.
28 The *C–s* plot for a one-step reaction is a trapezoid with area $\frac{1}{2}\Delta s\,[C(\mathrm{A}) + C(\mathrm{B})] = \frac{1}{2}C(\mathrm{A}) + \frac{1}{2}C(\mathrm{B})$, as $\Delta s = 1$. More elaborate *C–s* plots can be integrated by summing trapezoids, see ref. 9.
29 (*a*) See also ref. 66; (*b*) In the derivation of eqn. (3), we assume that the amounts of side-products are low, so that their complexities can be ignored.
30 (*a*) ^{11}C: no reports; (*b*) ^{13}C: D. Ziessow and M. Carroll, *Ber. Bunsenges. Phys. Chem.*, 1972, **76**, 61–64; (*c*) ^{14}C: U. Kosakowsky and R. Russow, *Isotopenpraxis*, 1986, **22**, 280–284.
31 (*a*) I. M. Abdelrehim, K. Pelhos, T. E. Madey, J. Eng and J. G. Chen, *J. Phys. Chem. B*, 1998, **102**, 9697–9707; (*b*) M. J. Lambregts, E. J. Munson, A. A. Kheir and J. F. Haw, *J. Am. Chem. Soc.*, 1992, **114**, 6875–6879.
32 M. Castiglioni, S. Deabate, E. Garrone, R. Giordano, B. Onida, G. Predieri and E. Sappa, *J. Cluster Sci.*, 1997, **8**, 381–405.
33 W. Büchele, H. Roos, H. Wanjek and H. J. Müller, *Catal. Today*, 1996, **30**, 33–39.
34 G. M. Whitman, *Br. Pat.* GB 616 671, 1949.
35 S. N. Orlik, Y. I. Pyatnitsky and G. G. Girushtin, *Ukr. Khim. Zh.*, 1992, **58**, 737–742.
36 Y. Nagase and T. Katou, *Chem. Lett.*, 2000, 436–437.
37 J.-M. Mas and P. Metivier, *Eur. Pat.* EP 398 774, 1990.
38 The reaction of **12** with a β-substituted acrolein affords a *ca.* 1:1 *Z*/*E* mixture, see: T. Ando and H. Ohsawa, *J. Chem. Ecol.*, 1993, **19**, 119–132.
39 S. Sakai and S. Takane, *J. Phys. Chem. A*, 1999, **103**, 2878–2882.
40 (*a*) E. Chanceller, M. L. García, A. Moyano, M. A. Pericàs and F. Serratosa, *Tetrahedron*, 1986, **42**, 1831–1839; (*b*) F. Serratosa, *Organic Chemistry in Action: The Design of Organic Synthesis*, Elsevier, Amsterdam, 1990, pp. 19–22
41 H. W. Whitlock, *J. Org. Chem.*, 1998, **63**, 7982–7989; note that eqn. (4) is our formulation of his theory, which was written in 'computerese'.
42 (*a*) K. Weissermel and H.-J. Arpe, *Industrial Organic Chemistry*, Wiley-VCH, New York, 1997; (*b*) L. S. Hegedus, *Transition Metals in the Synthesis of Complex Organic Molecules*, University Science Books, Mill Valley, CA, 2nd edn., 1999; (*c*) W. A. Nugent, T. V. RajanBabu and M. J. Burk, *Science*, 1993, 259, 479–483.
43 (*a*) B. M. Trost, *Pure Appl. Chem.*, 1988, **60**, 1615–1626; (*b*) *Templated Organic Synthesis*, eds. F. Diederich and P. J. Stang, Wiley–VCH, Weinheim, 2000.
44 (*a*) K. P. C. Vollhardt, *Angew. Chem., Int. Ed. Engl.*, 1984, **23**, 539–556; (*b*) W. Reppe, N. v. Kutepow and A. Magin, *Angew. Chem., Int. Ed. Engl.*, 1969, **8**, 727–733.
45 (*a*) P. L. Pauson, *Tetrahedron*, 1985, **41**, 5855–5860; (*b*) N. E. Schore, *Org. React.*, 1991, **40**, 1–90; (*c*) K. M. Brummond and J. L. Kent, *Tetrahedron*, 2000, **56**, 3263–3283.
46 (*a*) J. Bao, W. D. Wulff, V. Dragisich, S. Wenglowsky and R. G. Ball, *J. Am. Chem. Soc.*, 1994, **116**, 7616–7630; (*b*) see also: K. H. Dötz, *Angew. Chem., Int. Ed. Engl.*, 1984, **23**, 587–608.
47 G. Solomons and C. Fryhle, *Organic Chemistry*, Wiley, New York, 7th edn., 2000.
48 (*a*) G. H. Posner, J. P. Mallamo and A. Y. Black, *Tetrahedron*, 1981, **37**, 3921–3926; (*b*) see also: G. H. Posner and E. M. Shulman-Roskes, *J. Org. Chem.*, 1989, **54**, 3514–3515 and references cited therein.
