Complexity of synthetic reactions. The use of complexity indices to evaluate reactions, transforms and disconnections[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 based on fundamental mathematical entities (topological invariants) are used to calculate the changes that take place during synthetic reactions, such as the Aldol, Diels–Alder, Wittig, Reppe–Vollhardt, Pauson–Khand and House–Whitesides reactions. Such calculations provide a way to compare the power of various reactions to increase molecular complexity, and they furnish a 'yardstick' to gauge the state of the art of synthetic chemistry. Retrosynthetic transforms and disconnections are also evaluated, and in the latter case, topological simplification principles are derived. In this way we may reveal the 'unseen hand' of synthesis.

Introduction

Complexity is a subject of growing interest in many fields,[2] including chemistry.[3] Our approach to understanding the complexity of a system such as a molecule or a plan for its synthesis is to abstract it as a graph and then use the tools of graph theory, information theory and combinatorics to quantify the relationships in it.[4–14] Mathematical terms are defined intuitively; for rigorous definitions standard texts can be consulted[15] and a Prolegomenon to chemical graph theory is given in the Electronic supplementary information (ESI). Where possible, the corresponding chemical terms are used.

Based on the number of pairs of adjacent bonds η, the first general index of chemical complexity was $C(\eta,\varepsilon) = C(\eta) + C(\varepsilon)$, where $C(\eta)$ contains the information on size, connectivity and symmetry, and $C(\varepsilon)$ contains the information on elemental (including isotopic) composition.[4] In a more comprehensive approach, the 'all possible substructures' method for measuring molecular complexity and diversity has been introduced as a specific application of the general 'all possible subgraphs' method for characterizing the complexity of any system that can be abstracted as a graph.[11,14]

To facilitate the identification of equivalent substructures, we draw a chemical structure as its hydrogen-suppressed molecular graph $M(G)$, in which atoms other than H are represented by *points* (also called *vertices*) and bonds by *lines* (also called *edges*). They are similar to the abbreviated structural formulas commonly used by organic chemists in which H atoms and the bonds to them are omitted. Multiple bonds in alkenes and alkynes are represented by the corresponding *multiple lines*, and lone pairs of electrons on heteroatoms, carbanions or carbenes by *loops*. When needed, *labels* (colors, numbers, letters, symbols, *etc.*) on the points or lines can be used to convey a wide range of information, such as absolute configuration, isotopic composition, bond lengths or torsion angles.

Alkanes are abstracted as *trees*, graphs without cycles (including multiple lines and loops), and more specifically n-alkanes are represented by *paths* P_n. The cycloalkanes are abstracted as *cycles* C_n, for example, cyclohexane is represented by C_6. 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. Each chemical element is assigned a unique color, usually black (closed circle) or white (open circle) or abstract (*e.g.*, an asterisk * or the lack of one). (*N.B.*, alkanes are subsumed under heteroalkanes as the trivial case for which $h = 0$.)

Substructures of molecule M are represented by *subgraphs* $S_i(M)$ of the corresponding molecular graph M, which are connected (1-component) graphs that have all their points and lines in M. It follows from the definition that M is a subgraph of itself, and we also count methane, P_1. A substructure is usually not a stable molecule itself, owing to a deficit of H atoms; nevertheless, we name substuctures and the corresponding subgraphs according to the molecules they become when H atoms are added to any free valences. For example, butane has one butane subgraph, two propanes, three ethanes and four methanes. The subgraphs for aldol are enumerated in the companion paper.[1] A *spanning subgraph* contains all the points in a graph, but not necessarily all the lines, as in the case of a *spanning tree* of a cyclic graph.

Since chemical structures contain 3-dimensional information and graphs do not, a given subgraph does not necessarily represent a unique substructure, for example, C_6 can represent either chair or boat cyclohexane. When needed, graphs and subgraphs can be labeled with conformational information, for example, labeling the lines in C_6 with torsion angles differentiates the two conformations. However, when conformational information is included, identical moieties in a molecule or supramolecular complex may appear to be unique, and opportunities for synthetic efficiency based on symmetry may be missed. Therefore, we follow the common convention that neglects conformational complexity and count subgraphs and substructures on the basis of connectivity.[16] In this approximation there is a 1:1 correspondence between substructures and subgraphs. Since equilibrium conformations are determined by minimizing the total energy of a system, they are usually independent of the synthetic reactions used to construct the system. This approach is highly advantageous for synthetic analysis.[1]

A précis of the kinds of complexity is presented in the next section, and they are discussed in more detail in the companion

---

† Electronic supplementary information (ESI) available: a Prolegomenon to chemical graph theory; Table 13 and Figure 1, summarizing all possible disconnections of azacyclohexane **28**. See http://www.rsc.org/suppdata/nj/b2/b210843g/

paper.[1] For indices that measure complexity, continuously increasing (discrete) functions are preferred, although monotonically increasing ones are allowed. There are several other criteria for complexity indices, which have been discussed in detail;[12,13,17] examples are our criterion that they increase in homologous series[5] and Lehn's requirement that they be global in nature.[18] While topology can influence energetics, we do not explicitly consider thermodynamics in our model.

As robust indices of complexity for any graph G, we introduced the number of kinds of subgraphs $N_S(G)$ and the total number of subgraphs $N_T(G)$.[11] The former can also be thought of as the *diversity* of interactions or other relationships. In chemical terms they are the number of kinds of substructures N_S and the total number of substructures N_T, respectively; for butane $N_S = 4$ and $N_T = 10$. (When conformational information is included, N_T remains the same, but N_S may be larger. This is true for any other labeling, as well.) Rücker and Rücker have described a computer program for the calculation of N_S and N_T.[19] Bone and Villar[16] have introduced a similar approach based on induced subgraphs (*cf.* Prolegomenon).

When one cannot enumerate all possible subgraphs, there are selected subsets of them that can be used, such as induced subgraphs, spanning subgraphs, trees or paths. (Paths are assumed to be 'self-avoiding.') Tutte *et al.* used the number of spanning trees to gauge network complexity;[20] however, this index is insensitive to branching within a family of isomers, which is an important aspect of chemical complexity.[9,17] As a measure of topological complexity in polycyclic compounds, Corey and Cheng proposed the number of alternative paths,[21] which has the same limitation. The total number of heteroalkanes (labeled trees) T_T is a useful index,[12] which we compare to the total number of linear heteroalkanes (labeled paths) P_T and the number of kinds of these substructures, T_S and P_S, respectively. Table 1 summarizes index values for illustrative small molecules, a number of which are used in subsequent calculations. A numerical scale needs a 'zero point' and a good one arises naturally from the mathematics in this case: the complexity of dihydrogen is $C(H_2) = 0$ no matter which index is used, since its H-suppressed molecular graph is the null graph G_0.[13]

All of the indices in Table 1 increase monotonically as one goes up the homologous series of alkanes (P_n: methane, ethane, propane, butane, pentane, *etc.*), cycloalkanes (C_n: cyclopropane, cyclobutane, cyclopentane, *etc.*) and complete molecules (K_n: ethane, cyclopropane, tetrahedrane). (*Complete molecules* are represented by H-suppressed molecular graphs that are *complete graphs*, that is, each point is joined to every other point.[22]) Except for P_S, they increase monotonically

Table 1 Complexity indices for some useful small molecules

Molecule	Graph	N_T [N_T(lpe)]	N_S [N_S(lpe)]	T_T	T_S	P_T	P_S	η	$C(\eta,\varepsilon)$
Methane	● ($P_1 \cong K_1 \cong K_{1,0}$)	1	1	1	1	1	1	0	—[a]
Ethane	●–● ($P_2 \cong K_2 \cong K_{1,1}$)	3	2	3	2	3	2	0	—[a]
Propane	●–●–● ($P_3 \cong K_{1,2}$)	6	3	6	3	6	3	1	0.0
Butane	●–●–●–● (P_4)	10	4	10	4	10	4	2	2.0
Pentane	●–●–●–●–● (P_5)	15	5	15	5	15	5	3	7.5
Hexane	●–●–●–●–●–● (P_6)	21	6	21	6	21	6	4	12.0
Cyclopropane	△ ($C_3 \cong K_3$)	10	4	9	3	9	3	3	4.8
Cyclobutane	□ (C_4)	17	5	16	4	16	4	4	8.0
Cyclopentane	⬠ (C_5)	26	6	25	5	25	5	5	11.6
Cyclohexane	⬡ (C_6)	37	7	36	6	36	6	6	15.5
Azacyclohexane[b]		37 (59)	18 (31)	36	17	36	17	6	30.9
Cyclohexene		69	17	51	6	51	6	9	45.1
1,3-Cyclohexadiene		133	23	73	6	73	6	12	66.0
1,4-Cyclohexadiene		129	20	72	6	72	6	12	58.0
Benzene		261	25	105	6	105	6	15	69.4
Bicyclobutane	 ($K_4 - x$)	33	9	25	5	23	4	8	36.0
Tetrahedrane	 (K_4)	64	10	38	5	34	4	12	43.0
2-Methylpropane	●–●< ($K_{1,3}$)	11	4	11	4	10	3	3	4.8
2,2-Dimethylpropane	>●< ($K_{1,4}$)	20	5	20	5	15	3	6	15.5
Ethylene (ethene)	●=●	5	3	4	2	4	2	1	0.0
Propene	●=●–●	10	5	8	3	8	3	3	7.5
Allene	●=●=●	18	6	11	3	11	3	6	21.0
1-Butene	●=●–●–●	16	7	13	4	13	4	4	14.0
2-Butene (*E* or *Z*)	●–●=●–●	18	7	14	4	14	4	5	15.2
2-Methylpropene	●=●<	19	7	15	4	13	3	6	23.0
Butadiene	●=●–●=●	26	8	17	4	17	4	6	21.0
Methanol	○–●	3 (9)	3 (7)	3	3	3	3	0	2.0
Ethanol	○–●–●	6 (15)	5 (11)	6	5	6	5	1	2.8
Dimethyl ether	●–○–●	6 (18)	4 (10)	6	4	6	4	1	2.8
1-Propanol	○–●–●–●	10 (22)	7 (15)	10	7	10	7	2	7.2
Methyl ethyl ether	●–○–●–●	10 (28)	7 (17)	10	7	10	7	2	7.2
2-Propanol	○–●<	11 (26)	7 (15)	11	7	10	6	3	10.8
Formaldehyde	○=●	5 (17)	4 (10)	4	3	4	3	1	2.0
Acetaldehyde	○=●–●	10 (31)	7 (17)	8	5	8	5	3	10.3
Propanal	○=●–●–●	16 (46)	10 (24)	13	7	13	7	4	17.2
Acetone	○=●<	19 (58)	10 (24)	15	7	13	6	6	26.3
Acrolein	○=●–●=●	26 (74)	14 (32)	17	7	17	7	6	30.3
Glyoxal	○=●–●=○	26 (203)	10 (38)	17	6	17	6	6	25.0
Acetylene (ethyne)	●≡●	9	4	5	2	5	2	3	4.8
Carbon monoxide	○≡●	9 (32)	5 (16)	5	3	5	3	3	6.8

[a] Not defined.[4d] [b] The N atom is symbolized by *.

with size (number of points n) for the stars ($K_{1,n-1}$: methane, ethane, propane, 2-methylpropane, 2,2-dimethylpropane)[22] and with cyclicity (*e.g.*, butane, cyclobutane, bicyclobutane, tetrahedrane). They increase with bond order (ethane, ethene, ethyne) except for P_S and T_S. Finally, they increase with branching in isomeric molecules (*e.g.*, butane *vs.* 2-methylpropane and pentane *vs.* 2,2-dimethylpropane) except for P_T and the 'kinds of' indices P_S, T_S and N_S. While two indices may both increase monotonically with a complexity factor, one may be better for characterizing it, for example N_T more fully reflects cyclicity than T_T, since it explicitly includes cycles.

