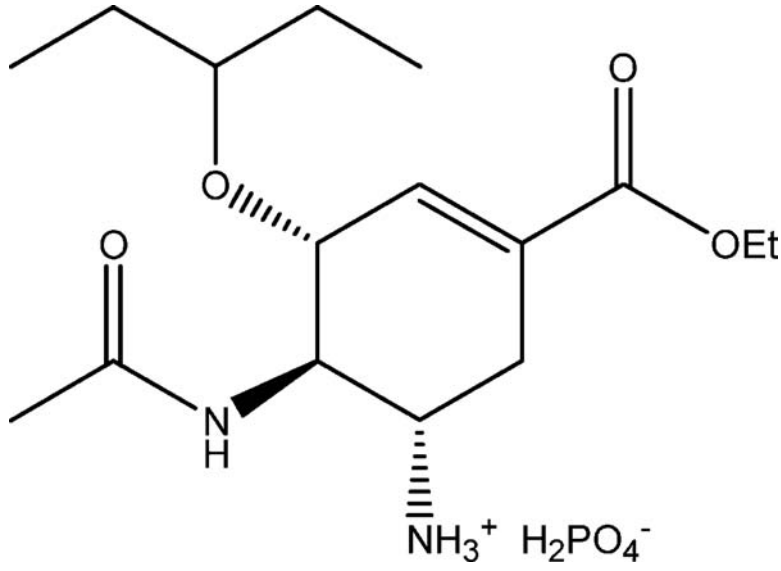
Organic Process Research & Development

Article

Global Green Chemistry Metrics Analysis Algorithm and Spreadsheets: Evaluation of the Material Efficiency Performances of Synthesis Plans for Oseltamivir Phosphate (Tamiflu) as a Test Case

John Andraos

Org. Process Res. Dev., **2009**, 13 (2), 161-185• DOI: 10.1021/op800157z • Publication Date (Web): 18 December 2008



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Organic Process Research & Development is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Organic Process Research & Development **2009**, *13,* 161–185

Global Green Chemistry Metrics Analysis Algorithm and Spreadsheets: Evaluation of the Material Efficiency Performances of Synthesis Plans for Oseltamivir Phosphate (Tamiflu) as a Test Case

John Andraos*

Department of Chemistry, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada

Abstract:

This work discloses an easy-to-use algorithm to evaluate the global material efficiency performance of any kind of synthesis plan regardless of complexity to a given target molecule according to green metrics criteria. The algorithm is robust and has been adapted to Excel spreadsheets for rapid calculation and graphing of the numerical results. In order to demonstrate the facile utility of this exceptional tool for process and synthetic chemists in the evaluation and ranking of synthetic performance, various synthesis plans for oseltamivir phosphate, a neuraminidase inhibitor used to treat the H5N1 influenza virus, have been investigated. In particular, six industrial syntheses and nine plans from academic groups have been thoroughly and rigorously evaluated according to kernel and global reaction mass efficiencies and *E*-factors, atom economy, and overall yield performances. In addition, all reported plans were evaluated according to new synthesis elegance parameters including fraction of sacrificial reagents by molecular weight, hypsicity (oxidation level) index, and number of target bonds made per reaction step. Target structure bond maps and profiles are introduced as convenient ways to visually describe synthetic strategy compactly. These powerful algorithms and visual aids can be used to immediately spot bottlenecks in a synthesis plan. Moreover, they allow deeper understanding and critiquing of synthesis plans, thereby assisting chemists in suggesting new directions for further optimization.

1. Introduction

The quest to achieve reliable and high-yielding syntheses of important target molecules for various purposes has been a preoccupation of chemists for nearly three centuries since Lavoisier's pillar discovery of the conservation of mass law for chemical reactions and its practical corollary of balancing any kind of chemical equation so that all reactant masses used and product masses formed are accounted for. Currently, the criterion of producing products with minimal waste material generated has been popularized under the rubrics of green chemistry and sustainable development. Over the last two decades these ideas have evolved initially from altruistic philosophies, then to well-documented real-world practical examples that illustrate the merits of waste reduction and efficient material consumption, and finally to the establishment of rigorous scientific disciplines in their own right that fully describe and account for all material and energy consumption. The emerging field of green metrics[1] is such a discipline that introduces sets of parameters that can be used to measure efficiency of a given process, to rank such a process against others according to some defined scales, and to set "goodness" criteria according to chosen threshold values of such parameters. It is precisely this transition from a qualitative "show and tell" approach that juxtaposes former "bad" plans to new and improved "green" ones on the basis of usually one criterion to a quantitative one that involves a thorough multivariable analysis of optimization. This new approach is currently accelerating the acceptance and respectability of green chemistry as a proper scientifically based discipline in the wider chemistry community.

In the field of synthetic organic chemistry the cornerstone material efficiency green metrics that are used to parametrize individual chemical reactions from the positive point of view (that is, from the point of view of target product produced with respect to input materials used) are reaction yield, atom economy (AE),[2] and reaction mass efficiency (RME).[3] The counter-green metric that describes waste production with respect to target product production is environmental impact factor, or *E*-factor.[4] The fundamental link between these metrics is of course fully balanced chemical equations that show chemical structures of all reactants and all products along with their molecular weights and masses. Recent reports[5] successfully unified these parameters and described succinctly material efficiency for individual chemical reactions. In doing so, it was necessary to define precisely what constitutes waste material. This implied that computations could be conducted under two broad regimes. Metrics determined at the kernel level describe

* Author to whom correspondence may be addressed. E-mail: jandraos@yorku.ca. Fax: 416-736-5936.

(1) Lapkin, A., Constable, D. J. C., Eds. *Green Chemistry Metrics: Measuring and Monitoring Sustainable Processes;* Wiley: Chichester, 2008.

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

(3) (a) Curzons, A. D.; Constable, D. J. C.; Mortimer, D. N.; Cunningham, V. L. *Green Chem.* **2001**, *3*, 1. (b) Constable, D. J. C.; Curzons, A. D.; Freitas dos Santos, L. M.; Geen, G. R.; Hannah, R. E.; Hayler, J. D.; Kitteringham, J.; McGuire, M. A.; Richardson, J. E.; Smith, P.; Webb, R. L.; Yu, M. *Green Chem.* **2001**, *3*, 7. (d) Constable, D. J. C.; Curzons, A. D.; Cunningham, V. L. *Green Chem.* **2002**, *4*, 521. (e) Steinbach, A.; Winkenbach, R. *Chem. Eng*, **2000**, *94*, April.

(4) (a) Sheldon, R. A. *ChemTech* **1994**, *24* (3), 38. (b) Sheldon, R. A. *C. R. Acad. Sci., Ser. IIc Chim.* **2002**, *3*, 541. (c) Sheldon, R. A. *Pure Appl. Chem.* **2001**, *72*, 1233. (d) Sheldon, R. A. *Green Chem.* **2007**, *9*, 1273.

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

10.1021/op800157z CCC: $40.75
Published on Web 12/18/2008

intrinsic chemical performance, and waste under these conditions therefore includes byproducts from the intended target reaction, side products from competing reactions involving the same set of reactants, and unreacted starting materials. Kernel metrics are especially useful in singling out in the early planning stages those reactions that are strategically designed to minimize byproduct formation and are therefore promising candidates for scale-up. Metrics determined at the global level describe, as the name suggests, the material efficiency of *all* of the input material consumed. Waste then includes, in addition to all of the materials itemized at the kernel level, all auxiliary materials such as solvents, ligands, catalysts, and additives used in the reaction operation, wash and extraction solvents used in the workup operation, and solvents, drying materials, and chromatographic materials used in purification operations. Global metrics give a true measure of waste production from all sources, and radial pentagons[6] were advanced as visual aids in determining good and poor performing reactions in synthesis plans, and pointing out which contributing metrics parameters were responsible for attenuations in the individual reaction RME values.

The work on individual reaction analysis was then extended to describe material efficiency for entire synthesis plans[7] which are composed of consecutive reaction steps in linear and/or convergent sequences along single and parallel branches as the case may be. Synthesis trees were introduced as another key visual aid in the analysis of material efficiency for synthesis plans, especially for tracking input materials, reagent stoichiometric coefficients, step count, reaction yields, and points of convergence for convergent plans with multiple branches. At the kernel level the order of importance of metrics that attenuate overall RME performance for a synthesis plan is step count, individual reaction yields, and atom economy. At the global level the strongest attenuator of RME is auxiliary material consumption. For short synthesis plans computations could be done using a handheld calculator; however, it became obvious that they were quite tedious and error prone for linear plans exceeding 10 steps and most convergent plans of any length. In order for widespread implementation of green chemistry practices as routine standard protocols by process and synthetic chemists, it was imperative to develop an easy-to-use spreadsheet format that automates computation of all of the key green metrics calculations described previously and includes internal check calculations that ensure correct results. This powerful tool has now been realized, and the present announcement means that the problem of global material efficiency analysis for individual chemical reactions and synthesis plans is now completely solved. We also address criticisms that final route selection decisions and rankings of plans that are based only on kernel metrics are flawed.

The structure of the present paper is as follows. We begin by disclosing the steps in the algorithm showing the relevant expressions. A global expression for the *E*-factor for synthesis plans is given which can be decomposed into its contributing *E*-factors from reaction byproducts, side products from competing side reactions, and unreacted starting materials (*E*-kernel), excess reagent consumption (*E*-excess), and auxiliary material consumption (*E*-auxiliaries). This expression is shown to be general and applicable to any kind of synthesis plan regardless of its complexity. A set of detailed instructions is provided in the Supporting Information on the use of the Excel spreadsheets including example template Excel files that may be readily used on any plan. This algorithm and associated spreadsheets are demonstrated for the analysis of six industrial and nine academic plans to the neuraminidase inhibitor oseltamivir phosphate marketed as Tamiflu and used to treat the H5N1 influenza virus. This compound was chosen as a model case since it is a modern pharmaceutical of high importance and interest, and its synthesis plans are well documented in the literature. The six industrial plans include the original Gilead route[8] and the five Roche plans: first-generation quinic acid route,[9] second-generation quinic acid route,[10] shikimic acid route,[11] Diels−Alder route,[12] and desymmetrization route.[13] The nine academic routes are the Corey,[14] Fang,[15] Fukuyama,[16] Kann,[17] Okamura−Corey,[18] and Trost[19] plans, and three routes from the Shibasaki group.[20,21] All plans are evaluated and ranked according to the various metrics discussed in prior works. Advantages and disadvantages of each method are fully discussed, and suggestions for future developments are put forward. In addition, all reported plans were

(6) Andraos, J.; Sayed, M. *J. Chem. Educ.* **2007**, *84*, 1004.

