Organic Process Research & Development **2006**, *10,* 212–240

On Using Tree Analysis to Quantify the Material, Input Energy, and Cost Throughput Efficiencies of Simple and Complex Synthesis Plans and Networks: Towards a Blueprint for Quantitative Total Synthesis and Green Chemistry

John Andraos*

Department of Chemistry, York University, Toronto, ON M3J 1P3 Canada

Abstract:

Synthetic plans or networks may be depicted as trees in a graph-theoretical sense. When drawn in a systematic way according to a defined convention key "green" metrics relating to the efficiency of performance of a synthesis to a target molecule may be easily obtained by inspection, that is, by a "connect-the-dots" approach. Example metrics include the cumulative and overall reaction mass efficiency (RME), the overall raw materials cost (RMC), and the fraction of total energy input directed to product (FTE). Throughout this paper kernel metrics are used to determine and compare the intrinsic efficiencies of synthetic plans since these depend directly on the nature of the chemical transformations and not on ancillary variables such as solvent usage, etc. Histograms of these metrics versus reaction stage allow for the easy determination of the mass-, cost-, and input energy-determining steps for a given synthesis plan. Other useful parameters that can be determined from a synthesis tree include the degree of convergence, the degree of asymmetry, the optimum time to complete a synthesis, and the degree of building to target structure with respect to reaction stage (molecular weight first moment). All of these metrics allow for easy comparison and ranking of synthetic plans. It is demonstrated that the tree analysis is robust and is applicable to any synthetic plan or network of any degree of complexity. The concept of "overall reaction yield" is shown to be applicable only to linear synthesis plans or networks and is replaced by the more general overall RME metric for syntheses involving mixed linear and convergent segments. The synthesis of the antibacterial agent triclosan is used as a tutorial exercise to introduce key concepts. Further example synthetic plans analyzed by the present tree analysis illustrating various plan types include quinine (Woodward−Doering−Rabe, Stork, Jacobsen, and Acharya−Kobayashi methods), sildenafil (asymmetric convergent), absinthin (symmetric convergent), papaverine (convergent using common intermediates), bupleurynol (multicomponent convergent), and polypeptide syntheses (Fischer, Bergmann−Zervas, Merrifield, azide, anhydride, and segment doubling methods). Example synthetic networks examined include industrial syntheses of veronal (5,5-diethylbarbituric acid) (complex branching to target node) and feedstock products derived from phthalic anhydride (complex branching from source node).

1. Introduction

Since the coining of the terms "atom economy"[1] and "environmental impact factor" or "*E*-factor"[2] the study of so-called "green metrics" to quantify the "greenness" of individual chemical reactions is now a well-established branch of green chemistry. Two recent works have extended these ideas and unified key concepts into a coherent whole.[3] In defining reaction metrics it is important to be precise in definitions and mathematical representations so as to avoid confusion and misinterpretation. The terms atom economy (AE) and *E*-factor based on molecular weight (E_{mw}) are quantities depending only on the molecular weights of reactants and products in a balanced chemical equation. (For the sake of expediency in carrying out computations for identification of material-efficient synthetic plans, molecular weights using atomic weights of most abundant isotopes of elements may be used as is done throughout this work. Errors in the resulting reaction metrics amount to less than one part in a thousand.) Therefore, one does not include solvent, catalyst, or any other molecules other than reactants and products in the definition of atom economy. On the other hand, the terms reaction mass efficiency (RME) and *E*-factor based on mass (E_m) are broader quantities based on the actual masses of reactants, solvents, catalysts, and other materials used in performing a given chemical reaction. They may or may not include molecules other than reactants and products, depending on whether these ancillary molecules are reclaimed or eliminated in a chemical process. If they are reclaimed or eliminated, then the resultant RME and E_m metrics are called *kernel* metrics since they depend on the intrinsic chemical performance of a reaction. That is, they depend on both molecular weights of reactants and products *and* the reaction yield. If, on the other hand, the ancillary molecules are not reclaimed or eliminated, then the full general definitions of RME and E_m apply (vide infra).

A number of key points were made in unifying ideas.[3] First, it was established that the general master equation for

* To whom correspondence should be addressed. E-mail: jandraos@yorku.ca. Fax: 1-416-736-5936.

(1) (a) Trost, B. M. *Science* **1991**, *254*, 1471. (b) Trost, B. M. *Acc. Chem. Res.* 2002, **35**, 695. (c) Trost, B. M. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 259. (d) Sheldon, R. A. *Pure Appl. Chem.*, **2001**, *72*, 1233.

(2) (a) Sheldon, R. A. *Chem. Ind. (London)* **1992**, 903. (b) Sheldon, R. A. *CHEMTECH* **1994**, *24*(3), 38. (c) Sheldon, R. A. *Chem. Ind. (London)* **1997**, 12.

(3) (a) Andraos, J. *Org. Process Res. Dev.* **2005**, *9*, 149. (b) Andraos, J. *Org. Process Res. Dev.* **2005**, *9*, 404.

Published on Web 02/14/2006
10.1021/op0501904 CCC: $33.50

reaction mass efficiency (RME),[4] given by eq 1, for a single chemical reaction is governed by four independent factors each ranging in value between 0 and 1: reaction yield (ϵ), atom economy (AE), reciprocal of stoichiometric factor (1/SF) taking into account excess reagents used, and materials recovery parameter (MRP) taking into account reaction solvent, reaction catalyst, and all other materials usage in the postreaction workup and purification stages. Each of these factors potentially act to attenuate the value of RME.

$$\mathrm{RME} = (\epsilon)(\mathrm{AE})\left(\frac{1}{\mathrm{SF}}\right)(\mathrm{MRP}) = (\epsilon)(\mathrm{AE})\left(\frac{1}{\mathrm{SF}}\right)\left(\frac{1}{1+\dfrac{\epsilon(\mathrm{AE})[c+s+\omega]}{(\mathrm{SF})(m_\mathrm{P})}}\right) \quad (1)$$

where c, s, and ω are the masses of reaction catalyst, reaction solvent, and all postreaction materials respectively, and m_P is the mass of the collected target product. It can be seen that the theoretical maximum value of RME is equal to 1 for a reaction run under stoichiometric conditions (SF = 1), with all catalysts, solvents, and other postreaction materials recovered or eliminated (MRP = 1), with reaction yield 100% (ϵ = 1), and producing no byproducts (AE = 1). The practical maximum value of RME is ϵ(AE) which is governed only by the instrinsic chemical performance of the reaction. The reaction yield and atom economy metrics are therefore kernel metrics. Second, it was shown that the environmental impact factor based on mass (E_m), commonly called the Sheldon *E*-factor, is related to the RME by the simple expression given in eq 2 which results as a consequence of the law of conservation of mass for a chemical reaction.

$$\mathrm{RME} = \frac{1}{1+E_\mathrm{m}} \quad (2)$$

An analogous relationship exists between AE and E_mw. Third, it was established that for comparisons of chemical performances of raw material efficiency to be made between individual reactions leading to the same target product, as for example oxidation of a given secondary alcohol by various oxidizing agents, or between different synthesis plans leading to the same target product, such as a complex pharmaceutical, it is sufficient to compare only RME values under best-case scenario conditions (i.e., reclaiming and/or elimination of solvents and other ancillary materials in workup and purification phases) based on the kernel metrics ϵ and AE. Such an RME is therefore a kernel RME. This assertion is valid since the other two metrics SF and MRP will necessarily attenuate RME and do not contribute to the intrinsic chemical behaviour of a given reaction. Generally, MRP will lower RME to a greater degree than 1/SF. These ideas were applied to a database of more than 400 named organic reactions in which minimum atom economies were calculated on the basis of generalized Markush chemical reactions including the determination of probability functions for the likelihood a given reaction would exceed a given threshold value of RME under a variety of constraints. Simple numerical algorithms for determining RMEs for linear and simple convergent synthetic plans were also determined.

The upshot of the previous probability analysis[3b] was that for a reaction to be called "green" it (a) must have an AE greater than 61.8% so that AE > E_mw and (b) must reclaim and/or eliminate solvents and other ancillary materials so that RME is also at least 61.8% and hence RME > E_m. *Both criteria must be satisfied*. Therefore, reactions with AE values of 100% and whose solvents are committed to waste are not "green". Similarly, reactions which eliminate solvents altogether but have AE values below 61.8% are also not "green". These rigorous criteria show that the achievement of truly "green" reactions is at best achievable with a 38% probability if the AE cutoff is set at 61.8%. A thorough survey of the database of named organic reactions shows that about 55% of them have a chance of meeting this target provided that they are carried out with a minimum of solvent and without using excess reagents.

This report introduces the depiction of simple and complex synthetic plans and networks as trees which greatly simplifies the determination of kernel RME values and avoids lengthy algebraic calculations. When drawn in a systematic way according to a defined convention a number of key green metrics may be easily obtained *by inspection*, that is, by a "connect-the-dots" approach. Example metrics include the cumulative and overall kernel reaction mass efficiency (RME), the overall kernel raw materials cost (RMC), and the fraction of total energy input directed to product (FTE). Histograms of these metrics versus reaction stage allow for the easy determination of the mass-, cost-, and input energy-determining steps for a given synthesis plan. Other new and useful parameters that can be determined from a synthesis tree include the degree of convergence, the degree of asymmetry, the minimum time to complete a synthesis, and the molecular weight moment to target structure with respect to reaction stage. All of these metrics allow for easy comparison and ranking of synthetic plans. It is demonstrated that the tree analysis is robust and is applicable to any synthetic plan or network of any degree of complexity. The concept of "overall reaction yield" is shown to be applicable only to linear synthesis plans or networks and is replaced by the more general overall kernel RME metric for syntheses involving mixed linear and convergent segments. A "pseudo-overall reaction yield" given by $\epsilon_{\mathrm{pseudo-overall}} = (\mathrm{RME})_{\mathrm{overall}}/(\mathrm{AE})_{\mathrm{overall}}$ may also be used for complex synthetic plans or networks.

We begin by introducing a convention for drawing a synthesis tree or network from a reaction scheme that can be applied to any situation. All metrics appearing in this work are defined and derived using a simple methodology based on these trees. The synthesis of the antibacterial agent triclosan is used as a tutorial exercise. Further illustrative

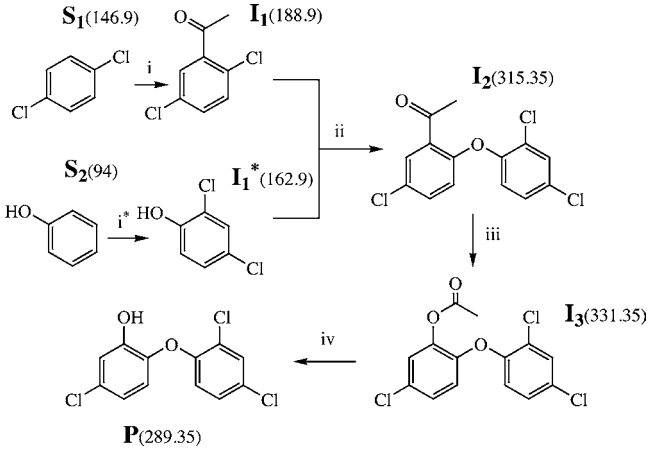
(4) There are other equivalent terms given by other authors for reaction mass efficiency. Eissen and Metzger use "mass index" which is the reciprocal of RME as defined in this paper (see Eissen, M.; Metzger, J. O. *Chem. Eur. J.* **2002**, *8*, 3580). Steinbach and Winkenbach use "balance yield" or "Bilanzausbeute" for RME (see Steinbach, A.; Winkenbach, R. *Chem. Eng.* **2000**, *April*, 94).

example synthetic plans analyzed by the present tree analysis illustrating various plan types include quinine (Woodward−Doering−Rabe, Stork, Jacobsen, and Acharya−Kobayashi methods), sildenafil (asymmetric convergent), absinthin (symmetric convergent), papaverine (convergent using common intermediate), bupleurynol (multicomponent convergent), and polypeptide syntheses (Fischer, Bergmann−Zervas, Merrifield, azide, anhydride, and segment doubling methods). Example synthetic networks examined include industrial syntheses of veronal (complex branching to target node) and feedstock products derived from phthalic anhydride (complex branching from source node).

2. Tree Construction

Trees are well-defined objects in graph theory[5] and have found application in a limited range of organic chemistry problems. The stick and wedge diagrams chemists draw depicting chemical structures are in fact graphs, where the labeled nodes represent atoms, and the lines or branches represent covalent bonds. The oldest application of graph theory to chemistry is the enumeration of structural isomers for a given molecular formula which was first investigated by the mathematician Arthur Cayley.[6] Related to this is the relationship, discovered by Oliver J. Lodge, for determining the number of rings and/or unsaturations for a given molecular formula which is based on the concept of valence.[7] This relationship is extremely powerful in narrowing down the structural possibilities for an unknown compound and can be correlated with chemical and spectroscopic evidence. Hendrickson depicted synthetic plans as trees and defined a convergence parameter based on the total number of possible paths from all input nodes to the final target node.[8] Bertz has used ideas in graph theory and information theory in organic synthesis to define molecular complexity.[9] Kier and Hall have quantified molecular connectivity in terms of valence indices and used it to correlate properties of compounds for quantitative structure activity modelling studies.[10]

Scheme 1[a]



[a] Reaction conditions: (i) acetyl chloride, $AlCl_3$ catalyst (94.3%); (i)* 2 Cl_2 (81%); (ii) 1/2 K_2CO_3, CuCl catalyst, xylenes (48.3%); (iii) 62.5% H_2O_2, 1/2 maleic anhydride, CH_2Cl_2 (91.3%); (iv) MeOH, 35% HCl catalyst (94.5%). Molecular weights in g/mol are given in parentheses.

In this work a synthetic plan such as that shown by the set of reactions for the synthesis of the antibacterial triclosan[11] in Scheme 1 may be depicted as a synthesis tree as shown in Figure 1. This paper describes the first application of graphical trees in the determination of the efficiencies of synthesis plans and networks that incorporate "green" metrics parameters. In constructing a synthesis tree, read from left to right, the following conventions are made: (a) the *x*-axis represents reaction stages; (b) the *y*-axis represents input structures entered as filled dots beginning at the origin; (c) for a given reaction step input structures are entered vertically with unit spacing; and (d) intermediate and final products are represented as open and shaded circles, respectively, whose coordinates are determined as the centroids of the dots corresponding to their preceding reactant input structures according to eq 3 until the final product is reached,

$$\text{centroid} = \frac{1}{2^{n-1}}\left[\sum_{j=0}^{n-1} a_{j+1}\binom{n-1}{j}\right] = \frac{1}{2^{n-1}}\left[\sum_{j=0}^{n-1} a_{j+1}\left(\frac{(n-1)!}{(j!)(n-1-j)!}\right)\right] \quad (3)$$

(5) (a) Bonchev D.; Rouvray, R. H. *Chemical Graph Theory: Introduction and Fundamentals*; Taylor & Francis: London, 2003. (b) Andrasfai, B. *Introductory Graph Theory*; Adam Hilger: Bristol, 1977. (c) Trinajstic, N. *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1992. (d) Wilson, R. J.; Watkins, J. J. *Graphs: An Introductory Approach*; Wiley: New York, 1990. For representative articles on applications of graph theory to chemistry see: Rouvray, D. H. *Endeavour* **1975**, *34*, 28; Rouvray, D. H. *Am. Scientist* **1973**, *61*, 729; Rouvray, D. H. *CHEMTECH* **1973**, 379; Rouvray, D. H. *Chem. Br.* **1977**, *13*, 52; Balaban, A. T.; Kennedy, J. W.; Quintas, L. V. *J. Chem. Educ.* **1988**, *65*, 304; and Bertz, S. H. *Discrete Appl. Math.* **1988**, *19*, 65.

(6) (a) Cayley, A. *Philos. Mag.* **1854**, *7*[4], 40. (b) Cayley, A. *Philos. Mag.* **1857**, *13*[4], 172. (c) Cayley, A. *Philos. Mag.* **1859**, *18*[4], 374. (d) Cayley, A. *Philos. Mag.* **1860**, *20*[4], 337. (e) Crum Brown, A. *Trans. R. Soc. Edinburgh* **1864**, *23*, 707. (f) Cayley, A. *Philos. Mag.* **1874**, *47*[4], 444. (g) Cayley, A. *Chem. Ber.* **1875**, *8*, 1056. (h) Cayley, A. *Rep. Br. Assoc. Adv. Sci.* **1875**, 257. (i) Cayley, A. *Philos. Mag.* **1877**, *3*[5] 34. (j) Cayley, A. *Am. J. Math.* **1881**, *4*, 266. (k) Henze, H. R.; Blair, C. M. *J. Am. Chem. Soc.* **1931**, *53*, 3077. (l) Polya, G. *Acta Math.* **1936**, *68*, 145.

(7) (a) Lodge, O. J. *Philos. Mag.* **1875**, *50*[4], 367. (b) Rouvray, D. H. *J. Chem. Educ.* **1975**, *52*, 768. In graph theoretical terms this quantity is known as the cyclomatic number. The general formula for the number of rings and/or unsaturations for molecular formula $C_aH_bX_cN_dO_eS_fP_gB_hSi_i$ (X = F, Cl, Br, I) is given by $^1/_2[2 + \sum_{j=1}^{k} n_j(v_j - 2)]$ where k is the number of element types in the molecular formula, n_j is the number of the jth element type (a for C, b for H, c for X, d for N, e for O, f for S, g for P, h for B, i for Si), and v_j is the valence of the jth element type (4 for C, 1 for H, 1 for X, 3 for N (amino groups), 4 for N+ (ammonium, nitro, and azoxy groups), 2 for O (alcohols, peroxides), 1 for O− (organic oxides), 2 for S (thiols, sulfides), 4 for S (sulfoxides, sulfinates, sulfites), 6 for S (sulfones, sulfates, sulfonates), 3 for P (phosphines, phosphinates), 5 for P (phosphine oxides, phosphonates), 6 for P (phosphates), 3 for B, and 4 for Si. The formula is also valid for structures of transient species such as carbocations, carbanions, and carbenes where the valence for C is set to 3, 3, and 2, respectively.

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

(9) (a) Bertz; S. H.; Sommer, T. J. In *Organic Synthesis: Theory and Applications*; Hudlicky, T., Ed.; JAI Press: Greenwich: Connecticut, 1993; Vol. 2, p 67 and references therein. (b) Bertz, S. H. *J. Am. Chem. Soc.* **1981**, *103*, 3599. (c) Bertz, S. H. *J. Am. Chem. Soc.* **1982**, *104*, 5801. (d) Bertz, S. H. *Bull. Math. Biol.* **1983**, *45*, 849. (e) Bertz, S. H. *New J. Chem.* **2003**, *27*, 860. (f) Bertz, S. H. *New J. Chem.* **2003**, *27*, 870.

(10) (a) Hall, L. H.; Kier, L. B. *J. Mol. Graphics Model.* **2001**, *20*, 4. (b) Hall, L. H.; Kier, L. B. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; Vol. 2, p 367 and references therein.

(11) Lourens, G. J. WO Patent 9910310, 1997.

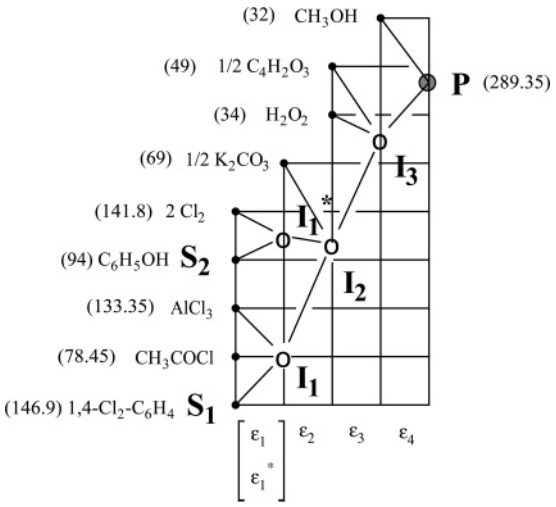


Figure 1. **Synthesis tree for the synthesis of triclosan according to reactions given in Scheme 1. Synthesis parameters: 9 inputs, 4 intermediates, 4 reaction stages, 5 reactions, 2 parallel reactions. Synthesis type: mixed linear and convergent. Molecular weights in grams per mole for input reactant and final product output nodes are given in parentheses. Reaction yields: $\epsilon_1 = 0.943$, $\epsilon_1^* = 0.81$, $\epsilon_2 = 0.483$, $\epsilon_3 = 0.913$, and $\epsilon_4 = 0.945$.**

where $n \geq 2$, n is the number of points corresponding to the number of reactant input structures, a_{j+1} is the ordinate of the $(j + 1)$th input, and $0! = 1$ by definition. Equation 3 is well-known in classical mechanics where it is used to calculate the center of mass coordinates of equal point masses along a straight line. In the case of a synthesis plan having multiple linear segments that converge, the synthesis tree is constructed with the longest branch beginning at the origin.

From Figure 1 example calculations illustrating the use of eq 3 to determine the ordinates of intermediate nodes follow. The ordinates of intermediates I_1, I_1*, I_2, I_3, and P as given in the last column of Table 1 are

$$\text{centroid}(I_1) = \frac{1}{2^{3-1}}\left[a_1\binom{2}{0} + a_2\binom{2}{1} + a_3\binom{2}{2}\right] = \frac{1}{4}[0(1) + 1(2) + 2(1)] = 1$$

$$\text{centroid}(I_1^*) = \frac{1}{2^{2-1}}\left[a_1\binom{1}{0} + a_2\binom{1}{1}\right] = \frac{1}{2}[3(1) + 4(1)] = 3.5$$

$$\text{centroid}(I_2) = \frac{1}{2^{3-1}}\left[a_1\binom{2}{0} + a_2\binom{2}{1} + a_3\binom{2}{2}\right] = \frac{1}{4}[1(1) + 3.5(2) + 5(1)] = 3.25$$

$$\text{centroid}(I_3) = \frac{1}{2^{3-1}}\left[a_1\binom{2}{0} + a_2\binom{2}{1} + a_3\binom{2}{2}\right] = \frac{1}{4}[3.25(1) + 6(2) + 7(1)] = 5.5625$$

$$\text{centroid}(P) = \frac{1}{2^{2-1}}\left[a_1\binom{1}{0} + a_2\binom{1}{1}\right] = \frac{1}{2}[5.5625(1) + 8(1)] = 6.78125$$

Note that the ordinates of I_1 and I_1* can be determined by inspection as the geometric centers of the preceding input nodes.