The least sophisticated index, P_S, has the most exceptions, and the most sophisticated index, N_T, has the least. Thus, N_T accounts for all structural features except heteroatoms. When a heteroatom has a lone pair of electrons (lpe), a loop can be added in addition to the color label, and N_T(lpe) increases along with N_S(lpe). Like $C(\eta,\varepsilon)$, N_S reflects symmetry (*cf.* butadiene, acrolein, glyoxal), which makes it an index of extrinsic complexity.[1] It accounts for heteroatoms and all other complexity factors except branching in a family of isomers and is the most robust diversity index. It should be noted that N_T, N_S, T_T and T_S are global,[18] since they contain spanning subgraphs. Thus, the 'all possible substructures' method for measuring molecular complexity and diversity is very general, but all individual indices have well-defined limitations, which must be borne in mind when it comes to applications.

Results and discussion

Synthetic reactions

Synthesis involves assembling the parts of a system to make the whole. The way they interact is what makes the whole more than the sum of its parts, and the complexity of the interactions is the *intrinsic complexity*; it is called *molecular complexity* when the system is a molecule. The way the parts are arranged can have a profound effect on the most efficient way to fabricate the system, and the complexity of construction is the *extrinsic complexity*; it is called *synthetic complexity* when the system is a molecule.[13] Currently, there are many factors that contribute to the synthetic complexity of a molecule or supramolecular complex: (i) structural details of the target (molecular complexity), (ii) synthetic reactions available (state of the art of synthetic chemistry), (iii) availability of starting materials (including economic considerations), (iv) methods of purification (state of the art of separations science) and (v) environmental constraints (legal or ethical). Eventually, when the state of the art becomes sophisticated enough, the structural details of the target will no longer matter. (Obviously, we are approaching this state asymptotically.) Wender *et al.* have discussed some of these issues in the context of the 'ideal' synthesis,[23] which we will define for our purposes in the next paper.[1] Hendrickson has treated computer-assisted synthesis design from his perspective on this concept, which features the use of construction reactions exclusively.[24] Hudlicky has emphasized the importance of practicality,[25] and Trost has highlighted atom economy,[26] which are especially important to industrial chemists.

Most of the above factors are outside the immediate control of the synthetic chemist and, once a target is chosen, they are all outside it. What he or she can control are the specific reactions selected, that is the *synthesis plan*. (*N.B.*, the starting materials are determined by the reactions used in the plan.) In this paper we focus on molecular complexity and how it changes during synthetic reactions, transforms and disconnections. In the subsequent paper we will examine how molecular complexity and synthetic complexity affect multistep syntheses.[1] The examples below were chosen to illustrate the principles involved and were kept simple enough that the calculations can be performed without a computer. Nevertheless, a computer program is very useful for checking the results.[19]

The yields of synthetic reactions are generally high,[27] since in most cases substantial development work has been done on them; nevertheless, they are rarely 100%. The drop in overall yield with the number of steps has been referred to as the 'arithmetic fiend.'[28] Reactions that rapidly build up a molecular skeleton are associated with large increases in molecular complexity, and they are prime candidates for development as general synthetic methods, as they tend to minimize the number of steps in a synthesis. As illustrated in the following paper,[1] it is virtually impossible to use only construction reactions—at the current state of the art.

In order to advance the state of the art, new synthetic reactions are developed every year and it is natural to ask, how do they compare to previous ones in terms of their power to efficiently construct molecules? One application of an intrinsic complexity index is to measure the change in molecular complexity during synthetic reactions. In this way an index of molecular complexity can serve as a yardstick to gauge the state of the art of synthetic chemistry; for example, $C(\eta,\varepsilon)$ was used to demonstrate that the Weiss–Cook reaction compares favorably with the Diels–Alder reaction in terms of its overall power to increase complexity.[4]

To calculate the change in complexity ΔC for a reaction, the sum of the C values for the reactants is subtracted from the sum for the products [eqn. (1)]. When two or more identical molecules are joined, the initial complexity is simply that of one of them, since there is only one reactant that can undergo yield-decreasing side reactions. On the other hand, identical components resulting from a disconnection are included in the calculation of ΔC, since they are generally not identical in the corresponding transform (*cf.* next section).

$$\Delta C(\text{reaction}) = \Sigma C(\text{products}) - \Sigma C(\text{reactants}) \qquad (1)$$

$$I_T(M_1 + M_2) = I_T(M_1) + I_T(M_2) \qquad (2)$$

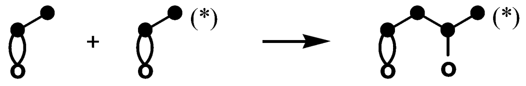
$$I_S(M_1 + M_2) = I_S(M_1) + I_S(M_2) - I_S(M_1,M_2) \qquad (3)$$

To calculate the sum of 'total' indices $I_T(M_1)$ and $I_T(M_2)$ for molecules M_1 and M_2, one simply adds them [eqn. (2)], and to calculate the sum of 'kinds of' indices $I_S(M_1)$ and $I_S(M_2)$, one adds them and then subtracts the number of duplicates $I_S(M_1,M_2)$ [eqn. (3)].[29] Similarity calculations have also been introduced into synthetic analysis and eqns. (2) and (3) are the denominators for the corresponding average and Tanimoto coefficients, respectively.[30] Lone pairs of electrons are not included, unless they are involved in a reaction or needed on heteroatoms.

Examples of useful synthetic reactions are analyzed below. Additional information about these reactions can be found in the literature cited and organic chemistry textbooks.[31]

The Aldol reaction is one of the most versatile, ranging from the simple, archetypal dimerization of acetaldehyde **1**[32] to the 'complex Aldol reactions' used to synthesize phorboxazole B.[33] A significant increase in complexity is expected, owing to a doubling in molecular size and a modest increase in diversity (●=○ + ●=○ → ●=○ + ●–○). Table 2 contains the values of selected complexity indices for the dimerization of **1** to aldol **2**. The molecules do not have high cyclicities (number of rings/number of atoms); therefore, the total number of substructures N_T can be approximated by the total number of heteroalkanes T_T: $\Delta N_T = 36 - 10 = 26$ and $\Delta T_T = 30 - 8 = 22$. Since the degree of branching is not high, the latter can be approximated by the total number of linear heteroalkanes, $\Delta P_T = 26 - 8 = 18$; however, the systematic error accumulates with each successive approximation and ΔP_T is not a particularly good estimate of ΔN_T. The same is true for the 'kinds of' indices: $\Delta N_S = 12$, $\Delta T_S = 8$ and $\Delta P_S = 5$. The number of

Table 2 Aldol reaction of **1** with itself or **1***



1 **1 (1*)** **2 (2*)**

Index[a]	**1 (1*)**	**1 + 1 (1 + 1*)**	**2 (2*)**	ΔC (ΔC*)
N_T	10 (10)	10 (20)	36 (36)	26 (16)
N_S	7 (8)	7 (11)	19 (23)	12 (12)
T_T	8 (8)	8 (16)	30 (30)	22 (14)
T_S	5 (6)	5 (8)	13 (17)	8 (9)
P_T	8 (8)	8 (16)	26 (26)	18 (10)
P_S	5 (6)	5 (8)	10 (14)	5 (6)
η	3 (3)	3 (6)	7 (7)	4 (1)
$C(\eta,\varepsilon)$	10.3 (12.3)	10.3 (22.6)	42.8 (46.1)	32.5 (23.5)

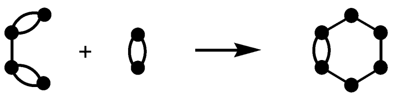
[a] See ref. 29 for sample calculations.

heteroatoms does not change during the course of the Aldol reaction, so that any terms added to C to account for them, as in $C(\eta,\varepsilon)$,[4] tend to cancel out upon the application of eqn. (1). Consequently, the 'total' indices, which are not affected by heteroatoms, are useful in cases such as this.