49 (*a*) S. H. Bertz, J. M. Cook, A. Gawish and U. Weiss, *Org. Synth.*, 1986, **64**, 27–38; (*b*) A. K. Gupta, X. Fu, J. P. Snyder and J. M. Cook, *Tetrahedron*, 1991, **47**, 3665–3710.
50 (*a*) B. M. Trost and R. I. Higuchi, *J. Am. Chem. Soc.*, 1996, **118**, 10 094–10 105; (*b*) C. C. Huval and D. A. Singleton, *J. Org. Chem.*, 1994, **59**, 2020–2024.
51 (*a*) P. A. Wender, A. J. Dyckman, C. O. Husfeld, D. Kadereit, J. A. Love and H. Rieck, *J. Am. Chem. Soc.*, 1999, **121**, 10 442–10 443; (*b*) P. A. Wender, C. O. Husfeld, E. Langkopf, J. A. Love and N. Pleuss, *Tetrahedron*, 1998, **54**, 7203–7220.
52 (*a*) F. G. West, C. Hartke-Karger, D. J. Koch, C. E. Kuehn and A. M. Arif, *J. Org. Chem.*, 1993, **58**, 6795–6803; (*b*) T. R. Barbee and K. F. Albizati, *J. Org. Chem.*, 1991, **56**, 6764–6773.
53 (*a*) N. Etkin, T. L. Dzwiniel, K. E. Schweibert and J. M. Stryker, *J. Am. Chem. Soc.*, 1998, **120**, 9702–9703; (*b*) K. E. Schwiebert and J. M. Stryker, *J. Am. Chem. Soc.*, 1995, **117**, 8275–8276.
54 J. B. Hendrickson and J. S. Farina, *J. Org. Chem.*, 1980, **45**, 3359–3361.
55 (*a*) G. Casiraghi, F. Zanardi, G. Rassu and P. Spanu, *Chem. Rev.*, 1995, **95**, 1677–1716; (*b*) H.-U. Blaser, *Chem. Rev.*, 1992, **92**, 935–952.
56 (*a*) K. C. Nicolaou, R. D. Groneberg, T. Miyazaki, N. A. Stylianides, T. J. Schulze and W. Stahl, *J. Am. Chem. Soc.*, 1990, **112**, 8193–8195; (*b*) K. C. Nicolaou, C. W. Hummel, E. N. Pitsinos, M. Nakada, A. L. Smith, K. Shibayama and H. Saimoto, *J. Am. Chem. Soc.*, 1992, **114**, 10 082–10 084; (*c*) see also: R. L. Halcomb, S. H. Boyer, M. D. Wittman, S. H. Olson, D. J. Denhart, K. K. C. Liu and S. J. Danishefsky, *J. Am. Chem. Soc.*, 1995, **117**, 5720–5749; (*d*) S. A. Hitchcock, S. H. Boyer, M. Y. Chu-Moyer, S. H. Olson and S. J. Danishefsky, *Angew. Chem., Int. Ed. Engl.*, 1994, **33**, 858–862.
57 (*a*) W. S. Knowles, *Angew. Chem., Int. Ed.*, 2002, **41**, 1999–2007; (*b*) R. Noyori, *Angew. Chem., Int. Ed.*, 2002, **41**, 2008–2022; (*c*) K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2002, **41**, 2024–2032; (*d*) M. Sawamura and Y. Ito, *Chem. Rev.*, 1992, **92**, 857–871; (*e*) F. A. Davis and B.-C. Chen, *Chem. Rev.*, 1992, **92**, 919–934; (*f*) J. K. Whitesell, *Chem. Rev.*, 1992, **92**, 953–964; (*g*) R. L. Halterman, *Chem. Rev.*, 1992, **92**, 965–994.
58 (*a*) S. Masamune, W. Choy, J. S. Petersen and L. R. Sita, *Angew. Chem., Int. Ed. Engl.*, 1985, **24**, 1–30; (*b*) K. Miyashita, H. Miyabe, K. Tai, H. Iwaki and T. Imanishi, *Tetrahedron*, 2000, **56**, 4691–4700; (*c*) T. Ziegler and G. Lemanski, *Eur. J. Org. Chem.*, 1998, 163–170; (*d*) D. A. Evans, M. G. Yang, M. J. Dart, J. L. Duffy and A. S. Kim, *J. Am. Chem. Soc.*, 1995, **117**, 9598–9599; (*e*) C. Martin, W. Prünck, M. Bortolussi and R. Bloch, *Tetrahedron: Asymmetry*, 2000, **11**, 1585–1592; (*f*) R. A. Fernandes and P. Kumar, *Tetrahedron Lett.*, 2000, **41**, 10 309–10 312.
59 (*a*) R. Baum, *Chem. Eng. News*, 1993, November 29, pp.50–51; (*b*) R. Scheffold, *Chimia*, 1989, **43**, 37–38; (*c*) For the case for synthesis, see: D. Seebach, *Angew. Chem., Int. Ed. Engl.*, 1990, **29**, 1320–1367; (*d*) E. J. Corey, *The Chemist*, 1989, July/August, pp. 3–5.