(7) (a) Andraos, J. *Org. Process Res. Dev.* **2006**, *10*, 212. (b) Andraos, J.; Izhakova, J. *Chim. Oggi/Int. J. Ind. Chem.* **2006**, *24* (6, Suppl.), 31. (c) Andraos, J. *Can. Chem. News* **2007**, *59* (4), 14. (d) Andraos, J. In *Green Chemistry Metrics: Measuring and Monitoring Sustainable Processes*; Lapkin, A., Constable, D. J. C., Eds.; Wiley: Chichester, 2008; Chapter 4. (e) Andraos, J. In *Handbook of Green Chemistry*; Nelson, W. M., Ed.; Oxford University Press: Oxford. Manuscript submitted.

(8) Rohloff, J. C.; Kent, K. M.; Postich, M. J.; Becker, M. W.; Chapman, H. H.; Kelly, D. E.; Lew, W.; Louie, M. S.; McGee, L. R.; Prisbe, E. J.; Schultze, L. M.; Yu, R. H.; Zhang, L. *J. Org. Chem.* **1998**, *63*, 4545.

(9) Karpf, M.; Trussardi, R. *J. Org. Chem.* **2001**, *66*, 2044.

(10) Federspiel, M.; Fischer, R.; Hennig, M.; Mair, H. J.; Oberhauser, T.; Rimmler, G.; Albiez, T.; Bruhin, J.; Estermann, H.; Gandert, C.; Göckel, V.; Götzö, S.; Hoffmann, U.; Huber, G.; Janatsch, G.; Lauper, S.; Röckel-Stäbler, O.; Trussardi, R.; Zwahlen, A. G. *Org. Process Res. Dev.* **1999**, *3*, 266.

(11) Harrington, P. J.; Brown, J. D.; Foderaro, T.; Hughes, R. C. *Org. Process Res. Dev.* **2004**, *8*, 86.

(12) Abrecht, S.; Karpf, M.; Trussardi, R.; Wirz, B. (F. Hoffmann-La Roche, AG). EP Patent 1127872, 2000.

(13) (a) Zutter, U.; Iding, H.; Wirz, B. (F. Hoffmann-La Roche AG). EP Patent 1146036, 2000. (b) Zutter, U.; Iding, H.; Spurr, P.; Wirz, B. *J. Org. Chem.* **2008**, *73*, 4895.

(14) (a) Yeung, Y. Y.; Hong, S.; Corey, E. J. *J. Am. Chem. Soc.* **2006**, *128*, 6310. (b) For Corey oxazaborolidinium catalyst synthesis, see Corey, E. J.; Shibata, T.; Lee, T. *J. Am. Chem. Soc.* **2002**, *124*, 3808.

(15) Shie, J. J.; Fang, J. M.; Wang, S. Y.; Tsai, K. C.; Cheng, Y. S. E.; Yang, A. S.; Hsiao, S. C.; Su, C. Y.; Wong, C. H. *J. Am. Chem. Soc.* **2007**, *129*, 11892.

(16) (a) Satoh, N.; Akiba, T.; Yokoshima, S.; Fukuyama, T. *Angew. Chem., Int. Ed.* **2007**, *46*, 5734. For MacMillan catalyst synthesis, see: (b) Ahrendt, K. A.; Borths, C. J.; MacMillan, D. W. C. *J. Am. Chem. Soc.* **2000**, *122*, 4243.

(17) (a) Bromfield, K. M.; Gradén, H.; Hagberg, D. P.; Olsson, T.; Kann, N. *Chem. Commun.* **2007**, 3183. (b) Gradén, H.; Hallberg, J.; Kann, N. *J. Comb. Chem.* **2004**, *6*, 783.

(18) Kipassa, N. T.; Okamura, H.; Kina, K.; Hamada, T.; Iwagawa, T. *Org. Lett.* **2008**, *10*, 815.

(19) (a) Trost, B. M.; Zhang, T. *Angew. Chem., Int. Ed.* **2008**, *47*, 3759. (b) For Trost ligand, see Trost, B. M.; Van Vranken, D. L.; Bingel, C. *J. Am. Chem. Soc.* **1992**, *114*, 9327. (c) Swift, G.; Swern, D. *J. Org. Chem.* **1967**, *32*, 511. (d) For Du Bois ligand synthesis, see: Espino, C. G.; Fiori, K. W.; Kim, M.; Du Bois, J. *J. Am. Chem. Soc.* **2004**, *126*, 15378.

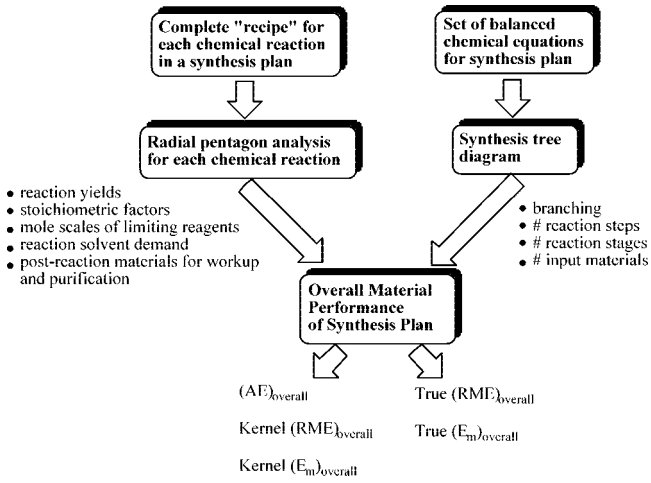


Figure 1. **General paradigm for determining key material efficiency green metrics beginning with reported original literature procedures for each reaction in a synthesis plan with fully balanced chemical equations.**

evaluated according to new synthesis elegance parameters including fraction of sacrificial reagents by molecular weight, hypsicity (oxidation level) index, and number of target bonds made per reaction step. Target structure bond maps and profiles are introduced as convenient ways to visually describe synthetic strategy compactly. These new insights help to link the aims of synthetic chemists who are routinely engaged in inventing new reaction methodologies and developing "elegant" synthesis plans to target molecules with the aims of process chemists who are engaged in developing short efficient plans that minimize waste, hazards, toxicities, and ultimately cost. The endeavors of both groups of chemists are mutually compatible and are clearly in line with the goals of green chemistry; hence, any distinctions made about optimization by these groups of chemists is purely artificial and has nothing to do with actually solving the problem of developing efficient synthesis plans to a target molecule, whether it be a natural product or a compound of commercial importance.

2. Algorithm

Figure 1 shows an overview paradigm flowchart of how a chemist would assess a synthesis plan to a given target molecule from a set of experimental procedures and balanced chemical equations for each and every step in the synthesis. The relevant amounts of materials used in a procedure and stoichiometric coefficients are inputted into the PENTAGON Excel spreadsheet according to the following appropriate categories: reagents, catalysts and ligands, reaction solvents, workup materials, and purification materials. The set of output parameters calculated are reaction scale and reaction yield with respect to limiting reagent, stoichiometric factor, mass of excess reagents, atom economy, kernel and global reaction mass efficiency, and total mass of auxiliary materials. A radial pentagon is drawn to depict the global RME for a given reaction according to eq 1.

$$\mathrm{RME} = (\varepsilon)(\mathrm{AE})\left(\frac{1}{\mathrm{SF}}\right)(\mathrm{MRP}) \tag{1}$$

where ε is the reaction yield with respect to the limiting reagent in a reaction given by the mole ratio of the target product collected and of the limiting reagent; AE is the reaction atom economy given by the ratio of the molecular weight of the target product to the sum of molecular weights of all reagents in the balanced chemical equation; SF is the stoichiometric factor, taking into account the use of excess reagents, and is given by

$$\mathrm{SF} = 1 + \frac{\sum \text{excess masses of all reagents}}{\sum \text{stoichiometric masses of all reagents}} \tag{2}$$

MRP is the material recovery parameter (taking into account the use of auxiliary materials) and is given by

$$\mathrm{MRP} = \frac{1}{1 + \dfrac{(\varepsilon)(\mathrm{AE})(c + s + \omega)}{m_{\mathrm{p}}(\mathrm{SF})}} \tag{3}$$

where m_{p} is the mass of the collected target product, c is the mass of catalyst used, s is the mass of reaction solvent, and ω is the mass of all other auxiliary materials used in the workup and purification phases of the reaction. The spreadsheet has two check calculations, one for the global RME using eq 1 and a simple quotient of the mass of target product collected versus the sum of all input material masses, and the other for the stoichiometric factor using eq 2 and the ratio $(\varepsilon)(\mathrm{AE})/\mathrm{RME}_{\mathrm{kernel}}$. In addition, intermediate RME values between the kernel (maximum RME) and global (minimum RME) values may be computed which would correspond to various scenario cases, depending on what materials may be selected for retrieval. These effects on RME performance are automatically graphed on the pentagon to visually assess the magnitude of the resultant gain in RME value. This process is repeated for each reaction step.

The next task is to construct a synthesis tree according to the convention previously described[7] to determine the number of branches, the number of reaction steps and stages, and the number of input materials. The data from the synthesis tree and the radial pentagons are then inputted into the LINEAR-kernel, LINEAR-complete, CONVERGENT-kernel, or CONVERGENT-complete spreadsheets depending on the type of synthesis plan. For ease of use, these spreadsheets are set up so that the input variables are entered in bold-faced cell boxes that are clearly labeled with appropriate headings. All other labeled columns have cells with embedded formulas already encoded so the user need not concern themselves with the mechanics of computation. The Supporting Information contains example template Excel files mentioned above, including a detailed set of instructions for their use particularly if rows are to be added

(20) (a) Fukuta, Y.; Mita, T.; Fukuda, N.; Kanai, M.; Shibasaki, M. *J. Am. Chem. Soc.* **2006**, *128*, 6312. (b) Mita, T.; Fukada, N.; Roca, F. X.; Kania, M.; Shibasaki, M. *Org. Lett.* **2007**, *9*, 259. (c) Yamatsugu, K.; Kamijo, S.; Suto, Y.; Kanai, M.; Shibasaki, M. *Tetrahedron Lett.* **2007**, *48*, 1403. (d) Personal communication from Prof. M. Shibasaki, June 2008.