All chemical equations in the plan are balanced with appropriate stoichiometric coefficients. A distinction is made here between reaction steps and reaction stages. A reaction stage may be composed of a single reaction step or at least two parallel reaction steps run simultaneously. In Figure 1 it can be seen that the first stage is composed of two parallel reactions. This designation is important in spotting points of convergence in a synthesis plan or network, in determining

Table 1. **Summary of molecular weights, scales, masses, and coordinates for all nodes in triclosan synthesis tree shown in Figure 1**

Node	MW (g/mol)	Scale (moles)	Mass (g)	Co-ordinates
P	289.35	x	$289.35x$	$\left(4, \frac{217}{32}\right) = (4, 6.78125)$
CH_3OH	32	$\frac{x}{\epsilon_4}$	$\frac{32x}{\epsilon_4}$	(3,8)
I_3	331.35	$\frac{x}{\epsilon_4}$	$\frac{331.35x}{\epsilon_4}$	$\left(3, \frac{89}{16}\right) = (3, 5.5625)$
$1/2\ C_4H_2O_3$	49	$\frac{x}{\epsilon_3\epsilon_4}$	$\frac{49x}{\epsilon_3\epsilon_4}$	(2,7)
H_2O_2	34	$\frac{x}{\epsilon_3\epsilon_4}$	$\frac{34x}{\epsilon_3\epsilon_4}$	(2,6)
I_2	315.35	$\frac{x}{\epsilon_3\epsilon_4}$	$\frac{315.35x}{\epsilon_3\epsilon_4}$	$\left(2, \frac{13}{4}\right) = (2, 3.25)$
$1/2\ K_2CO_3$	69	$\frac{x}{\epsilon_2\epsilon_3\epsilon_4}$	$\frac{69x}{\epsilon_2\epsilon_3\epsilon_4}$	(1,5)
I_1^*	162.9	$\frac{x}{\epsilon_2\epsilon_3\epsilon_4}$	$\frac{162.9x}{\epsilon_2\epsilon_3\epsilon_4}$	$\left(1, \frac{7}{2}\right) = (1, 3.5)$
I_1	188.9	$\frac{x}{\epsilon_2\epsilon_3\epsilon_4}$	$\frac{188.9x}{\epsilon_2\epsilon_3\epsilon_4}$	(1,1)
$2\ Cl_2$	141.8	$\frac{x}{\epsilon_1^*\epsilon_2\epsilon_3\epsilon_4}$	$\frac{141.8x}{\epsilon_1^*\epsilon_2\epsilon_3\epsilon_4}$	(0,4)
S_2 (phenol)	94	$\frac{x}{\epsilon_1^*\epsilon_2\epsilon_3\epsilon_4}$	$\frac{94x}{\epsilon_1^*\epsilon_2\epsilon_3\epsilon_4}$	(0,3)
$AlCl_3$	133.35	$\frac{x}{\epsilon_1\epsilon_2\epsilon_3\epsilon_4}$	$\frac{133.35x}{\epsilon_1\epsilon_2\epsilon_3\epsilon_4}$	(0,2)
CH_3COCl	78.45	$\frac{x}{\epsilon_1\epsilon_2\epsilon_3\epsilon_4}$	$\frac{78.45x}{\epsilon_1\epsilon_2\epsilon_3\epsilon_4}$	(0,1)
S_1 (1,4-dichlorobenzene)	146.9	$\frac{x}{\epsilon_1\epsilon_2\epsilon_3\epsilon_4}$	$\frac{146.9x}{\epsilon_1\epsilon_2\epsilon_3\epsilon_4}$	(0,0)

the optimum time required to complete a synthesis, and in appropriate resource and time management in the planning of a synthesis. Each reaction step has an associated reaction yield with respect to the limiting reagent (ϵ), which is the multiplicative product of the intrinsic chemical yield, the workup yield, and the purification yield, an associated reaction time (t), and an associated energy input (Ψ). A fundamental assumption underlying the tree construction is that each intermediate product collected is entirely committed as a reactant in the successive step. Essentially the tree traces the mass attenuation of input materials, or throughput, until the final product is reached. The reaction sequence is akin to a set of successive sieves each with a given mesh size corresponding to a reaction yield that will dictate how much material passes from one reaction stage to the next. The coordinates of the final product are important in determining the degree of convergence of the plan with respect to the total number of input reactants. The next section describes kernel green metrics and key parameters that can be deduced from the shape and connectivity of the synthesis tree that are useful in describing the efficiency of a synthesis plan.

3. Kernel Green Metrics

3a. Materials Usage Analysis. In determining the overall kernel RME for a synthesis plan from its synthesis tree one needs to determine the ratio of the output product mass to the sum of all input reactant masses. The calculation is carried out by determining what input masses are required for all reactants so that a target mass of product is obtained given the experimental reaction yields for each reaction step. From the synthesis tree this is easily achieved by defining a target scale for the final product in moles, x = mass of target product/MW target product, and working backwards toward reactant inputs following the lines connecting the dots to determine the required scales at each intermediate and input node. The scale at a given node is given by the quotient of the final product scale and the multiplicative product of the reaction yields corresponding to the reaction steps connecting that node to the target product node as traced by the joining lines in the synthesis tree. At each node the corresponding mass in grams is obtained by multiplying the scale in moles at that node by the molecular weight in grams per mole of the corresponding chemical structure. Table 1 summarizes the data for the triclosan synthesis tree shown in Figure 1. Note that in the convergent first stage the scales for aluminum trichloride, acetyl chloride, and 1,4-dichlorobenzene in the Friedel−Crafts reaction depend on the ϵ_1 reaction yield; whereas, the scales for phenol and chlorine in the chlorination reaction depend on the ϵ_1* reaction yield. The kernel overall RME for the triclosan synthesis plan is then given by eqs 4a−b.

$$\mathrm{RME} = \frac{289.35x}{S} \tag{4a}$$

where, S, the total mass of reagents used is given by

$$S = x\left[\frac{32}{\epsilon_4} + \frac{49+34}{\epsilon_3\epsilon_4} + \frac{69}{\epsilon_2\epsilon_3\epsilon_4} + \frac{133.35+78.45+146.9}{\epsilon_1\epsilon_2\epsilon_3\epsilon_4} + \frac{94+141.8}{\epsilon_1^*\epsilon_2\epsilon_3\epsilon_4}\right]$$

$$= x\left[\frac{32}{\epsilon_4} + \frac{83}{\epsilon_3\epsilon_4} + \frac{69}{\epsilon_2\epsilon_3\epsilon_4} + \frac{358.7}{\epsilon_1\epsilon_2\epsilon_3\epsilon_4} + \frac{235.8}{\epsilon_1^*\epsilon_2\epsilon_3\epsilon_4}\right] \tag{4b}$$

Using reported yields in the patent[11] of $\epsilon_1{}^* = 0.81$,$\epsilon_1 = 0.943$, $\epsilon_2 = 0.483$, $\epsilon_3 = 0.913$, and $\epsilon_4 = 0.945$ eqs 4a and 4b yield 0.1517 for the value of the kernel RME for this synthesis. The number of terms in the denominator of the RME expression correspond to the number of reaction steps in the synthesis plan.[13] The corresponding overall kernel E-factor based on mass using eq 2 is 5.59 g waste per g triclosan. In general the kernel overall RME is given by

$$\mathrm{RME} = \frac{px}{\sum_j^M \frac{xr_j}{\epsilon_M\epsilon_{M-1}\ldots\epsilon_j}} = \frac{p}{\sum_j^M \frac{r_j}{\epsilon_M\epsilon_{M-1}\ldots\epsilon_j}} \tag{5}$$

where p and r_j are the molecular weights of target product and jth reactant, respectively, the reaction yields correspond to the connecting paths in the synthesis tree between the jth reactant and the final product, and M is the number of reaction steps. Note that if solvents, catalysts, excess reagents, and postreaction materials are included for each step then the complete RME expression becomes

$$\mathrm{RME} = \frac{px}{\sum_j^M \frac{xr_j}{\epsilon_M\epsilon_{M-1}\ldots\epsilon_j} + \sum_j^M (s_j + c_j + \phi_j + \omega_j)} \tag{6}$$

where $\sum_j s_j$ is the sum of masses of all solvents, $\sum_j c_j$ is the sum of masses of all catalysts used, $\sum_j \phi_j$ is the sum of masses of all excess reagents, and $\sum_j \omega_j$ is the sum of masses of all postreaction materials in the workup and purification phases for all M reactions in the plan. Equation 6 reduces to eq 5 when all of these extra terms are set to zero. The kernel overall RME expressions given by eqs 4 and 5 represent best-case scenarios and are the ones used to gauge the intrinsic chemical performances of synthetic plans. When all reaction yields in eqs 4 and 5 are set to 1 (i.e., optimum reaction reaction conditions of 100%) the kernel overall RME expressions collapse to the overall atom economy as expected. In the case of the triclosan synthetic plan this value is 37% from eqs 4a,b.

$$\lim_{\epsilon_j \to 1} \mathrm{RME} = \mathrm{AE} = \frac{289.35}{32+83+69+358.7+235.8} = \frac{289.35}{778.5} = 0.3717 \tag{7}$$

The corresponding E-factor based on molecular weight is $E_{\mathrm{mw}} = (1/AE) - 1 = 1.69$.

(12) Ebel, E.; Bell, J.; Fries, A.; Kasey, C.; Berkebile, J. M. *J. Chem. Educ.* **1947**, *24*, 449.

(13) This is true for synthesis plans consisting of reactions that each have at least one reactant input. In general the number of terms in the denominator of the RME expression is equal to the total number of reactions in the plan minus the number of reactions that do not involve additional reactant inputs such as rearrangements or other unimolecular transformations.

Since the synthesis plan has a point of convergence it is not possible to define an "overall yield" for the entire synthesis by simply multiplying the respective reaction yields as would be correct for a truly linear synthesis. This can be deduced by observing that there is no common yield factor that both the denominator and numerator can be multiplied by so that all fractions in the denominator disappear. In the case of a linear plan this would be possible and thus the resulting numerator would be the multiplicative product of the molecular weight of the target product, the scale of the entire synthesis, and a reaction yield factor given by $\epsilon_1\epsilon_2...\epsilon_M$ commonly referred to as the "overall yield". Previous to this work "overall yields" for complex syntheses were commonly determined either by erroneously multiplying all reaction yields in a plan, or more correctly by multiplying reaction yields in the longest branch of a plan usually corresponding, in the language of graph theoretical trees, to the root of its synthesis tree. In general for any synthesis plan of any degree of complexity the present work shows that the overall kernel RME, which incorporates reaction yields and atom economies according to the connectivity of reaction inputs, intermediates and final target product is indeed the best measure of its material efficiency. Alternatively, one may define a "pseudo-overall yield" as

$$\epsilon_{\text{pseudo-overall}} = \frac{\text{RME}_{\text{overall kernel}}}{\text{AE}_{\text{overall}}} \tag{8}$$

For a linear plan the pseudo-overall yield as defined above is numerically close to the multiplicative product of the reaction yields. The difference between the two values diminishes as the reaction yields approach 1 as would be expected. However, for complex plans with several converging branches this alternative definition becomes less useful.

A great advantage of the synthesis tree approach is that kernel RME and AE metrics and mass of waste production may be determined between any two reaction steps or stages in a synthesis plan by following the connecting paths between the relevant nodes. This allows for cumulative kernel metrics to be determined as well as overall metrics. For example, in the triclosan plan the second stage consists of a coupling reaction between intermediates I_1 and I_1*. From the reaction scales for the appropriate nodes as given in Table 1 the kernel RME for this reaction is

$$\begin{aligned}(\text{RME})_2 &= \frac{\text{mass}_{\text{output},2}}{\text{mass}_{\text{input},2}} \\ &= \frac{\dfrac{315.35x}{\epsilon_3\epsilon_4}}{\dfrac{69x}{\epsilon_2\epsilon_3\epsilon_4} + \dfrac{162.9x}{\epsilon_2\epsilon_3\epsilon_4} + \dfrac{188.9x}{\epsilon_2\epsilon_3\epsilon_4}} \\ &= \frac{315.35\epsilon_2}{69 + 162.9 + 188.9} \\ &= 0.7494\epsilon_2 \\ &= (\text{AE})_2\epsilon_2 \\ &= 0.3620 \end{aligned} \tag{9}$$

and the kernel E_m is

$$E_{m,2} = \frac{1}{(\text{RME})_2} - 1 = \frac{1}{0.3620} - 1 = 1.78 \tag{10}$$

The minimum mass of waste generated in grams in this step is

$$\begin{aligned}\bar{w}_2 &= \text{mass}_{\text{input},2} - \text{mass}_{\text{output},2} \\ &= \frac{x}{\epsilon_2\epsilon_3\epsilon_4}[69 + 162.9 + 188.9] - \frac{x}{\epsilon_3\epsilon_4}315.35 \\ &= \frac{x}{\epsilon_3\epsilon_4}\left[\frac{420.8}{\epsilon_2} - 315.35\right] \\ &= 644.28x \end{aligned} \tag{11}$$

The cumulative kernel RME from reaction stage 1 to reaction stage 2 is given by

$$\begin{aligned}&(\text{RME})_{1\to2} \\ &= \frac{\text{mass}_{\text{output},1\to2}}{\text{mass}_{\text{input},1\to2}} \\ &= \frac{\dfrac{315.35x}{\epsilon_3\epsilon_4}}{\dfrac{(146.9 + 78.45 + 133.35)x}{\epsilon_1\epsilon_2\epsilon_3\epsilon_4} + \dfrac{(94 + 141.8)x}{\epsilon_1^*\epsilon_2\epsilon_3\epsilon_4} + \dfrac{69x}{\epsilon_2\epsilon_3\epsilon_4}} \\ &= \frac{315.35}{\dfrac{358.7}{\epsilon_1\epsilon_2} + \dfrac{235.8}{\epsilon_1^*\epsilon_2} + \dfrac{69}{\epsilon_2}} \\ &= 0.2057 \end{aligned} \tag{12}$$

the corresponding cumulative kernel E_m is

$$E_{m,1\to2} = \frac{1}{(\text{RME})_{1\to2}} - 1 = 3.86 \tag{13}$$

and the cumulative mass of waste generated in grams is at least

$$\begin{aligned}\bar{w}_{1\to2} &= \text{mass}_{\text{input},1\to2} - \text{mass}_{\text{output},1\to2} \\ &= \frac{x}{\epsilon_3\epsilon_4}\left[\frac{358.7}{\epsilon_1\epsilon_2} + \frac{235.8}{\epsilon_1^*\epsilon_2} + \frac{69}{\epsilon_2} - 315.35\right] \\ &= 1411.43x \end{aligned} \tag{14}$$

When all reaction yields are set to 1 (i.e., 100%) the expression given in eq 12 collapses to the cumulative atom economy between stages 1 and 2

$$\begin{aligned}(\text{AE})_{1\to2} &= \frac{315.35}{358.7 + 235.8 + 69} \\ &= \frac{315.35}{663.5} \\ &= 0.4753 \end{aligned} \tag{15}$$

These formulas are entirely consistent with Eissen's recent algebraic analysis of cumulative atom economies for linear

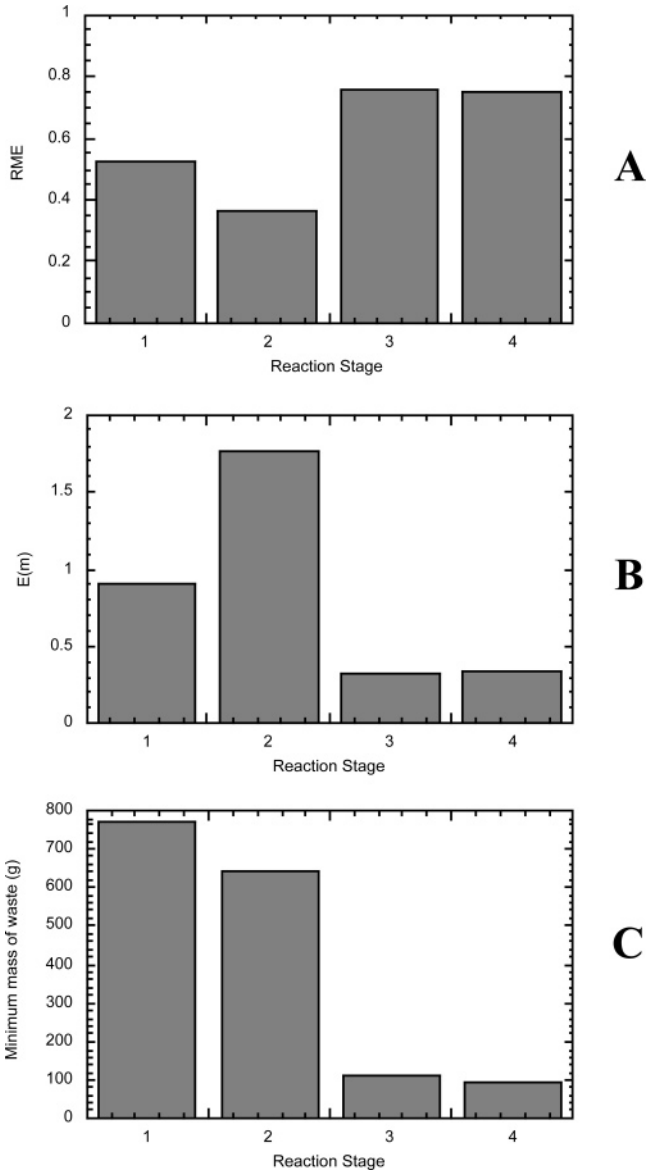


Figure 2. **(a) Histogram showing kernel RME values as function of reaction stage for triclosan synthesis. (b) Histogram showing kernel E_m values as a function of reaction stage for triclosan synthesis. (c) Histogram showing minimum mass of waste generated as function of reaction stage for triclosan synthesis (based on 1 mole scale). Reaction yields used as given in refs 11 and 12. See Scheme 1 for synthetic plan or Figure 1 for synthesis tree.**

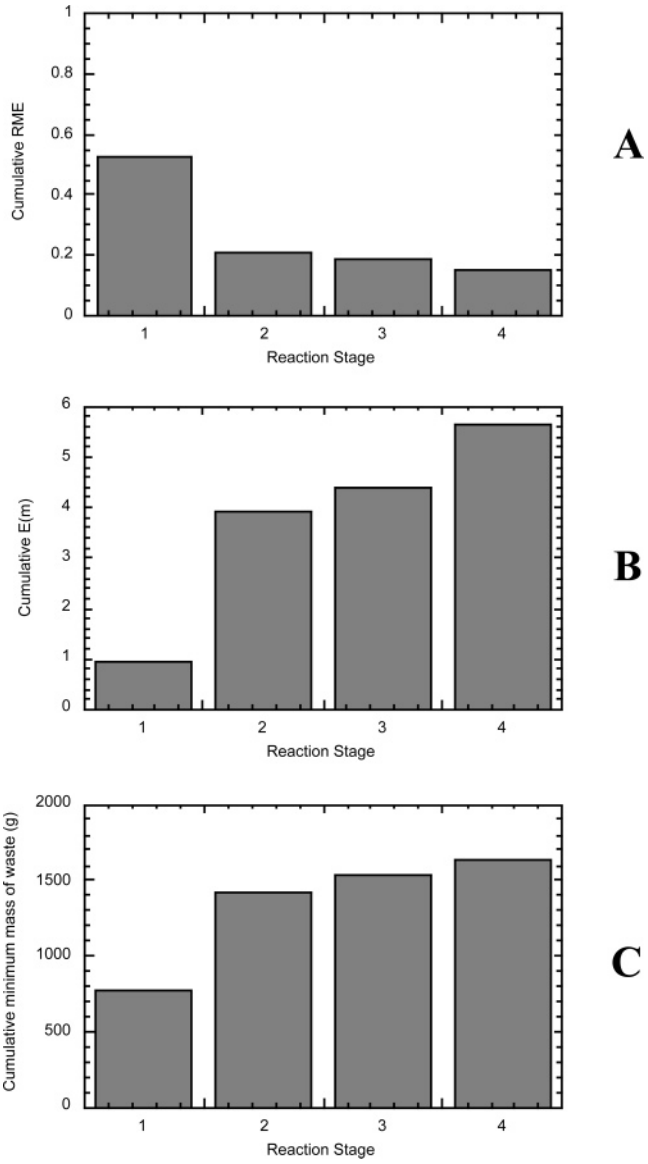


Figure 3. **(a) Graph showing cumulative kernel RME as a function of reaction stage for triclosan plan. (b) Graph showing cumulative kernel E_m as a function of reaction stage for triclosan plan. (c) Graph showing cumulative minimum mass of waste generated as a function of reaction stage (based on 1 mole scale). Reaction yields used as given in refs 11 and 12. See Scheme 1 for synthetic plan or Figure 1 for synthesis tree.**

sequences.[14] The synthesis tree analysis presented here is more general and advantageous in that it provides a pictorial representation of a synthesis plan, it can be easily extended to complex nonlinear plans, and it incorporates reaction yields into the determination of RME which is a truer measure of synthesis efficiency than AE alone.

Figure 2 shows distribution plots of kernel RME and E_m metrics and minimum mass of waste generated as a function of reaction stage. Figure 3 shows plots of the corresponding cumulative kernel RME and E_m metrics and cumulative mass of waste generated as function of reaction stage. From these graphical displays it is possible to spot the waste determining step or stage (i.e., the step or stage producing the highest proportion of all the waste). There are two aspects to this. One can assign the coupling between intermediates I_1 and I_1* in stage 2 as the waste-determining step and stage on the basis of its highest kernel E_m (lowest kernel RME) as shown in Figure 2B. In Figure 3B it is observed that the transition between reaction stages 1 and 2 shows the greatest difference in the cumulative kernel E_m magnitude. A similar observation can be made from Figure 3C with respect to cumulative minimum mass of waste. On the other hand reaction stage 1 is composed of two parallel reactions, acetylation of 1,4-dichlorobenzene and chlorination of phenol, which necessarily take place at the highest scale because they occur at the very beginning of the synthesis plan. As shown in Figure 2C the combined minimum mass

(14) Eissen, M.; Mazur, R.; Quebbemann, H. G.; Pennemann, K. H. *Helv. Chim. Acta* **2004**, *87*, 524.

of waste from stage 1 is highest, and so this stage contributes the greatest proportion of all the waste produced in the entire synthesis. Note that the height of the bar corresponding to stage 1 is the same in Figures 2C and 3C.