When the starting materials are different, the initial complexity is higher and ΔC is lower. The values in parentheses in Table 2 are for the directed,[34] labeled Aldol reaction of **1** with acetaldehyde-2-^{13}C **1***.[35] The changes in molecular complexity are $\Delta N_T = 36 - (10 + 10) = 16$, $\Delta T_T = 30 - (8 + 8) = 14$ and $\Delta P_T = 26 - (8 + 8) = 10$, which are significantly lower than for the unlabeled analog. The changes in molecular diversity, $\Delta N_S = 12$, $\Delta T_S = 9$ and $\Delta P_S = 6$, are essentially the same as above. As long as the carbonyl group is the single most important functional group for organic synthesis, the Aldol reaction will be a mainstay.[34] It has been prominent in natural product synthesis[33,36] and it played a key role in Prinzbach *et al.*'s pagodane route to dodecahedrane.[37] The discovery of long-range steric effects by Danishefsky *et al.*[36] underscores the importance of global topological indices.[18] Asymmetric Aldol reactions are currently an active frontier.[33,34,36] (The effect of chirality on complexity indices is discussed in the companion paper.[1])

The prototypical Diels–Alder reaction of butadiene **3** and ethylene **4** to afford cyclohexene **5** is examined in Table 3.[38] Complexity is expected to increase significantly, owing to the formation of a new ring. When reactions are compared, the number of molecules joined must be the same, as this factor has a major effect (*vide infra*). The increase in molecular complexity for this Diels–Alder reaction, $\Delta N_T = 38$, is more than twice that for the labeled Aldol reaction ($\Delta N_T = 16$), whereas the increase in molecular diversity, $\Delta N_S = 9$, is essentially the same ($\Delta N_S = 12$). The approximation based on trees (paths) is a good one for the total number, $\Delta T_T = \Delta P_T = 30$, but not for the number of kinds, $\Delta T_S = \Delta P_S = 2$, since the latter are not sensitive to multiple bonds. We do not attach much meaning to the difference between 9 and 12; on the other hand, the difference between 38 and 16 is significant. There have been a plethora of applications of this reaction, including iterative[39] and tandem processes;[40] the 'domino Diels–Alder reaction' was used by Paquette and colleagues to initiate the first synthesis of dodecahedrane.[41] Intramolecular,[42] asymmetric,[43,44] and hetero-Diels–Alder reactions (*vide infra*)[44–46] are active areas of research.

Table 3 Diels–Alder reaction of **3** and **4**



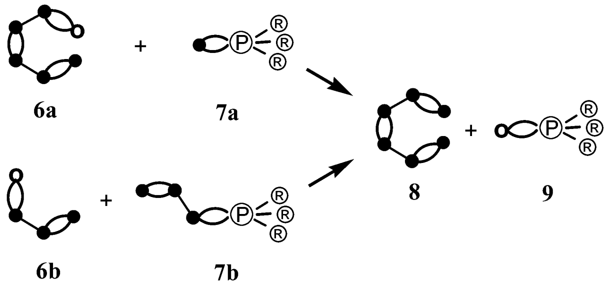
3 **4** **5**

Index	**3**	**4**	**3 + 4**	**5**	ΔC
N_T	26	5	31	69	38
N_S	8	3	8	17	9
T_T, P_T	17	4	21	51	30
T_S, P_S	4	2	4	6	2
η	6	1	7	9	2
$C(\eta)$	21.0	0.0	21.0	45.1	24.1

The Wittig reaction of aldehyde **6a** or **6b** with phosphorane **7a** or **7b**, respectively, to afford 1,3,5-hexatriene **8** and phosphine oxide **9** is analyzed in Table 4.[47] Carbon-oxygen and carbon-phosphorus double bonds in the reactants are exchanged for carbon-carbon and phosphorus-oxygen double bonds in the products (●=○ + ●= Ⓟ → ●=● + ○= Ⓟ); consequently, we expect the change in diversity to be negative. The residues (R) on phosphorus are usually phenyl groups, which are inert under the standard reaction conditions; therefore, this group is abstracted as a single vertex Ⓡ. Methylenephosphorane **7a** (Ⓡ$_3$Ⓟ=●) is a unique case that has the same complexity as the phosphine oxide by-product **9** (Ⓡ$_3$Ⓟ=○), since one color is merely substituted for another in their molecular graphs. For the reaction of **6a** with **7a**, the complexity changes are $\Delta N_T = \Delta T_T = \Delta P_T = 0$, $\Delta N_S = -14$ and $\Delta T_S = \Delta P_S = -4$, and our qualitative assessment is confirmed.

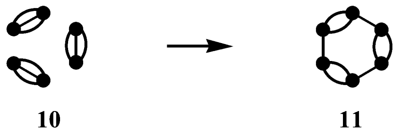
A more 'typical' Wittig reaction is that of acrolein **6b** and allylidenephosphorane **7b**.[47] Molecular complexity decreases in this case, $\Delta N_T = -36$ and $\Delta T_T = -24$. (*N.B.*, P_T is not sensitive to branching, which is an important factor in this case.) Except for the least sensitive index, $\Delta P_S = -4$, molecular diversity decreases here even more than in the previous case: $\Delta N_S = -18$ and $\Delta T_S = -8$. In contrast to the previous example, where the substrate was controlling, C(**6a**) > C(**7a**), the complexity calculation in this case is dominated by the phosphorane, C(**7b**) > C(**6b**), which is usually taken for granted. All Wittig reactions share the drawback that they are not atom-economical.[26] For reactions involving double bonds it should be noted that the corresponding E- and Z-stereoisomers have the same complexity, since they have the same connectivity. Wittig reactions have been applied to the synthesis of many insect phermones,[48] and if one focuses on the substrate, the increase in complexity is substantial (*vide infra*).

Table 4 Wittig reactions of **6a** with **7a** and **6b** with **7b**



Index	**6a (6b)**	**7a (7b)**	**8**	**9**	ΔC**a** (ΔC**b**)
N_T	95 (26)	36 (141)	95	36	0 (−36)
N_S	34 (14)	14 (42)	19	14	−14 (−18)
T_T	48 (17)	28 (83)	48	28	0 (−24)
T_S	11 (7)	10 (20)	6	10	−4 (−8)
P_T	48 (17)	19 (50)	48	19	0 (0)
P_S	11 (7)	7 (13)	6	7	−4 (−4)
η	11 (6)	10 (15)	11	10	0 (0)
$C(\eta,\varepsilon)$	72.0 (30.3)	53.0 (103.1)	58.1	53.0	−13.9 (−22.3)

Table 5 Reppe–Vollhardt cyclotrimerization of **10**

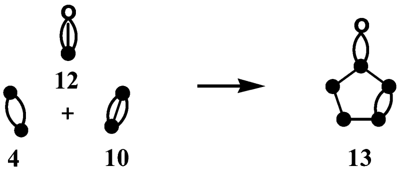


Index	10	11	ΔC
N_T	9	261	252
N_S	4	25	21
T_T, P_T	5	105	100
T_S, P_S	2	6	4
η	3	15	12
$C(\eta)$	4.8	69.4	64.6

The Reppe–Vollhardt reaction,[49] the cyclotrimerization of alkynes on a transition metal template, is analyzed in Table 5 for the archetypal conversion of acetylene **10** to benzene **11** (Pd or Pt catalyst).[50] Molecular complexity escalates rapidly with the number of rings and double bonds in a molecule, owing to the combinatorial nature of these indices, which results in dramatic increases for all the 'total' indices: $\Delta N_T = 252$ and $\Delta T_T = \Delta P_T = 100$. The change in diversity, $\Delta N_S = 21$, is also significant. (*N.B.*, T_S and P_S are not sensitive to multiple bonds, which are important in this reaction.) When the alkyne must be prepared *via* a multistep procedure, as in Hecht and Fréchet's seminal dendrimer synthesis based on this reaction,[51] the synthesis plan is highly convergent and reflexive.[1] This reaction was used in tandem with an intramolecular Diels–Alder reaction in the classic synthesis of estrone by Funk and Vollhardt,[52] and it has been applied in many other imaginative ways.[53]

The Pauson–Khand reaction, which forges an alkene, an alkyne and carbon monoxide into a five-membered ring containing a carbon-carbon double bond, is catalyzed by Co and other transition metals.[54] For the prototypical reaction of ethylene **4**, acetylene **10** and carbon monoxide **12** that gives 2-cyclopentenone **13** (Table 6),[55] we have $\Delta N_T = 122$ and $\Delta N_S = 42$. (Again, the approximations are not useful, owing to multiple bonds and cyclicity.) The total number of atoms is the same for the Aldol, Diels–Alder, Reppe–Vollhardt and Pauson–Khand reactions discussed here; however, the number of heteroatoms is not the same in every case. As it is independent of heteroatoms, N_T is useful for such situations. Based on this index, the last two reactions are the most powerful, in large part because three molecules are joined rather than two (*cf.* next section). Double-barreled Pauson–Khand reactions have been used to generate six new carbon-carbon bonds and four new five-membered rings in spectacular one-flask syntheses by the Keese and Cook groups.[56] This reaction has become a standard method for cyclopentenone synthesis, as the list of metals that mediate it keeps growing[57] and chiral catalysts provide high levels of asymmetric induction.[58]

Table 6 Pauson–Khand reaction of **4**, **10** and **12**

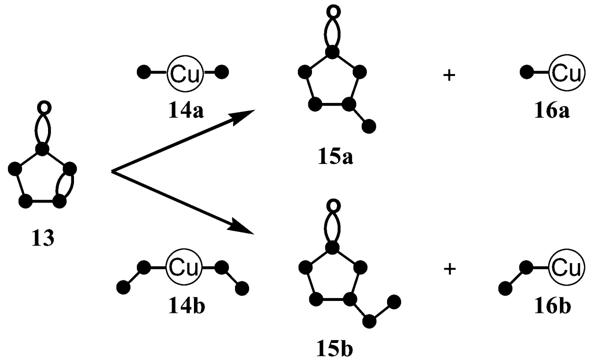


Index	4	10	12	4 + 10 + 12	13	ΔC
N_T	5	9	9	23	145	122
N_S	3	4	5	8	50	42
T_T	4	5	5	14	80	66
T_S	2	2	3	4	15	11
P_T	4	5	5	14	62	48
P_S	2	2	3	4	11	7
η	1	3	3	7	13	6
$C(\eta,\varepsilon)$	0	4.8	6.8	11.6	92.1	80.5

The House–Whitesides reaction,[59] the conjugate addition of organocuprates R_2CuLi to α,β-unsaturated carbonyl compounds, is illustrated for 2-cyclopentenone **13** in Table 7.[60–63] In order to keep the calculations as simple as possible, the starting cuprate **14a** (R = Me) or **14b** (R = Et) is treated as the monomer, the aggregation state in tetrahydrofuran solvent,[64] rather than the dimer, which is present in ether.[65] The changes in complexity indices for the overall reaction of **13** with **14a** to give 3-methylcyclopentanone **15a** are $\Delta N_T = -26$, $\Delta T_T = 7$, $\Delta N_S = -4$ and $\Delta T_S = 10$. This is an interesting case in which the indices based on heteroalkanes diverge significantly from those based on all possible substructures, since the carbon-carbon double bond in the reactant contributes more to the latter than the tertiary methyl substituent in the product. As the group added becomes more complex, its contribution predominates, and all these indices increase in the 'typical' House–Whitesides reaction of **13** with **14b** to give 3-ethylcyclopentanone **15b**: $\Delta N_T = 16$, $\Delta T_T = 36$, $\Delta N_S = 12$ and $\Delta T_S = 17$.