(21) For Shibasaki ligand synthesis, see: (a) Mita, T.; Fujimori, I.; Wada, R.; Wen, J.; Kanai, M.; Shibasaki, M. *J. Am. Chem. Soc.* **2005**, *127*, 11252. (b) Matsumoto, S.; Yabu, K.; Kanai, M.; Shibasaki, M. *Tetrahedron Lett.* **2002**, *43*, 2919. (c) Hamashima, Y.; Kanai, M.; Shibasaki, M. *J. Am. Chem. Soc.* **2000**, *122*, 7412. (d) Banaag, A. R.; Tius, M. A. *J. Am. Chem. Soc.* **2007**, *129*, 5328.

or subtracted as needed and if more than two branches are required for convergent plans. The LINEAR spreadsheets are defaulted to a 10-step plan, and the CONVERGENT spreadsheets are defaulted to a plan with two 10-step branches. The synthesis performance spreadsheets have check calculations for the balancing of all chemical equations, kernel RME and *E*-factor metrics, molecular weight first-moment building-up parameter, and hypsicity or oxidation level index.[1]

The algorithm for the computation of overall or global AE, RME, and *E*-factor for any synthesis plan of any degree of complexity is given by the following sequence of computations beginning with the last step ($j = N$) and working toward the first step ($j = 1$) using the synthesis tree connectivities as a guide:

(1) Set the basis mole scale of the target product, *x*, to be 1 mol.

(2) Enter the number of branches (*b*), reaction stages (*N*), number of reaction steps (*M*), number of input materials (*I*), and molecular weight of final target molecule (p_n).

(3) For each branch in the plan enter the sum of the molecular weights for reagents and byproducts in each step, and the sequence of molecular weights of all intermediate products.

(4) For each step in each branch enter the following parameters from the radial pentagon analysis: mass of excess reagents, SF, reaction yield, mass of reaction solvent, total mass of workup materials, total mass of purification materials, mass of catalyst and ligand, and mole scale of limiting reagent as reported in the literature procedure.

(5) Determine the *E*-factor based on molecular weights, $(E_{mw})_j$, using

$$(E_{mw})_j = \frac{(\sum \mathrm{MW}_{byproducts})}{p_j} \tag{4}$$

(6) Determine $(AE)_j$ using

$$(AE)_j = \frac{1}{1 + (E_{mw})_j} \tag{5}$$

(7) Determine the reaction scales for each reagent as prescribed by the synthesis tree using

$$\frac{x}{\prod_k^{n \to j} \varepsilon_k} \tag{6}$$

(8) Determine kernel RME and *E*-kernel (byproducts and unreacted starting materials contribution) for any given step using $(RME)_{kernel,j} = (AE)_j \varepsilon_j$ and

$$(E_{kernel})_j = \frac{1}{(RME)_{kernel,\,j}} - 1 \tag{7}$$

(9) Determine the kernel mass of reagents used in any given step using

$$\left(\frac{x}{\prod_k^{n \to j} \varepsilon_k}\right)\left(\sum \mathrm{MW}_{reagents}\right) \tag{8}$$

(10) Determine kernel mass of waste in any given step due to byproducts and unreacted starting materials contribution using

$$\bar{w}_{kernel,j} = \frac{p_j}{(AE)_j}\left(\frac{x}{\prod_k^{n \to j} \varepsilon_k}\right)[1 - \varepsilon_j (AE)_j] \tag{9}$$

(11) Determine mass of waste due to excess reagents in any given step using

$$\bar{w}_{excess,j} = \frac{p_j}{(AE)_j}\left(\frac{x}{\prod_k^{n \to j} \varepsilon_k}\right)[(SF)_j - 1] \tag{10}$$

(12) Determine mass of waste due to auxiliary materials with appropriate scale correction factor for any given step using

$$\bar{w}_{auxiliary,\,j} = \left(\frac{x}{\prod_k^{n \to j} \varepsilon_k}\right)\left(\frac{c_j + s_j + \omega_j}{x_j^*}\right) \tag{11}$$

(13) Determine the total mass of waste in any given step using

$$\begin{aligned}\bar{w}_{total,\,j} &= \bar{w}_{kernel,\,j} + \bar{w}_{excess\,,\,j} + \bar{w}_{auxiliary,\,j} \\ &= \left(\frac{p_j}{(AE)_j}\right)\left(\frac{x}{\prod_k^{n \to j} \varepsilon_k}\right)[(SF)_j - \varepsilon_j (AE)_j] + \left(\frac{x}{\prod_k^{n \to j} \varepsilon_k}\right) \times \left(\frac{c_j + s_j + \omega_j}{x_j^*}\right)\end{aligned} \tag{12}$$

For the overall performance of a synthesis plan:

(14) Determine $(E_{mw})_{overall}$ by summing molecular weights of all byproducts in the plan and dividing by the molecular weight of the final target product.

(15) Determine $(AE)_{overall}$ using

$$(AE)_{overall} = \frac{1}{1 + (E_{mw})_{overall}} \tag{13}$$

(16) Determine $\bar{w}_{total} = \sum_j^n \bar{w}_{total,j}$ by summing all terms found in step 13.

(17) Determine overall *E*-factor using

$$E_{total} = \frac{1}{p_n}\sum_j \left(\frac{1}{\prod_k^{n \to j} \varepsilon_k}\right)\left[\left(\frac{p_j}{(AE)_j}\right)[(SF)_j - \varepsilon_j (AE)_j] + \frac{c_j + s_j + \omega_j}{x_j^*}\right] \tag{14}$$

(18) Determine overall RME using

$$(\mathrm{RME})_{\mathrm{total}} = \frac{1}{1 + E_{\mathrm{total}}} \tag{15}$$

(19) Determine overall yield by multiplying reaction yields along longest branch of synthesis tree.

Symbol definitions:

x is moles of target product in the synthesis plan.

$\prod_{k}^{n \to j} \varepsilon_k$ is the multiplicative chain of reaction yields connecting the target product node to the reactant nodes for step j as per synthesis tree diagram read from right to left (i.e., in the direction $n \to j$).

p_j is the molecular weight of product of step j.

ε_j is the reaction yield with respect to limiting reagent for step j.

$(\mathrm{AE})_j$ is the atom economy for step j.

$(\mathrm{SF})_j$ is the stoichiometric factor for step j that accounts for all excess reagents used in that step.

$c_j + s_j + \omega_j$ is the sum of masses of auxiliary materials used in step j, namely the mass of catalyst, mass of reaction solvent, and mass of all other post-reaction materials used in the workup and purification phases.

x_j^* is the experimental mole scale of limiting reagent in step j as reported in an experimental procedure.

The summations run over the n reaction steps. A step begins with an isolated intermediate and ends with the following next isolated intermediate. The algorithm essentially determines the true mass throughput from all input material masses to the final mass of target product, assuming that all product material collected in any reaction is entirely committed as a reagent in the next subsequent reaction step. The strategy of carrying out the computation in the reverse sense so that the mole scales of all input reagents and intermediates are normalized relative to the mole scale of the target product greatly facilitates the computation and avoids potential errors due to matching of scales of convergent branches if the computation were to be done in the forward sense. This observation is due to the fact that the target product node is the only node that is common to all branches in the synthesis tree, and so it becomes obvious to use the mole scale of the target product as the reference scale for the entire plan. The value of a synthesis tree diagram is readily apparent in the determination of the mole scale at any node since it is obtained directly by following the connectivity path between that node and the target product node and noting the corresponding chain of reaction yield values linking the two nodes.

The overall E-factor given by eq 14 in step 17a may be partitioned into its three components as shown explicitly below:

$$E_{\mathrm{total}} = E_{\mathrm{byproducts\&unreactedreagents}} + E_{\mathrm{excessreagents}} + E_{\mathrm{auxiliaries}} \tag{16}$$

where

$$E_{\mathrm{byproducts\&unreacted\ reagents}} = E_{\mathrm{kernel}} = \frac{1}{p_n} \sum_j \left(\frac{1}{\prod_k^{n \to j} \varepsilon_k} \right) \left(\frac{p_j}{(\mathrm{AE})_j} \right) [1 - \varepsilon_j (\mathrm{AE})_j] \tag{17a}$$

$$E_{\mathrm{excessreagents}} = \frac{1}{p_n} \sum_j \left(\frac{1}{\prod_k^{n \to j} \varepsilon_k} \right) \left(\frac{p_j}{(\mathrm{AE})_j} \right) [(\mathrm{SF})_j - 1] \tag{17b}$$

$$E_{\mathrm{auxiliaries}} = \frac{1}{p_n} \sum_j \left(\frac{1}{\prod_k^{n \to j} \varepsilon_k} \right) \left(\frac{c_j + s_j + \omega_j}{x_j^*} \right) \tag{17c}$$

This formula is useful in determining the precise distribution of waste components for the entire plan. It should be noted that since auxiliary materials represent the bulk of the mass of materials used in any given reaction it is evident that, even before doing any computations, the relative contributors to E-total are E-auxiliaries $\gg$ E-excess $>$ E-kernel. The above universal expression for the E-factor may be compared with the more complex case-by-case formulas for mass intensity (MI) presented in a recent report[22] that were derived by a forward sense computation of various synthesis plan types. These were based on a rederivation of prior work[5,7] using a "stoichiometric ratio" parameter instead of the more natural "stoichiometric difference" parameter to depict excess reagent consumption. Expressions for MI are immediately related to the E-factor and RME through the simple expression given in eq 18.

$$\mathrm{MI} = \frac{1}{\mathrm{RME}} = E + 1 \tag{18}$$

Equation 18 is true for any set of conditions common to all three variables.

Four other metrics that are also useful in ranking the efficiency of synthesis plans according to design and strategy elements are molecular weight first-moment building-up parameter, μ_1, fraction of sacrificial reagents by molecular weight, f(sac), hypsicity or oxidation level index (HI), and number of target bonds made per reaction step (B/M). The molecular weight first moment[7] is determined from eq 19.

(22) Auge, J. *Green Chem.* **2008**, *10*, 225. A recent note (Andraos, J. *Green Chem.* **2008**, submitted) has clarified a key error in this paper with respect to the description of a variable that is related to the excess mass of reagent consumption that would otherwise invalidate the formulas presented there. Other errors in the implementation of the author's algorithm on his chosen illustrative example were also corrected in this note. The expression for the E-factor in eqs 17a−c given in the present work is simpler as it is completely general and does not require separate case-by-case formulas that depend on the type of synthesis plan.

$$\mu_1 = \frac{\sum_j [\mathrm{MW}_{\text{intermediate product, } j} - \mathrm{MW}_{\text{target product}}]}{N+1} \quad (19)$$

This parameter essentially tracks the difference in molecular weight between any given synthesis intermediate and the target product over the course of a synthesis plan. Good syntheses are characterized by large negative μ_1 values since they use low molecular weight starting materials that are then used to produce intermediate products that in turn progressively increase in molecular weight with minimal overshoots above the target product molecular weight until the final target product is reached.

The fraction of sacrificial reagents by molecular weight tracks those reagents whose atoms never get incorporated in the final target structure and is found from eq 20.