3b. Materials Cost Analysis. From the denominator in eq 4a which represents the total mass of input reactants one can also obtain the kernel minimum raw materials cost (RMC) function in $ per mole directly by inspection. Thus,

$$\frac{\mathrm{RMC}}{x} = \frac{32\$_{\mathrm{MeOH}}}{\epsilon_4} + \frac{49\$_{\mathrm{MA}} + 34\$_{\mathrm{H_2O_2}}}{\epsilon_3\epsilon_4} + \frac{69\$_{\mathrm{K_2CO_3}}}{\epsilon_2\epsilon_3\epsilon_4} + \frac{141.8\$_{\mathrm{Cl_2}} + 94\$_{\mathrm{PhOH}}}{\epsilon_1^*\epsilon_2\epsilon_3\epsilon_4} + \frac{133.35\$_{\mathrm{AlCl_3}} + 78.45\$_{\mathrm{AC}} + 146.9\$_{1,4-\mathrm{DCB}}}{\epsilon_1\epsilon_2\epsilon_3\epsilon_4} \quad (16)$$

where the $ symbols represent the unit costs of the input reactants on a per gram basis (MA = maleic anhydride, AC = acetyl chloride, DCB = dichlorobenzene). In general the kernel minimum RMC function is obtainable from the general denominator in eq 5:

$$\frac{\mathrm{RMC}}{x} = \sum_j^M \frac{\$_j r_j}{\epsilon_M \epsilon_{M-1} \cdots \epsilon_j} \quad (17)$$

where r_j is the molecular weight of the jth input reagent. All other material and nonmaterial costs associated with the manufacture of the target product can be directly added to the right-hand side of eq 17 to obtain the true total RMC. From eq 16 it is possible to construct pie graphs as shown in Figure 4A−C, for the triclosan synthesis that depict the kernel fractional costs of each input reagent (cost distribution by input reagent) and the kernel fractional costs of each reaction step or stage (cost distribution by reaction step or stage), respectively. One may be able to easily determine the minimum cost-determining reagent (in this case, chlorine), minimum cost-determining reaction step (in this case, the chlorination of phenol), and the minimum cost-determining reaction stage for the synthesis plan (in this case, stage 1) at the kernel level. In effect the overall kernel RME may be interpreted as the fraction of the kernel RMC directed to making the target product. Therefore, the actual amount of money spent on raw materials that is directed to producing the target product is just the multiplicative product of the overall RME and RMC, and conversely the raw materials cost that is directed to producing waste is $(1 - \mathrm{RME_{overall}}) \times$ RMC. An apparent general realization is that the costs of input reactants in the early stages of a plan must be kept as low as possible since the reaction scales in these stages are highest and attenuation of material resources is inevitable as the plan proceeds in the forward direction toward the target product. Dramatic examples of this will be illustrated in Section 6 where competing synthetic plans to common complex target structures are compared. Throughout this paper unit prices for chemicals based on the largest scale available from the Aldrich 2003−2004 Catalogue were used for illustrative RMC calculations since Aldrich Chemical Co. is widely recognized as the first-choice supplier of fine chemicals to university laboratories in North America. However, it is acknowledged that prices of bulk chemicals depend on many factors including scale. Prices used in this report are meant only to demonstrate how RMC calculations can be done from synthesis trees.

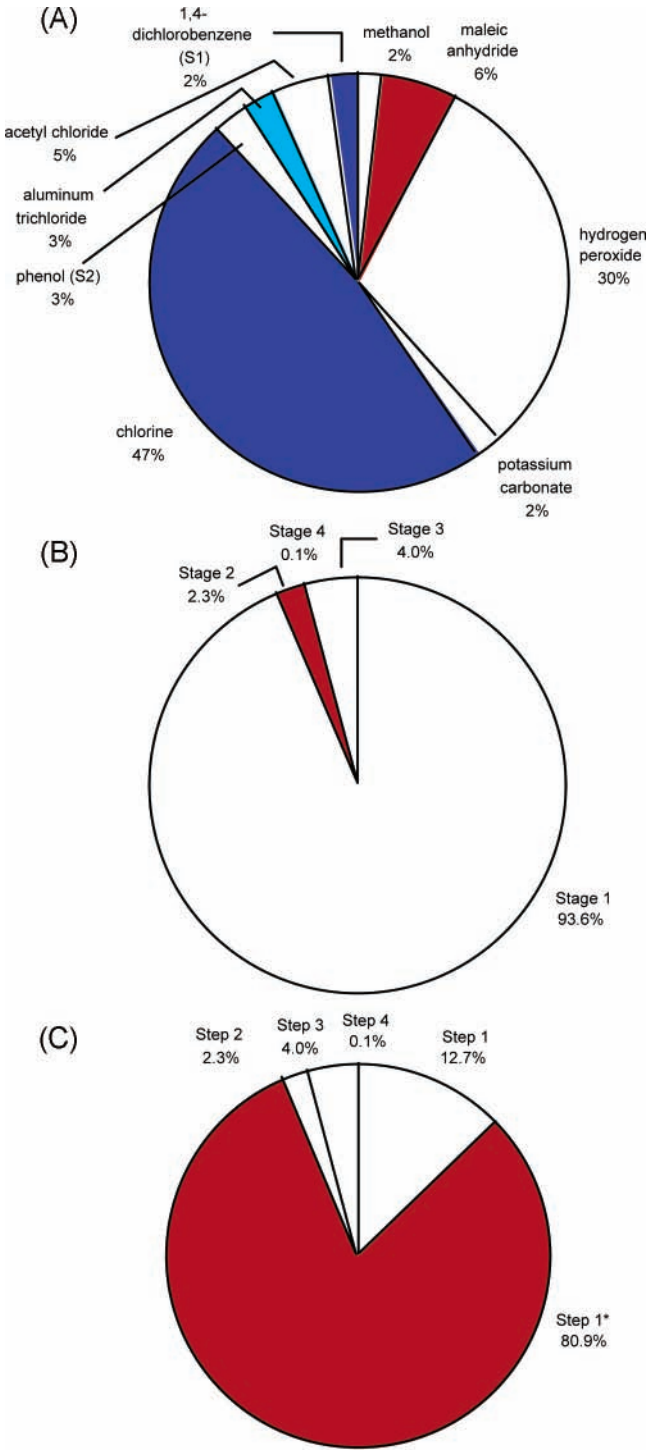


Figure 4. **(A) Pie chart showing the fractional cost distribution by input reagent on a per mole basis. (B) Pie chart showing the fractional minimum RMC distribution by reaction stage for the synthesis of 1 mole triclosan. (C) Pie chart showing the fraction minimum RMC distribution by reaction step for the synthesis of 1 mole triclosan. Unit costs used in the calculations were taken from an Aldrich 2003−2004 Catalogue in Canadian dollars based on prices for largest unit quantities available. See Scheme 1 for synthetic plan or Figure 1 for synthesis tree.**

3c. Energy Usage Analysis. If each step in a synthesis plan has an associate input energy, Ψ, then the total energy input for the entire synthesis is

$$\Psi_{\text{total}} = \sum_{j}^{M} \Psi_j \tag{18}$$

This input energy includes energy consumption during the reaction and all postreaction phases such as heating, cooling, and operating under a pressure exceeding 1 atm. Basically, input energy will be required if any reaction operations are performed above or below standard temperature and pressure conditions of 25 °C and 1 atm. Clark[15] has described experimental determinations of energy inputs for chemical reactions using a domestic electricity meter set up in series with the laboratory power supply. A useful energy metric that is characteristic of how well the input energy is utilized in a synthesis plan is the fraction of the total energy input that is directed to making the target product. For the *j*th reaction in a synthesis plan with input energy Ψ_j the amount of input energy directed toward product is

$$\Psi_{\text{product},j} = (\text{RME})_j \Psi_j \tag{19}$$

and the amount of input energy directed toward producing waste is

$$\Psi_{\text{waste},j} = [1 - (\text{RME})_j]\Psi_j \tag{20}$$

Therefore, the total amount of input energy for a synthesis plan that is directed to the target product is

$$\Psi_{\text{product}} = \sum_{j}^{M} (\text{RME})_j \Psi_j \tag{21}$$

and that directed toward producing waste is

$$\Psi_{\text{waste}} = \sum_{j}^{M} [1 - (\text{RME})_j]\Psi_j \tag{22}$$

The respective product forming and waste forming input energy fractions are given by

$$\Phi_{\text{product}} = \text{FTE} = \frac{\sum_{j}^{M} (\text{RME})_j \Psi_j}{\sum_{j}^{M} \Psi_j} \tag{23a}$$

and

$$\Phi_{\text{waste}} = \frac{\sum_{j}^{M} [1 - (\text{RME})_j]\Psi_j}{\sum_{j}^{M} \Psi_j} \tag{23b}$$

(15) Gronnow, M. J.; White, R. J.; Clark, J. H.; Macquarrie, D. J. *Org. Process Res. Dev.* **2005**, *9*, 516.

Equations 23a and 23b show that these parameters are weighted quantities and that the materials kernel green metric RME parameters for each reaction are the weighting factors.

This provides a strong connection between the materials and energy usage metrics for a synthesis plan as it tells us how well the input energy is partitioned toward making a desired product. In accordance with Clark's recent analysis[15] in which he proposes an energy metric as the reaction input energy per mole of target product formed in units of kWh/mol we may rewrite eq 19 for the *j*th reaction in a synthesis plan as

$$\frac{\Psi_{\text{product},j}}{\dfrac{x}{\epsilon_j \epsilon_{j+1} \cdots \epsilon_M}} = \left[\prod_{j}^{M} \epsilon_j\right]\left(\text{RME}\right)_j \Psi_j / x \tag{24}$$

and eq 23a for all reactions in the entire plan as

$$\frac{\Phi_{\text{product}}}{x} = \frac{\sum_{j}^{M} (\text{RME})_j \Psi_j}{x \sum_{j}^{M} \Psi_j} \tag{25}$$

4. Tree Parameters

In characterizing a synthesis plan, the geometric shape of its corresponding synthesis tree offers possible measures such as degree and rate of convergence toward making the final target product and degree of asymmetry. The first qualitative description of convergence in synthesis plans was given by Velluz and co-workers[16a] and described more quantitatively by others.[9a,16b] These parameters naturally will depend on the number of reactions, *M*, the number of reaction stages, *N*, the number of reactant input structures required to build the product chemical structure, *I*, and the number of intermediate structures along the synthesis path. The convention described earlier in constructing synthesis trees by determining the coordinates of intermediate products and ultimately the final target product using the method of centroids now becomes apparent.

We first note the following fundamental relationships for linear and convergent synthesis plans with the following designations: I = number of reactant inputs, M = number of reactions, N = number of reaction stages, L = number of parallel reactions, G = number of stages with parallel reactions.

For linear synthesis plans,

Height of tree $= I - 1$

Width of tree $= N = M$

Number of intermediate products $= N - 1 = M - 1$

(16) (a) Velluz, L.; Valls, J.; Mathieu, J. *Angew. Chem., Int. Ed. Engl.* **1967**, *6*, 778. (b) Wender, P.; Miller, B. L. In *Organic Synthesis: Theory and Applications*; Hudlicky, T., Ed.; JAI Press: Greenwich: Connecticut, 1993; Vol. 2, p 27 and references therein.

Number of convergent reactions = 0

$G = 0$

$L = 0$

and for convergent synthesis plans:

Height of tree = $I - 1$

Width of tree = $N < M$

Number of intermediate products ≥ N

Number of intermediate products = $M - 1$

$G \leq$ Number of intermediate products $- N + 1$

(equality means that no more than two parallel reactions occur in each stage having parallel reactions, otherwise more than two parallel reactions occur)

Number of points of convergence = Number of branches attached to main reduced tree consisting of vertexes representing all intermediate structures and final product structure along root of tree.

For plans with parallel reactions with (i) at least one stage having more than 2 parallel reactions, then $L > 2G$ and $M > N + G$, and (ii) all reaction stages having parallel reactions with no more than 2 parallel reactions, then $L = 2G$ and $M = N + G$.

Figure 5 shows the same synthesis tree for triclosan as given before in Figure 1 with some modifications made and added geometric parameters superimposed. The input reactants are aligned vertically along the *y*-axis in unit intervals to clearly show the height dimension of the tree. The point designated as P_{mcr} describes the coordinates of the target product if it had been made in a hypothetical single multicomponent reaction (MCR) by using all input reactants. This describes the most convergent synthesis possible given the input structures used to build up the target structure. The more reaction steps and reaction stages are concatenated the more this limit is achievable practically. The point designated as *P* represents the coordinates of the target product for a given synthesis plan. The abscissa is equal to the number of reaction stages and the ordinate by the method of centroids is related to the number of reactant inputs and the shape and connectivity of the synthesis tree. The triangles $R_1R_2P_{mcr}$ and R_1R_2P share a common base which represents the height of the tree. An appropriate measure of degree of convergence may be made with respect to a hypothetical *I*-component MCR (*I* is the total number of input reactants). Hence, we may take the ratio of the angles θ_P and θ_{mcr} subtended at P and P_{mcr}, respectively, to represent this. In terms of the synthesis plan parameters the degree of convergence, δ, is then

$$\delta = \frac{\arctan\left(\frac{I - 1 - \langle y \rangle}{N}\right) + \arctan\left(\frac{\langle y \rangle}{N}\right)}{2 \arctan\left(\frac{I - 1}{2}\right)} \qquad (26)$$

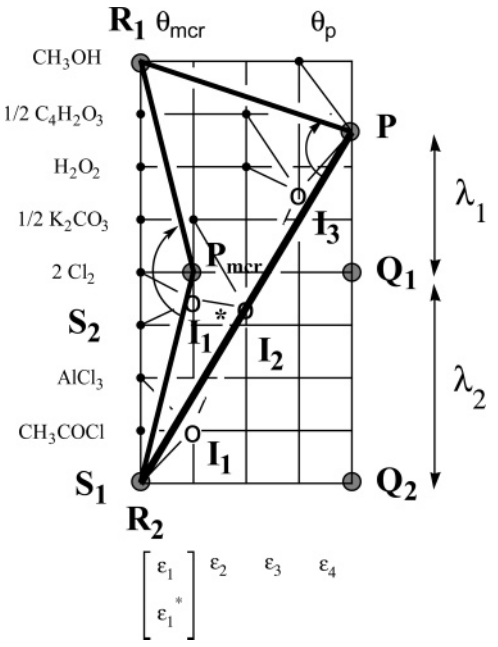


Figure 5. **Synthesis tree for the synthesis of triclosan according to reactions given in Scheme 1. Coordinates of P_{mcr} denote position of final product if it were formed in a hypothetical single step nine-component reaction corresponding to a reaction with maximal convergence using the given input reagents. Angles subtended at points P_{mcr} ($\theta_{mcr} = \angle R_1P_{mer}R_2$) and P ($\theta_p = \angle R_1PR_2$) and line segment lengths as shown are used to calculate the degree of convergence, the relative rate of convergence, and the asymmetry parameter. See text for details.**

where *I* is the number of input reactant structures, *N* is the number of reaction stages, $\langle y \rangle$ is the ordinate of the target point P, and $0 < \delta < 1$. A completely convergent plan has $\delta = 1$ where all input materials are used to reach the target in a single reaction stage as in a multicomponent reaction. A plan with $\delta = 0$ corresponds to one that is the most linear possible, consisting of sequential transformations over one or more reaction stages with no input reagents incorporated into the starting substrate. An example of such a plan represented as a horizontal line with intermediate nodes is a sequence of rearrangements and/or intramolecular cyclizations. For the case of the triclosan synthesis plan in Scheme 1 $\delta = 0.5029$. The relative rate of convergence, ρ_{rel}, for a given synthesis is given by the ratio of slopes of the lines R_2P and R_2P_{mcr}

$$\rho_{rel} = \frac{\rho_{actual}}{\rho_{I-MCR}} = \frac{\frac{\langle y \rangle}{N}}{\frac{I - 1}{2}} = \frac{2\langle y \rangle}{N(I - 1)} \qquad (27)$$

The units of ρ_{actual} (corresponding to the experimental synthesis plan) and ρ_{I-MCR} (corresponding to the hypothetical single step MCR using the *I* reactant inputs of the experimental synthesis plan) are number of reactant inputs per reaction stage. A completely convergent plan, such as a multicomponent single stage synthesis, has $\rho_{rel} = 1$ and a completely linear plan as that described above has $\rho_{rel} = 0$. For the case of the triclosan synthesis plan in Scheme 1 $\rho = 0.4238$. An asymmetry parameter, β, ranging between

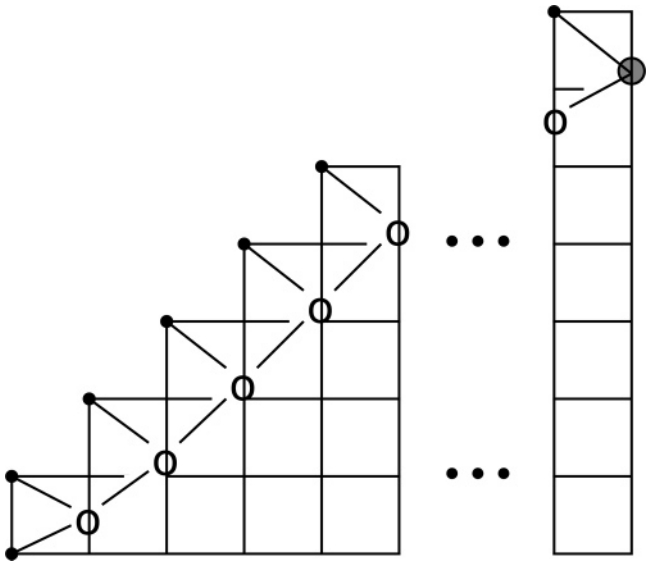


Figure 6. **Simple linear synthesis plan with *I* input reactants and *N* reaction stages involving sequential addition of an input reactant at each reaction stage so that $I = N + 1$.**

0 and 1 may be defined as

$$\beta = \frac{\bar{P}\bar{Q}_1}{\bar{Q}_1\bar{Q}_2} = \frac{\lambda_2}{\lambda_1} = \frac{\langle y\rangle - \frac{I-1}{2}}{\frac{I-1}{2}} \tag{28}$$

where $\beta = 0$, ($\langle y\rangle = (I - 1)/2$) represents a completely symmetric plan and $\beta = 1$,($\langle y\rangle = I - 1$) represents a completely asymmetric plan. This parameter essentially tells us the degree of skewness of the triangle R_1R_2P compared to $R_1R_2Q_1$ where both involve the same number of input reactants and number of reaction stages. Note that the hypothetical one-step MCR plan represents the most convergent plan possible for the *I* reactant inputs involved and is completely symmetric as expected. For the case of the triclosan synthesis plan in Scheme 1 $\beta = 0.6953$.

For a simple linear sequence beginning with a bimolecular reaction and thereafter involving one input structure added sequentially in turn to each intermediate product as depicted in Figure 6 the above three parameters as functions of the number of reaction stages are given by

$$\delta = \frac{\arctan\left(\frac{N - 1 + 2^{-N}}{N}\right) + \arctan\left(\frac{1 - 2^{-N}}{N}\right)}{2\arctan\left(\frac{N}{2}\right)} \tag{29}$$

$$\rho_{rel} = \frac{2(N-1) + 2^{1-N}}{N^2} \tag{30}$$

$$\beta = \frac{2(N-1) + 2^{1-N}}{N} - 1 \tag{31}$$

As the number of reaction stages becomes very large $\delta \rightarrow 0.25$ (minimum possible degree of convergence), $\rho_{rel} \rightarrow 0$ (minimum possible relative rate of convergence), and $\beta \rightarrow 1$ (complete asymmetry). If two parallel equal length sequences as shown in Figure 6 converge to a target product, then it can be shown that

$$\delta = \frac{\arctan\left(\frac{3N}{2} - \frac{1}{2} + 2^{-N}}{N+1}\right) + \arctan\left(\frac{\frac{3}{2} + \frac{N}{2} - 2^{-N}}{N+1}\right)}{2\arctan\left(N + \frac{1}{2}\right)} \tag{32}$$

$$\rho_{rel} = \frac{2\left(\frac{3N}{2} - \frac{1}{2} + 2^{-N}\right)}{(N+1)(2N+1)} \tag{33}$$

$$\beta = \frac{3N - 1 + 2^{1-N}}{2N+1} - 1 \tag{34}$$

As the number of reaction stages becomes very large $\delta \rightarrow 0.46$ (minimum degree of convergence for two parallel reaction sequences), $\rho_{rel} \rightarrow 0$ (minimum possible relative rate of convergence for two parallel reaction sequences), and $\beta \rightarrow 0.5$. These results can be generalized to ℓ parallel equal length converging sequences:

$$\delta = \frac{\arctan\left(\frac{(\ell-1)N}{2} + \frac{\ell-3}{2} + 2^{-N}}{N+1}\right) + \arctan\left(\frac{\frac{\ell+1}{2} + \frac{(\ell-1)N}{2} - 2^{-N}}{N+1}\right)}{2\arctan\left(\frac{\ell N + \ell - 1}{2}\right)} \tag{35}$$

$$\rho_{rel} = \left(\frac{\frac{(\ell+1)N}{2} + \frac{\ell-3}{2} + 2^{-N}}{N+1}\right)\left(\frac{2}{\ell N + \ell - 1}\right) \tag{36}$$

$$\beta = \frac{2\left(\frac{(\ell+1)N}{2} + \frac{\ell-3}{2} + 2^{-N}\right)}{\ell N + \ell - 1} - 1 \tag{37}$$

The limiting values of these parameters for large *N* are $\delta \rightarrow 1/2\pi[\arctan((\ell+1)/2) + \arctan((\ell-1)/2)]$, $\rho_{rel} \rightarrow 0$, and $\beta \rightarrow 1/\ell$. Figure 7A depicts graphically the relationship given by eq 35 which illustrates the points that for any number of parallel converging sequences the degree of convergence diminishes with increasing number of reaction stages, and that for long parallel sequences that converge to a target structure late in a plan the limiting degree of convergence increases. Figure 7B illustrates the second point more directly.