For pedagogical reasons, we show the overall reaction to the ketone, the final product after aqueous work-up, rather than the initial reaction to the enolate.[66] This allows the retron,[21] which triggers the retrosynthetic transform, to be the β-alkyl (or β-aryl) ketone, rather than the corresponding enolate. Catalytic organocuprate reactions, such as the Kharasch reaction,[60] can be atom-economical,[26] but stoichiometric ones that use both R groups on Cu are rare.[67] For valuable R groups a non-transferred second ligand can be used,[68] as in the Cu-based synthesis of prostaglandins.[69] As illustrated here, the Pauson–Khand/House–Whitesides combination is a natural one.[70] The value of ΔC(tandem) for the overall reaction of **4**, **10**, **12** and **14b** to **15b** and **16b** is simply the sum of the values for the individual reactions: $\Delta C(\text{tandem}) = \Delta C(\text{reaction 1}) + \Delta C(\text{reaction 2}) = 122 + 16 = 138$ for $C = N_T$.

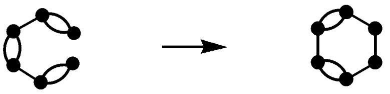
Table 7 House–Whitesides reactions of **13** with **14a** and **14b**[a]



Index[b]	14a (14b)	15a (15b)	16a (16b)	ΔCa (ΔCb)
N_T	6 (15)	122 (170)	3 (6)	−26 (16)
N_S	4 (8)	47 (65)	3 (5)	−4 (12)
T_T	6 (15)	90 (125)	3 (6)	7 (36)
T_S	4 (8)	26 (35)	3 (5)	10 (17)
P_T	6 (15)	58 (73)	3 (6)	−7 (2)
P_S	4 (8)	12 (14)	3 (5)	0 (0)
η	1 (3)	12 (13)	0 (1)	−2 (−2)
$C(\eta,\varepsilon)$	2.8 (11.1)	86.2 (96.6)	2.0 (2.8)	−6.7 (−3.8)

[a] Each cuprate has a Li counterion, which is not shown. The work-up is addition of methanol.[66] [b] For **13** see Table 6.

Table 8 [3,3]-Sigmatropic rearrangement of **8**



8 **17**

Index	**8**	**17**	ΔC
N_T	95	133	38
N_S	19	23	4
T_T, P_T	48	73	25
T_S, P_S	6	6	0
η	11	12	1
$C(\eta)$	58.1	66.0	7.9

Table 10 Summary of complexity changes for synthetic reactions.[a]

Reaction	ΔN_T	ΔT_T	ΔP_T	$\Delta\eta$	ΔN_S	$\Delta C(\eta,\varepsilon)$
Aldol (labeled)	16 e	14 f	10 e	1 d	11 d	23.5 d
Diels–Alder	38 c	30 d	30 c	2 c	9 e	24.1 c
Wittig (unique)	0 g	0 h	0 g	0 f	−14 i	−13.9 h
Wittig (typical)	−36 j	−24 j	0 g	0 f	−18 j	−22.3 i
Reppe–Vollhardt	252 a	100 a	100 a	12 a	21 b	64.6 b
Pauson–Khand	122 b	66 b	48 b	6 b	42 a	80.5 a
House–Whitesides (Me)	−26 h	7 g	−7 i	−2 h	−4 g	−6.7 g
House–Whitesides (Et)	16 e	36 c	2 f	−2 h	12 c	−3.8 f
[3,3]-Sigmatropic	38 c	25 e	25 d	1 d	4 f	7.9 e
Reduction	−32 i	−15 i	−15 j	−3 j	−10 h	−29.6 j

[a] Key: a #1, b #2, c #3, d #4, e #5, f #6, g #7, h #8, i #9, j #10.

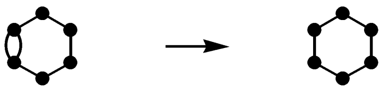
Applications of cuprate conjugate addition are legion;[71] nevertheless, the asymmetric version of this reaction needs further development.[72]

The [3,3]-sigmatropic rearrangement of *Z*-1,3,5-hexatriene **8** to 1,3-cyclohexadiene **17** is examined in Table 8.[73] Molecular complexity increases markedly, $\Delta N_T = 38$ and $\Delta T_T = \Delta P_T = 25$, owing to the formation of a new ring. Molecular diversity does not change significantly, $\Delta N_S = 4$. For a rearrangement to result in a significant increase in complexity, it must create a new ring or more highly substituted skeletal atom. Sigmatropic rearrangements have gained in stature, as they have been used in tandem with important synthetic reactions (*e.g.*, Aldol, Diels–Alder, Wittig, *etc.*)[40] to further extend their reach and for 'chirality transfer.'[74] Thus, the [3,3]-sigmatropic rearrangement of a thiono imidazolide was a key step in Nicolaou *et al.*'s synthesis of calicheamicin γ_1.[75]

The catalytic hydrogenation of cyclohexene **5** to cyclohexane **18** is analyzed in Table 9.[76] There is a substantial decrease in molecular complexity, $\Delta N_T = -32$. The approximations based on T_T and P_T are not good ones, since the cyclicity of the product is half that of the starting material. (*N.B.*, double bonds can be counted as '2-cycles.') The decrease in diversity is also significant, $\Delta N_S = -10$. The complexity of dihydrogen is zero, and the catalyst is both a starting material and a product, so that its complexity cancels out upon application of eqn. (1). In general, it is advisable to avoid steps in which complexity decreases, as they are associated with 'excess complexity;'[1] however, this negative aspect is more than outweighed when reduction also accomplishes a major strategic goal, such as setting the correct stereochemistry. Thus, asymmetric hydrogen transfers to double bonds (C=C, C=N, C=O, *etc.*) are extremely important for the synthesis of chiral molecules.[77]

Table 10 summarizes the values of ΔN_T, ΔT_T, ΔP_T, $\Delta\eta$, ΔN_S and $\Delta C(\eta,\varepsilon)$ for the above reactions except for the unlabeled Aldol reaction, which is the only one that has identical reactants and is therefore not directly comparable to the rest (*cf.* reflexivity[1]). According to all six indices, the two most powerful reactions (#1, #2) are Reppe–Vollhardt and Pauson–Khand, with the caveat that the order is reversed for the extrinsic indices, ΔN_S and $\Delta C(\eta,\varepsilon)$. The main reason for their high rankings is the fact that three components are joined (*vide infra*).

Table 9 Reduction of **5**



5 **18**

Index	**5**	**18**	ΔC
N_T	69	37	−32
N_S	17	7	−10
T_T, P_T	51	36	−15
T_S, P_S	6	6	0
η	9	6	−3
$C(\eta)$	45.1	15.5	−29.6

The two reactions tied for #3 according to ΔN_T, Diels–Alder cycloaddition and [3,3]-sigmatropic rearrangement, are also ranked highly by ΔT_T (#4, #5), ΔP_T (#3, #4), $\Delta\eta$ (#3, #4) and $\Delta C(\eta,\varepsilon)$ (#3, #5), but slightly lower by ΔN_S (#5, #6, respectively). These two reactions are very similar electronically, and it is not surprising that they have similar values of topological indices, including complexity indices.

According to ΔT_T and ΔN_S, the #3 reaction is the 'typical' House–Whitesides reaction (R = Et), which is (tied for) #5 according to ΔN_T. Cyclicity is a significant factor in this reaction, so that ΔT_T is not as good as ΔN_T. A different order for the extrinsic complexity indices is not unexpected, and in any event the difference in rank is not large. The 'typical' House–Whitesides and Aldol reactions, which join two different components, have good scores, as expected for construction reactions.

The two least powerful reactions according to ΔN_T (#9, #10), double bond reduction and 'typical' Wittig reaction, are also the lowest according to ΔT_T (#9, #10) and $\Delta C(\eta,\varepsilon)$ (#10, #9) and among the lowest according to ΔP_T (#10, #7), $\Delta\eta$ (#10, #6) and ΔN_S (#8, #10, respectively). While the overall complexity change for the Wittig reaction is usually negative, the change for the substrate is very positive, for example $\Delta C(\text{substrate}) = N_T(\mathbf{8}) - N_T(\mathbf{6b}) = 69$, since a double bond is created. [This is counterbalanced by the change for the P-containing molecules, $\Delta C(\text{P}) = N_T(\mathbf{9}) - N_T(\mathbf{7b}) = -105$. *N.B.*, $-105 + 69 = -36$.]

In many cases, the simplest complexity index, η,[4] gives a good approximation of the most robust one, N_T, suggesting that applications of the former are fundamentally sound.[6,12]

Retrosynthetic transforms and disconnections

A retrosynthetic *transform* τ is conceptually the exact reverse of a synthetic reaction σ, and *vice versa*. The series of transforms from the target to the starting materials is the *retrosynthetic scheme*,[21] and the microscopic reverse is the *synthetic scheme* or *synthesis plan*. The complexity change during a transform can be calculated by using eqn. (1) and bearing in mind that a starting structure for the transform is a product of the reaction, and *vice versa*. Then, the complexity of the transform is the negative of the complexity of the corresponding synthetic reaction, $C(\tau) = -C(\sigma)$. Therefore, when selecting transforms, the most powerful one has the largest negative value of ΔC.

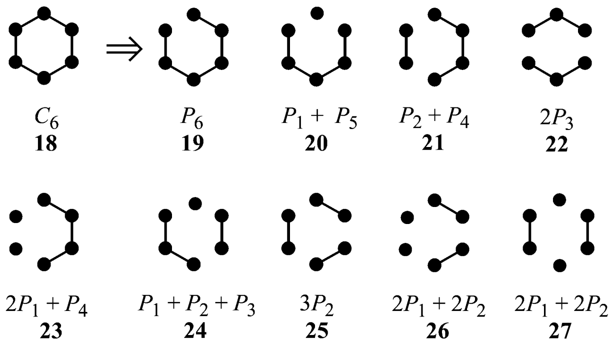
A *disconnection* δ is derived from a transform τ by cutting it down to the 'bare bones' of the skeletal (sigma) bonds broken or formed and the shortest paths (if any) joining their termini.[78] It does not include ancillary changes in bonding, for example δ(Diels–Alder) breaks two sigma bonds separated by one sigma bond in a six-membered ring, but the pi bond that is broken in τ(Diels–Alder) is not included, as it does not contribute to skeletal simplification. We include the 'shortest paths' in order to differentiate intermolecular and intramolecular versions of a reaction, since the length of the tether can influence the outcome of the latter.[79] (*N.B.*, disconnection has been used synonymously with transform;[80] however, we use the former with its original topological meaning.[21])

An effective disconnection simplifies the molecular skeleton by breaking it into two or more pieces or by reducing the number of rings in it. At the current state of the art, it is sometimes necessary to make bonds or rearrange them during retrosynthetic analysis for strategic reasons. Bonds are broken in the corresponding synthetic reaction, as in the final step in the synthesis of substituted cyclopentenones *via* sequential σ (Pauson–Khand)–σ(House–Whitesides)–σ(retro-Diels–Alder).[70] Substituted α,ω-diesters[62] and α,ω-aldehydic esters[63] have been prepared by cleavage of the corresponding α-methoxycarbonylcycloalkanones and α-hydroxycycloakanones, respectively, which resulted from tandem cuprate addition-enolate functionalization.