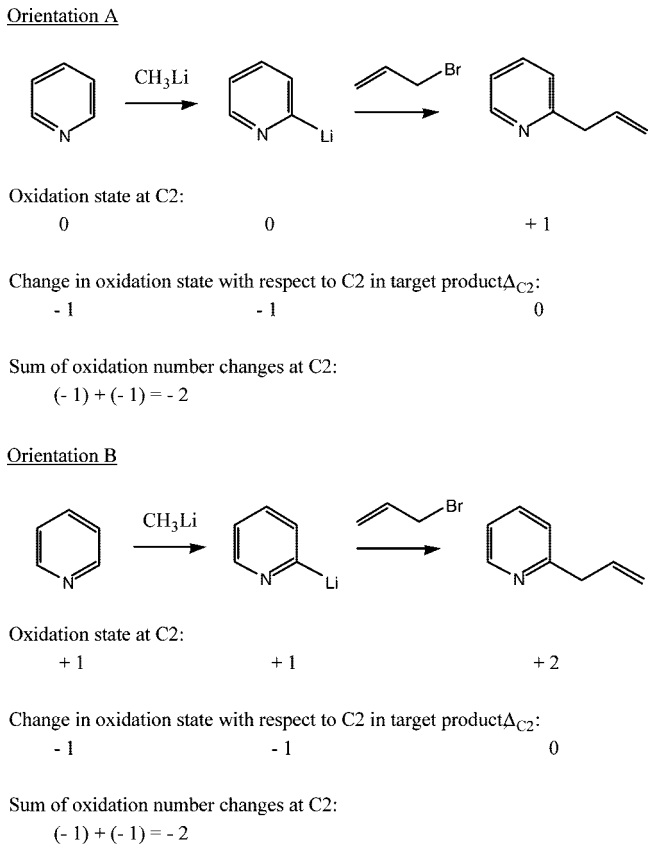
$$f(\mathrm{sac}) = 1 - \frac{\sum \mathrm{MW}_{\text{reagents in whole or in part ending up in target product}}}{\sum \mathrm{MW}_{\text{all reagents}}} = 1 - \frac{(\mathrm{AE})_{\text{overall}} \sum \mathrm{MW}_{\text{reagents in whole or in part ending up in target product}}}{\mathrm{MW}_{\text{target product}}} \quad (20)$$

This parameter is more probing than atom economy in that it is intimately connected to the dissection of a target structure that traces the origin of each atom back to its corresponding starting materials. This target bond-forming mapping exercise as will be illustrated later is very powerful in determining the number of target bonds made, the location of such bonds in the target structure, and at what reaction step they were made. Essentially, atoms in the target structure are grouped in subsets, and these are correlated with the sets of atoms in the starting material structures. This method allows by a simple process of elimination the identification of those reagents used whose atoms never get incorporated into the target structure and are therefore considered sacrificial in their purpose. Sacrificial reagents include those that serve as protecting groups, those that change the electronic states of key atoms (usually via redox reactions so that skeletal building bond-forming reactions are possible), those that are used to control stereochemistry such as chiral auxiliary groups that impart spatial discrimination in the attack of reagents on a structure, those that are used in substitution reactions to switch poor leaving groups into better ones, and those that are reducing or oxidizing agents that are used in *subtractive* redox reactions, that is, reactions in which oxygen atoms or hydrogen atoms are *removed* from a structure rather than additive ones that contribute such atoms to the target structure. All of these reaction types would be considered "nonproductive" with respect to the actual building of target structure bonds. Clearly, the overall goal is to minimize f(sac). The ratio of the number of target bonds made to the number of reaction steps, B/M, and target bond-forming profiles as a function of reaction stage are useful in probing synthesis strategy. Low ratios and gaps in the profiles are directly linked to the use of sacrificial reagents. Clearly, good syntheses utilize reaction steps that achieve as many target bonds as possible in a single step. The distribution of target bonds formed shows pictorially not only whether a plan has gaps of "nonproductive" activity but also if the building-up activity is evenly spread over the entire plan or is localized in the early, middle, or late stages of a plan.

The hypsicity or oxidation level index originates from a concept put forward by Hendrickson[23] that suggested the minimization, or at best the elimination, of redox reactions in the design of a synthesis strategy so that a so-called isohypsic condition (HI = 0) could be achieved. This idea is consistent with the observation that redox reactions as a reaction class are the worst performing in terms of atom economy as previously noted in an extensive survey of a database of named organic reactions.[5b] A caveat is that an HI of zero can also be achieved by a fortuitous algebraic cancelation of increases and decreases in oxidation level changes regardless of reaction type. The method of determining HI for a synthesis plan has been described elsewhere,[7d] and its implementation in the LINEAR and CONVERGENT spreadsheets is also given in the Supporting Information. The basic idea is that, once all of the target bond-forming steps are mapped out as described in the preceding paragraph, the set of atoms involved in those bond-forming steps are automatically determined. The oxidation state of each of these atoms is tracked in the reverse sense from the final target structure back to the originating starting material structure. Just as how the molecular weight first-moment parameter was determined, the difference between the oxidation state of a given atom in an intermediate and the oxidation state of that same atom in the target structure is noted. Essentially the oxidation states of the atoms in the target structure are the reference states. For a given atom these differences are summed, and the process is repeated for all other atoms involved in target bond-forming steps. Then an overall sum of oxidation number changes is determined over all such atoms, and then this sum is divided by the number of reaction stages along the longest branch plus one. Again, the extra stage accounts for the zeroth stage of starting materials. A key point to keep in mind in determining the oxidation states of atoms is that they depend directly on Lewis and resonance structures, and so the "active" or most important canonical resonance structures of reactants are used as the basis of the determination of oxidation states of the relevant atoms in target bond-forming steps. Also, the orientation of π bonds in aromatic structures should be preserved from structure to structure throughout the course of the synthesis plan until the final target structure is reached to maintain consistency in the absolute values of the relevant oxidation states. An example is the C2 position of pyridine which can have oxidation states of +1 or 0, depending on the orientation of π bonds in the aromatic ring. The example shown in Scheme 1 illustrates this point as the same overall change in oxidation state at C2 is obtained for both π bond orientations. What is important is that the *changes* in oxidation states of the set of

(23) Hendrickson, J. B. *J. Am. Chem. Soc.* **1971**, *93*, 6847.

Scheme 1. **Example tracking of oxidation number changes at a given atom involved in target bond-forming reactions**



atoms in bond-forming steps are properly documented over the course of the synthesis plan.

Good synthesis strategies have HI = 0 under the condition that redox reactions are completely eliminated since they are the most wasteful class of organic reactions, or have HI < 0 for monotonically increasing oxidation states where additive oxidation reactions are utilized, or have HI > 0 for monotonically decreasing oxidation states where additive reduction reactions are utilized. The latter two cases generally correlate with high atom economical "green" routes having low fractions of sacrificial reagents. Hypsicity profiles are also an excellent tool to visualize the "ups" and "downs" of oxidation level changes and become necessary in picking out these conditions. Conclusions drawn only on the basis of the magnitude and sign of HI are insufficient.

3. Oseltamivir Phosphate Syntheses

The algorithm presented above has been tested on the 15 reported synthesis plans for the neuraminidase inhibitor, oseltamivir phosphate, which was discovered 10 years ago[24] and whose synthesis plans have recently been reviewed.[25] To avoid repetition, the synthesis schemes for all of the plans will not be presented here; however, the Supporting Information contains all of the schemes showing balanced chemical equations, the corresponding synthesis trees, and the sets of radial pentagons. Scheme 2 gives an overall synthesis map of starting materials used by various routes. All industrial plans are linear, and the Fang synthesis is the only convergent plan.

3.1. Industrial Plans. Schemes S1−S6 (Supporting Information) show respectively the Gilead azide, Roche quinic acid azide-free first generation (G1), Roche quinic acid azide-free second generation (G2), Roche shikimic acid azide-free third generation (G3), Roche Diels−Alder fourth generation (G4), and Roche desymmetrization fifth generation (G5) plans. In the Supporting Information (SI) Figures A1−A6 show the respective synthesis trees, and the sets of radial pentagons are shown in Figures B1−B6 (SI).

3.2. Academic Plans. Schemes S7−S15 (SI) show respectively the Corey, Shibasaki G1, Shibasaki G2, Shibasaki G3, Fukuyama, Kann, Trost (long), Fang, and Okumara−Corey plans. Figures A7−A15 (SI) show the corresponding synthesis trees, and Figures B7−B15 (SI) show the sets of radial pentagons.

3.3. Material Efficiency Analysis. Table 1 summarizes all of the green metrics parameters for all 15 plans ranked in descending order of kernel mass of waste. The Roche third-generation (G3) plan is the best performer at the kernel level with the highest kernel RME (11.5%), lowest kernel mass of waste (3.1 kg waste/mol oseltamivir phosphate), and highest overall yield (39% over 13 steps) and has the lowest molecular weight fraction of sacrificial reagents (45.7%). The Roche fourth-generation (G4) plan and the short Trost plan starting from the lactone, 6-oxa-bicyclo[3.2.1]oct-3-en-7-one, as starting material have the shortest number of steps (9) and involve the least number of input materials (17). The Roche G4 plan has the highest atom economy (23.9%) followed closely by the Roche G3 (21.0%), Gilead (20.5%), and Roche G2 (20.3%) plans. The Fang and Shibasaki second- and third-generation plans involve the greatest degree of building up from low molecular weight starting materials at −167.02, −163.86, and −167.26 g/mol/stage, respectively. The Roche G5 and Shibasaki G3 plans have the highest number of target bonds made per reaction step at 1.36 and 1.27, respectively.

Table 2 summarizes the complete set of *E*-factors for all plans according to the three contributions given in eqs 17a−c. The details of auxiliary material consumption were completely documented for the industrial plans; however, this was not so for the academic plans, and so the best case scenario lower limits for the *E*-aux contributions and therefore for *E*-total could be determined for them. Solvent consumption and mass of packing material used in chromatographic purification procedures were by far the most neglected pieces of information not given in experimental procedures. Overall, the Roche G3 plan ranks as the most material efficient with the least value for *E*-total at 230.9 mass units of waste per mass unit of oseltamivir phosphate. It should be noted that the ordering of the plans by *E*-total or true mass of waste per mole of target product in Table 2 is slightly different from that given in Table 1, based on kernel metrics. However, the same plans appear at the tops and bottoms of both lists,

(24) (a) Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, W. G.; Stevens, R. C. *J. Am. Chem. Soc.* **1997**, *119*, 681. (b) Kim, C. U.; Lew, W.; Williams, M. A.; Wu, H.; Zhang, L.; Chen, X.; Escarpe, P. A.; Mendel, D. B.; Laver, W. G.; Stevens, R. C. *J. Med. Chem.* **1998**, *41*, 2451.

(25) (a) Shibasaki, M.; Kanai, M. *Eur. J. Org. Chem.* **2008**, 1839. (b) Farina, V.; Brown, J. D. *Angew. Chem., Int. Ed.* **2006**, *45*, 7330.