60 (*a*) *Green Chemical Syntheses and Processes*, ACS Symposium Series 767, eds. P. T. Anastas, L. G. Heine and T. C. Williamson, ACS, Washington, D.C., 2000; (*b*) G. W. V. Cave, C. L. Raston and J. L. Scott, *Chem. Commun.*, 2001, 2159–2169; (*c*) M. A. Matthews, *Pure Appl. Chem.*, 2001, **73**, 1305–1308; (*d*) J. H. Clark, *Pure Appl. Chem.*, 2001, **73**, 103–111; (*e*) V. P. Bui, T. V. Hansen, Y. Stenstrøm and T. Hudlicky, *Green Chem.*, 2000, **2**, 263–265.

61 (*a*) P. T. Anastas, J. D. Nies and S. C. DeVito, in *Benign by Design*, ACS Symposium Series 577, eds. P. T. Anastas and C. A. Farris, ACS, Washington, D.C., 1994, pp. 166–184; (*b*) J. B. Hendrickson, in *Green Chemistry: Designing Chemistry for the Environment*, ACS Symposium Series 626, eds. P. T. Anastas and T. C. Williamson, ACS, Washington, D.C., 1996, pp. 214–231.

62 D. A. Evans, D. M. Fitch, T. E. Smith and V. J. Cee, *J. Am. Chem. Soc.*, 2000, **122**, 10 033–10 046.

63 (*a*) R. B. Merrifield, *Angew. Chem., Int. Ed. Engl.*, 1985, **24**, 799–810; (*b*) G. Barany, N. Kneib-Cordonier and D. G. Mullen, *Int. J. Pept. Protein Res.*, 1987, **30**, 705–739.

64 (*a*) L. Velluz, J. Valls and G. Nominé, *Angew. Chem., Int. Ed. Engl.*, 1965, **4**, 181–200; (*b*) L. Velluz, J. Valls and J. Mathieu, *Angew. Chem., Int. Ed. Engl.*, 1967, **6**, 778–789.