The three parameters δ, ρ_{rel}, and β may be used to rank the performances of different synthetic plans to a common target structure rather succinctly. As a test case a number of combinations of synthesis plans involving four reactant input structures were evaluated. These tree shapes were first constructed and studied by Cayley.[6b] The results are summarized in Table 2. For completeness well-known graph theoretic metrics such as branching index,[5] Wiener index,[5] Hendrickson's convergence parameter,[8] and Randic connectivity index for hydrocarbons[17] were also evaluated for the tree diagrams. These indexes depend only on the relative *connectivity* of the tree nodes. They do not take into account the relative *positions* of the nodes, i.e., precise coordinates

(17) (a) Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609. (b) Randic, M. *J. Mol. Graphics Model.* **2001**, *20*, 19.

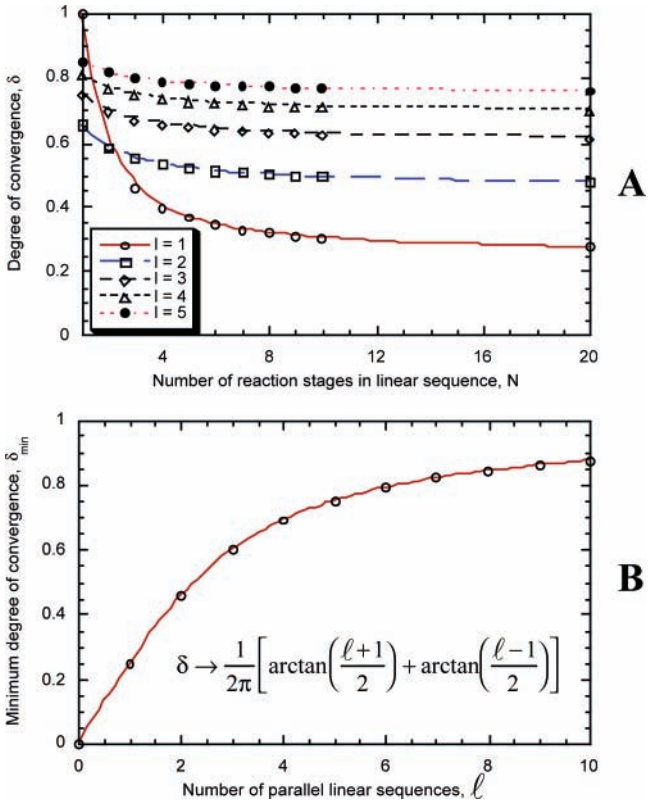


Figure 7. **(A) Relationship between degree of convergence and number of stages in linear sequence, *N*, as a function of number of parallel linear sequences of equal length according to eq 35. (B) Relationship between minimum degree of convergence and number of parallel linear sequences of equal length.**

and distances between nodes. The branching index, BI, is given by

$$\mathrm{BI} = \frac{1}{2}\sum_{i=1}^{v} d_i(d_i - 1) \tag{38}$$

where d_i is the degree or valency of vertex i and v is the total number of vertexes in the graph. The Wiener index, W, or path number, is half the sum of the off-diagonal elements of the distance matrix,[5] **D**, for the graph given by

$$W = \frac{1}{2}\sum_{ij} D_{ij}, \quad i \neq j \tag{39}$$

where D_{ij} is the length of the shortest path (minimum number of edges) between vertex i and j. The Hendrickson convergence parameter is the sum of the number of edges (paths) connecting each substrate node in the synthesis tree to the terminal node represented by the target product. Overcounting of paths is permissible. The Randic branching index, ^1X, is given by

$$^1\mathrm{X} = \sum_{ij} \frac{1}{\sqrt{d_i d_j}}, \quad i \neq j \tag{40}$$

where d_i is the degree or valency of vertex i, d_j is the degree or valency of vertex j, and the sum is taken over edges ij bounded by adjacent vertexes i and j.

It is observed that no sensible correlations can be made between these traditional parameters and the tree shapes since different tree shapes have the same index values. For example, trees II, III, V, VI, and VIII have the same branching index of 7; trees II and V have the same Wiener index of 28; trees III and VII have the same Hendrickson parameter of 8; and trees VI and VIII have the same Randic index of 3.181. What is needed is a parameter or set of parameters that uniquely describe each tree shape so that unambiguous ranking is possible. The δ, ρ, and β parameters presented here satisfy this criterion and hence better capture the behaviours of the synthesis trees. On going from left to right, the asymmetry parameter increases as the degree of convergence decreases as expected. Moreover, the δ convergence parameter allows for ranking of the trees that is intuitively consistent with their shapes. For example, the Hendrickson convergence parameter for trees III and VIII only differ by one unit, yet it is obvious that the former plan is far more symmetric and convergent than the latter. These two tree shapes have often been used to juxtapose a convergent and a linear plan.[9] The last row of Table 2 yields interesting comparisons of the kernel RMEs as functions of the reaction yield as shown in Figure 8 under the simplifying assumption that all reactant input molecular weights are identical and equal to r and all reaction yields are identical and equal to ϵ. It is observed that the more convergent the plan is, the more linear is the dependence of the kernel RME on the reaction yield.

The key trends to increase convergence are to decrease the number of reaction stages, to increase the number of parallel reactions per stage, and to increase the number of reactant input components per reaction. The last strategy of increasing the frequency of multicomponent reactions[18] is particularly effective if such reactions gravitate toward the end stages in a synthesis plan.

5. Synthesis Planning and Management

When planning a synthesis it is often the case that several plans are considered and evaluated on the basis of various criteria in order to determine the most optimal and cost-effective approach. Evaluations based on the synthesis tree method described here can be quite helpful in this decision making. The synthesis trees not only facilitate computation of key convergence properties and green metrics relating to

(18) (a) Posner, G. H. *Chem. Rev.* **1986**, *86*, 831. (b) Ugi, I.; Dömling, A.; Hörl, W. *Endeavour* (New Series) **1994**, *18*(3), 115. (c) Lombardo, M.; Trombini, C. *Seminars in Organic Synthesis;* Società Chimica Italiana: Milan, 1998; p 7. (d) Ugi, I. *Proc. Estonian Acad. Sci. Chem.* **1998**, *47*, 107. (e) Ugi, I.; Dömling, A.; Werner, B. *J. Heterocycl. Chem.* **2000**, *37*, 647. (f) Kappe, C. O. *Acc. Chem. Res.* **2000**, *33*, 879. (g) Dömling, A. *Curr. Opin. Chem. Biol.* **2000**, *4*, 318. (h) Ugi, I. *Pure Appl. Chem.* **2001**, *73*, 187. (i) Dömling, A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 306. (j) Weber, L. *Drug Discovery Today* **2002**, *7*, 143. (k) Weber, L. *Curr. Med. Chem.* **2002**, *9*, 1241. (l) Orru, R. V. A.; de Greef, M. *Synthesis* **2003**, 1471. (m) von Wangelin, A.; Neumann, H.; Gördes, D.; Klaus, S.; Strübing, D.; Beller, M. *Chem. Eur. J.* **2003**, *9*, 4286. (n) Ugi, I.; Werner, B.; Dömling, A. *Molecules* **2003**, *8*, 53. (o) Simon, C.; Constantieux, T.; Rodriguez, J. *Eur. J. Org. Chem.* **2004**, 4957. (p) Ramachary, D. B.; Barbas, C. F., III. *Chem. Eur. J.* **2004**, *10*, 5323. (q) Ramon, D. J.; Yus, M. *Angew. Chem., Int. Ed.* **2005**, *44*, 1602. (r) Weber, L.; Katrin, I.; Almstetter, M. *Synlett* **1999**, 366. (s) Armstrong, R. W.; Combs, A. P.; Tempest, P. A.; Brown, S. D.; Keating, T. A. *Acc. Chem. Res.* **1996**, *29*, 123. (t) Dömling, A. *Chem. Rev.* **2006**, *106*, 17.

Table 2. **Summary of parameters for all combinations of synthesis trees involving four input reagents**

	Type I	Type II	Type III	Type IV	Type V	Type VI	Type VII	Type VIII	Type IX
Branching index, BI[a]	6	7	7	6	7	7	8	7	8
Wiener index, W[b]	16	28	48	29	28	46	44	46	68
Hendrickson parameter, H[c]	4	5	8	6	7	7	8	9	10
Randic connectivity index, ^1X[d]	2	2.561	3.126	4.952	2.561	3.181	2.621	3.181	3.719
Number of reactions, M	1	2	3	2	2	3	3	3	4
Number of reaction stages, N	1	2	2	2	2	3	3	3	4
Number of parallel reactions, L	0	0	2	0	0	0	0	0	0
Co-ordinates of final product, P	$\left(1,\frac{3}{2}\right)$	$\left(2,\frac{3}{2}\right)$	$\left(2,\frac{3}{2}\right)$	$\left(2,\frac{15}{8}\right)$	(2,2)	$\left(3,\frac{15}{8}\right)$	(3,2)	$\left(3,\frac{17}{8}\right)$	$\left(4,\frac{17}{8}\right)$
θ_p (degrees)	112.620	73.740	73.740	72.510	71.565	52.561	52.125	51.571	40.319
Degree of convergence with respect to 4-component single-step reaction, δ	1	0.6548	0.6548	0.6438	0.6355	0.4667	0.4628	0.4579	0.3580
Rate of convergence, ρ_{actual}	$\frac{3}{2}=1.5$	$\frac{3}{4}=0.75$	$\frac{3}{4}=0.75$	$\frac{15}{16}=0.938$	1	$\frac{15}{24}=0.625$	$\frac{2}{3}=0.667$	$\frac{17}{24}=0.708$	$\frac{17}{32}=0.531$
ρ_{4-mcr}	$\frac{3}{2}=1.5$	$\frac{3}{2}=1.5$	$\frac{3}{2}=1.5$	$\frac{3}{2}=1.5$	$\frac{3}{2}=1.5$	$\frac{3}{2}=1.5$	$\frac{3}{2}=1.5$	$\frac{3}{2}=1.5$	$\frac{3}{2}=1.5$
Relative rate of convergence, ρ_{rel}	1	$\frac{1}{2}=0.5$	$\frac{1}{2}=0.5$	$\frac{5}{8}=0.625$	$\frac{2}{3}=0.667$	$\frac{5}{12}=0.417$	$\frac{4}{9}=0.444$	$\frac{17}{36}=0.472$	$\frac{17}{48}=0.354$
Asymmetry parameter, β	0	0	0	$\frac{1}{4}=0.25$	$\frac{1}{3}=0.333$	$\frac{1}{4}=0.25$	$\frac{1}{3}=0.333$	$\frac{5}{12}=0.417$	$\frac{5}{12}=0.417$
Kernel RME[e]	$\frac{p}{r}\left(\frac{\varepsilon}{4}\right)$	$\frac{p}{r}\left(\frac{\varepsilon^2}{3\varepsilon+1}\right)$	$\frac{p}{r}\left(\frac{\varepsilon^2}{4}\right)$	$\frac{p}{r}\left(\frac{\varepsilon^2}{2(\varepsilon+1)}\right)$	$\frac{p}{r}\left(\frac{\varepsilon^2}{\varepsilon+3}\right)$	$\frac{p}{r}\left(\frac{\varepsilon^3}{2\varepsilon^2+\varepsilon+1}\right)$	$\frac{p}{r}\left(\frac{\varepsilon^3}{\varepsilon^2+2\varepsilon+1}\right)$	$\frac{p}{r}\left(\frac{\varepsilon^3}{2+\varepsilon+\varepsilon^2}\right)$	$\frac{p}{r}\left(\frac{\varepsilon^4}{\varepsilon^3+\varepsilon^2+\varepsilon+1}\right)$

[a] See refs 5a,c and eq 38. [b] See refs 5a,c and eq 39. [c] See ref 8. [d] See ref 17 and eq 40. [e] Assuming all input MW are identical and equal to *r*, all reaction yields are identical and equal to ϵ, and product MW is equal to *p*.

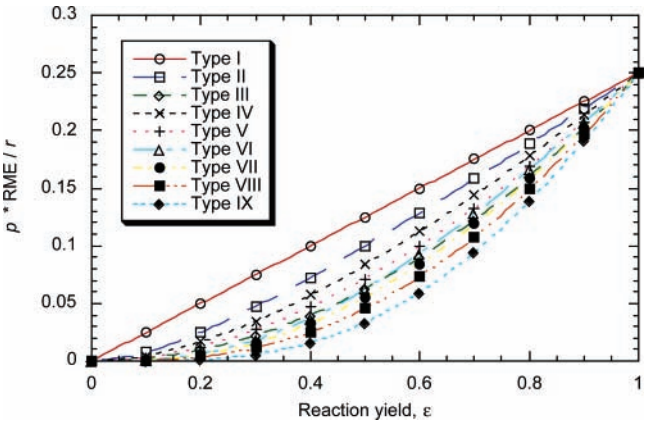


Figure 8. **Relationship showing dependence of reaction yield factor in kernel RME expression as a function of reaction yield for the synthesis plans shown in Table 2.**

materials and energy usage and cost, but also they provide insights into the management of a plan as it is to be implemented. One of the key features is that a synthesis tree reveals the order in which chemical reactions should be carried out. For example, in the triclosan plan in Figure 1 the first convergent stage suggests that in order to save time the Friedel−Crafts and chlorination reactions should be carried out simultaneously. Such prudent planning not only allows for a faster production of the target compound but also potentially saves on labour costs as well. It is possible, therefore, to write a general expression for the total optimum time to complete a synthesis plan as

$$T_{\text{total}} = \sum_{k}^{N} \max\{t_k\} \tag{41}$$

where for reaction stage *k* the reaction time taken is the maximum of the reaction times for reactions taking place simultaneously in that stage. It is obvious that the reaction taking the longest time in a reaction stage involving parallel reactions will govern the length of time taken during that stage. Each of the reaction times in eq 41 takes into account the actual reaction time itself including workup and purifica-

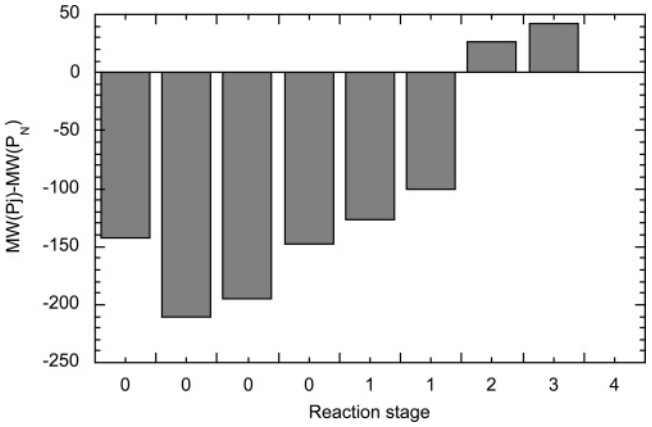


Figure 9. **Triclosan synthesis molecular weight profile based in Scheme 1 and Figure 1 for molecular weights of substrate starting materials and intermediate products relative to molecular weight of triclosan. MW(P_N) = 289.35 g/mol; μ_1 = −171.03 g/mol per reaction stage. The bars for the zeroth stage represent 1,4-dichlorobenzene, acetyl chloride, phenol, and 2 equiv of chlorine, respectively; for the first stage, intermediates I_1* and I_1, respectively; for the second stage, intermediate I_2; and for the third stage, intermediate I_3.**

tion procedures. The specific expression for the given triclosan plan is

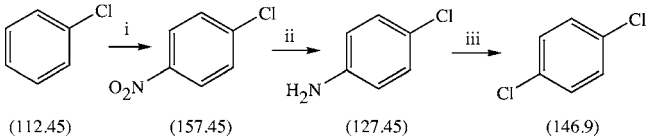
$$T_{\text{total}} = \max\{t_1, t_{1*}\} + t_2 + t_3 + t_4 \quad (42)$$

A second evaluation that can be made, by analogy with plots that trace the building up of molecular complexity of intermediate products as the synthesis proceeds,[9,19] is to plot the molecular weight of starting substrates and intermediate products as a function of reaction stage. Figure 9 shows such a plot for the triclosan synthesis discussed. Reaction stages with multiple points indicate that these stages involve parallel reactions, and thus the synthesis plan has points of convergence. From this plot we may derive using the method of moments a parameter that describes the net building up of structure from the set of initial input and intermediate structures toward the final target product. The molecular weight first moment per reaction stage about the target product molecular weight, MW(P_N), in units of grams per mole per reaction stage is given by

$$\mu_1 = \left(\frac{1}{N+1}\right)\Big[\sum_{\text{stages}}[\text{MW(intermediates)}] + \sum_{\text{branches}}[\text{MW(starting materials)}] + \text{MW(target product)} - [\text{total number of intermediates and starting materials}] \cdot \text{MW(target product)}\Big] \quad (43)$$

where N is the number of reaction stages in the synthesis plan and the starting materials correspond to those inputs at the beginning stage of each branch. The zeroth stage representing the starting substrates for the longest branch or root of the synthesis tree is accounted for by the extra stage in the denominator. If a reaction stage has parallel reactions and therefore consists of more than one intermediate product being formed in that stage, then each of their respective molecular weights are included in the first summed term. The second summed term accounts for molecular weights of input starting materials at the beginning of each branch provided they contribute to the structure of the immediately resulting product. A positive value for μ_1 indicates an overall net loss in MW per reaction stage (net degradation), and a negative value indicates an overall net gain in MW per reaction stage (net building up). The larger the magnitude of the first moment the greater is the effect of degradation or building up. Applying these ideas to the synthesis tree in Figure 1, the contributing molecular weights of intermediates are 188.9, 162.9, 315.35, and 331.35 g/mol for I_1, I_1*, I_2, and I_3, respectively; and those of input starting materials are 146.9 and 78.45 g/mol for 1,4-dichlorobenzene and acetyl chloride, respectively, for the beginning of the main branch, and 94 and 141.8 g/mol, respectively, for phenol and 2 equiv of chlorine for the beginning of the parallel branch. Note that aluminum trichloride is not included as a starting input material in the main branch since it does not contribute to the structure of the immediate product I_1. Substitution of the appropriate molecular weights of the above 9 species including that of triclosan and $N = 4$ into eq 43 yields a molecular weight first moment per reaction stage of $\mu_1 = -855.15/5 = -171.03$ g/mol indicating that the triclosan synthesis plan in Figure 1 involves an overall net gain of about 171 g/mol per reaction stage. Good synthesis plans are characterized by fewer reaction stages, the frequent occurrence of convergent reaction stages (i.e., parallel reactions), and large negative molecular weight first moments per reaction stage.

Scheme 2[a]



[a] Reaction conditions: (i) HNO_3, HZSM-5 zeolite (96%); (ii) 3 H_2, Raney Ni, EtOH (97%); (iii) $NaNO_2$, 2 HCl, CH_3CN, then heat (96%). See ref 20. Molecular weights in g/mol are given in parentheses.

The ready availability of input reagents plays a major role in synthesis planning and consequently has profound effects on the shapes of synthesis trees and their associated parameters. For example, if instead 2,4-dichlorophenol (I_1* in Figure 1 and Scheme 1) is purchased and 1,4-dichlorobenzene (S_1) is made by the sequence shown in Scheme 2, then the resulting tree is as shown in Figure 10. Table 3 summarizes and compares key kernel reaction metrics and tree parameters for the two synthesis plans in Figures 1 and 10.

On comparing the two synthetic plans it is observed that the second is linear with a greater number of reaction steps, stages, and inputs, a slightly lower overall kernel RME, a lower degree of convergence and relative rate of convergence, and a higher degree of asymmetry. Consistent with the fact that the second plan proceeds over nearly twice the number of reaction stages as the first but with nearly equal overall summed differences between molecular weights of target product and intermediates and starting inputs, its MW first moment is more positive, indicating that there is a lower

(19) Whitlock, H. W. *J. Org. Chem.* **1998**, *63*, 7982.

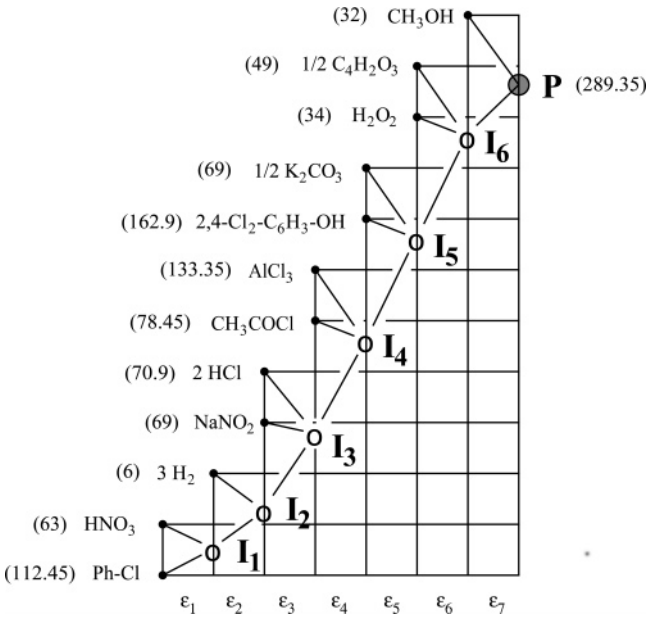


Figure 10. **Synthesis tree for the synthesis of triclosan using 2,4-dichlorophenol and chlorobenzene as a starting materials. Synthesis parameters: 12 inputs, 6 intermediates, 7 reaction stages, 7 reactions, 0 parallel reactions. Synthesis type: linear. Molecular weights in grams per mole for input reactant and final product output nodes are given in parentheses. Reaction yields: $\epsilon_1 = 0.96$, $\epsilon_2 = 0.97$, $\epsilon_3 = 0.96$, $\epsilon_4 = 0.943$, $\epsilon_5 = 0.483$, $\epsilon_6 = 0.913$, and $\epsilon_7 = 0.945$. Intermediates: I_1 is 1-chloro-4-nitrobenzene, I_2 is 4-chloroaniline, I_3 is 1,4-dichlorobenzene, and I_4, I_5, and I_6 correspond to structures I_1, I_2, and I_3 respectively from Figure 1.**

degree of building up to the target structure. In terms of raw materials cost, the second plan turns out to be 3 times cheaper than the first route on a per gram basis as shown by the respective RMC values. Moreover, the RMC for 1,4-dichlorobenzene determined for Scheme 2 is $0.089 CAD per gram using Aldrich 2003−2004 prices, which can be compared directly to what Aldrich sells this product for at $ 0.021 CAD per gram. This observation may simply mean that this product is produced more cheaply by another route, such as direct chlorination of benzene and then separating the various chorinated products by distillation, or that the unit prices for materials used in Scheme 2 may be found to be cheaper from other suppliers.[21] A reviewer has noted that it may reflect economies of scale, cheaper labour, or overstock and lack of demand. Nevertheless, this example demonstrates that a shorter synthesis does not necessarily always translate into a cheaper synthesis and that market unit prices of chemicals are determined by a complex of economic factors beyond core raw materials costs of their progenitors such as demand, taxes, and costs due to ancillary materials and solvents used in the process during workup and purification, equipment maintenance and depreciation, safety and regulatory compliance, energy consumption, waste treatment and disposal, and labour. Anderson has discussed these other factors in the context of the pharmaceutical industry beyond the material throughput analysis described here.[22] Laird has discussed cost estimates for new molecules using an ad hoc equation in an editorial.[23] More broadly, it points to the fact that no preset assumptions should be made from the outset and that a full metrics analysis must be done for each competing plan before drawing any conclusions on the merits of one over another.