There is a one-to-one correspondence between reactions and transforms, but not between transforms and disconnections. Thus, for any given transform there is one disconnection; however, for a given disconnection there may be a number of transforms, for example, the disconnection corresponding to $C_6 \Rightarrow 3P_2$ has several transforms, corresponding to the (sometimes hypothetical) reactions $\tau_1 \Leftrightarrow$ σ(ethyne cyclotrimerization), $\tau_2 \Leftrightarrow$ σ(2 ethyne + ethene cycloaddition), $\tau_3 \Leftrightarrow$ σ(ethyne + 2 ethene cycloaddition) and $\tau_4 \Leftrightarrow$ σ(ethene cyclotrimerization). The first two are variations of σ(Reppe–Vollhardt).[49] Disconnections allow the synthesis planner to focus on the underlying topological changes in a retrosynthetic scheme and the corresponding synthesis plan. These concepts are developed more fully in the subsequent paper.[1]

The one- through four-bond disconnections of a six-membered ring (C_6) are calculated in Table 11. The smallest simplification is given by the one-bond disconnection $C_6 \Rightarrow P_6$ ($\Delta N_T = -16$), which corresponds to intramolecular alkylation,[31] and the largest by the four-bond ones, **18** ⇒ **26** and **18** ⇒ **27**, subsumed under $C_6 \Rightarrow 2P_1 + 2P_2$ ($\Delta N_T = -29$).[81]

Table 11 Selected disconnections of **18**



Index[a]	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**, **27**
ΔN_T	−16	−21	−24	−25	−25	−27	−28	−29
ΔN_S	−1	−2	−3	−4	−3	−4	−5	−5
ΔT_T, ΔP_T	−15	−20	−23	−24	−24	−26	−27	−28
ΔT_S, ΔP_S	0	−1	−2	−3	−2	−3	−4	−4
$\Delta\eta$	−2	−3	−4	−4	−4	−5	−6	−6
$\Delta C(\eta)$	−3.5	−8.0	−13.5	−15.5	−13.5	−15.5	—[b]	—[b]

[a] Index values for **18** are summarized in Table 1. [b] Not defined.[4d]

The former corresponds to the Dötz–Wulff reaction of a Fischer metal-carbene complex with two equivalents of alkyne.[82] (The metal-carbene complex contributes both P_1 fragments, one from the carbene moiety and one from a CO ligand; note that the O atom is not involved in the disconnection.[83]) The latter corresponds to Liebeskind *et al.*'s convergent *p*-quinone synthesis from two alkynes, which can be different, and two molecules of CO.[84]

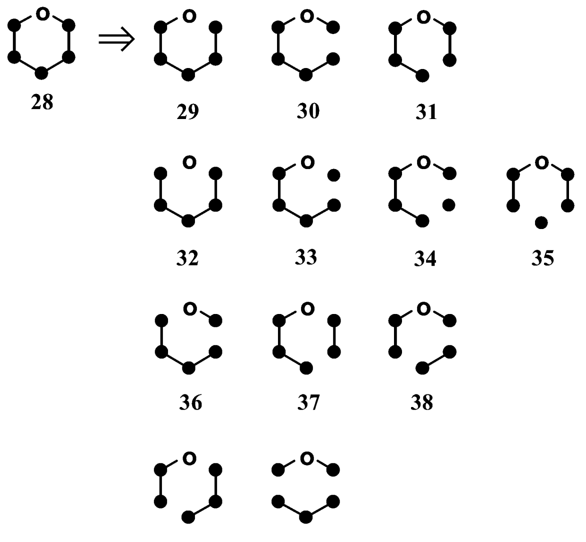
In order of increasing simplification, the two-bond disconnections are $C_6 \Rightarrow P_1 + P_5$ ($\Delta N_T = -21$), $C_6 \Rightarrow P_2 + P_4$ ($\Delta N_T = -24$) and $C_6 \Rightarrow 2P_3$ ($\Delta N_T = -25$). The first one is relatively rare,[83] nevertheless, a corresponding reaction is Wender and Eck's double-barreled conjugate addition of a 1,5-biscuprate to a β-chloro-α,β-unsaturated ketone.[85] Another example is the reaction of phosphorane **7a** with pyrylium salts.[86] The middle one corresponds to σ(Diels–Alder), which is arguably the most famous textbook example.[21,31] The last one is also rare;[83] a relevant example is the reaction of phosphorane **7b** with α,β-unsaturated ketones.[87] More work in this area is justified, since the two-bond disconnections all have significant (negative) values of ΔC.

The three-bond disconnections are $C_6 \Rightarrow 2P_1 + P_4$ ($\Delta N_T = -25$), which does not correspond to a general synthetic reaction, $C_6 \Rightarrow P_1 + P_2 + P_3$ ($\Delta N_T = -27$), which corresponds to σ(Dötz),[82] and $C_6 \Rightarrow 3P_2$ ($\Delta N_T = -28$), which corresponds to σ(Reppe–Vollhardt).[49] In general, the three-bond disconnections are ranked above the two-bond ones and below the four-bond ones, so that the major simplifying factor is the number of resulting components; however, there can be gray areas at the borders. According to the intrinsic (molecular) complexity indices (N_T, T_T and η), $C_6 \Rightarrow 2P_3$ gives the same simplification as $C_6 \Rightarrow 2P_1 + P_4$. The extrinsic (synthetic) complexity indices [N_S, T_S and $C(\eta)$], which are sensitive to symmetry, suggest that the former is slightly better; however, the symmetry in a disconnection is not necessarily present in the transform from which it is derived. Consequently, we prefer to use intrinsic complexity indices for these calculations. Although there is some degeneracy in ΔC for all of the indices, the order established by ΔN_T is never contradicted; therefore, we believe that our conclusions about the order of simplification are very secure.

Simplification increases with the number of components of equal complexity that result from a disconnection (*simplifying principle 1*),[88] for example, $C_6 \Rightarrow 6P_1$, while not realistic,[81] is better than $C_6 \Rightarrow 3P_2$, which in turn is better than $C_6 \Rightarrow 2P_3$. As vouchsafed by the preceding calculations, $C_6 \Rightarrow P_3 + P_3$ is better than $C_6 \Rightarrow P_2 + P_4$, which is better than $C_6 \Rightarrow P_1 + P_5$, and also $C_6 \Rightarrow P_2 + P_2 + P_2$ is better than $C_6 \Rightarrow P_1 + P_2 + P_3$, which is better than $C_6 \Rightarrow P_1 + P_1 + P_4$. Therefore, for alternative disconnections of a target into a given number of components, simplification increases as their complexities approach equality, and the maximum is obtained when they are equal (*simplifying principle 2*). *Inter alia*, these simplifying principles are in harmony with the heuristic of convergence.[1] A gray area occurs when a smaller number of components of equal complexity is compared to a larger number of unequal complexity, as discussed in the preceding paragraph. It is not always possible to make the components equal, such as when there is an odd number of atoms in a ring. Thus, for a five-membered ring the best three-bond disconnection is $C_5 \Rightarrow P_1 + P_2 + P_2$, which corresponds to σ(Pauson–Khand).

Table 12 summarizes the one- and two-bond disconnections of a six-membered ring that contains a heteroatom: azacyclohexane (piperidine) **28**. (All possible disconnections are summarized in the ESI.) The 'total' indices are not sensitive to heteroatoms unless they have at least one lone pair of electrons (lpe), which is the case here. Based on the intrinsic index N_T(lpe) and the extrinsic ones, N_S(lpe), N_S and T_S, the best one-bond disconnection is **28** ⇒ **29**, that is, breaking the heteroatom-carbon bond gives the greatest simplification

Table 12 One- and two-bond disconnections of azacyclohexane **28**.[a]



Index[b]	29	30	31	36	37	38	39	40
ΔN_T	−16	−16	−16	−24	−24	−24	−25	−25
ΔN_S	−7	−5	−6	−12	−11	−11	−12	−12
ΔT_T, ΔP_T	−15	−15	−15	−23	−23	−23	−24	−24
ΔT_S, ΔP_S	−6	−4	−5	−11	−10	−10	−11	−11
$\Delta\eta$	−2	−2	−2	−4	−4	−4	−4	−4
$\Delta C(\eta,\varepsilon)$	−11	−11	−11	−26.9	−23.7	−23.7	−28.1	−28.1
ΔN_T(lpe)	−32	−28	−26	−44	−42	−40	−44	−43
ΔN_S(lpe)	−14	−9	−10	−23	−20	−19	−22	−22

[a] The lone pair of electrons (lpe) on N (open circle) is represented by a loop, not shown in the graphs above, as it is not used in all calculations. [b] Index values for **28** are summarized in Table 1.

(*simplifying principle 3*). According to the heuristic rules for strategic bonds,[21] this is the best one-bond disconnection, because of the relative ease of formation of heteroatom-carbon bonds with current synthetic methodology. Our analysis gives the same result, but is purely topological.