65 (*a*) J. B. Hendrickson, *J. Am. Chem. Soc.*, 1977, **99**, 5439–5450; (*b*) see also: J. B. Hendrickson, *Acc. Chem. Res.*, 1986, **19**, 274–281.

66 S. Warren, *Organic Synthesis: The Disconnection Approach*, Wiley, Chichester, 1982, p. 352.

67 S. I. Zavjalov, A. G. Zavozin, O. V. Dorofeeva and G. I. Ezhova, *Khim.-Farm. Zh.*, 1986, **20**, 329–334.

68 (*a*) S. T. Anisfeld and P. T. Lansbury, *J. Org. Chem.*, 1990, **55**, 5560–5562; (*b*) S. T. Cohen-Anisfeld and P. T. Lansbury, *J. Am. Chem. Soc.*, 1993, **115**, 10 531–10 537; (*c*) see also: D. Vetter, D. Tumelty, S. K. Singh and M. A. Gallop, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 60–63; (*d*) Z.-G. Wang, X. Zhang, D. Live and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2000, **39**, 3652–3656

69 (*a*) R. B. Woodward, *Pure Appl. Chem.*, 1968, **17**, 519–547; (*b*) R. B. Woodward, *Pure Appl. Chem.*, 1971, **25**, 283–304; (*c*) R. B. Woodward, *Pure Appl. Chem.*, 1973, **33**, 145–177; (*d*) A. Eschenmoser and C. E. Wintner, *Science*, 1977, **196**, 1410–1420; (*e*) R. B. Woodward, in *Frontiers of Bioorganic Chemistry and Molecular Biology*, eds. Y. A. Ovchinnikov and M. N. Kolosov, Elsevier, Amsterdam, 1979, pp. 519–547; (*f*) A. Eschenmoser, *Chem. Heritage*, 2001, **19**, 12–13 and 46–49.

70 (*a*) R. Davis and K. G. Untch, *J. Org. Chem.*, 1979, **44**, 3755–3759; (*b*) R. E. Donaldson, J. C. Saddler, S. Byrn, A. T. McKenzie and P. L. Fuchs, *J. Org. Chem.*, 1983, **48**, 2167–2188; (*c*) R. Noyori and M. Suzuki, *Angew. Chem., Int. Ed. Engl.*, 1984, **23**, 847–876.

71 (*a*) D. L. Boger and K. Machiya, *J. Am. Chem. Soc.*, 1992, **114**, 10 056–10 058; (*b*) M. A. Tius, X. Gu and J. Gomez-Galano, *J. Am. Chem. Soc.*, 1990, **112**, 8188–8189; (*c*) A. S. Kende, K. Kawamura and R. J. DeVita, *J. Am. Chem. Soc.*, 1990, **112**, 4070–4072.

72 D. L. Boger, W. Chai and Q. Jin, *J. Am. Chem. Soc.*, 1998, **120**, 7220–7225.

73 (*a*) C. W. Lindsley, L. K. Chan, B. C. Goess, R. Joseph and M. D. Shair, *J. Am. Chem. Soc.*, 2000, **122**, 422–423; (*b*) S. L. Schreiber, *Science*, 2000, **287**, 1964–1969; (*c*) J. Y. Roberge, X. Beebe and S. J. Danishefsky, *J. Am. Chem. Soc.*, 1998, **120**, 3915–3927.

74 K. C. Nicolaou, K. R. Reddy, G. Skokotas, F. Sato and X.-Y. Xiao, *J. Am. Chem. Soc.*, 1992, **114**, 7935–7936.

75 (*a*) Intramolecular Diels–Alder approach: L. Mandell, D. E. Lee and L. F. Courtney, *J. Org. Chem.*, 1982, **47**, 610–615; (*b*) Tandem Aldol/Aldol approach: P. DeCapite, *Dissertation*, 1984, Brandeis University; (*c*) For successful non-reflexive syntheses, see: intermolecular Diels–Alder: G. Vidari, S. Ferriño and P. A. Grieco, *J. Am. Chem. Soc.*, 1984, **106**, 3539–3548; (*d*) Intramolecular Diels–Alder: T. K. M. Shing and Q. Jiang, *J. Org. Chem.*, 2000, **65**, 7059–7069.