Table 3. **Summary of reaction metrics and synthesis tree parameters for triclosan synthesis plans shown in Figures 1 and 10**

	Figure 1 synthesis tree	Figure 10 synthesis tree
Kernel Reaction Metrics		
AE	0.3717	0.3288
E_{mw}	1.69	2.04
RME	0.152	0.1370
E_m	5.59	6.30
$\epsilon_{pseudo-overall}$ ($\epsilon_{overall}$)	0.408	0.417 (0.351)
number of reaction inputs, I	9	12
number of reaction steps, M	5	7
number of reaction stages, N	4	7
μ_1 (g/mol per reaction stage)	−171.03	−108.99
RMC[a] ($/mol)	333.95 ($1.15/g)	111.28 ($0.38/g)
Tree Parameters		
P coordinate	(4,217/32)	(7,20301/2048)
θ_p (deg)	76.411	63.601
θ_{mcr} (deg)	151.928	159.390
degree of convergence, δ	0.503	0.399
ρ_{actual}	1.695	1.416
ρ_{I-mcr}	4	5.5
relative rate of convergence, ρ_{rel}	0.424	0.258
asymmetry, β	0.695	0.802

[a] Based on unit costs ($/g) taken from an Aldrich 2003−2004 Catalogue in Canadian dollars using prices for the largest unit listed in the catalogue: methanol, 0.0108; maleic anhydride, 0.1047; hydrogen peroxide, 0.1888 (based on 30 wt % solution); potassium carbonate, 0.0462; chlorine, 0.6110; phenol, 0.047; aluminum trichloride, 0.0639; acetyl chloride, 0.0643; 1,4-dichlorobenzene, 0.0212; 2,4-dichlorophenol, 0.0563; hydrochloric acid, 0.0381 (based on 37 wt % solution); sodium nitrite, 0.0662; hydrogen, 0.2875 (assuming pressure of lecture bottle is 1800 psi); nitric acid, 0.0292 (based on 70 wt % solution); and chlorobenzene, 0.0129.

6. Example Synthesis Plans

To test further the robustness of these new methodologies in a variety of scenarios, literature synthetic plans for quinine,[24] sildenafil,[25] absinthin,[26] papaverine,[27] and bupl-

(20) (a) Jayasuriya, K.; Damavarapu, R. U.S. Patent 5,946,638, 1998. (b) Winans, C. F. *J. Am. Chem. Soc.* **1939**, *61*, 3564. (c) Brackman W.; Smit, P. J. *Recl. Trav. Chim. Pays-Bas* **1966**, *85*, 857.

(21) Using a reaction yield of 15% for production of 1,4-dichlorobenzene by direct chlorination of benzene as given in Faith, W. L.; Keyes, D. B.; Clark, R. L. *Industrial Chemicals*, 3rd ed., Wiley: New York, 1966; p 261, and again using Aldrich prices for benzene ($0.0196 per gram) and chlorine ($0.6110 per gram), the resulting RMC for this product now becomes $2.125 CAD/g.

(22) Anderson, N. G. *Org. Process Res. Dev.* **2004**, *8*, 260.

(23) Laird, T. *Org. Process Res. Dev.* **2005**, *9*, 125.

(24) (a) Woodward, R. B.; Doering, W. E. *J. Am. Chem. Soc.* **1945**, *67*, 860. (b) Rabe, P. *Chem. Ber.* **1911**, *44*, 2088. (c) Rabe, P.; W. Huntenburg, W.; Schultze, A.; Volger, G. *Chem. Ber.* **1931**, *64*, 2487. (d) G. Stork, G.; Niu, D.; Fujimoto, A.; Koft, E. R.; Balkovec, J. M.; Tata, J. R.; Dake, G. R. *J. Am. Chem. Soc.* **2001**, *123*, 3239. (e) Raheem, I. T.; Goodman, S. N.; Jacobsen, E. N. *J. Am. Chem. Soc.***2004**, *126*, 706. (f) Igarashi, J.; Katsukawa, M.; Wang, Y.-G.; Acharya, H. P.; Kobayashi, Y. *Tetrahedron Lett.* **2004**, *45*, 3783. (g) See Kaufman, T. S.; Ruveda, E. A. *Angew. Chem., Int. Ed.* **2005**, *44*, 854 for a general review.

(25) (a) Dunn, P. J.; Galvin, S.; Hettenbach, K. *Green Chem.* **2004**, *6*, 43. (b) Dale, D. J.; Dunn, P. J.; Golightly, C.; Hughes, M. L.; Levett, P. C.; Pearce, A. K.; Searle, P. M.; Ward, G.; Wood, A. S. *Org. Process Res. Dev.* **2000**, *4*, 17.

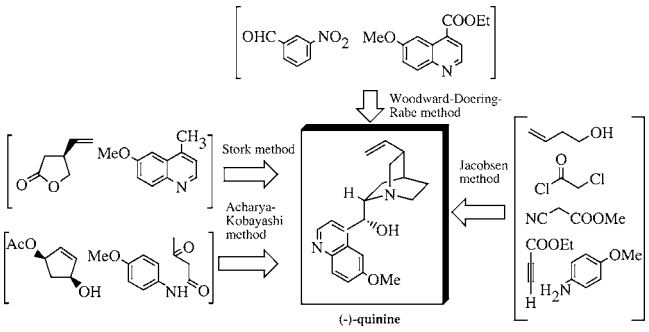
(26) Zhang, W.; Luo, S.; Fang, F.; Chen, Q.; Hu, H.; Jia, X.; Zhai, H. *J. Am. Chem. Soc.* **2005**, *127*, 18.

Table 4. **Summary of reaction metrics and synthesis tree parameters for quinine synthesis plans**

	Woodward−Doering−Rabe (linear)	Stork (convergent)	Jacobsen (convergent)	Acharya−Kobayashi (convergent)
	Kernel Reaction Metrics			
AE	0.0844	**0.0864**	0.0669	0.0529
E_{mw}	10.84	10.57	13.96	17.89
RME	6.34×10^{-5}	$\mathbf{1.21 \times 10^{-2}}$	8.42×10^{-3}	6.00×10^{-3}
E_m	1.58×10^{4}	8.18×10^{1}	1.18×10^{2}	1.66×10^{2}
$\epsilon_{pseudo-overall}$ ($\epsilon_{overall}$)	7.50×10^{-4} (2.94×10^{-4})	$\mathbf{1.40 \times 10^{-1}}$	1.26×10^{-1}	1.13×10^{-1}
number of reaction inputs, *I*	36	**33**	38	51
number of reaction steps, *M*	19	**18**	21	28
Number of reaction stages, *N*	19	17	**16**	21
μ_1 (g per mole per reaction stage)	−60.51	+47.11	**−114.71**	−57.14
RMC[a] ($ CAD/g)	3509.80	1819.69	**684.82**	5.64×10^{7}
	Tree Parameters			
P coordinate	(19,33.669)	(17,30.338)	(16,35.283)	(21,48.561)
θ_p (deg)	64.572	66.320	71.732	70.535
θ_{mer} (deg)	173.459	172.847	173.811	175.419
degree of convergence, δ	0.372	0.384	**0.413**	0.402
ρ_{actual}	1.772	1.785	2.205	2.312
ρ_{I-mcr}	17.5	16	18.5	25
relative rate of convergence, ρ_{rel}	0.101	0.112	**0.119**	0.0925
asymmetry, β	0.924	**0.896**	0.907	0.942

[a] Based on unit costs ($ per gram) taken from an Aldrich 2003−2004 Catalogue in Canadian dollars using prices for the largest unit listed in the catalogue.

Scheme 3



eurynol[28] were analyzed accordingly. Full chemical schemes with balanced chemical equations and synthesis trees are given in the Supporting Information. The plans range in complexity from simple linear to complex mixed linear and convergent. Tables 4−7 summarize key kernel reaction metrics and tree parameters for each of the total syntheses examined.

(i) Quinine. The synthesis of quinine is considered the quintessential classic in the art of total synthesis.[24g] Scheme 3 summarizes the starting materials for four routes beginning with the historical Woodward−Doering−Rabe plan to the racemic mixture and including three recent modern stereoselective routes to the (−) isomer. As can be seen from Table 4 the most material-efficient synthesis is that of Stork with an overall kernel RME of 1.2% and a pseudo-overall yield of 14% in 17 reaction stages. This is mainly because it has the fewest reaction inputs, reaction stages, and reaction steps and has the highest average reaction yield per step of 86%. The kernel overall RME values for the Woodward−Doering−Rabe and Stork plans of 0.0063% and 1.2% parallel previous determinations[3a] of overall RME values of 0.0032% and 0.16%, respectively, when stoichiometric factors are taken into account for each step. However, in terms of raw material cost the Jacobsen plan wins out at $685 CAD/g because the five starting materials are the cheapest of all the routes. It also happens to be the most convergent plan with a 41% degree of convergence though all of the routes are for the most part dominated by long linear sequences as shown by their similar low relative rate of convergence of about 0.1 and high asymmetry. There are three points of convergence in the Jacobsen plan compared with one each for the Stork and Acharya−Kobayashi plans and none for the Woodward−Doering−Rabe plan. The Acharya−Kobayashi plan has by far the longest route (21 reaction stages) and begins with a very expensive reagent, *cis*-4-acetoxy-2-cyclo-penten-1-ol, which costs about $970 CAD/g. Since it appears at the beginning of the synthesis tree, amplification of material cost is inevitable, and this puts the cost to produce 1 g of quinine by this route at an astronomical value of $56 million dollars! It is interesting to compare the RMC figures calculated here with the Aldrich price of $4.13 CAD/g for the racemic product at 90% purity. This, of course, reflects the likelihood that the Aldrich product is obtained by extraction from natural sources and not by a total synthesis route. Indeed this supposition is consistent with the observation that up to 13% of the dry weight of bark of *Cinchona ledgeriana* trees is quinine, a value exceeding the RMEs of all four total syntheses, and represents the major commercial source of this material.[29] The Stork plan could be more competitive with the Jacobsen plan in terms of RMC if it

(27) (a) Pictet, A.; Gams, A. *Compt. Rend.* **1909**, *149*, 210. (b) Pictet, A.; Gams, A. *Chem. Ber.* **1909**, *42*, 2943. (c) Kropp, W.; Decker, H. *Chem. Ber.* **1909**, *42*, 1184. (d) Wahl, H. *Bull. Chim. Soc. Fr.* **1950**, 680. (e) Kindler, K.; Peschke, K. *Arch. Pharm. (Weinheim, Ger.)* **1934**, *272*, 236. (f) Pal, B. C. *J. Sci. Ind. Res.* **1958**, *17A*, 270. (g) Redel, J.; Bouteville, A.; *Bull. Chim. Soc. Fr.* **1949**, 443. (h) Dean, F. H. Unpublished results from Ontario Research Foundation.

(28) (a) Antunes, L. M.; Organ, M. G. *Tetrahedron Lett.* **2003**, *44*, 6805. (b) Ghasemi, H.; Antunes, L. M.; Organ, M. G. *Org. Lett.* **2004**, *6*, 2913.

Table 5. **Summary of reaction metrics and synthesis tree parameters for convergent sildenafil and absinthin synthesis plans**

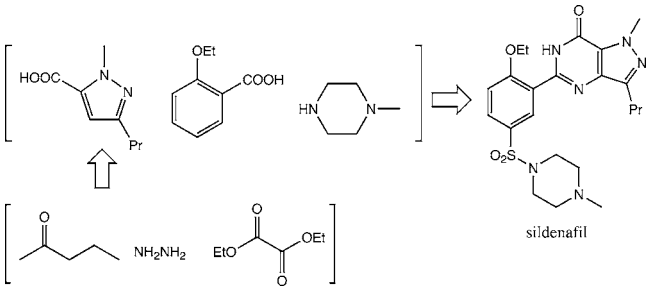
	Sildenafil[a] (convergent)	Absinthin[b] (pseudo-convergent)
	Kernel Reaction Metrics	
AE	0.523 (0.402)	0.156
E_{mw}	0.91 (1.48)	5.41
RME	0.374 (0.175)	0.0311
E_m	1.67 (4.71)	31.17
$\epsilon_{pseudo-overall}$ ($\epsilon_{overall}$)	0.715 (0.435)	0.199
number of reaction inputs, I	10 (15)	22 (15)
number of reaction steps, M	7 (11)	14 (10)
Number of reaction stages, N	5 (9)	10 (10)
μ_1(g/mol per reaction stage)	−422.52 (−369.81)	−315.71 (+35.08)
	Tree Parameters	
P coordinate	(5, 7.423)	(10, 19.697)
	(9, 12.405)[a]	(10, 12.711)[b]
θ_p (deg)	78.381 (64.090)	70.509 (59.153)
θ_{mcr} (deg)	154.942 (163.740)	169.119 (163.740)
degree of convergence, δ	0.506 (0.391)	0.417 (0.361)
ρ_{actual}	1.486 (1.378)	1.970 (1.271)
ρ_{1-mcr}	4.5 (7)	10.5 (7)
relative rate of convergence, ρ_{rel}	0.330 (0.197)	0.188 (0.182)
asymmetry, β	0.651 (0.772)	0.876 (0.816)

[a] Values in parentheses pertain to an extended synthesis tree which includes the synthesis of the starting 1-methyl-3-propyl-1*H*-pyrazole-carboxylic acid. [b] Values in parentheses pertain to a true linear synthesis tree configuration where the stoichiometric coefficients of all reactant inputs are doubled in the branch leading to 2 equiv of monomer I_4 which then undergo Diels−Alder dimerization in the fifth reaction stage.

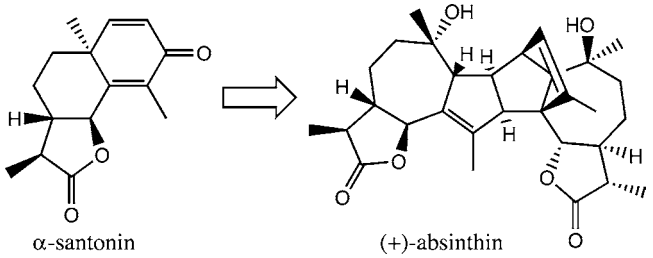
had been possible to synthesize the required (−)-4-vinyl-γ-butyrolactone stereoselectively instead of making the racemic mixture and discarding half the material in a resolution step.[30] Curiously, a synthesis of the opposite (+)-stereoisomer by stereoselective radical cyclization has been achieved.[32] In terms of tracking the building-up character of the syntheses to the target quinine, the Jacobsen plan comes out on top with a net building of 115 g/mol per reaction stage. Although the Stork plan is the most material efficient overall, this trend is offset by it having a net *loss* of 47 g/mol per reaction stage over the entire synthesis. This is due to the frequent use of bulky protecting groups which momentarily increase the molecular weights of intermediates along the way by a significant margin over the target threshold of 324 g/mol.

(ii) Sildenafil. The industrial synthesis of sildenafil (Viagra) has been highlighted as a milestone in "green" synthesis design in the pharmaceutical industry because it has a reported low environmental impact factor based on mass of 6 kg of waste per kg of product.[25] The reduction in the E_m value was achieved largely by reducing or eliminating solvent usage and by optimizing reaction yields. A previous detailed determination of the "green" metrics for this synthesis corroborated this finding.[3a] The kernel E_m and RME values given in Table 5 show that the maximum performance of this synthesis plan has yet to be reached with values of 1.7 kg waste/kg product and 37%, respectively. With respect to the shape of the synthesis tree the plan exhibits a high degree of convergence of 51% since the convergent step occurs in the penultimate step in the plan. It needs to be pointed out that these figures apply to a synthesis beginning with 1-methyl-3-propyl-1*H*-pyrazole-5-carboxylic acid as the key starting material (see Scheme 4). If, however, the synthesis of this material is taken into account and the resulting longer synthesis tree reanalyzed, the kernel overall RME drops to 18% from 37%, the kernel overall E_m increases to 4.7 from 1.7 kg waste/kg product, and the pseudo-overall yield drops to 44% from 72%. As expected with a longer linear branch, the degree of convergence drops to 39% from 51%, the asymmetry increases to 0.77 from 0.65, the relative rate of convergence to the target product drops from 0.33 to 0.20, and the MW first moment increases from −423 to −370 g/mol per reaction stage, indicating a lower degree of building is going on in the revised plan.

Scheme 4



Scheme 5



(iii) Absinthin. The recent synthesis of (+)-absinthin from α-santonin (see Scheme 5) provides an interesting case study to test ideas presented here. Its synthesis plan involves a four-step linear sequence to a key intermediate which undergoes

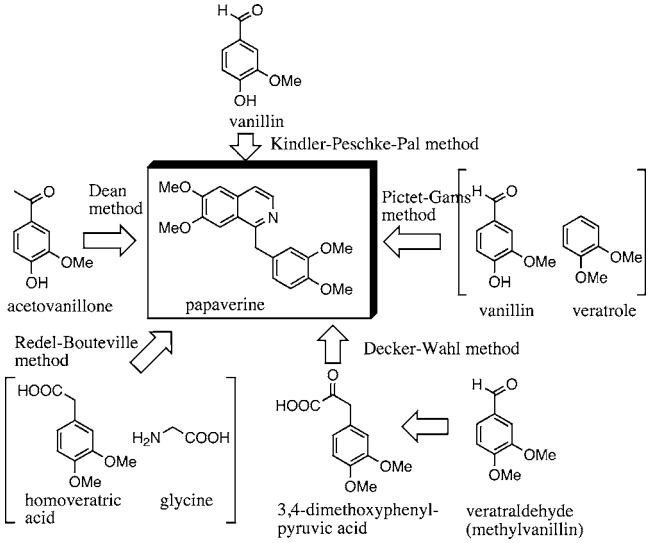
(29) Leete, E. *Acc. Chem. Res.* **1969**, *2*, 159.
(30) (a) Kondo, K.; Mori, F. *Chem. Lett.* **1974**, 741. (b) Ishibashi, F.; Taniguchi, E. *Phytochemistry* **1998**, *49*, 613.
(31) Bertz, S. H. In *Complexity in Chemistry: Introduction and Fundamentals*; Taylor & Francis: London, 2003; Chapter 3, p 91.
(32) Villar, F.; Kolly-Kovac, T.; Equey, O.; P. Renaud, P. *Chem. Eur. J.* **2003**, *9*, 1566.

dimerization in a Diels−Alder fashion. This is followed by a deprotection step and four more steps which serve to epimerize the two hydroxy groups. The implication is that if β-santonin were used as a starting material instead of α-santonin (+)-absinthin would presumably be obtained stereoselectively, thus eliminating the last four steps. However, this supposition needs to be verified experimentally. An analysis of such a shortened plan predicts that its overall kernel RME should increase to 5% from 3% and its overall kernel E_m should decrease by 12.9 g waste/g product, assuming that reaction yields to the Diels−Alder adduct are the same as those for α-santonin.

The plan is designated as pseudo-convergent because the two converging branches are identical and therefore lead to the same structure before the converging dimerization step. There are therefore two ways to analyze the synthesis plan. One analysis is based on a tree with two identical branches converging in the Diels−Alder step followed by a further linear sequence; the other is based on a complete linear tree configuration where the stoichiometric coefficients of all reactant inputs are doubled in the branch leading up to the Diels−Alder product. Table 5 summarizes the results of metrics and tree parameters for this pseudo-convergent synthesis plan based on the two kinds of synthesis tree depictions. Both analyses yield identical results for the kernel reaction metrics as they should; however, they differ in the molecular weight moment and in the tree parameters. The convergent plan exhibits a net building up of 316 g/mol per reaction stage whereas the linear plan shows a net degradation of 35 g/mol per reaction stage. The convergent plan is more asymmetric and has a higher degree of convergence and a slightly higher relative rate of convergence. Both tree depictions appear to adequately describe the synthesis so it is not possible to say which tree configuration is more correct than the other.

(iv) Papaverine. The opiate alkaloid, papaverine, from *Papaver somniferum* is an anti-spasmodic, vasodilator, and smooth muscle relaxant. Its total synthesis has been studied since Pictet and Gams early work in 1909 and has since been followed up by various industrial syntheses up till the early 1950s using important industrial commodities as vanillin, acetovanillone, veratraldehyde (methylvanillin), and homoveratric acid as starting materials (see Scheme 6). Table 6 summarizes the results of the present treatment to five synthetic plans for this natural product. All are convergent plans except for the linear Redel−Bouteville plan. The Kindler−Peschke−Pal plan is the most efficient material performer with the highest kernel RME of 15% and pseudo-overall yield of 55% in eight stages, and greatest building up of 280 g/mol per reaction stage. It also happens to be the cheapest method with the lowest RMC of 45 cents per gram since it begins with very cheap starting materials and has very high reaction yields per step with no step below 79% yield. The runner-up in material performance is the Dean method with a kernel RME of 8% and pseudo-overall yield of 40% also in eight stages. The Redel−Bouteville plan has the highest atom economy at 32%, but since its pseudo-overall reaction yield is lower at 13%, its resulting kernel

Scheme 6



RME drops to 4%. The Dean method utilizes an expensive and toxic reagent early in the second stage, thallium(III)-nitrate, to transform an acetophenone to a methyl aryl acetate in a redox reaction. The high cost of this reagent puts the overall RMC to produce papaverine by this method to $22/g, the second highest of the five methods. In order for this method to be competitive with the Kindler−Peschke−Pal method, it becomes imperative to either replace it with a cheaper reagent to effect the same transformation or, failing that, to find an appropriate cheap and efficient recycling reaction that recycles the thallium(I) byproduct back to thallium(III) reagent.