Scheme 2. **Synthesis map showing various starting materials used for the synthesis of oseltamivir phosphate**

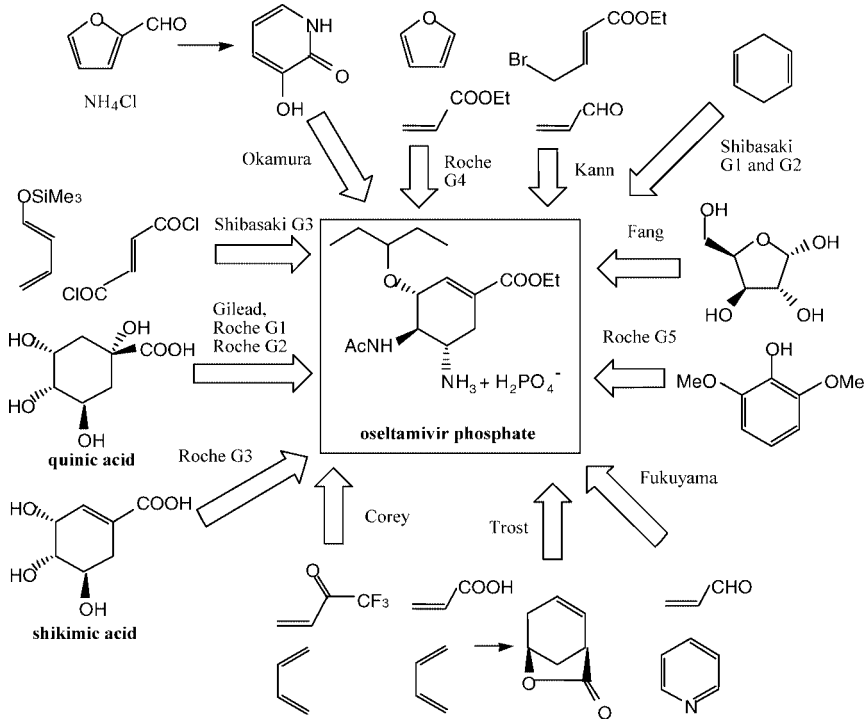


Table 1. **Summary of metrics for syntheses of oseltamivir phosphate arranged in ascending order of kernel waste production**

plan	year	type	N^a	M^b	I^c	$\mu_1{}^d$	β^e	δ^f	f(sac)^g	B/M^h	HI^i	% overall yield	% AE	% Kernel RME	kernel mass of waste (kg)^j
Roche (shikimic acid route - G3)	1999, 2004	linear	13	13	19	−102.68	0.861	0.345	**0.457**	0.77	+0.14	**39.0**	21.0	**11.5**	**3.1**
Roche (quinic acid route - G2)	1999, 2004	linear	14	14	20	−93.57	0.868	0.338	0.467	0.71	+0.13	21.9	20.3	9.0	4.2
Trost (short)	2008	linear	**9**	**9**	**17**	−35.40	0.865	0.397	0.630	1	−0.2	29.9	16.1	5.6	6.9
Corey	2006	linear	11	11	**17**	−153.68	0.843	0.361	0.743	0.91	−3	22.4	17.2	5.5	7.2
Roche (desymmetrization route - G5)	2000	linear	11	11	24	−110.48	0.779	**0.440**	0.677	**1.36**	+1	25.6	13.8	5.3	7.3
Trost (long)	2008	linear	12	12	21	−99.92	0.892	0.397	0.690	0.83	−1.54	16.2	13.4	4.0	9.8
Roche (quinic acid route - G1)	2001	linear	12	12	21	−131.23	0.862	0.378	0.625	0.75	+0.15	7.6	18.5	3.2	12.5
Fang	2007	convergent	17	18	35	**−167.02**	0.950	0.379	0.671	0.67	+1.44	13.4	12.0	3.1	12.7
Gilead	1998	linear	12	12	21	−135.85	0.867	0.377	0.607	0.83	+0.39	6.3	20.5	2.7	15.0
Fukuyama	2007	linear	13	13	22	−133.48	0.875	0.369	0.638	0.85	−2.64	5.5	15.9	2.4	16.4
Roche (Diels−Alder route - G4)	2000	linear	**9**	**9**	17	−142.56	**0.756**	0.432	0.505	**1.11**	−0.5	1.1	**23.9**	1.5	24.4
Okamura-Corey	2008	linear	13	13	25	−124.76	0.893	0.386	0.754	0.69	−0.43	2.6	16.8	1.3	32.0
Kann	2007	linear	15	15	25	−61.26	0.880	0.363	0.788	0.60	−0.56	3.4	11.8	0.9	47.4
Shibasaki G2	2007	linear	16	16	32	**−163.86**	0.914	0.385	0.837	0.69	−0.71	4.5	10.0	0.8	47.9
Shibasaki G3	2007	linear	11	11	23	**−167.26**	0.880	0.406	0.481	**1.27**	−0.33	1.4	16.1	0.6	73.6
Shibasaki G1	2006	linear	15	15	34	−132.10	0.919	0.403	0.766	0.80	−1.19	1.4	9.6	0.3	150.3

[a] Number of reaction stages. [b] Number of reaction steps. [c] Number of input materials. [d] Molecular weight first-moment building-up parameter (g/mol/reaction stage). [e] Degree of asymmetry. [f] Degree of convergence. [g] Fraction of sacrificial reagents by molecular weight. [h] Number of target bonds made per reaction step. [i] Hypsicity index. [j] Basis is 1 mol of oseltamivir phosphate target product.

thus reinforcing the notion that plan performance analysis done at the kernel level is good enough to pick out both the best-performing candidates and the least efficient ones from a large set of plans. It is unlikely that a plan ranked very low at the kernel metrics level will suddenly skyrocket to the top of the list at the global metrics level of analysis, or conversely that a highly ranked plan at the kernel level will end up at the bottom of the global metrics

Table 2. **Summary of *E*-factors for syntheses of oseltamivir phosphate arranged in ascending order of total waste production**

plan	*E*-kernel	*E*-excess	*E*-auxiliaries	*E*-total	total mass of waste (kg)[a]
Roche (shikimic acid route - G3)	**7.7**	**24.6**	**198.6**	**230.9**	**94.7**
Roche (quinic acid route - G2)	10.1	30.0	267.7	307.9	126.2
Roche (quinic acid route - G1)	30.6	71.1	755.5	857.2	351.4
Roche (desymmetrization route - G5)	17.8	68.4	847.4	933.6	382.8
Gilead	36.7	91.5	808.6	936.7	384.0
Fang	31.0	274.8	>2275.1	>2580.9	>1058
Trost (short)[b]	16.8	141.5	>2527.1	>2685.4	>1101
Trost (long)[b]	23.8	144.5	>2690.5	>2858.7	>1172
Corey[c]	17.5	208.8	>3056.5	>3282.9	>1346
Fukuyama[d]	40.0	163.4	>3843.0	>4046.5	>1659
Roche (Diels−Alder route - G4)	66.8	181.2	>4855.6	>5103.5	>2092
Kann	115.5	285.9	>13238.0	>13639.5	>5592
Shibasaki G1[e]	366.6	3772.8	>12055.0	>16194.4	>6640
Shibasaki G2[e]	116.8	1279.9	>18817.8	>20214.5	>8288
Okamura-Corey	78.0	439.8	>21926	>22444	>9202
Shibasaki G3	179.5	1554.1	>24805.8	>26539.4	>10881

[a] Basis is 1 mol of oseltamivir phosphate target product. [b] Includes synthesis of Trost ligand and Du Bois catalyst. [c] Includes synthesis of oxazaborolidinium catalyst. [d] Includes synthesis of MacMillan catalyst. [e] Includes synthesis of Shibasaki ligand.

ranking. The more plans there are for a given target, the better the overall assessment will be. Of course, the best possible decisions and assessments can only be made if absolutely all of the material consumption is properly documented in experimental procedures; otherwise at best only upper limits on global RME and lower limits on *E*-total can be made. This is probably the most serious problem that is hampering the implementation of green metrics as a standard decision-making tool in plan optimization and route selection. The key observation in Tables 1 and 2 is that good attributes are scattered over a number of plans that span the entire list. What is suggested by these ranking results is that the next proposed synthesis plan for oseltamivir phosphate should strive to have 9 or fewer reaction steps, to use 17 or fewer input materials, to have an atom economy higher than 24%, to have a kernel RME higher than 12%, to have an overall yield higher than 39%, to have an *E*-total less than 230 mass units of waste per mass unit of target product, to produce less than 95 kg of total waste material per mole of oseltamivir, to have an *f*(sac) value less than 46%, to have a first-moment building-up parameter lower than −167 g/mol/stage, and to produce at least 1.3 target bonds per reaction step. These parameters frame the current challenge faced by chemists in finding the true optimum synthesis of oseltamivir phosphate.

In determining the overall global RME and *E*-factors for the Trost, Corey, Fukuyama, and Shibasaki first- and second-generation plans, the syntheses of the Trost ligand, Du Bois catalyst, oxazaborolidinium catalyst, MacMillan catalyst, and Shibasaki ligand were accounted for, respectively. This was done by first calculating the mole scale of catalyst or ligand required for the particular step by multiplying the mole scale of that reaction as determined from the synthesis tree of oseltamivir phosphate by the mol % of catalyst or ligand as given in the radial pentagon analysis for that step. This result became the target scale of the catalyst or ligand and was used as the normalizing scale for its synthesis tree from which the appropriately scaled masses of input reagent and auxiliary materials were determined for its production. The overall RME performance for oseltamivir phosphate was then found from eq 21.

$$\mathrm{RME}_{\mathrm{overall}}^{\text{oseltamivir phosphate}} = \mathrm{mass}_{\text{oseltamivir phosphate}} / [\left(\sum \mathrm{mass}_{\mathrm{reagents}} + \sum \mathrm{mass}_{\mathrm{auxiliaries}}\right)_{\text{oseltamivir phosphate}} + \left(\sum \mathrm{mass}_{\mathrm{reagents}} + \sum \mathrm{mass}_{\mathrm{auxiliaries}}\right)_{\text{ligand or catalyst}}] \quad (21)$$

where the mass of oseltamivir phosphate corresponded to 1 mol of material. The synthesis plans and trees and radial pentagons for the production of these ligands and catalysts are also given in the Supporting Information in Figures C1−C5, D1−D5, and E1−E5, respectively.