76 N. E. Schore, B. E. La Bella, M. J. Knudsen, H. Hope and X. J. Xu, *J. Organomet. Chem.*, 1984, **272**, 435–446 ; the (low) yield was not optimized.

77 E. Asada and T. Takiguchi, *Br. Pat.* GB 1 520 513, 1978; A one-step conversion has not been reported, nevertheless, it should be possible given the information herein.

78 H. P. Figeys and M. Gelbcke, *Bull. Soc. Chim. Belg.*, 1974, **83**, 381–392.

79 N. D. Lazo, J. L. White, E. J. Munson, M. Lambregts and J. F. Haw, *J. Am. Chem. Soc.*, 1990, **112**, 4050–4052.

80 (*a*) I. Fleming and S. K. Ghosh, *J. Chem. Soc., Perkin Trans. 1*, 1998, 2733–2747 (nonactin, S_4); (*b*) P. A. Bartlett, J. D. Meadows and E. Ottow, *J. Am. Chem. Soc.*, 1984, **106**, 5304–5311 (nonactin, S_4); (*c*) P. Wipf and C. P. Miller, *J. Am. Chem. Soc.*, 1992, **114**, 10 975–10 977 (westiellamide, C_3); (*d*) A. B. Smith III, C. M. Adams, S. A. Kozmin and D. V. Paone, *J. Am. Chem. Soc.*, 2001, **123**, 5925–5937 (cylindrocyclophane A, C_2); (*e*) T. R. Hoye, P. E. Humpal and B. Moon, *J. Am. Chem. Soc.*, 2000, **122**, 4982–4983 (cylindrocyclophane A, C_2); (*f*) A. B. Smith III, S. A. Kozmin and D. V. Paone, *J. Am. Chem. Soc.*, 1999, **121**, 7423–7424 (cylindrocyclophane F, C_2); (*g*) R. S. Coleman and E. B. Grant, *J. Am. Chem. Soc.*, 1995, **117**, 10 889–10 904 (phleichrome, calphostin A, C_2); (*h*) D. L. Boger and J.-H. Chen, *J. Am. Chem. Soc.*, 1993, **115**, 11 624–11 625 (sandramycin, C_2); (*i*) D. L. Boger and C. M. Baldino, *J. Am. Chem. Soc.*, 1993, **115**, 11 418–11 425 (isochrysohermidin, C_2); (*j*) J. D. White, T. R. Vedananda, M.-C. Kang and S. C. Choudhry, *J. Am. Chem. Soc.*, 1986, **108**, 8105–8107 (aplasmomycin, C_2); (*k*) E. J. Corey, D. H. Hua, B.-C. Pan and S. P. Seitz, *J. Am. Chem. Soc.*, 1982, **104**, 6816–6820 (aplasmomycin, C_2).

81 S. Alvarez and F. Serratosa, *J. Am. Chem. Soc.*, 1992, **114**, 2623–2630.

82 D. H. R. Barton, A. M. Deflorin and O. E. Edwards, *J. Chem. Soc.*, 1956, 530–534.

83 (*a*) O. L. Chapman, M. R. Engel, J. P. Springer and J. C. Clardy, *J. Am. Chem. Soc.*, 1971, **93**, 6696–6698. See also ref. 73(a) and the following references: (*b*) I. R. Baxendale, A.-L. Lee and S. V. Ley, *Synlett*, 2001, 1482–1484; (*c*) M. R. Iyer and G. K. Trivedi, *Bull. Chem. Soc. Jpn.*, 1992, **65**, 1662–1664; (*d*) A. Nishiyama, H. Eto, Y. Terada, M. Iguchi and S. Yamamura, *Chem. Pharm. Bull.*, 1983, **31**, 2834–2844.