The tree analyses of the Kindler−Peschke−Pal and Dean methods highlight a useful material efficient synthetic strategy used in both plans; namely, the synthesis of a key intermediate which is partitioned into two pathways which in turn converge at a later stage in the synthesis. Synthetic plans exhibiting this strategy have been termed "reflexive".[32] In the Kindler−Peschke−Pal plan the key intermediate is homoveratronitrile which is split into two paths; one part is reduced to the corresponding amine, and the other part is hydrolyzed to the corresponding acid. Both of these transformations occur in parallel steps in the fifth reaction stage. These products are then recombined in the sixth stage to produce homoveratrylhomoveratramide. From the synthesis tree it is possible to determine the partitioning ratio for the transformations homoveratronitrile to homoveratric acid and homoveratronitrile to homoveratramine to be 1:0.79 or 56% and 44%, respectively. In the Dean plan the key intermediate is methyl 3,4-dimethoxyphenylacetate. Rather than partitioning this material into two different paths where both parts are transformed to intermediates which are later recombined, one portion is saved and left untransformed until a later reaction stage. With this modification, about 64% of the aryl acetate produced in the second stage is committed to the next steps leading to homoveratramine, and 36% is saved for reaction with homoveratramine in the sixth reaction stage to produce homoveratrylhomoveratramide.

With respect to the tree parameters, as expected the only linear plan (Redel−Bouteville) has the lowest degree of

Table 6. **Summary of reaction metrics and synthesis tree parameters for papaverine synthesis plans**

	Pictet−Gams	Decker−Wahl[b]	Redel−Bouteville	Kindler−Peschke−Pal	Dean[c]
		Kernel Reaction Metrics			
AE	0.136	0.197	**0.317**	0.274	0.199
E_{mw}	6.37	4.03	2.15	2.65	4.03
RME	5.21×10^{-3}	2.75×10^{-2}	4.04×10^{-2}	$\mathbf{1.51 \times 10^{-1}}$	7.96×10^{-2}
E_m	191.12	35.37	23.75	5.63	11.57
$\epsilon_{pseudo-overall}$ ($\epsilon_{overall}$)	0.038	0.138 (0.106)	0.127 (0.0764)	**0.550**	0.400
number of reaction inputs, I	18	20 (13)	**11**	15	12 (8)
number of reaction steps, M	11	12 (9)	**8**	13	10 (8)
number of reaction stages, N	8	9 (9)	8	8	8 (8)
μ_1(g/mol per reaction stage)	−223.18	−83.10 (+86.40)	−29.96	**−280.23**	−212.89 (−212.89)
RMC[a] ($ CAD/g)	29.04	4.72	8.17	**0.45**	22.05
		Tree Parameters			
P coordinate	(8, 15.408)	(9, 17.675) (9, 10.747)	(8, 8.731)	(8, 11.227)	(8, 8.965) (8, 5.215)
θ_p (deg)	73.812	71.389 (57.981)	56.514	73.647	62.528 (45.678)
θ_{mcr} (deg)	166.580	167.982 (161.075)	157.380	163.740	159.390 (148.109)
degree of convergence, δ	0.443	0.425 (0.360)	0.359	**0.450**	0.392 (0.308)
ρ_{actual}	1.926	1.964 (1.194)	1.091	1.403	1.121 (0.652)
ρ_{I-mcr}	8.5	9.5 (6)	5	7	5.5 (3.5)
relative rate of convergence, ρ_{rel}	**0.227**	0.207 (0.199)	0.218	0.200	0.204 (0.186)
asymmetry, β	0.813	0.861 (0.791)	0.746	**0.604**	0.630 (0.490)

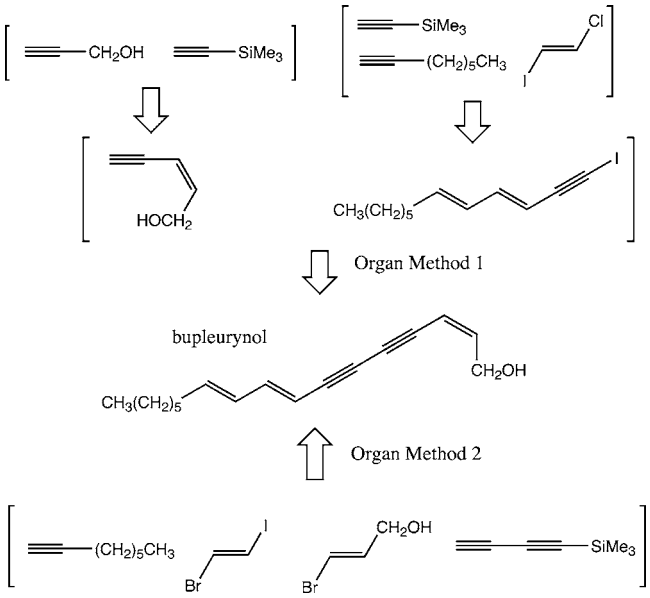
[a] Based on unit costs ($/g) taken from an Aldrich 2003−2004 Catalogue in Canadian dollars using prices for the largest unit listed in the catalogue. [b] Values in parentheses pertain to a true linear synthesis tree configuration where the stoichiometric coefficients of all reactant inputs are doubled in the branch leading to 2 equiv of intermediate I_3 (3,4-dimethoxyphenylpyruvic acid) which then undergoes amidation in the fourth reaction stage. [c] Values in parentheses pertain to an alternative synthesis tree configuration where the short branch leading to intermediate I_2 (methyl 3,4-dimethoxyphenylacetate) is repeated horizontally along the reaction stage axis rather than vertically along the reactant input axis.

convergence. The most material efficient Kindler−Peschke−Pal plan is also the most convergent with a degree of convergence of 45% and is the most symmetric. Interestingly, the original Pictet−Gams synthesis has the highest relative rate of convergence at 0.23 and second highest degree of convergence at 44% despite its lowest kernel RME rating of 0.5% and 4% pseudo-overall yield among all the plans.

The Decker−Wahl plan has two identical branches that converge in the fourth stage in a pseudo-dimerization of 3,4-dimethoxyphenylpyruvic acid in the presence of ammonia. As for the absinthin plan an analysis of a modified linear Decker−Wahl plan was also carried out. Consistent with the findings of the linearized and convergent absinthin plans, the linearized Decker−Wahl plan showed a net degradation of 86 g/mol per reaction stage compared to a net building up of 83 g/mol per reaction stage in the convergent version. The degree of convergence decreased to 36% in the linear plan from 43% in the convergent plan. Again, the convergent plan is more asymmetric and has a slightly higher relative rate of convergence. Kernel reaction metrics for both versions are identical.

(v) Bupleurynol. The two recent syntheses of bupleurynol[28] by convergent strategies (see Scheme 7) is an ideal case to test the merits of the synthesis tree parameters presented in this work. Table 7 summarizes the relevant data

Scheme 7



for a convergent plan (Method 1) consisting of two branches of six and seven reaction steps, respectively, which then converge to the target product in the final eighth stage, and for a single-stage eight-component plan (Method 2) involving the sequential addition of reagents without isolation of any

Table 7. **Summary of reaction metrics and synthesis tree parameters for convergent bupleurynol synthesis plans**

	method 1 (convergent)	method 2A (single-stage MCR convergent)	method 2B (3-stage convergent)	method 2C (6-stage convergent)
		Kernel Reaction Metrics		
AE	0.0805	0.1832	0.1832	0.1832
E_{mw}	11.42	4.46	4.46	4.46
RME	0.0177	0.0788	a	a
E_m	55.66	11.71	a	a
$\epsilon_{pseudo-overall}$ ($\epsilon_{overall}$)	0.219	0.43	a	a
number of reaction inputs, I	24	8	8	8
number of reaction steps, M	15	1	4	7
number of reaction stages, N	8	1	3	6
μ_1(g per mole per reaction stage)	−97.02	−183.1	−29.62	−33.79
		Tree Parameters		
P coordinate	(8, 18.523)	(1, 3.5)	(3, 5.5)	(6, 5.828)
θ_p (deg)	95.871	148.109	87.955	55.219
θ_{mcr} (deg)	170.061	148.109	148.109	148.109
degree of convergence, δ	0.564	1	0.594	0.373
ρ_{actual}	2.315	3.5	1.83	0.971
ρ_{I-mcr}	11.5	3.5	3.5	3.5
relative rate of convergence, ρ_{rel}	0.201	1	0.524	0.278
asymmetry, β	0.611	0	0.571	0.665

[a] Insufficient reaction yield data available to make determination.

intermediate products. This second method is a rare example in which the hypothetical scenario discussed in section 4 of having all necessary building block inputs assembled in one multicomponent reaction stage has been achieved experimentally. By definition its degree of convergence is 100%, and it is completely symmetric with a β value of 0. In terms of its material efficiency its kernel RME is about 4.5 times larger than the first convergent synthesis plan.[33] It should be noted that this novel demonstration of increased material performance in a one-pot synthesis was performed on a scale of less than 2 mmol, and that a key condition for the whole synthesis to work was the proviso that the byproducts produced in any one operation did not adversely affect the chemistry of another. The challenge of scalability using this approach remains an open problem, but the results given in this and earlier work[3] support the contention that at least the idea of designing highly efficient synthetic plans by incorporating single-stage simultaneous or sequential multicomponent approaches is on the right track.

Figure 11 shows diagrams of modified synthesis trees for Method 2 that depict variations, depending on the isolation of intermediate products along the way to the target product. This is a nice demonstration of the ideas put forward in section 4 on various tree parameters. As can be readily seen, more isolations of products translate into more reaction stages. This in turn distorts the original symmetric eight-component tree so that the degrees of convergence and asymmetry decrease and increase, respectively. From the materials performance point of view it is expected that the overall kernel RME values should progressively decrease from a value of 8% unless fortuitously all reactions afford

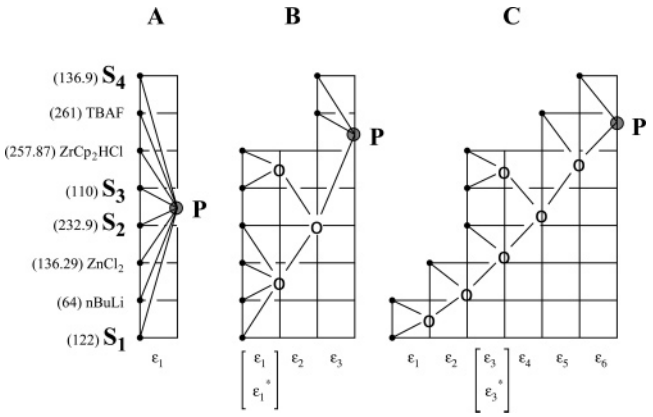


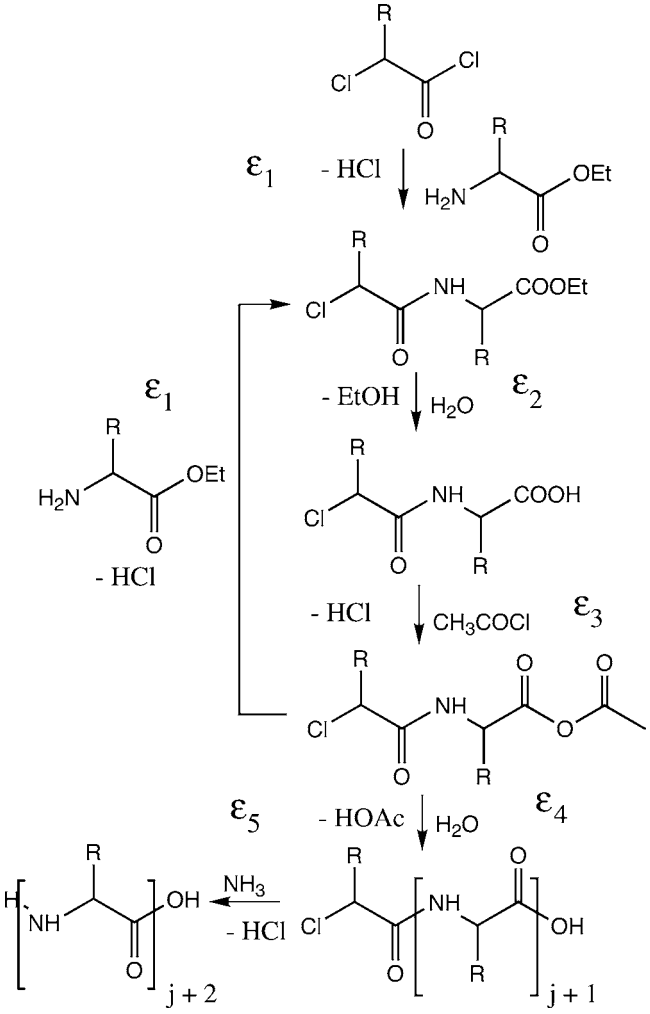
Figure 11. **Bupleurynol synthesis variations for Method 2: (A) "Single-stage MCR convergent", a sequential eight-component coupling single-stage sequence without isolation of intermediates (coordinate of *P*, (1, 3.5); $\rho_{rel} = 1$; $\beta = 0$; $\gamma = 1$); (B) "Three-stage convergent", a sequential eight-component coupling three-stage sequence with isolation of three intermediates (coordinate of *P*, (3, 5.5); $\rho_{rel} = 0.524$; $\beta = 0.571$; $\delta = 0.594$); (C) "Six-stage convergent", a sequential eight-component coupling six-stage sequence with isolation of all intermediates (coordinate of *P*, (6, 5.828); $\rho_{rel} = 0.278$; $\beta = 0.665$; $\delta = 0.373$). See Supporting Information, Part 3, for synthesis schemes and intermediate structures (S_1 = trimethylsilylbutadiyne, S_2 = *trans*-1-bromo-2-iodoethene, S_3 = 1-octyne, and S_4 = *cis*-3-bromo-2-propen-1-ol).**

quantitative reaction yields. Since no isolations of products were conducted by the Method 2 reaction, yields for the relevant steps are not available to make RME calculations possible. The overall atom economy determination of 18%, however, remains unchanged as expected.

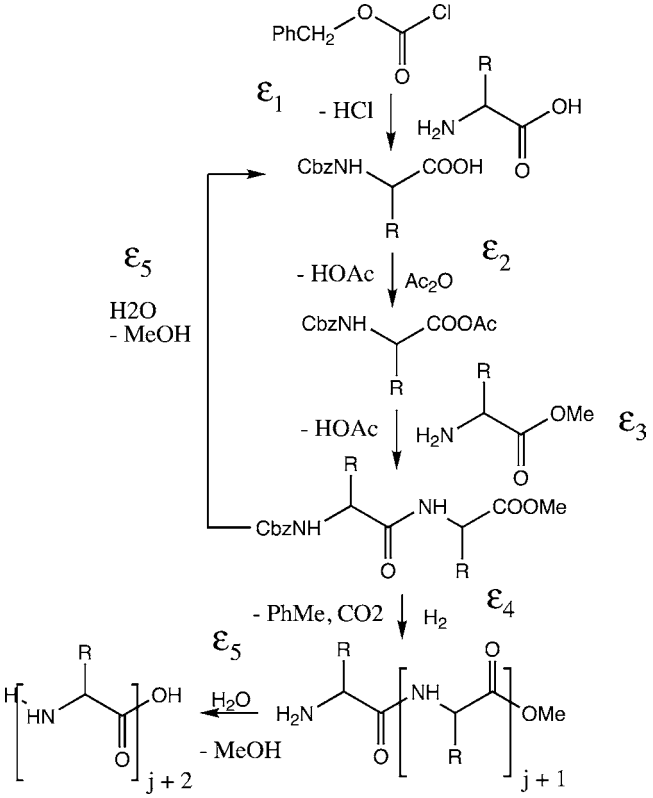
(vi) Polypeptide Synthesis. The tree analysis method also can handle synthesis plans that involve repeated cycles or loops. As an example several types of polypeptide synthesis methodologies have been examined including: the Fischer[34] (Scheme 8), Bergmann−Zervas[35] (Scheme 9), Merrifield[36]

(33) The claim in ref 28b that the "overall yield" of the single-stage sequential multicomponent synthesis of bupleurynol is 7 times greater than the first convergent synthesis is based on the erroneous determination of "overall yield" as the multiplicative product of all reactions in both branches of the earlier synthesis. The kernel RME values given in Table 7 give the correct material performances for both synthesis plans.

Scheme 8. **(Fischer method)**



Scheme 9. **(Bergmann−Zervas method)**



(Scheme 10), azide method[37] (Scheme 11), anhydride method[37,38] (Scheme 12), and segment doubling strategies[39] (Scheme 13). Synthesis trees for all methods are given in the Supporting Information. For the sake of simplicity in comparing the methods according to their intrinsic linking performance from monomer to oligomers at the common denominator level, the following assumptions were made: (1) kernel reaction metrics were determined for the synthesis of polypeptides of one type of amino acid such as polyglycine, polyalanine, etc., and (2) all R groups of amino acids are assumed to be protected as necessary.

Table 8 summarizes key kernel metrics expressions obtained from the trees for each method. Tables 9 and 10 summarize limiting AE values for each method as a function of amino acid R group size for polypeptides of infinite chain length, and kernel overall RME values for the synthesis of octamers using representative averaged literature reaction yields, respectively. In terms of atom economy performance the ranking of the methods in descending order is: azide, anhydride, Fischer, Bergmann−Zervas, segment doubling, and Merrifield. In terms of overall kernel RME performance the ranking in descending order changes to: segment doubling, anhydride, Fischer, Merrifield, Bergmann−Zervas, and azide. The results of the kernel RME ranking parallel the ascending order of number of different reaction yield parameters involved: segment doubling (3), anhydride (4), Fischer (5), Merrifield (5), Bergmann−Zervas (5), and azide (6). For the synthesis of an 8-mer peptide nine reactions are required by the segment doubling method, 16 by the anhydride method, 23 by the Fischer method, 17 by the Merrifield method, 23 by the Bergmann−Zervas method, and 24 by the azide method. This is a clear demonstration of the intuitive observation that an increased number of different reactions involved will strongly attenuate the mass efficiency of product. The highly efficient segment doubling strategy

(34) (a) Fischer, E.; Fourneau, E. *Chem. Ber.* **1901**, *34*, 2868. (b) Fischer, E. *Chem. Ber.* **1903**, *36*, 2982.

(35) Bergmann, M.; Zervas, L. *Chem. Ber.* **1932**, *65*, 1192.

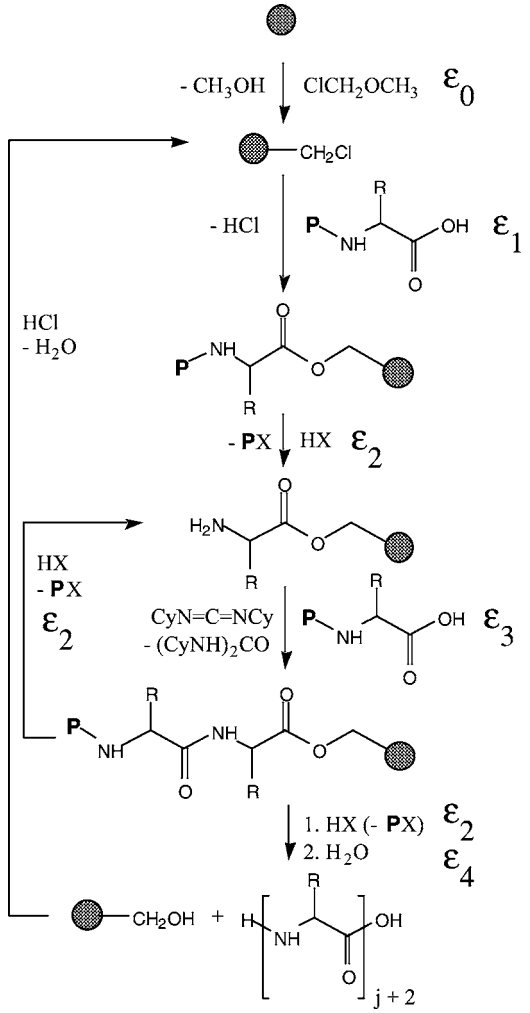
(36) (a) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149. (b) Merrifield, R. B. *Science* **1965**, *150*, 178. (c) Merrifield, R. B. *Angew. Chem., Int. Ed.* **1985**, *24*, 799. (d) Merrifield, R. B. *Pure Appl. Chem.* **1978**, *50*, 643. (e) Merrifield, R. B. *Science* **1986**, *232*, 341. (f) Merrifield, R. B. In *Peptides*; Gutte, B., Ed.; Academic Press: San Diego, 1995; p 93. (g) Loffet, A. *React. Polym.* **1994**, *22*, 165. (h) Okuda, T. *Naturwissenschaften* **1968**, *55*, 209. (i) Rich, D. H.; Singh, J. in *The Peptides. Analysis, Synthesis, Biology*; Gross, E., Meienhofer, J., Eds. Academic Press: New York, 1979; Vol. 1, p 241. (j) Sheehan, J. C.; Hess, G. P. *J. Am. Chem. Soc.* **1955**, *77*, 1067.

(37) (a) Curtius, T. *Chem. Ber.* **1902**, *35*, 3226. (b) Lloyd-Williams, P.; Albericio, F.; Girald, E. *Chemical Approaches to the Synthesis of Peptides and Proteins*, CRC Press: Boca Raton, FL, 1997. (c) Heinzel, W.; Verlander, M. S. *Ullmann's Encyclopedia of Industrial Chemistry*, 5th ed.; VCH: Weinheim, 1991; Vol. 19, p 159. (d) Meienhofer, J. In *The Peptides. Analysis, Synthesis, Biology*; Gross, E., Meienhofer, J., Eds.; Academic Press: New York, 1979; Vol. 1, p 197. (e) Klausner, Y. S.; Bodanszky, M. *Synthesis* **1974**, 549.