The performance of the Trost and Okamura modification of the Corey plan from different starting materials are good illustrative examples of the dilemma of comparing various plans beginning from different starting materials to a common target structure. Obviously, the performance of the Trost plan beginning from the bicyclic lactone and the Okamura plan beginning from the pyridone derivative are expected to be better than from acrylic acid and furfural, respectively as shown by the data in Tables 1 and 2. Clearly, the 3-hydroxy-2-pyridone, which is the structural tautomer of 2,3-dihydroxypyridine, is not a "readily available starting material" and needs to be synthesized. Trost makes a claim that the bicyclic lactone is commercially available but unfortunately gives no indication as to the source neither in the paper nor in the supplementary experimental information.[19a] A check of the latest 2007−2008 Aldrich catalogue, for example, which is by far the leading supplier of chemicals to academic laboratories, does not list this compound. It is important to trace the origins of all materials required in a synthesis to truly readily available materials so that comparisons are as fair as possible, especially if a claim of a significant advance in synthetic efficiency or "greenness" is made. Readily available starting materials are usually first-generation industrial feedstock compounds from the petrochemical industry. This one-sided open-ended comparison is a characteristic feature of the problem of ranking several plans that begin from different

Table 3. **Summary of bottlenecks for syntheses of oseltamivir phosphate**

plan	low reaction yields (≤60%)	low atom economies (≤60%)	high excess reagent usage	high auxiliary material usage	high number of reaction steps	high number of input materials	high MW fraction of sacrificial reagents	low number of target bonds per reaction step	other disadvantages
Roche (shikimic acid route - G3)		steps 8 and 6	steps 1 and 5	steps 6 and 5					step 5 is worst performer; sacrificial $SOCl_2$ used to carry out dehydration in step 1; bulky sacrificial 1,3-dimethylbarbituric acid used to carry out dideallyation in step 14
Roche (quinic acid route - G2)		steps 9 and 7	steps 2 and 6	steps 7, 6, 4, and 3					step 6 is worst performer; sacrificial acetonide group made in step 1; sacrificial Vilsmeier reagent used in dehydration in step 4; bulky sacrificial 1,3-dimethylbarbituric acid used to carry out dideallyation in step 14
Roche (quinic acid route - G1)	step 4	steps 10, 6, 4, and 7	steps 5, 6, and 2	steps 4 and 1					step 4 is worst performer; sacrificial $SOCl_2$ used in dehydration in step 4; sacrificial imination in step 10
Roche (desymmetrization route - G5)		steps 6, 8, 2, 11, 1, and 9	step 7	step 8					hazardous sodium azide used in step 10; sacrificial mesyl group used in step 1 to improve leaving group ability; sacrifical NBS used in step 3 to effect substitution in next step; demethylation in step 6 is a degradation not a synthesis step
Fang		steps 12, 14, 4, and 3	steps 14, 16, 5, and 3	step 14	yes	yes			step 14 is worst performer because of excess reagent consumption; protecting groups used in steps 1, 2, 6, and 7
Gilead	steps 11 and 4	steps 6, 4, and 7	steps 5, 6, and 2	steps 4 and 1					step 4 is worst performer; hazardous sodium azide reagent used in steps 8 and 10; sacrificial acetonide group made in step 1; sacrificial $SOCl_2$ used in dehydration in step 4; sacrificial PMe_3 used in aziridation in step 9
Trost (long)		steps 3, 6, 10, 7, 5, and 2	steps 12, 9, and 8	steps 8 and 5					steps 8 and 5 are worst performers; two-thirds of reactions have poor atom economies mainly because of the use of high molecular weight reagents to transfer single atoms to the target structure, for example, trimethylsilylphthalimide and 2-trimethylsilylethane sulfonylamide are used to transfer nitrogen atoms in steps 4 and 7 respectively
Corey	step 10	steps 3, 7, 5, 9, and 2	steps 10 and 11	step 8			yes: Me_3SiOTf and iodine in step 3; NBS in step 6		step 8 is worst performer; chiral Corey−Shibata−Bakshi-type catalyst in first step needs to be synthesized

Table 3. **Continued**

plan	low reaction yields (60%)	low atom economies (60%)	high excess reagent usage	high auxiliary material usage	high number of reaction steps	high number of input materials	high MW fraction of sacrificial reagents	low number of target bonds per reaction step	other disadvantages
Fukuyama	step 3	steps 5, 4, and 9	steps 12 and 10	step 3					step 3 is worst performer; chiral MacMillan catalyst in step 2 needs to be synthesized; subtractive reduction in step 4 (*N*-Boc deprotection); bulky sacrificial 1,3-dimethylbarbituric acid used to carry out breakdown of allylcarbamate in step 13
Roche (Diels−Alder route - G4)	steps 2 and 3	steps 7, 4, and 3	steps 5 and 3	step 2					step 2 is worst performer
Okamura−Corey	step 7	steps 1, 6, 7, and 11	steps 12, 4, 13, and 5	step 3			yes: chlorine in step 1; *tert*-butyldimethylsilyl chloride in step 2; nosyl chloride in step 3;		step 3 is worst performer
Kann	steps 6 and 14	steps 2, 9 and 13	step 14, 9, and 3	steps 9 and 13			yes: PPh_3 in step 1; $Fe_2(CO)_9$ in step 3; $Ph_3C^+ PF_6^-$ in step 4; (1R, 2S)-*trans*-2-phenylcyclohexanol in step 5; HPF_6 in step 7; PPh_3 in step 13	yes	steps 9 and 13 are worst performers; separation of diastereomers required in step 6; (1R, 2S)-*trans*-2-phenylcyclohexanol and metallocarbonyl are used as a sacrificial chiral controlling groups in steps 5 and 8 respectively; sacrificial oxidation in step 9 to deprotect metallocarbonyl moiety; sacrificial oxidation in step 10; hazardous sodium azide reagent used in step 11
Shibasaki G2	steps 14 and 15	steps 6, 1, 12, 13, 5, and 3	steps 15, 16, and 9	steps 15 and 11	yes	yes	yes: 3,5-dinitrobenzoyl chloride in step 3; iodine in step 6; diethylphosphoroyl cyanide in step 11; PPh_3 and DEAD in steps 12 and 13;		steps 15 and 11 are worst performers; chiral ligand in step 4 needs to be synthesized; sacrificial oxidation in steps 1 and 10; hazardous sodium azide reagent used in step 2
Shibasaki G3	steps 11, 10, 9, and 1	steps 8, 1, 6, 7, and 5	steps 9, 3, and 6	steps 5, 1, and 3					step 5 is worst performer followed closely by steps 1 and 3; sacrificial oxidation in step 5; *N*-bromosuccimide, PPh_3 and DEAD are sacrificial reagents used in steps 6 and 8, respectively
Shibasaki G1	steps 15, 14, and 11	steps 1, 12, 3, 5, 6, 10, 8, 9, and 7	step 15	steps 7, 9, and 10	yes	yes	yes: 3,5-dinitrobenzoyl chloride in step 3; NBS in step 8; PPh_3 in step 6; PPh_3 and DEAD in step 10		steps 7 and 9 are worst performers; chiral ligand in step 4 needs to be synthesized; sacrificial oxidation in step 1; hazardous sodium azide reagent used in step 2; lowest overall yield; lowest overall atom economy; highest overall waste production

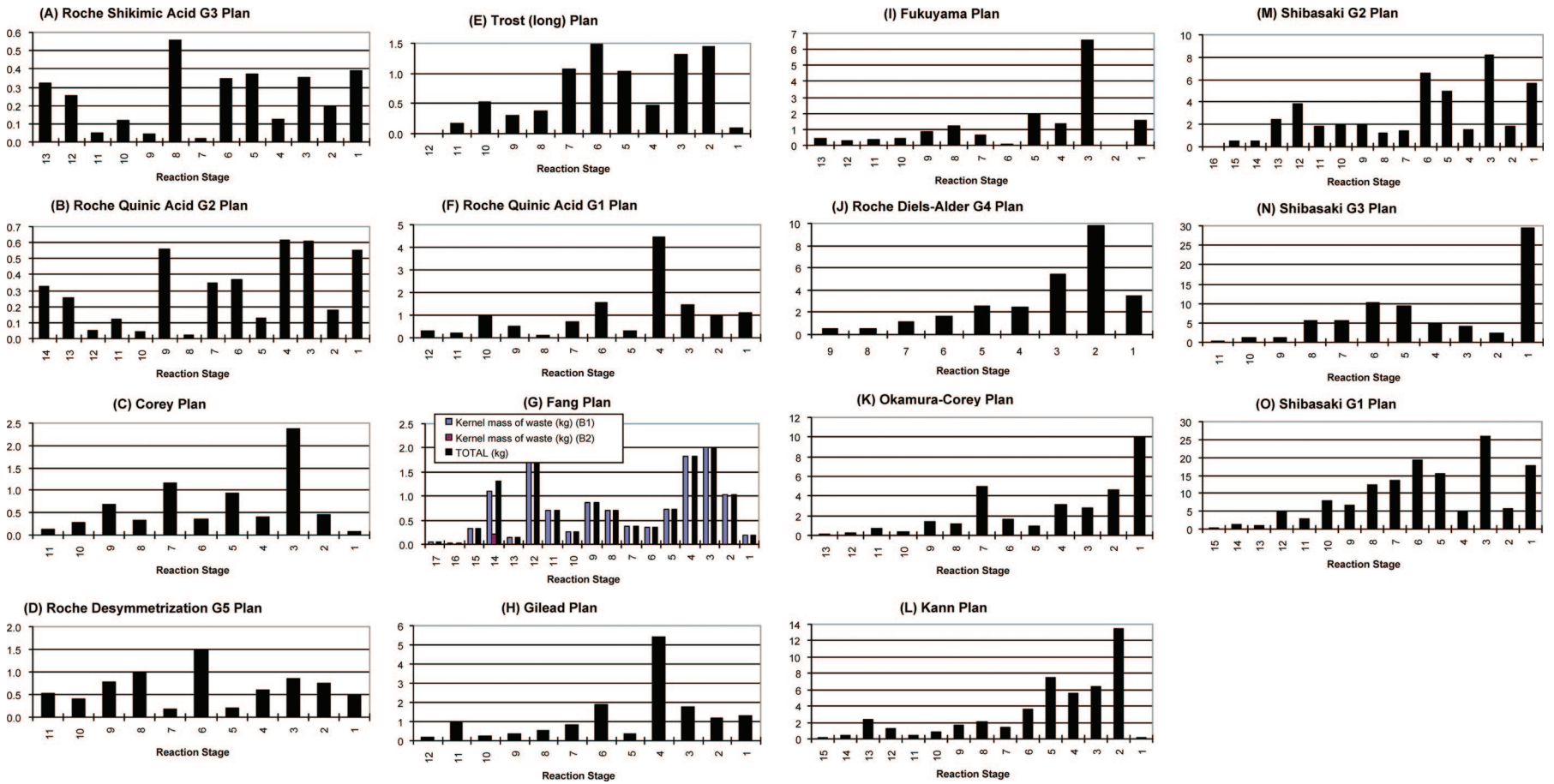


Figure 2. **Kernel waste distribution profiles for various plans to oseltamivir phosphate: (A) *E*-kernel = 7.7; (B) *E*-kernel = 10.1; (C) *E*-kernel = 17.5; (D) *E*-kernel = 17.8; (E) *E*-kernel = 23.8; (F) *E*-kernel = 30.6; (G) *E*-kernel = 31.0; (H) *E*-kernel = 36.7; (I) *E*-kernel = 40.0; (J) *E*-kernel = 66.8; (K) *E*-kernel = 78.0; (L) *E*-kernel = 115.5; (M) *E*-kernel = 116.8; (N) *E*-kernel = 179.5; (O) *E*-kernel = 366.6. Ordinate is kernel mass of waste (kg).**