The two-bond disconnections that give **32**–**35** by breaking adjacent bonds are better than the one-bond disconnections of **28**, but are not discussed in detail here, since they do not correspond to general synthetic reactions.[89]

According to ΔN_T(lpe) and ΔN_S(lpe), the hetero-Diels–Alder disconnections of **28** are ranked **36** > **37** > **38** (by product number), and ΔN_S and ΔT_S also identify **36** as the best. This order can be understood by applying two of the simplifying principles discussed above: **36** and **37** involve breaking heteroatom-carbon bonds and are therefore ranked above **38** (simplifying principle 3), and **36** also makes the complexities of the two halves as equal as possible (simplifying principle 2). All three of these disconnections are possible for the aza-Diels–Alder reaction[45] and a great deal of effort has been spent optimizing the substituents on the diene and dienophile in order to electronically facilitate all of them.[45,46]

According to ΔN_T(lpe), disconnection **28** ⇒ **39** is better than **28** ⇒ **40**, which is consistent with simplifying principle 3. They are comparable to the aza-Diels–Alder reaction in simplifying power. The other possible disconnections of **28** can be treated similarly and the results are consistent with the three principles of simplification (*cf.* ESI).

It has been conjectured that topology is the 'unseen hand' of organic synthesis,[11] and mathematically characterizing disconnections is one way to make it manifest. There are still powerful disconnections (*e.g.*, $C_6 \Rightarrow 2P_3$ and $C_6 \Rightarrow 2P_1 + P_4$) that do not correspond to general synthetic reactions and they suggest promising areas for research. Multicomponent reactions are beginning to receive the attention they deserve from the viewpoint of topological efficiency.[56,90]

Conclusion

Bradley observed in the *New Scientist*, "It is rather astonishing how little chemists really know about designing the shortest and simplest synthesis of a new molecule."[91] Molecular complexity considerations are a promising approach to the problem of optimal synthetic simplicity and efficiency. The design and use of reactions that substantially increase molecular complexity and disconnections that dramatically decrease it is one aspect, which is aided by indices of complexity. Another aspect is the minimization of excess complexity in synthetic routes, which is discussed in the following paper.[1]

Acknowledgements

The author is grateful to C. Rücker and R. Barone for constructive criticism, T. Sommer for useful suggestions and W.F. Wright for valuable literature searches.

References

1 Contribution 19 in a series on applications of discrete mathematics to chemistry. For 20, see the following paper: S. H. Bertz, *New. J. Chem.*, 2003, DOI: 10.1039/b210844p . For 18, see S. H. Bertz, *Chem. Commun.*, 2003, 1000–1001 . For 17, see S. H. Bertz, in *Complexity: Introduction and Fundamentals*, eds. D.H. Rouvray and D. Bonchev, Taylor and Francis, London, 2003, pp. 91–156.

2 (*a*) M. Gell-Mann, *The Quark and the Jaguar: Adventures in the Simple and the Complex*, W.H. Freeman, New York, 1994; (*b*) J. L. Casti, *Complexification*, Harper-Collins, New York, 1994; (*c*) M. M. Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos*, Touchstone-Simon & Schuster, New York, 1992; (*d*) *Interdisciplinary Approaches to Nonlinear Complex Systems*, eds. H. Haken and A. Mikhailov, Springer-Verlag, Berlin, 1993; (*e*) R. Badii and A. Politi, *Complexity: Hierarchical Structures and Scaling in Physics*, Cambridge University Press, Cambridge, 1997; (*f*) J. T. Bonner, *The Evolution of Complexity by Means of Natural Selection*, Princeton University Press, Princeton, NJ, 1988.

3 (*a*) G. Nicolis and I. Prigogine, *Exploring Complexity*, Freeman, New York, 1989; (*b*) *From Simplicity to Complexity in Chemistry—and Beyond*, eds. A. Müller, A. Dress and F. Vögtle, Vieweg, Braunschweig, 1996; (*c*) *From Simplicity to Complexity: Information–Interaction–Emergence*, eds. K. Mainzer, A. Müller and W.G. Saltzer, Vieweg, Braunschweig, 1998; (*d*) K. Mainzer, *Thinking in Complexity: The Complex Dynamics of Matter, Mind, and Mankind*, Springer-Verlag, Berlin, 3rd edn., 1997; (*e*) Y. Bar-Yam, *Dynamics of Complex Systems*, Addison-Wesley, Reading, MA, 1997.

4 (*a*) S. H. Bertz, *J. Am. Chem. Soc.*, 1981, **103**, 3599–3601; (*b*) The invariant η is the number of *connections* or pairs of adjacent lines. $C(\eta,\varepsilon) = C(\eta) + C(\varepsilon)$, where $C(\eta) = 2\eta\log_2\eta - \Sigma_i\ \eta_i\log_2\eta_i$ and $C(\varepsilon) = \varepsilon\log_2\varepsilon - \Sigma_j\ \varepsilon_j\log_2\varepsilon_j$; η_i is the number of connections in the ith class of equivalent connections, η is the total number of connections, ε_j is the number of atoms of element (or isotope) j, and ε is the total number of atoms; (*c*) When heteroatoms (or isotopes) are not involved, $C(\varepsilon) = 0$ and $C(\eta,\varepsilon) = C(\eta)$; (*d*) The values of $C(\eta,\varepsilon)$ and $C(\eta)$ are not defined for methane and (unlabeled) ethane.

5 (*a*) S. H. Bertz, *J. Chem. Soc., Chem. Commun.*, 1981, 818–820; (*b*) S. H. Bertz, *Bull. Math. Biol.*, 1983, **45**, 849–855.

6 S. H. Bertz, *J. Am. Chem. Soc.*, 1982, **104**, 5801–5803.

7 S. H. Bertz, in *Chemical Applications of Topology and Graph Theory*, ed. R. B. King, Elsevier, Amsterdam, 1983, pp. 206–221.

8 (*a*) S. H. Bertz, *J. Chem. Soc., Chem. Commun.*, 1984, 218–219; (*b*) S. H. Bertz, *J. Chem. Soc., Chem. Commun.*, 1986, 1627–1628.

9 S. H. Bertz, *Discrete Appl. Math.*, 1988, **19**, 65–83.

10 S. H. Bertz and T. J. Sommer, in *Organic Synthesis: Theory and Applications*, ed. T. Hudlicky, JAI Press, Greenwich, CT, 1993, vol. 2, pp. 67–92.

11 (*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 vol. 306, eds. T.H. Pierce and B.A. Hohne, American Chemical Society, Washington, DC, 1986, pp. 169–175.

12 S. H. Bertz and W. F. Wright, *Graph Theory Notes of New York (NY Acad. Sci.)*, 1998, XXXV, pp. 32–48.

13 S. H. Bertz and C. M. Zamfirescu, *Commun. Math. Comput. Chem. (MATCH)*, 2000, **42**, 39–70.

14 S. H. Bertz, *Chem. Commun.*, 2001, 2516–2517.

15 (*a*) F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA, 1969; (*b*) N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, FL, 2nd edn., 1992; (*c*) K. A. Ross and C. R. B. Wright, *Discrete Mathematics*, Prentice-Hall, Upper Saddle River, NJ, 4th edn., 1999; (*d*) D. I. A. Cohen, *Basic Techniques of Combinatorial Theory*, Wiley, New York, 1978.

16 R. G. A. Bone and H. O. Villar, *J. Comput. Chem.*, 1997, **18**, 86–107.

17 D. Bonchev, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 934–941.

18 J.-M. Lehn, *Supramolecular Chemistry: Concepts and Perspectives*, VCH, Weinheim, 1995.

19 (*a*) G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 314–320; (*b*) G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1457–1462.

20 (*a*) R. L. Brooks, C. A. B. Smith, A. H. Stone and W. T. Tutte, *Duke Math. J.*, 1940, **7**, 312–340; (*b*) For chemical examples see: T. J. N. Brown, R. B. Mallion, P. Pollak and A. Roth, *Discrete Appl. Math.*, 1996, **67**, 51–66.

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 The series of complete graphs K_n ($n = 1, 2, 3, \ldots$) and star graphs $K_{1,n-1}$ ($n = 1, 2, 3, \ldots$) are infinite; however, the corresponding homologous series of hydrocarbons are limited by the valence of carbon to $n \leq 4$ and $n \leq 5$, respectively. Methane $K_1 \cong K_{1,0}$ and ethane $K_2 \cong K_{1,1}$ are (trivial) members of both series.

23 (*a*) P. A. Wender, S. T. Handy and D. L. Wright, *Chem. Ind.*, 1997, 765–769; (*b*) see also: 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.

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 T. Hudlicky, *Chem. Rev.*, 1996, **96**, 3–30.

26 (*a*) B. M. Trost, *Science*, 1991, **254**, 1471–1477; (*b*) B. M. Trost, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 259–281.

27 P. L. Fuchs, *Tetrahedron*, 2001, **57**, 6855–6875.

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

29 *E.g.*, there are $N_{\mathrm{S}} = 7$ kinds of subgraphs for the acetaldehyde graph M_1 (○=●–●, ○=●, ○–●–●, ○–●, ●–●, ○, ●) and $N_{\mathrm{S}} = 8$ for the [2-C]-labeled acetaldehyde graph M_2 (○=●–★, ○=●, ○–●–★, ○–●, ●–★, ○, ★, ●) ; both have $N_{\mathrm{T}} = 10$. Then, there are $N_{\mathrm{T}}(M_1 + M_2) = 10 + 10 = 20$ total subgraphs and $N_{\mathrm{S}}(M_1 + M_2) = 7 + 8 - 4 = 11$ kinds of subgraphs for the 2-component graph $M_1 + M_2$, since there are $N_{\mathrm{S}}(M_1, M_2) = 4$ duplicates (●, ○, ○–●, ○=●). Trees and paths are treated analogously.

30 (*a*) R. Barone, M. Petitjean, C. Baralotto, P. Piras and M. Chanon, *J. Phys. Org. Chem.*, 2003, **16**, 9–15; (*b*) see also: R. Barone and M. Chanon, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 269–272; (*c*) M. Chanon, R. Barone, C. Baralotto, M. Juillard and J. B. Hendrickson, *Synthesis*, 1998, 1559–1583.

31 M. B. Smith and J. March, *March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*, Wiley, New York, 5th edn., 2000.

32 (*a*) A. Wurtz, *Bull. Soc. Chim. Fr.*, 1872, **17**, 436–442; (*b*) A. Wurtz, *Ber. Deutsch. Chem. Ges.*, 1872, **5**, 326–327; (*c*) A. Wurtz, *Compt. Rend.*, 1872, **74**, 1361–1367; see also ref. 79(a).