84 (*a*) T. S. McDermott, A. A. Mortlock and C. H. Heathcock, *J. Org. Chem.*, 1996, **61**, 700–709; (*b*) R. M. Borzilleri, S. M. Weinreb and M. Parvez, *J. Am. Chem. Soc.*, 1995, **117**, 10 905–10 913; (*c*) see also: A. G. M. Barrett, M. L. Boys and T. L. Boehm, *J. Org. Chem.*, 1996, **61**, 685–699.

85 R. W. Scott, J. Epperson and C. H. Heathcock, *J. Org. Chem.*, 1998, **63**, 5001–5012.

86 (*a*) C. Li, E. Lobkovsky and J. A. Porco, *J. Am. Chem. Soc.*, 2000, **122**, 10 484–10 485 (torreyanic acid); (*b*) T. R. Kelly and R. L. Xie, *J. Org. Chem.*, 1998, **63**, 8045–8048 (taspin); (*c*) R. L. Parsons and C. H. Heathcock, *Synlett*, 1996, 1168–1170 (tantazole B); (*d*) J. M. Schkeryantz, J. C. G. Woo and S. J. Danishefsky, *J. Am. Chem. Soc.*, 1995, **117**, 7025–7026 (gypsetin); (*e*) M. J. Bishop and M. A. Ciufolini, *J. Am. Chem. Soc.*, 1992, **114**, 10 081–10 082 (kuanoniamine D); (*f*) J. D. White, M. A. Avery, S. C. Choudhry, O. P. Dhingra, B. D. Gray, M.-C. Kang, S.-C. Kuo and A. J. Whittle, *J. Am. Chem. Soc.*, 1989, **111**, 790–792 (boromycin); (*g*) see ref. 84(a) (haliclonadiamine); (*h*) see ref. 85 (petrosin B).

87 S. H. Bertz and T. J. Sommer, *Chem. Commun.*, 2003, 1000–1001.

88 S. J. Danishefsky, J. T. Randolph, J. Y. Roberge, K. F. McClure and R. B. Ruggeri, *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)*, 1994, **35**, 977–978.

89 (*a*) S. R. Magnuson, *Tetrahedron*, 1995, **51**, 2167–2213; (*b*) A. G. M. Barrett, D. C. Braddock, P. D. de Koning, A. J. P. White and D. J. Williams, *J. Org. Chem.*, 2000, **65**, 375–380; (*c*) N. Feuerbacher and F. Vögtle, *Top. Curr. Chem.*, 1998, **197**, 1–18.

90 Stars with n branches can be symbolized $K_{1,n}$ and one-directional and two-directional syntheses correspond to the trivial stars $K_{1,1}$ and $K_{1,2}$, respectively. Chemically, n is limited by the valence of the atom involved in the branchpoint. See also ref. 1.

91 (*a*) G. R. Newkome, C. N. Moorefield and F. Vögtle, *Dendrimers and Dendrons: Concepts, Syntheses, Applications*, Wiley–VCH, Weinheim, 2001; (*b*) see also: D. A. Tomalia, H. Baker, J. Dewald, M. Hall, G. Kallos, S. Martin, J. Roeck, J. Ryder and P. Smith, *Polym. J. (Tokyo)*, 1985, **17**, 117–132.

92 S. Hecht and J. M. J. Fréchet, *J. Am. Chem. Soc.*, 1999, **121**, 4084–4085.

93 (*a*) E. T. Kaiser, H. Mihara, G. A. Laforet, J. W. Kelly, L. Walters, M. A. Findeis and T. Sasaki, *Science*, 1989, **243**, 187–192; (*b*) E. T. Kaiser, *Acc. Chem. Res.*, 1989, **22**, 47–54.

94 J. C. Hendrix, K. J. Halverson and P. T. Lansbury, *J. Am. Chem. Soc.*, 1992, **114**, 7930–7931.

95 R. B. Woodward, in *Perspectives in Organic Chemistry*, ed. A. Todd, Interscience, New York, 1956, pp. 155–184.