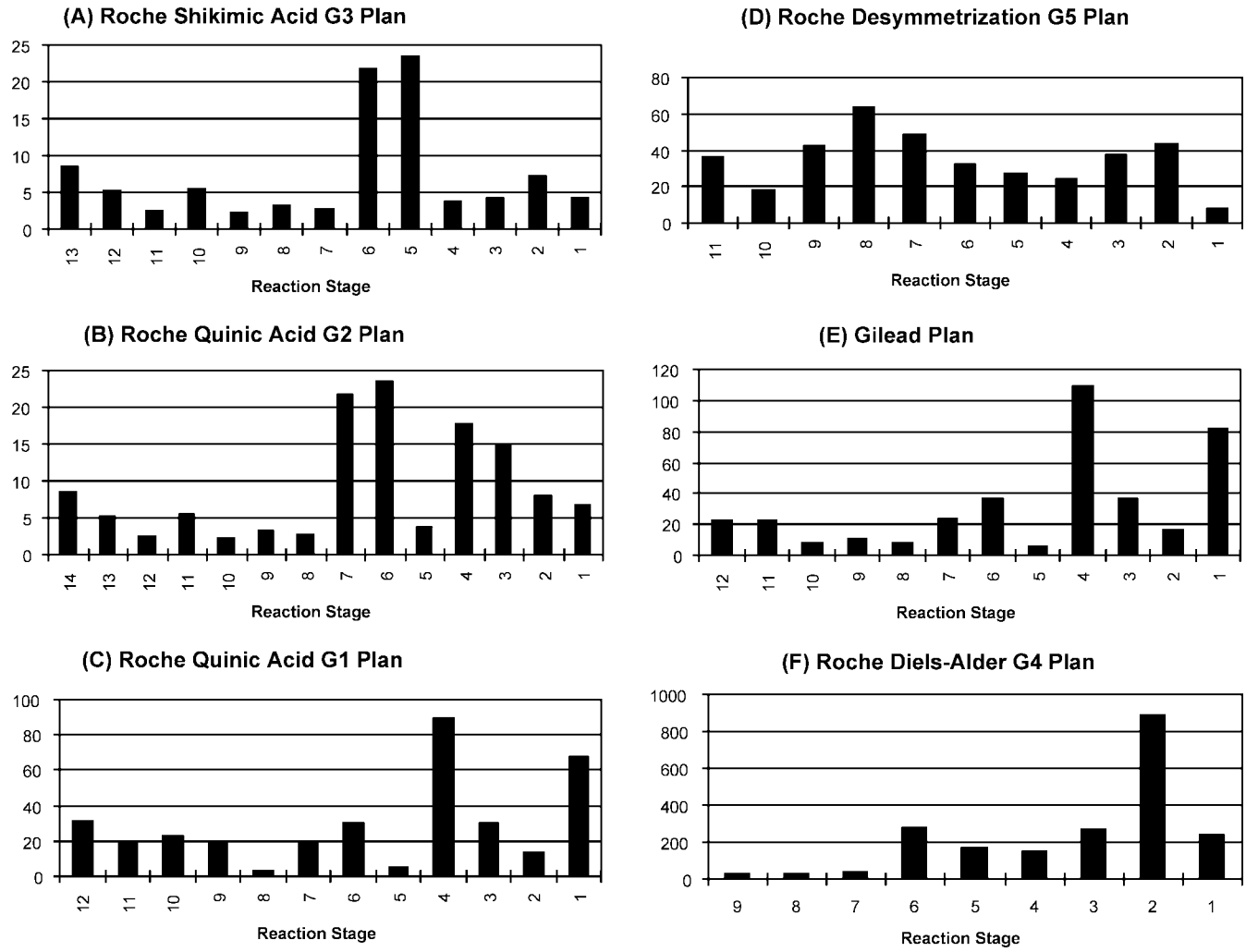


Figure 3. **Total waste distribution profiles for various industrial plans to oseltamivir phosphate: (A) *E*-total = 230.9; (B) *E*-total = 307.9; (C) *E*-total = 857.2; (D) *E*-total = 933.6; (E) *E*-total = 936.7; (F) *E*-total = 5103.5. Ordinate is total mass of waste (kg).**

starting materials but lead to a common target, and thus it leaves such rankings open to debate. It is very rare to find a comparison of several synthetic routes that begin from the same starting material and end up at the same target molecule. Nevertheless, the data shown in Tables 1 and 2 clearly show that the long and short versions of the Trost plan are very competitive with the top industrial plans and are worthy of consideration for scale-up development.

Table 3 summarizes the bottlenecks in each of the 15 plans and Figure 2 shows the kernel mass of waste profiles for all plans including their corresponding *E*-kernel performances. Since all material consumption was disclosed for the industrial plans, Figures 3 and 4 show the true total mass of waste and auxiliary material mass profiles, respectively, including their *E*-total and *E*-auxiliary performances. It is observed that the shapes of the distributions in Figures 3 and 4 are virtually identical as are the magnitudes of the ordinate scales. This implies that the contribution from byproducts, side products, and unreacted starting materials to the overall mass of waste produced is insignificant. The dwarfing effect of auxiliary mass profiles in Figure 4 when compared to their kernel waste distribution profiles in Figure 2 is quite typical and is entirely consistent with the conclusion that auxiliary materials such as solvents, extraction washes, etc. contribute the lion share of waste materials. Not surprisingly, it is in this category of waste production where the greatest efforts have been made to minimize overall waste for a scaled-up synthesis process.

3.4. Synthetic Efficiency Analysis. Figure 5 lists the target bond-forming maps for all 15 plans. These diagrams show which target bonds are made, their constituent atoms, and at what steps these bonds are made. Table 4 summarizes the various strategies used to install the four groups on the six-membered ring of oseltamivir phosphate. The plans may be divided into two broad categories with respect to the six-membered ring core: those that involve ring construction and those that begin with starting materials having preformed rings. For the former class, we observe the following strategies:

(a) The Corey, Fukuyama, Shibasaki G3, Roche G4, Trost (long), and Okamura plans all follow a Diels−Alder strategy;

(b) the Okamura plan involves a protected 2-pyridone derivative in a Diels−Alder reaction which is then followed by cleavage of the resulting [2.2.2] bicyclic intermediate;

(c) the Fukuyama plan involves a dihydropyridine derivative in a Diels−Alder reaction followed by cleavage of the resulting [2.2.2] bicyclic intermediate;

(d) the Kann plan uses a tandem Michael−Wittig reaction;

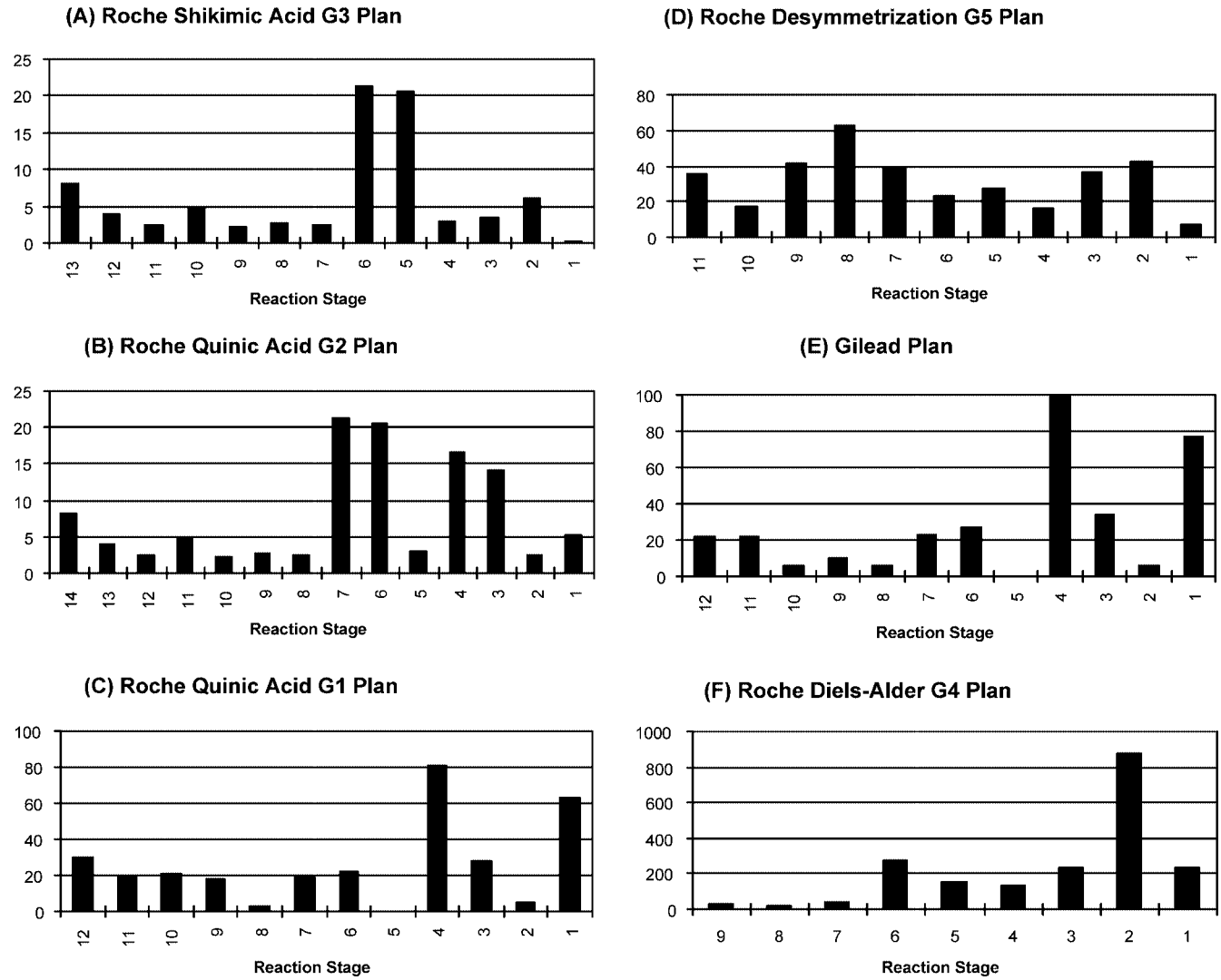


Figure 4. **Auxiliary mass distribution profiles for various industrial plans to oseltamivir phosphate: (A) *E*-auxiliary = 198.6; (B) *E*-auxiliary = 267.7; (C) *E*-auxiliary = 755.5; (D) *E*-auxiliary = 847.5; (E) *E*-auxiliary = 808.6; (F) *E*-auxiliary = 4855.6. Ordinate is mass of auxiliaries (kg).**

(e) the Fang plan uses an intramolecular Horner−Wittig−Emmons reaction.