33 (*a*) D. A. Evans, D. M. Fitch, T. E. Smith and V. J. Cee, *J. Am. Chem. Soc.*, 2000, **122**, 10 033–10 046; (*b*) see also: D. A. Evans, H. P. Ng and D. L. Rieger, *J. Am. Chem. Soc.*, 1993, **115**, 11 446–11 459; (*c*) K. Nishide, K. Shibata, T. Fujita, T. Kajimoto, C.-H. Wong and M. Node, *Heterocycles*, 2000, **52**, 1191–1201; (*d*) I. Paterson, K. R. Gibson, L. E. Keown, R. D. Norcross, R. M. Oballa and D. J. Wallace, in *Current Trends in Organic Synthesis*, eds. C. Scolastico and F. Nicotra, Kluwer Academic-Plenum, New York, 1999, pp. 299–306.

34 Reviews: (*a*) C. Palomo, M. Oiarbide and J. M. García, *Chem. Eur. J.*, 2002, **8**, 36–44 (general); (*b*) S. Saito and H. Yamamoto, *Chem. Eur. J.*, 1999, **5**, 1959–1962 (directed); (*c*) R. Mahrwald, *Chem. Rev.*, 1999, **99**, 1095–1120 (Lewis acids); (*d*) S. Takayama, G. J. McGarvey and C.-H. Wong, *Chem. Soc. Rev.*, 1997, **26**, 407–415 (enzymes); (*e*) H. Gröger, E. M. Vogl and M. Shibasaki, *Chem. Eur. J.*, 1998, **4**, 1137–1141. For an extensive list of reviews, see ref. 31.

35 (*a*) 2-^{13}C: N. Ouwerkerk, M. Steenweg, M. de Ruijter, J. Brouwer, J. H. van Boom, J. Lugtenburg and J. Raap, *J. Org. Chem.*, 2002, **67**, 1480–1489; (*b*) see also for 1-^{13}C: D. S. Muggli, J. T. McCue and J. L. Falconer, *J. Catal.*, 1998, **173**, 470–483.

36 (*a*) C. B. Lee, Z. Wu, F. Zhang, M. D. Chappell, S. J. Stachel, T.-C. Chou, Y. Guan and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2001, **123**, 5249–5259; (*b*) C. R. Harris, S. D. Kuduk, A. Balog, K. Savin, P. W. Glunz and S. J. Danishefsky, *J. Am. Chem. Soc.*, 1999, **121**, 7050–7062.

37 J.-P. Melder, R. Pinkos, H. Fritz, J. Wörth and H. Prinzbach, *J. Am. Chem. Soc.*, 1992, **114**, 10 213–10 231.

38 G. M. Whitman, *Br. Pat.* GB 616 671, 1949; for an extensive list of reviews, see ref. 31.

39 (*a*) P. Vogel, *Curr. Org. Chem.*, 1998, **2**, 255–280; (*b*) J. D. Winkler, *Chem. Rev.*, 1996, **96**, 167–176; (*c*) P. R. Ashton, G. R. Brown, N. S. Isaacs, D. Giuffrida, F. H. Kohnke, J. P. Mathias, A. M. Z. Slawin, D. R. Smith, J. F. Stoddart and D. J. Williams, *J. Am. Chem. Soc.*, 1992, **114**, 6330–6353; (*d*) E. Block and D. Putman, *J. Am. Chem. Soc.*, 1990, **112**, 4072–4074.

40 (*a*) T.-L. Ho, *Tandem Organic Reactions*, Wiley, New York, 1992; (*b*) For tandem Diels–Alder/Aldol, see ref. 42(a); (*c*) For tandem Reppe–Vollhardt/Diels–Alder, see ref. 52; (*d*) For tandem House–Whitesides/Wittig–Horner, see ref. 61; (*e*) For tandem Michael/Michael/sigmatropic rearrangement, see: P. J. Parsons and M. Stefinovic, *Synlett*, 1993, 931–932; (*f*) For many tandem processes including the Diels–Alder reaction, see: K. C. Nicolaou, T. Montagnon and S. A. Snyder, *Chem. Commun.*, 2003, 551–564.

41 (*a*) L. A. Paquette, R. J. Ternansky, D. W. Balogh and G. Kentgen, *J. Am. Chem. Soc.*, 1983, **105**, 5446–5450; (*b*) see also: L. A. Paquette and M. J. Wyvratt, *J. Am. Chem. Soc.*, 1974, **96**, 4671–4673; (*c*) D. McNeil, B. R. Vogt, J. J. Sudol, S. Theodoropulos and E. Hedaya, *J. Am. Chem. Soc.*, 1974, **96**, 4673–4674.

42 (*a*) G. Bélanger and P. Deslongchamps, *J. Org. Chem.*, 2000, **65**, 7070–7074; (*b*) E. Marsault, A. Toró, P. Nowak and P. Deslongchamps, *Tetrahedron*, 2001, **57**, 4243–4260.

43 (*a*) D. Carmona, M. Pilar-Lamata and L. A. Oro, *Coord. Chem. Rev.*, 2000, **200–202**, 717–772; (*b*) H. B. Kagan and O. Riant, *Chem. Rev.*, 1992, **92**, 1007–1019.

44 (*a*) K. A. Jorgensen, *Angew. Chem., Int. Ed.*, 2000, **39**, 3558–3588; (*b*) M. P. Doyle, I. M. Phillips and W. Hu, *J. Am. Chem. Soc.*, 2001, **123**, 5366–5367; (*c*) M. Martín, M. M. Afonso, A. Galindo and J. A. Palenzuela, *Synlett*, 2001, 117–119.

45 P. Buonora, J.-C. Olsen and T. Oh, *Tetrahedron*, 2001, **57**, 6099–6138.

46 (*a*) M. Behforouz and M. Ahmadian, *Tetrahedron*, 2000, **56**, 5259–5288; (*b*) F. Pautet, P. Nebois, Z. Bouaziz and H. Fillion, *Heterocycles*, 2001, **54**, 1095–1138.

47 (*a*) These reactions have not been reported; nevertheless, **7b** reacts with a β-substituted acrolein to give a *ca.* 1:1 Z/E mixture, see ref. 48(c); (*b*) For an extensive list of reviews, see ref. 31; (*c*) In principle the Z/E ratio in **6a** can be adjusted so that the product mixture from **6a** and **7a** is the same as that from **6b** and **7b**.

48 (*a*) J. S. Yadav, M. Y. Valli and A. R. Prasad, *Pure Appl. Chem.*, 2001, **73**, 1157–1162; (*b*) E. A. Couladouros and A. P. Mihou, *Tetrahedron Lett.*, 1999, **40**, 4861–4862; (*c*) T. Ando and H. Ohsawa, *J. Chem. Ecology*, 1993, **19**, 119–132; (*d*) H. J. Bestmann and N. E. Gunawardena, *Synthesis*, 1992, 1239–1241; (*e*) W. S. Leal, Y. Kuwahara and T. Suzuki, *Agric. Biol. Chem.*, 1989, **53**, 875–878; (*f*) S. C. Jain, D. E. Dussourd, W. E. Connor, T. Eisner, A. Guerrero and J. Meinwald, *J. Org. Chem.*, 1983, **48**, 2266–2270.

49 (*a*) W. Reppe, N. v. Kutepow and A. Magin, *Angew. Chem., Int. Ed. Engl.*, 1969, **8**, 727–733; (*b*) K. P. C. Vollhardt, *Angew. Chem., Int. Ed. Engl.*, 1984, **23**, 539–556.

50 (*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.

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

52 (*a*) R. L. Funk and K. P. C. Vollhardt, *J. Am. Chem. Soc.*, 1977, **99**, 5483–5484; (*b*) R. L. Funk and K. P. C. Vollhardt, *J. Am. Chem. Soc.*, 1979, **101**, 215–217.

53 (*a*) M. J. Eichberg, R. L. Dorta, D. B. Grotjahn, K. Lamottke, M. Schmidt and K. P. C. Vollhardt, *J. Am. Chem. Soc.*, 2001, **123**, 9324–9337; (*b*) C. Eickmeier, D. Holmes, H. Junga, A. J. Matzger, F. Scherhag, M. Shim and K. P. C. Vollhardt, *Angew. Chem., Int. Ed.*, 1999, **38**, 800–804; (*c*) C. Eickmeier, H. Junga, A. J. Matzger,

F. Scherhag, M. Shim and K. P. C. Vollhardt, *Angew. Chem., Int. Ed. Engl.*, 1997, **36**, 2103–2108; (*d*) C. Saá, D. D. Crotts, G. Hsu and K. P. C. Vollhardt, *Synlett.*, 1994, 487–489; (*e*) D. B. Grotjahn and K. P. C. Vollhardt, *Synthesis*, 1993, 579–605.

54 (*a*) K. M. Brummond and J. L. Kent, *Tetrahedron*, 2000, **56**, 3263–3283; (*b*) N. E. Schore, *Org. React.*, 1991, **40**, 1–90 ; see also ref. 57(a).

55 H. Becker, J. Henkelmann, M. Kiefer and T. Preiss, *Ger. Pat.* DE 10 012 553, 2001.

56 (*a*) M. Thommen, A. L. Veretenov, R. Guidetti-Grept and R. Keese, *Helv. Chim. Acta*, 1996, **79**, 461–476; (*b*) S. G. Van Ornum and J. M. Cook, *Tetrahedron Lett.*, 1996, **37**, 7185–7188; (*c*) S. G. Van Ornum and J. M. Cook, *Tetrahedron Lett.*, 1997, **38**, 3657–3658; (*d*) H. Cao, S. G. Van Ornum and J. M. Cook, *Tetrahedron Lett.*, 2000, **41**, 5313–5316; (*e*) S. G. Van Ornum, M. M. Bruendl, H. Cao, M. Reddy, D. S. Grubisha, D. W. Bennett and J. M. Cook, *J. Org. Chem.*, 2000, **65**, 1957–1971; (*f*) H. Cao, J. Flippen-Anderson and J. M. Cook, *J. Am. Chem. Soc.*, 2003, **125**, 3230–3231.