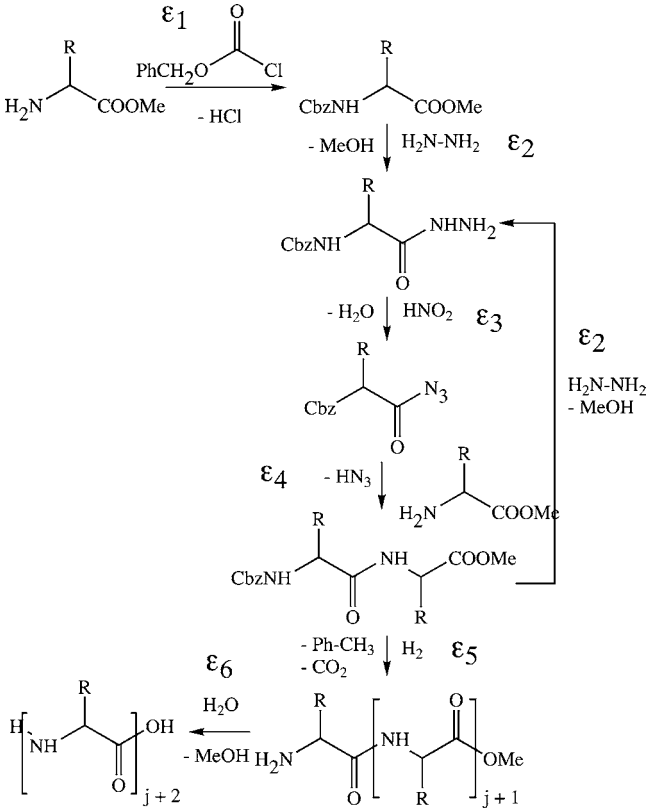
(38) (a) Meienhofer, J. In *The Peptides. Analysis, Synthesis, Biology*; Gross, E., Meienhofer, J., Eds.; Academic Press: New York, 1979; Vol. 1, p 263. (b) See ref 35g. (c) Wieland, H.; Bernhard, H. *Ann. Chem.* **1951**, *572*, 190. (d) Boissonnas, R. A. *Helv. Chim. Acta* **1951**, *34*, 874. (e) Vaughan, J. R., Jr. *J. Am. Chem. Soc.* **1951**, *73*, 3547.

(39) (a) Zhang, J.; Moore, J. S.; Xu, Z.; Aguirre, R. A. *J. Am. Chem. Soc.* **1992**, *114*, 2273. (b) Wender, P. A.; Jessop, T. C.; Pattabiraman, K.; Pelkey, E. T.; VanDeusen, C. L. *Org. Lett.* **2001**, *3*, 3229.

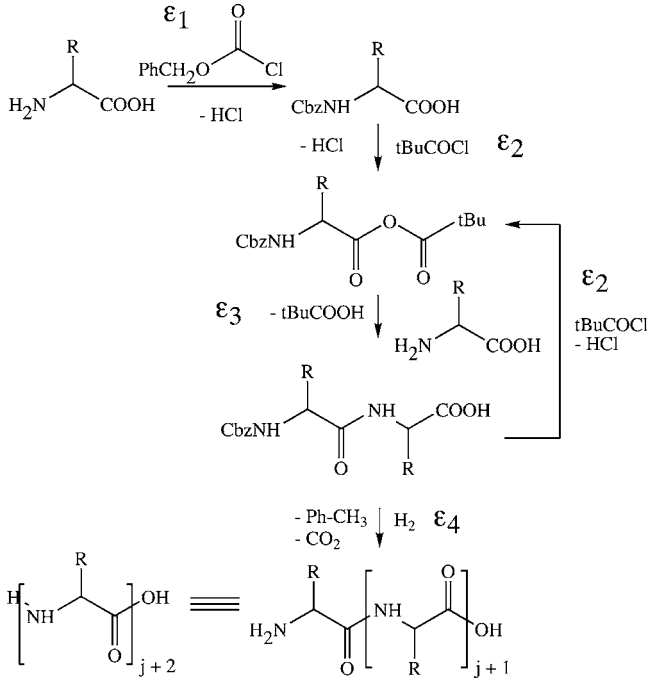
Scheme 10. **(Merrifield method)**



Scheme 11. **(Azide method)**



Scheme 12. **(Anhydride method)**



significantly reduces the number of steps but also has very high reaction yields exceeding 80%. A drawback with this method is that oligomers of intermediate lengths other than 2-mers, 4-mers, 8-mers, 16-mers, etc. cannot be synthesized. Although the azide method has the best atom economy, it is the worst RME performer because it involves the highest number of reactions, and about half of them are of modest yields (73−79% average yield). Simple calculations show that reaction yields have to exceed 80% before "good" RME results can be achieved by this method.

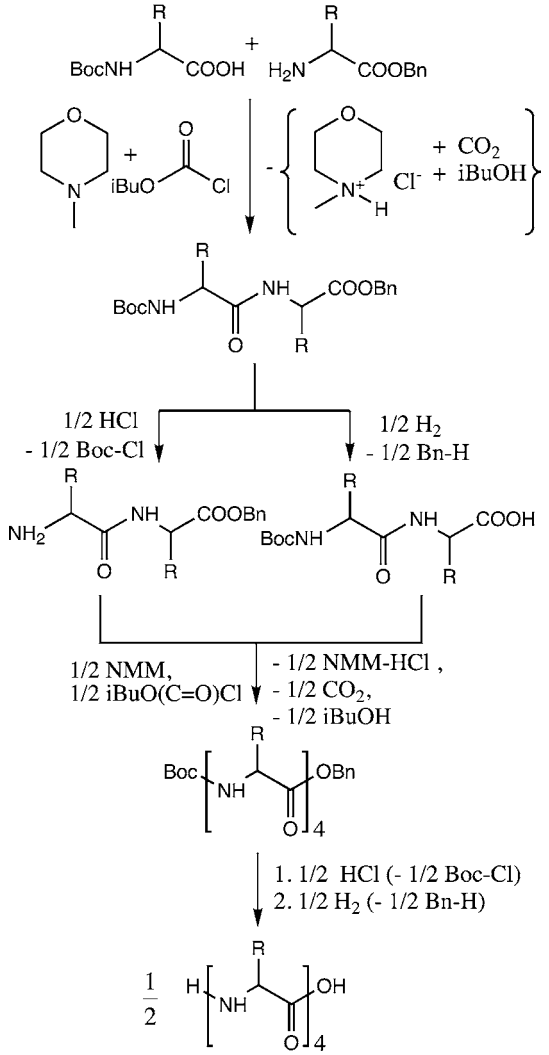
In terms of recycling potential the Merrifield synthesis has the possibility of recovery of the protecting group and the conversion of dicyclohexylurea (DCU) back to dicyclohexyldiimide (DCC).[40] For the synthesis of an *m*-mer polypeptide the accumulated mass of dicyclohexylurea waste collected by this method is given by

$$\text{mass}_{\text{DCU}} = 224\left(\frac{x}{\epsilon_2\epsilon_3\epsilon_4}\right)\left[\sum_{j=0}^{m-2}\frac{1}{(\epsilon_2\epsilon_3)^j}\right] = 224\left(\frac{x}{\epsilon_2\epsilon_3\epsilon_4}\right)\left[\frac{1-(\epsilon_2\epsilon_3)^{1-m}}{1-(\epsilon_2\epsilon_3)^{-1}}\right] \quad (44)$$

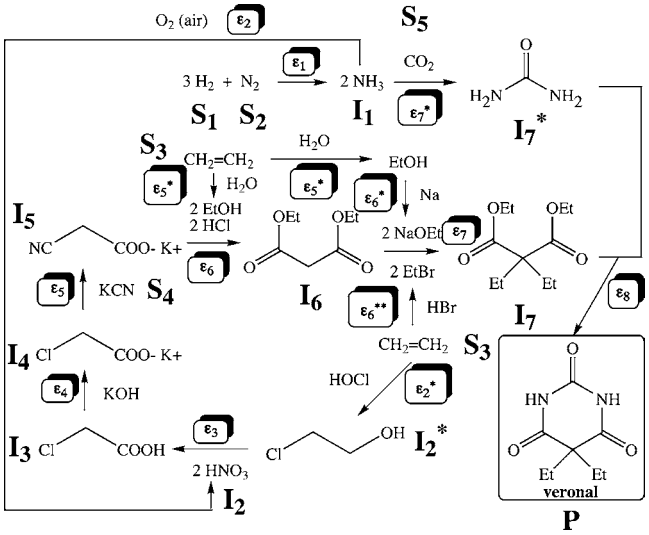
where the yields ϵ_2, ϵ_3, and ϵ_4 correspond to the amino group deprotection, DCC coupling step, and final hydrolysis from the polymer support, respectively, and x is the number of moles of final product. If these yields are set to unity,

(40) (a) Yamazaki, N. JP 10330344, 1998. (b) Fujibayashi, R. JP 08231491, 1995. (c) Hussenet, P.; Le Goff, P.; Sennyey, G. EP 723955, 1996. (d) Stevens, C. L.; Singhal, G. H.; Ash, A. B. *J. Org. Chem.* **1967**, *32*, 2895. (e) Smith, M.; Moffatt, J. G.; Khorana, H. G. *J. Am. Chem. Soc.* **1958**, *80*, 6204. (f) Amiard, G.; Heymes, R. *Bull. Soc. Chim. Fr.* **1956**, 1360.

Scheme 13. **(Segment doubling method, tetramer synthesis shown)**

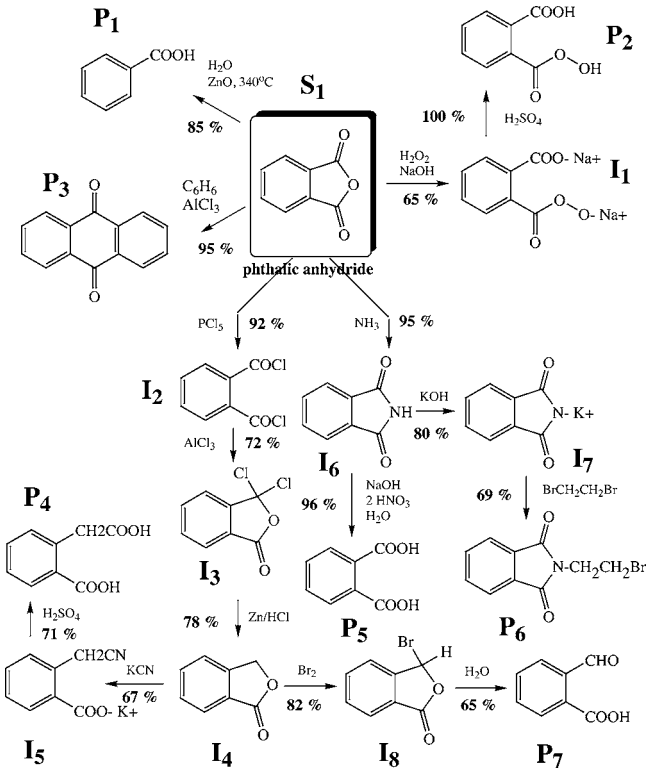


Scheme 14



then eq 44 reduces to the value of the minimum mass of waste DCU produced which is $224x(m - 1)$ grams. When expressed as a fraction of the total waste produced in the synthesis it can be shown that under these conditions

Scheme 15



that

$$\frac{\text{mass}_{\text{DCU}}}{\text{mass}_{\text{total waste}}} = \frac{224(m-1)}{m(P+X+224)-206} \qquad (45a)$$

when *N*-ethylmorpholine and 1-hydroxybenzotriazole are recovered in the DCC coupling step, and that

$$\frac{\text{mass}_{\text{DCU}}}{\text{mass}_{\text{total waste}}} = \frac{224(m-1)}{m(P+X+474)-456} \qquad (46b)$$

when *N*-ethylmorpholine and 1-hydroxybenzotriazole are not recovered in the DCC coupling step. For the synthesis of an infinitely long polypeptide these fractions tend to minimum values of 40−62% for the former case and 28−37% for the latter case depending on the X and P groups chosen in the synthesis. This is clear evidence that recycling DCU back to DCC is beneficial to the overall green performance of this method and would make it competitive with the anhydride method. Similar analysis of the accumulated 2,2-dimethylpropanoic acid waste product in the anhydride method shows that it accounts for a minimum of 74% of the total mass of waste produced for an infinitely long polypeptide when all reaction yields are set to unity, a best-case scenario situation. In the Bergmann−Zervas, azide, and anhydride methods the benzyloxycarbonyl(Cbz) group is reduced to toluene and carbon dioxide so that retrieval of toluene is probably the only viable option for recovery of materials. In the segment doubling strategy the *tert*-butoxycarbonyl protecting group and *N*-methylmorpholine can be recovered, and isobutyl alcohol may be converted back to isobutyl chloroformate for reuse. The Fischer method has the least opportunities for

Table 8. **Summary of kernel reaction metrics relationships for various polypeptide methodologies**[a]

Fischer method

$$(RME)_m = \frac{[18+m(R+56)](\varepsilon_1\varepsilon_2\varepsilon_3)^{m-1}(\varepsilon_4\varepsilon_5)}{R+111.9+\left(\sum_{j=0}^{m-2}(\varepsilon_1\varepsilon_2\varepsilon_3)^j\right)(R+102+\varepsilon_1[18+78.45\varepsilon_2])+(\varepsilon_1\varepsilon_2\varepsilon_3)^{m-1}(18+17\varepsilon_4)}$$

$$(AE)_m = \frac{18+m(R+56)}{m(R+198.45)-51.55}$$

$$(AE)_{m\to\infty} = \frac{R+56}{R+198.45}$$

Bergmann-Zervas method

$$(RME)_m = \frac{[18+m(R+56)](\varepsilon_2\varepsilon_3\varepsilon_5)^{m-1}(\varepsilon_1\varepsilon_4)}{D}$$

where

$$D = R+244.45+\left(\sum_{j=0}^{m-2}(\varepsilon_2\varepsilon_3\varepsilon_5)^j\right)\varepsilon_1(102+\varepsilon_2(R+88))+$$

$$18(\varepsilon_1\varepsilon_2\varepsilon_3)\left(\varepsilon_4(\varepsilon_2\varepsilon_3\varepsilon_5)^{m-2}+\sum_{j=0}^{m-3}(\varepsilon_2\varepsilon_3\varepsilon_5)^j\right)+2(\varepsilon_1\varepsilon_2\varepsilon_3)(\varepsilon_2\varepsilon_3\varepsilon_5)^{m-2}$$

$$(AE)_m = \frac{18+m(R+56)}{m(R+208)+38.45}$$

$$(AE)_{m\to\infty} = \frac{R+56}{R+208}$$

Merrifield method[b]

Case I: Recovery of N-ethylmorpholine and 1-hydroxybenzotriazole in DCC step

$$(RME)_m = \frac{[18+m(R+56)](\varepsilon_2\varepsilon_3)^{m-1}(\varepsilon_1\varepsilon_2\varepsilon_4)}{D}$$

where

$$D = (P+R+73)+\left(\sum_{j=0}^{m-2}(\varepsilon_2\varepsilon_3)^j\right)\varepsilon_1(X+1+\varepsilon_2(P+R+279))+\varepsilon_1(\varepsilon_2\varepsilon_3)^{m-1}(18\varepsilon_2+X+1)$$

X is molecular weight of X group (Br, Cl, CF_3COO, or $HN(CH_2)_4N$), P is molecular weight of protecting group (Cbz, tBOC, or fMOC)

$$(AE)_m = \frac{18+m(R+56)}{m(P+R+X+280)-188}$$

$$(AE)_{m\to\infty} = \frac{R+56}{P+R+X+280}$$

Case II: Counting N-ethylmorpholine and 1-hydroxybenzotriazole in DCC step as part of the waste

$$(RME)_m = \frac{[18+m(R+56)](\varepsilon_2\varepsilon_3)^{m-1}(\varepsilon_1\varepsilon_2\varepsilon_4)}{D}$$

where

$$D = (P+R+73)+\left(\sum_{j=0}^{m-2}(\varepsilon_2\varepsilon_3)^j\right)\varepsilon_1(X+1+\varepsilon_2(P+R+529))+\varepsilon_1(\varepsilon_2\varepsilon_3)^{m-1}(18\varepsilon_2+X+1)$$

$$(AE)_m = \frac{18+m(R+56)}{m(P+R+X+530)-438}$$

$$(AE)_{m\to\infty} = \frac{R+56}{P+R+X+530}$$

Azide method

$$(RME)_m = \frac{[18+m(R+56)](\varepsilon_2\varepsilon_3\varepsilon_4)^{m-1}(\varepsilon_1\varepsilon_5\varepsilon_6)}{D}$$

where

$$D = R+258.45+\left(\sum_{j=0}^{m-2}(\varepsilon_2\varepsilon_3\varepsilon_4)^j\right)\varepsilon_1(32+47\varepsilon_2+\varepsilon_2\varepsilon_3(R+88))+(\varepsilon_2\varepsilon_3\varepsilon_4)^{m-1}\varepsilon_1(2+18\varepsilon_5)$$

$$(AE)_m = \frac{18+m(R+56)}{m(R+167)-111.45}$$

$$(AE)_{m\to\infty} = \frac{R+56}{R+167}$$

Anhydride method

$$(RME)_m = \frac{[18+m(R+56)](\varepsilon_2\varepsilon_3)^{m-1}(\varepsilon_1\varepsilon_4)}{D}$$

where

$$D = R+244.45+\left(\sum_{j=0}^{m-2}(\varepsilon_2\varepsilon_3)^j\right)\varepsilon_1(120.45+\varepsilon_2(R+74))+2(\varepsilon_2\varepsilon_3)^{m-1}\varepsilon_1$$

$$(AE)_m = \frac{18+m(R+56)}{m(R+194.45)+52}$$

$$(AE)_{m\to\infty} = \frac{R+56}{R+194.45}$$

Segment Doubling method[c]

$$(RME)_\alpha = \frac{[18+m(R+56)](\varepsilon_1\varepsilon_2\varepsilon_2^*)^\alpha}{D}$$

where

$$D = \gamma_1(\varepsilon_2+\varepsilon_2^*)^{\alpha-1}+\gamma_2(\varepsilon_1\varepsilon_2\varepsilon_2^*)^{\alpha-1}+\gamma_3\left[\sum_{j=0}^{\alpha-2}(\varepsilon_2+\varepsilon_2^*)^j(\varepsilon_1\varepsilon_2\varepsilon_2^*)^{\alpha-j-2}\right]$$

$\gamma_1 = 2R+575.45$, $\gamma_2 = \varepsilon_1(2\varepsilon_2^*+36.45)$, $\gamma_3 = \varepsilon_1(2\varepsilon_2^*+36.45\varepsilon_2+237.45\varepsilon_2\varepsilon_2^*)$,

$m = 2^\alpha$, and α is the number of cycles ($\alpha \geq 2$)

$$(AE)_\alpha = \frac{18+2^\alpha(R+56)}{2^\alpha(R+425.675)-237.45}$$

$$(AE)_{\alpha\to\infty} = \frac{R+56}{R+425.675}$$

[a] See Schemes 7−12 for correspondence between reaction yields in formulas and reaction steps for each method. [b] Excluding functionalization of polymer support step (zeroth step). [c] Restricted to 2-mers, 4-mers, 8-mers, 16-mers, etc. by this method.

recycling. Incorporation of recycling reactions and recovery of byproducts in the above analysis will necessarily result in improved green performances for these methodologies.

7. Example Synthesis Networks

The method of synthesis trees may be used to facilitate the analysis of complex reaction plans or networks involving common starting materials that are used in different branches in convergent or divergent senses. Two illustrative examples given here are a single route to veronal (Scheme 14) involving a complex web of branching from various source nodes to a common target node (convergent sense), and various routes to feedstocks made from phthalic anhydride (Scheme 15) involving branching from a single source node (divergent sense).

The complex reaction network to synthesize veronal (5,5-diethylbarbituric acid) may be converted into a synthesis tree (see Figure 12) which greatly simplifies the entire scheme. At once we can see that it consists of 8 reaction stages, 18 reaction inputs, and 15 reactions. Reaction stages 2, 5, 6, and 7 involve 2, 3, 4, and 2 parallel reactions, respectively, that can be run concurrently to save process time. The expression for the overall kernel RME may be written by inspection by following the node connections as

$$(\mathrm{RME})_{\mathrm{overall}} = \frac{184}{S} \qquad (47a)$$

Table 9. **Summary of limiting AE values for various polypeptides of infinite chain length of one kind of amino acid**[a]

amino acid	MW R group	Fischer	Bergmann−Zervas	Merrifield[b]	azide method	anhydride method	segment doubling
Gly	1	0.286	0.273	0.115	0.339	0.292	0.134
Ala	15	0.333	0.318	0.139	0.390	0.339	0.161
Ser	31	0.379	0.364	0.165	0.439	0.386	0.191
Val	43	0.410	0.394	0.184	0.471	0.417	0.211
Thr	45	0.415	0.399	0.187	0.476	0.422	0.215
Cys	47	0.420	0.404	0.190	0.481	0.427	0.218
Leu	57	0.442	0.426	0.205	0.504	0.449	0.234
Ile	57	0.442	0.426	0.205	0.504	0.449	0.234
Asn	58	0.445	0.429	0.206	0.507	0.452	0.236
Asp	59	0.447	0.431	0.208	0.509	0.454	0.237
Lys	72	0.473	0.457	0.226	0.536	0.480	0.257
Gln	72	0.473	0.457	0.226	0.536	0.480	0.257
Glu	73	0.475	0.459	0.227	0.538	0.482	0.259
Met	75	0.479	0.463	0.230	0.541	0.486	0.262
His	81	0.490	0.474	0.238	0.552	0.497	0.270
Phe	91	0.508	0.492	0.251	0.570	0.515	0.285
Arg	100	0.523	0.506	0.262	0.584	0.530	0.297
Tyr	107	0.534	0.517	0.271	0.595	0.541	0.306
Trp	130	0.566	0.550	0.298	0.626	0.573	0.335

[a] Calculated using AE expressions given in Table 8. [b] X = Br (79.9), P = Cbz (135), and *N*-ethylmorpholine and 1-hydroxybenzotriazole are recovered in the DCC coupling step.