57 (*a*) P. L. Pauson, *Tetrahedron*, 1985, **41**, 5855–5860 (Co); (*b*) E. Negishi, S. J. Holmes, J. M. Tour and J. A. Miller, *J. Am. Chem. Soc.*, 1985, **107**, 2568–2569 (Zr); (*c*) K. Tamao, K. Kobayashi and Y. Ito, *J. Am. Chem. Soc.*, 1988, **110**, 1286–1288 (Ni); (*d*) K. Doyama, T. Joh, T. Shiohara and S. Takahashi, *Bull. Chem. Soc. Jpn.*, 1988, **61**, 4353–4360 (Rh); (*e*) A. J. Pearson and R. A. Dubbert, *J. Chem. Soc., Chem. Commun.*, 1991, 202–203 (Fe); (*f*) R. B. Grossman and S. L. Buchwald, *J. Org. Chem.*, 1992, **57**, 5803–5805 (Ti); (*g*) C. Mukai, M. Uchiyama and M. Hanaoka, *J. Chem. Soc., Chem. Commun.*, 1992, 1014–1015 (Mo); (*h*) T. R. Hoye and J. A. Suriano, *J. Am. Chem. Soc.*, 1993, **115**, 1154–1156 (W); (*i*) T. Kondo, N. Suzuki, T. Okada and T. Mitsudo, *J. Am. Chem. Soc.*, 1997, **119**, 6187–6188 (Ru); (*j*) T. Shibata and K. Takagi, *J. Am. Chem. Soc.*, 2000, **122**, 9852–9853 (Ir).

58 (*a*) F. A. Hicks and S. L. Buchwald, *J. Am. Chem. Soc.*, 1996, **118**, 11 688–11 689 (Ti); (*b*) N. Jeong, B. K. Sung and Y. K. Choi, *J. Am. Chem. Soc.*, 2000, **122**, 6771–6772; (*c*) see also refs. 54(a) and 57(j).

59 H. O. House, W. L. Respess and G. M. Whitesides, *J. Org. Chem.*, 1966, **31**, 3128–3141.

60 (*a*) Me$_2$CuLi: M. T. Reetz and A. Kindler, *J. Organometal. Chem.*, 1995, **502**, C5–C7 ; This study used catalytic Cu, which acts *via* the cuprate, since the Grignard reagent is in excess (ref. 59); See also refs. 61–63; (*b*) Et$_2$CuLi: S. Kurozumi, T. Toru, T. Tanaka, S. Miura, M. Kobayashi and S. Ishimoto, *US Pat.* 4 009 196, 1977.

61 O. Piva and S. Comesse, *Eur. J. Org. Chem.*, 2000, 2417–2424.

62 R. G. Salomon and M. F. Salomon, *J. Org. Chem.*, 1975, **40**, 1488–1492.

63 J. W. Benbow, R. Katoch, B. L. Martinez and S. B. Shetzline, *Tetrahedron Lett.*, 1997, **38**, 4017–4020.

64 (*a*) S. H. Bertz, *J. Am. Chem. Soc.*, 1990, **112**, 4031–4032; (*b*) A. Gerold, J. T. B. H. Jastrzebski, C. M. P. Kronenburg, N. Krause and G. van Koten, *Angew. Chem., Int. Ed. Engl.*, 1997, **36**, 755–757.

65 R. G. Pearson and C. D. Gregory, *J. Am. Chem. Soc.*, 1976, **98**, 4098–4104.

66 The organocopper(I) by-product would be hydrolyzed during the usual aqueous work-up to the corresponding alkane and a Cu(I) salt. In principle, methanol could be used to protonate the enolate, but leave the organocopper(I) species intact at low temperature; see: E. Piers and J. M. Chong, *Can. J. Chem.*, 1988, **66**, 1425–1429.

67 S. H. Bertz, G. Dabbagh and L. M. Williams, *J. Org. Chem.*, 1985, **50**, 4414–4415.

68 S. H. Bertz, M. Eriksson, G. Miao and J. P. Snyder, *J. Am. Chem. Soc.*, 1996, **118**, 10 906–10 907 and leading references cited therein to other approaches.

69 (*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.

70 (*a*) N. E. Schore, *Synth. Commun.*, 1979, **9**, 41–47; (*b*) For an asymmetric version, see: X. Verdaguer, J. Vázquez, G. Fuster, V. Bernardes-Génisson, A. E. Greene, A. Moyano, M. A. Pericàs and A. Riera, *J. Org. Chem.*, 1998, **63**, 7037; (*c*) See also ref. 90(f) for bond cleavage to a medium-sized ring.

71 (*a*) G. H. Posner, *An Introduction to Synthesis Using Organocopper Reagents*, Wiley-Interscience, New York, 1980; (*b*) *Organocopper Reagents–A Practical Approach*, ed. R.J.K. Taylor, Oxford University Press, Oxford, 1994, pp. 27–72; (*c*) *Modern Organocopper Chemistry*, ed. N. Krause, Wiley-VCH, Weinheim, 2002; (*d*) for an extensive list of reviews, see ref. 31.

72 (*a*) S. H. Bertz, G. Dabbagh and G. Sundararajan, *J. Org. Chem.*, 1986, **51**, 4953–4959; (*b*) B. E. Rossiter and N. M. Swingle, *Chem. Rev.*, 1992, **92**, 771–806; (*c*) For recent progress, see: N. Krause, in *Organic Synthesis Highlights IV*, ed. H.-G. Schmalz, Wiley-VCH, Weinheim, 2000, pp. 182–186.

73 S. Sakai and S. Takane, *J. Phys. Chem. A*, 1999, **103**, 2878–2882.

74 (*a*) S. D. Burke, J. L. Buchanan and J. D. Rovin, *Tetrahedron Lett.*, 1991, **32**, 3961–3964; (*b*) K. Mikami, K. Takahashi and T. Nakai, *J. Am. Chem. Soc.*, 1990, **112**, 4035–4037; (*c*) L. H. Zhang and J. M. Cook, *J. Am. Chem. Soc.*, 1990, **112**, 4088–4090.

75 (*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*) see also: 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.

76 (*a*) Homogeneous: M. Castiglioni, S. Deabate, E. Garrone, R. Giordano, B. Onida, G. Predieri and E. Sappa, *J. Cluster Sci.*, 1997, **8**, 381–405; (*b*) Heterogeneous: M. S. Eisen and T. J. Marks, *J. Am. Chem. Soc.*, 1992, **114**, 10 358–10 368.

77 (*a*) R. Noyori and T. Ohkuma, *Angew. Chem., Int. Ed.*, 2001, **40**, 40–73; (*b*) G. Zassinovich, G. Mestroni and S. Gladiali, *Chem. Rev.*, 1992, **92**, 1051–1069; (*c*) P. Antognazza, T. Benincori and F. Sannicolò, in *Seminars in Organic Synthesis (23rd Summer School 'A. Corbella')*, Societa Chimica Italiana, Rome, 1998, pp. 197–224.

78 In the case of multistep retrosynthetic schemes, other skeletal bonds can also be included to maintain continuity from structure to structure, see the next paper in this issue: S. H. Bertz, *New. J. Chem.*, DOI: 10.1039/b210844p.

79 (*a*) Aldol: A. T. Nielsen and W. J. Houlihan, *Org. React.*, 1968, **16**, 47–58; (*b*) Diels–Alder: J. R. Stille and R. H. Grubbs, *J. Org. Chem.*, 1989, **54**, 434–444 and see also ref. 42; (*c*) Wittig: F. Yvergnaux, Y. Le Floc'h, R. Grée and L. Toupet, *Tetrahedron Lett.*, 1989, **30**, 7393–7396; (*d*) Reppe–Vollhardt: see ref. 49(b); (*e*) Pauson–Khand: N. E. Schore and M. C. Croudace, *J. Org. Chem.*, 1981, **46**, 5436–5438 and see also ref. 54; (*f*) House–Whitesides: see ref. 85.

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

81 The five-bond ($C_6 \Rightarrow 4P_1 + P_2$) and six-bond ($C_6 \Rightarrow 6P_1$) disconnections result in greater simplification ($\Delta N_T = -30$ and -31, respectively), but they are not realistic for organic synthesis.

82 (*a*) J. Bao, W. D. Wulff, V. Dragisich, S. Wenglowsky and R. G. Ball, *J. Am. Chem. Soc.*, 1994, **116**, 7616–7630. We reserve 'Dötz reaction' for those cases that involve a double bond in conjugation with the metal-carbene bond, *e.g.*, the preparation of naphthalene derivatives from phenyl carbene complexes and 1 equiv of alkyne, see; (*b*) K. H. Dötz, *Angew. Chem., Int. Ed. Engl.*, 1984, **23**, 587–608; (*c*) K. H. Dötz and M. Popall, *Tetrahedron*, 1985, **41**, 5797–5802.

83 (*a*) N. E. Schore, *Chem. Rev.*, 1988, **88**, 1081–1119; (*b*) M. Lautens, W. Klute and W. Tam, *Chem. Rev.*, 1996, **96**, 49–92.

84 (*a*) L. S. Liebeskind, J. P. Leeds, S. L. Baysdon and S. Iyer, *J. Am. Chem. Soc.*, 1984, **106**, 6451–6453; (*b*) S. Iyer and L. S. Liebeskind, *J. Am. Chem. Soc.*, 1987, **109**, 2759–2770; (*c*) L. S. Liebeskind, S. L. Baysdon, M. S. South, S. Iyer and J. P. Leeds, *Tetrahedron*, 1985, **41**, 5839–5853.

85 P. A. Wender and S. L. Eck, *Tetrahedron Lett.*, 1977, **18**, 1245–1248.

86 G. Märkl, *Angew. Chem., Int. Ed. Engl.*, 1962, **1**, 511–511.

87 G. Büchi and H. Wüest, *Helv. Chim. Acta*, 1971, **54**, 1767–1776.

88 *Simplifying principle 1* also applies when the ratio of complexities of the components is constant.

89 For example, disconnection **32** corresponds to the reaction of 1,5-diiodopentane with the ammonia equivalent, trifluoroacetamide: P. A. Harland, P. Hodge, W. Maughan and E. Wildsmith, *Synthesis*, 1984, 941–943.

90 (*a*) G. H. Posner, *Chem. Rev.*, 1986, **86**, 831–844; (*b*) S. H. Bertz, J. M. Cook, A. Gawish and U. Weiss, *Org. Synth.*, 1986, **64**, 27–38; (*c*) I. Ugi, A. Dömling and W. Hörl, *Endeavour*, 1994, **18**, 115–122; (*d*) H. J. M. Gijsen and C.-H. Wong, *J. Am. Chem. Soc.*, 1995, **117**, 7585–7591; (*e*) H. Bienaymé, C. Hulme, G. Oddon and P. Schmitt, *Chem. Eur. J.*, 2000, **6**, 3321–3329; (*f*) G. H. Posner, Q. Wang, B. A. Halford, J. S. Elias and J. P. Maxwell, *Tetrahedron Lett.*, 2000, **41**, 9655–9659.

91 D. Bradley, *New Scientist*, 1997, November 15, pp. 40–43.