Table 10. **Summary of RME values for various 8-mer oligopeptides of one kind of amino acid**[a]

amino acid	MW R group	Fischer	Bergmann−Zervas	Merrifield[b]	azide method	anhydride method	segment doubling
Gly	1	0.0234	0.0081	0.0126	0.00397	0.101	0.136
Ala	15	0.0266	0.0095	0.0150	0.00462	0.118	0.162
Ser	31	0.0297	0.0109	0.0175	0.00528	0.134	0.190
Val	43	0.0318	0.0119	0.0193	0.00573	0.145	0.209
Thr	45	0.0321	0.0120	0.0196	0.00580	0.147	0.212
Cys	47	0.0324	0.0122	0.0199	0.00587	0.148	0.216
Leu	57	0.0338	0.0129	0.0213	0.00620	0.156	0.231
Ile	57	0.0338	0.0129	0.0213	0.00620	0.156	0.231
Asn	58	0.0340	0.0130	0.0214	0.00623	0.157	0.232
Asp	59	0.0341	0.0130	0.0215	0.00626	0.158	0.234
Lys	72	0.0358	0.0139	0.0232	0.00666	0.167	0.252
Gln	72	0.0358	0.0139	0.0232	0.00666	0.167	0.252
Glu	73	0.0359	0.0139	0.0234	0.00669	0.168	0.253
Met	75	0.0362	0.0140	0.0236	0.00674	0.169	0.256
His	81	0.0369	0.0144	0.0243	0.00691	0.173	0.264
Phe	91	0.0379	0.0150	0.0255	0.00718	0.180	0.277
Arg	100	0.0389	0.0155	0.0266	0.00740	0.185	0.288
Tyr	107	0.0395	0.0156	0.0273	0.00757	0.189	0.297
Trp	130	0.0415	0.0169	0.0297	0.00807	0.201	0.323

[a] Calculated using RME expressions given in Table 8. See Supporting Information Part 3 for references to reaction yields for each method. [b] X = Br (79.9), P = Cbz (135), and *N*-ethylmorpholine and 1-hydroxybenzotriazole are recovered in the DCC coupling step.

where

$$S = \frac{44}{\epsilon_7^*\epsilon_8} + (28+6)\left(\frac{1}{\epsilon_1\epsilon_7^*\epsilon_8} + \frac{3/2}{\epsilon_1\epsilon_2\ldots\epsilon_8}\right) + \frac{161.8}{\epsilon_6^{**}\epsilon_7\epsilon_8} + \frac{28}{\epsilon_7\epsilon_8}\left(\frac{2}{\epsilon_6^{**}} + \frac{2}{\epsilon_5^*\epsilon_6^*} + \frac{2}{\epsilon_5^*\epsilon_6} + \frac{1}{\epsilon_2^*\epsilon_3\epsilon_4\epsilon_5\epsilon_6}\right) + \frac{46}{\epsilon_6^*\epsilon_7\epsilon_8} + \frac{36}{\epsilon_5^*\epsilon_7\epsilon_8}\left(\frac{1}{\epsilon_6^*} + \frac{1}{\epsilon_6}\right) + \frac{70.9}{\epsilon_6\epsilon_7\epsilon_8} + \frac{65}{\epsilon_5\epsilon_6\epsilon_7\epsilon_8} + \frac{56}{\epsilon_4\epsilon_5\epsilon_6\epsilon_7\epsilon_8} + \frac{52.45}{\epsilon_2^*\epsilon_3\epsilon_4\epsilon_5\epsilon_6\epsilon_7\epsilon_8} + \frac{168}{\epsilon_2\epsilon_3\epsilon_4\epsilon_5\epsilon_6\epsilon_7\epsilon_8} \quad (47b)$$

Using the reaction yield parameters given in Figure 12 eqs 47a,b reduce to

$$(\mathrm{RME})_{\mathrm{overall}} = \frac{184}{2225.19} = 0.0827 \quad (48)$$

and the overall atom economy is

$$(\mathrm{AE})_{\mathrm{overall}} = \frac{184}{1017.15} = 0.1809 \quad (49)$$

When a common intermediate or starting material is required in more than one place in a synthesis plan, its mass partitioning ratio for each branch may be directly determined from the reaction scales at the relevant nodes in the synthesis tree. For example, the intermediate ammonia (I_1 and I_6***) is required in two branches in Figure 12, one to produce nitric acid in the second stage and the other to produce urea in the seventh stage. Using reaction yields given in the Supporting Information the corresponding ammonia mass partitioning ratio is

$$\frac{(\mathrm{mass\ NH_3})_{\mathrm{2nd\ stage}}}{(\mathrm{mass\ NH_3})_{\mathrm{7th\ stage}}} = \frac{\dfrac{(3/2)(2)(17)}{\epsilon_8\epsilon_7\epsilon_6\epsilon_5\epsilon_4\epsilon_3\epsilon_2}}{\dfrac{(2)(17)}{\epsilon_8\epsilon_7^*}} = \frac{(3/2)\epsilon_7^*}{\epsilon_7\epsilon_6\epsilon_5\epsilon_4\epsilon_3\epsilon_2} = 2.59 \quad (50)$$

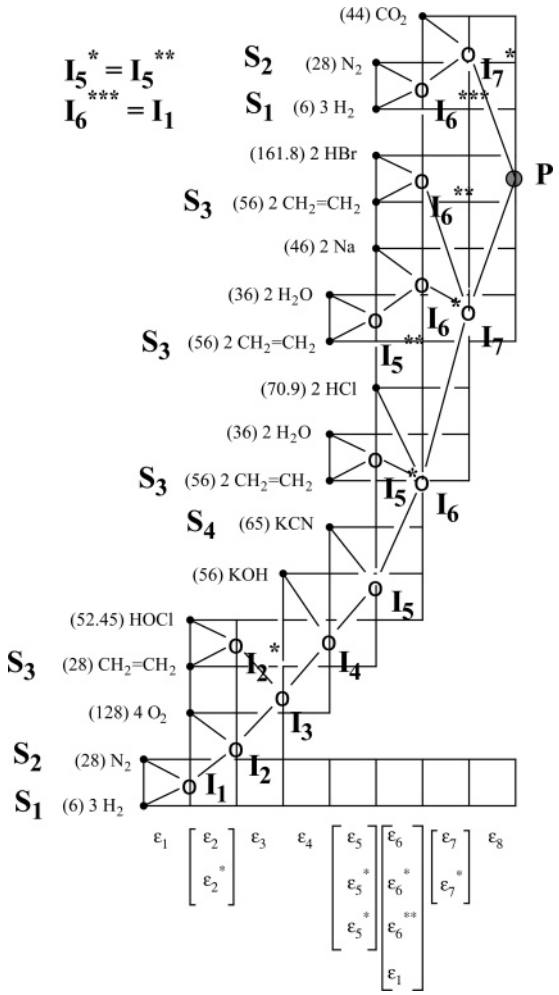


Figure 12. **Synthesis tree for the synthesis of veronal according to reactions given in Scheme 14. Synthesis parameters: 18 inputs, 14 intermediates, 8 reaction stages, 15 reactions, 11 parallel reactions. Synthesis type: mixed linear and multi-convergent. Molecular weights in grams per mole for input reactant and final product output nodes are given in parentheses. Reaction yields: $\epsilon_1 = 0.66$, $\epsilon_2 = 1$, $\epsilon_2^* = 0.88$, $\epsilon_3 = 0.84$, $\epsilon_4 = 1$, $\epsilon_5 = 0.85$, $\epsilon_5^* = 1$, $\epsilon_6 = 0.83$, $\epsilon_6^* = 1$, $\epsilon_6^{**} = 1$, $\epsilon_7 = 0.84$, $\epsilon_7^* = 0.80$, $\epsilon_8 = 0.75$. (See Supporting Information, Part 3, for references to reaction yields.)**

or 72% directed toward the second stage and 28% directed toward the seventh stage. Similarly, the mass partitioning ratio of ethylene starting material required in stages 2, 5, and 6 to produce ethylene chlorohydrin (EC), ethanol leading to diethylmalonate (DEM), ethanol leading to sodium ethoxide (SE), and ethylbromide (EB) is

$$(CH_2=CH_2)^{EC}_{\text{2nd stage}}:(CH_2=CH_2)^{DEM}_{\text{5th stage}}:(CH_2=CH_2)^{SE}_{\text{5th stage}}:(CH_2=CH_2)^{EB}_{\text{6th stage}}$$

$$= \frac{28}{\epsilon_8\epsilon_7\epsilon_6\epsilon_5\epsilon_4\epsilon_3\epsilon_2^*}:\frac{2(28)}{\epsilon_8\epsilon_7\epsilon_6\epsilon_5^*}:\frac{2(28)}{\epsilon_8\epsilon_7\epsilon_6^*\epsilon_5^*}:\frac{2(28)}{\epsilon_8\epsilon_7\epsilon_6^{**}}$$

$$= \frac{28}{0.3285}:\frac{2(28)}{0.4968}:\frac{2(28)}{0.5985}:\frac{2(28)}{0.5985}$$

$$= 85.24:112.72:93.57:93.57$$

$$= 22.1\%:29.3\%:24.3\%:24.3\% \quad (51)$$

The molecular weight profile shown in Figure 13 indicates

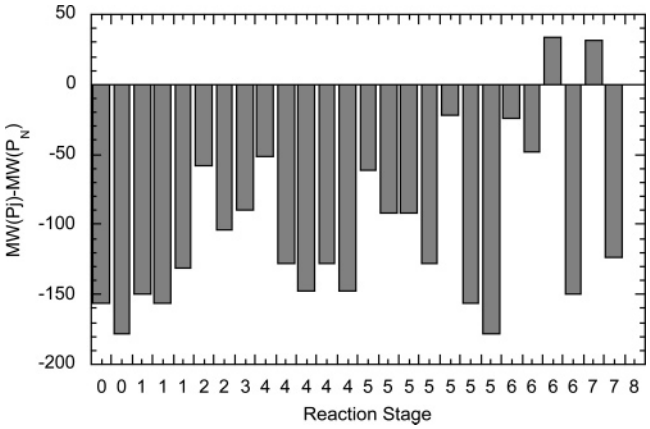


Figure 13. **Veronal synthesis molecular weight profile based in Scheme 14 and Figure 12 for molecular weights of substrate starting materials and intermediate products relative to molecular weight of veronal. MW(P_N) = 184 g/mol; $\mu_1 = -292.84$ g/mol per reaction stage.**

that the plan involves a building up of about 300 g/mol per reaction stage. This high value is indicative of the fact that small molecular weight materials are used as starting materials. In terms of the shape of the synthesis plan its degree of convergence is high at 53%, its asymmetry index is 0.591, and its rate of convergence is 1.7 reaction inputs per reaction stage.

Scheme 15 shows a divergent reaction network which may represent a repertoire of chemical products that a particular chemical company may manufacture from phthalic anhydride. Following the same procedure as before it is possible to construct synthesis trees for each product as shown in Figure 14 in order to determine, for example, the most material efficient route, the amount of starting phthalic anhydride (PA) required to produce target masses of all products from a single batch, and for the key intermediates, phthalimide (I_6) and phthalide (I_4), the respective mass branching ratios leading to the sets of products P_5 and P_6, and P_4 and P_7, respectively. The kernel RME values for products in descending order are anthraquinone (AQ) (73%), benzoic acid (BA) (62%), phthalic acid (PA2) (45%), *N*-(2-bromoethyl)phthalimide (NBEP) (37%), monoperphthalic acid (MPA) (30%), homophthalic acid (HPA) (7%), and phthalaldehydic acid (PAA) (5%). If 1 kg of each product is to be synthesized according to this network, then the mass of starting phthalic anhydride required is given by

$$(\text{mass})_{\text{PA}} = \text{MW}_{\text{PA}}\left[\frac{x_{\text{BA}}}{0.85} + \frac{x_{\text{MPA}}}{(0.65)(1)} + \frac{x_{\text{AQ}}}{0.95} + \frac{x_{\text{HPA}}}{(0.92)(0.72)(0.78)(0.67)(0.71)} + \frac{x_{\text{PA2}}}{(0.95)(0.96)} + \frac{x_{\text{NBEP}}}{(0.95)(0.8)(0.69)} + \frac{x_{\text{PAA}}}{(0.92)(0.72)(0.78)(0.82)(0.65)}\right] \quad (52)$$

where the *x* variables represent the mole scales of each target product equal to $\text{mass}_{\text{product}}/\text{MW}_{\text{product}}$. Since 1000 g are required for each product, eq 52 reduces to 12.44 kg of required PA when the respective molecular weights of each product are substituted. The mass partition ratio for producing

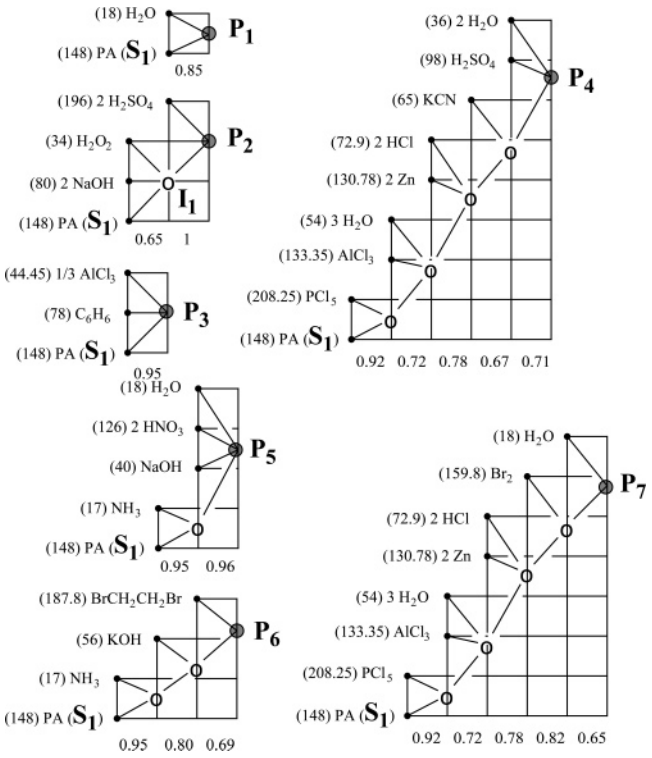


Figure 14. **Synthesis trees pertaining to the divergent synthesis network for phthalic anhydride given in Scheme 15. Reaction yields for each step are given along the *x*-axis of each tree diagram.**

homophthalic acid (P_4) and phthalaldehydic acid (P_7) from phthalide is

$$\frac{\text{mass}_{P4}}{\text{mass}_{P7}} = \frac{\dfrac{180x_{P4}}{(0.71)(0.67)}}{\dfrac{150x_{P7}}{(0.65)(0.82)}} = 1.345\left(\frac{x_{P4}}{x_{P7}}\right) \tag{53}$$

where x_{P4} and x_{P7} represent the mole scales of the target products. Similarly, for producing phthalic acid (P_5) and *N*-(2-bromoethyl)-phthalimide (P_6) from phthalimide the mass partition ratio is

$$\frac{\text{mass}_{P5}}{\text{mass}_{P6}} = \frac{\dfrac{166x_{P5}}{(0.96)}}{\dfrac{253.9x_{P6}}{(0.69)(0.8)}} = 0.376\left(\frac{x_{P5}}{x_{P6}}\right) \tag{54}$$

6. Conclusions

A new methodology based on graphical tree analysis is introduced to determine kernel green metrics and important tree parameters that quantitatively characterize synthetic plans and networks. Using a variety of literature examples it has been shown that the method is robust as it can handle synthesis plans or networks of any degree of complexity. Raw materials, input energy, and cost throughput efficiencies are readily evaluated using a simple "connect-the-dots" approach. Reaction mass efficiency incorporating number of steps, reaction yields, and individual reaction atom economies is the best descriptor of intrinsic material efficiency of a synthetic plan. For reaction optimization the number of steps and magnitude of reaction yields are generally the strongest controllers of the magnitude of the overall RME. The overall kernel RME is a key metric from which the overall kernel minimum raw material cost (RMC) is directly determined. The fraction of total input energy directed to product (FTE) is found to be a weighted average of reaction energy inputs where the weights are the kernel reaction mass efficiencies for individual reactions. Histograms depicting kernel metrics as functions of reaction stage or step give visual descriptions of the performance of a given synthesis plan. Minimum waste and minimum RMC determining reaction stages and steps may be easily identified so as to screen out poorer performing plans in the early stages of synthesis development. These criteria can greatly facilitate efforts to optimize synthesis plans to important targets.

As is already the common practice, it is advisable to begin with cheapest materials that structurally resemble as closely as possible the intended target. Failing a feasible true single-stage simultaneous or sequential multicomponent reaction with all starting materials in hand, this means that linear sequences to get to the target must be as short as possible. The strategy is to devise a retrosynthetic plan that divides the target structure into components that each can be made by their own linear sequences in a parallel fashion and that these are brought together in a convergent step in the late stages of the plan in the form of a multicomponent reaction. This results in a convergent synthesis with a corresponding high degee of convergence. The strength of the tree analysis method is in the clear and standardized depiction of synthesis plans from which several metrics gauging the efficiency of production of a given target product can be deduced without recourse to lengthy computations. Moreover, different plans to the same target may be compared and analyzed critically. Advantages and disadvantages for competing plans may be quantified precisely and unambiguously. Such parametrizations reveal strengths and weaknesses of a given plan and serve to point out how it can be further optimized, or else abandoned for a completely different plan. It is imperative to do a thorough analysis of all kernel metrics and tree parameters to avoid the pitfall of improperly characterizing a given synthesis plan. All of this serves to improve resource management in the synthesis of important chemical commodities.

Since the synthesis tree approach essentially chronicles the manufacturing origin of each chemical in a sort of genealogical progression, it raises the important question of how far back one should go with a metrics determination with starting progenitor compounds.[41] Environmentally conscientious firms may also consider how companies dispose of their generated wastes before purchasing reagents and intermediates. It is difficult to assign numerical values to these considerations. In principle one should go all the way back to the feedstocks from the petroleum and bulk chemicals industries to do a complete analysis as far as possible. However, a second more complex corollary question is where

(41) These points were brought to my attention by John Kindervater, Environmental Consultant for Eli Lilly and Company at the Canada-US Joint Workshop on Innovative Chemistry in Clean Media held in Montreal, Canada, May 20−21, 2004.

to draw the line of responsibility for checking to see if a purchased starting material was also produced in a "green" way. In effect there is an "inheritance" of production built into a total metrics analysis. This is problematic when buying materials from suppliers who may not wish to disclose their manufacturing processes or synthetic plans for legitimate proprietary reasons.

It is advocated that at least at the level of all purchased starting materials, the reporting of kernel green metrics and characteristic synthesis plan parameters be included along with yield per step and full physical and spectroscopic data for all intermediate structures, as part of the general practice and protocol for reporting on the total syntheses of new chemical targets in the literature, particularly if a claim of "greenness" is made.

A software adaptation of this tree analysis to total synthesis design and optimization is currently being developed.

Acknowledgment

Dr. Floyd H. Dean is thanked for helpful suggestions and stimulating discussions, particularly for drawing my attention to industrial syntheses of papaverine. Ms. Joy McCourt is thanked for carrying out initial literature searches on and analysis of the synthesis of triclosan.

Supporting Information Available

Full chemical synthesis plans and synthesis trees for the synthesis of quinine, sildenafil, absinthin, papaverine, and bupleurynol; various polypeptide methodologies; and synthesis trees for veronal and phthalic anhydride synthesis networks. Detailed calculations of minimum RMC and molecular weight first moments per reaction stage are also given for each plan. This information is available free of charge via the Internet at http://www.pubs.acs.org.

List of Symbols and Definitions

α	number of cycles in segment doubling method for polypeptide synthesis
a_{j+1}	ordinate of $(j + 1)$th reactant input in synthesis plan
AE	atom economy
BI	branching index
β	asymmetry parameter
c_j	mass of catalyst for jth reaction
δ	degree of convergence relative to single step MCR
d_i	degree of vertex i in graph
D_{ij}	distance between vertex i and vertex j
$\$_j$	unit cost of the jth input reactant on a per gram basis
ϵ_j	reaction yield for jth reaction step
$\epsilon_{\text{pseudo-overall}}$	pseudo-overall reaction yield
E_{m}	environmental impact factor based on mass
E_{mw}	environmental impact factor based on molecular weight
FTE	fraction of total energy input
Φ_{product}	fraction of total energy input that is directed to forming product
Φ_{waste}	fraction of total energy input that is directed to forming waste
ϕ_j	mass of excess reagent in jth reaction
G	number of stages with parallel reactions
H	Hendrickson convergence parameter
I	number of reactant input structures in a synthesis plan
ℓ	number of parallel linear sequences
L	number of parallel reactions
μ_1	first molecular weight moment
M	number of reaction steps in a synthesis plan
MCR	multicomponent reaction
m	polypeptide chain length
m_{P}	mass of target product
MRP	materials recovery parameter
MW(P_j)	molecular weight of jth intermediate product in synthesis plan
MW(P_N)	molecular weight of target product in synthesis plan
n	number of points arranged vertically that correspond to reactant input structures from which centroids of intermediate products are calculated
N	number of reaction stages in a synthesis plan
p	molecular weight of target product P in synthesis plan
P	coordinates of target product in synthesis plan
P	molecular weight of protecting group used in Merrifield polypeptide synthesis
P_{mcr}	coordinates of target product in a single step MCR
Ψ	total input energy in kWh consumed in synthesis plan
Ψ_{product}	total input energy in kWh consumed in synthesis plan that is directed toward forming product
Ψ_{waste}	total input energy in kWh consumed in synthesis plan that is directed toward forming waste
Ψ_j	total input energy in kWh for jth reaction in synthesis plan
$\Psi_{\text{product},j}$	input energy in kWh for jth reaction that is directed toward forming product
$\Psi_{\text{waste},j}$	input energy in kWh for jth reaction that is directed toward forming waste
ρ_{rel}	relative rate of convergence
R	molecular weight of R groups in amino acids
r_j	molecular weight of jth reactant in synthesis plan
RMC	raw material cost
RME	reaction mass efficiency
s_j	mass of solvent for jth reaction
SF	stoichiometric factor
t_j	reaction time for jth reaction
T_{total}	total optimal reaction time for entire synthesis
θ_{P}	angle subtended at point P in synthesis tree
θ_{mer}	angle subtended at point P_{mcr} in synthesis tree
υ	total number of vertexes in graph
W	Wiener index
$\bar{w}_j$	mass of waste in jth reaction step

x	number of moles of target product in synthesis plan (defines target scale of P)
X	molecular weight of X group in Merrifield polypeptide synthesis
^{1}X	Randic branching index
$\langle y \rangle$	ordinate of point P in synthesis plan
ω_j	mass of all postreaction materials in workup and purification phases for jth reaction

Received for review October 6, 2005.

OP0501904