For the latter class, we observe the following strategies:

(a) the Roche G1, Roche G2, and Gilead plans begin with quinic acid;

(b) the Roche G3 plan begins with shikimic acid;

(c) the Roche G5 plan uses 2,6-dimethoxyphenol;

(d) the Shibasaki G1 and G2 plans begin with 1,4-cyclohexadiene;

(e) the short Trost plan begins with a bicyclic lactone, 6-oxa-bicyclo[3.2.1]oct-3-en-7-one.

Quinic acid and shikimic acid are good precursors because the ring is preformed and there are stereogenic centers which may be exploited to advantage. Shikimic acid is preferred over quinic acid because it saves a dehydration step which is necessary for the installation of the double bond in the ring. The Fang strategy uses the carbohydrate D-xylose as a source of stereogenic centers, however epimerization is required. It is common to exploit carbohydrates or other natural products from the chiral pool as starting materials for targets especially when the target structure has a significant number of contiguous stereogenic centers.

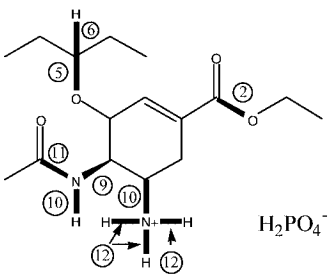
From the target bond maps in Figure 5 it is possible to examine the pairwise similarities of target bonds constructed. The fraction of matching bonds between two general structures A and B may be defined as a kind of similarity index and is given by eq 22.

$$f_{\text{match}} = \frac{C}{C+D} = \frac{C}{B_A + B_B - C} \tag{22}$$

where, C is the number of target bonds common to both structures, B_A is the number of target bonds in structure A, B_B is the number of target bonds in structure B, and D is the sum of noncommon or different bonds in both structures. An $f_{\text{match}} = 1$ is a complete matching of all target bonds in both structures, whereas, an $f_{\text{match}} = 0$ means that no target bonds are common between the two structures and therefore implies that the two plans have strategies that are radically different. Table 5 summarizes such an analysis of the 105 unique pairwise comparisons for the 15 plans. Focusing on the highest values of f_{match} we observe the following patterns:

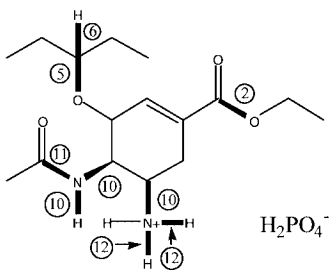
(a) The Gilead, Roche G2, and Roche G3 have identical target bonds ($f_{\text{match}} = 1$).

Gilead azide plan



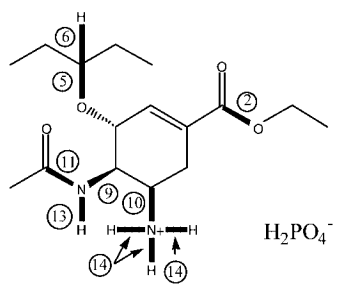
quinic acid (1), ethanol (2), 3-pentanone (5), BH_3 Me_2S (6), sodium azide (8 -> 9), sodium azide (10), ammonium chloride (10), acetic anhydride (11), phosphoric acid (12), dihydrogen (12)

Roche quinic acid G1 plan



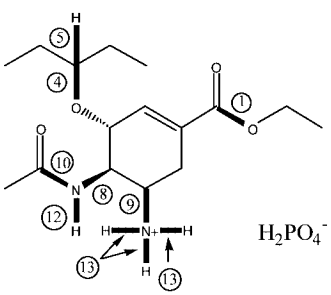
quinic acid (1), ethanol (2), 3-pentanone (5), BH_3 Me_2S (6), allylamine (8), 2 eq. allylamine (10), acetic anhydride (11), phosphoric acid (12), ethanolamine (12)

Roche quinic acid G2 plan



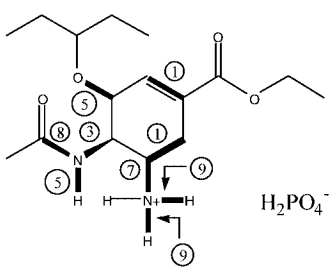
quinic acid (1), ethanol (2), 3-pentanone (5), triethylsilane (6), t-butylamine (8 -> 9), diallylamine (10), acetic anhydride (11), trifluoroacetic acid (13), 1,3-dimethylbarbituric acid (14), phosphoric acid (14)

Roche Shikimic Acid G3 Plan



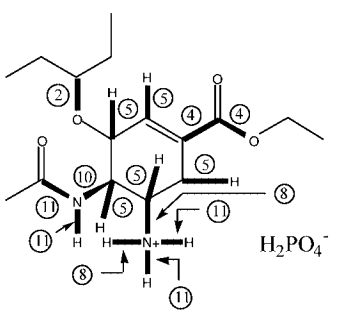
shikimic acid (1), ethanol (1), 3-pentanone (4), triethylsilane (5), t-butylamine (7 -> 8), diallylamine (9), acetic anhydride (10), trifluoroacetic acid (12), phosphoric acid (13), 1,3-dimethylbarbituric acid (13)

Roche Diels-Alder G4 Plan



furan (1), ethyl acrylate (1), diphenylphosphoryl azide (3), 3-pentanol (5), allylamine (7), acetic anhydride (8), ethanolamine (9), phosphoric acid (9)

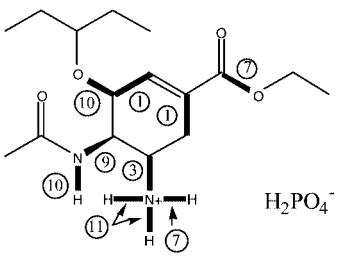
Roche Desymmetrization G5 Plan



3-pentanol (1), 2,6-dimethoxyphenol (2), carbon monoxide (4), ethanol (4), dihydrogen (5), water (6), diphenylphosphoryl azide (8), sodium azide (10), dihydrogen (11), acetic anhydride (11), hydrobromic acid (11), phosphoric acid (11)

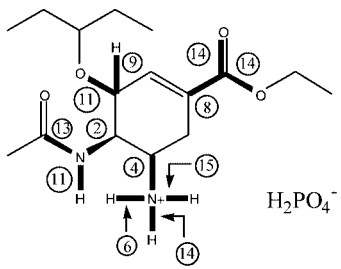
Figure 5. **Target structure maps for various plans to oseltamivir phosphate showing which target bonds are made and at what reaction step. List of reagents used that in whole or in part end up in the target structure are shown below each structure map.**

Corey Plan



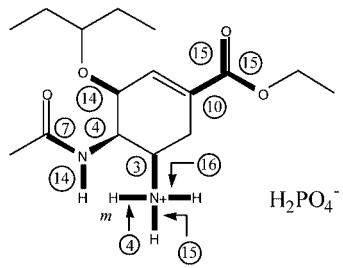
1,3-butadiene (1), trifluoroethyl acrylate (1), ammonia (2 -> 3), N-bromoacetamide (8 -> 9), ethanol (7), 3-pentanol (10), water (11), phosphoric acid (11)

Shibasaki G1 Plan



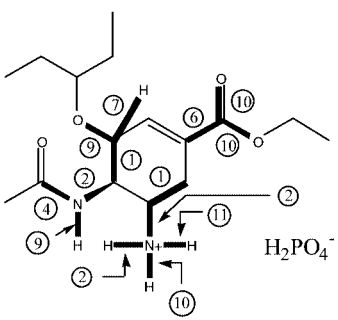
1,3-cyclohexadiene (1), sodium azide (2), trimethylsilylcyanide (8), 3-pentanol (11), trimethylsilylazide (4), water (6), $LiHAl(O^tBu)_3$ (9), acetic anhydride (13), 2 eq. water (14), ethanol (14), phosphoric acid (15)

Shibasaki G2 plan



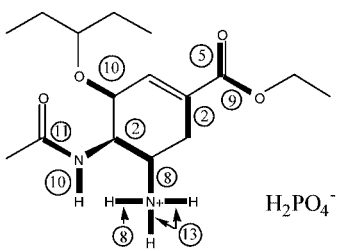
1,3-cyclohexadiene (1), sodium azide (3), trimethylsilyl azide (4), water (4), thioacetic acid (7), lithium cyanide (10), 3-pentanol (14), ethanol (15), water (15), phosphoric acid (16)

Shibasaki G3 Plan



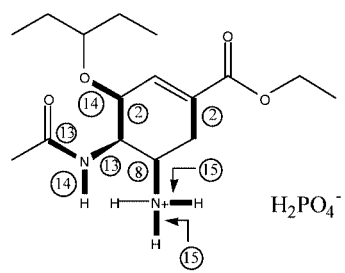
1-(trimethylsilyloxy)-1,3-butadiene (1), fumaryl chloride (1), t-butanol (2), 2 eq. trimethylsilyl azide (2), acetic anhydride (4), trimethylsilyl cynanide (6), $LiAlH(O^tBu)_3$ (7), 3-pentanol (9), 2 eq. water (10), ethanol (10), phosphoric acid (11)

Fukuyama Plan



pyridine (1), acrolein (2), sodium periodate (5), ammonia (6 -> 8), allyl alcohol (8), sodium ethoxide (9), 3-pentanol (10), acetic anhydride (12), 1,3-dimethylbarbituric acid (13), phosphoric acid (13)

Kann Plan



acrolein (1), ethyl 4-bromo-2-butenoate (1), NH_2-Boc (8), sodium azide (11 -> 13), acetic anhydride (13), 3-pentanol (14), water (15), phosphoric acid (15)

Figure 5 Continued.

(b) The Roche G2-Roche G3, Roche G2-Gilead, and Roche G2-Roche G1 pairs of plans have f_{match} equal to 0.90.

(c) The Corey and Trost-long plans have f_{match} equal to 0.91.

(d) The Shibasaki G1 and G2 plans have f_{match} equal to 0.92.

(e) The Shibasaki G1 and G3 plans have f_{match} equal to 0.86.

(f) The Shibasaki G2 and G3 plans have f_{match} equal to 0.79.

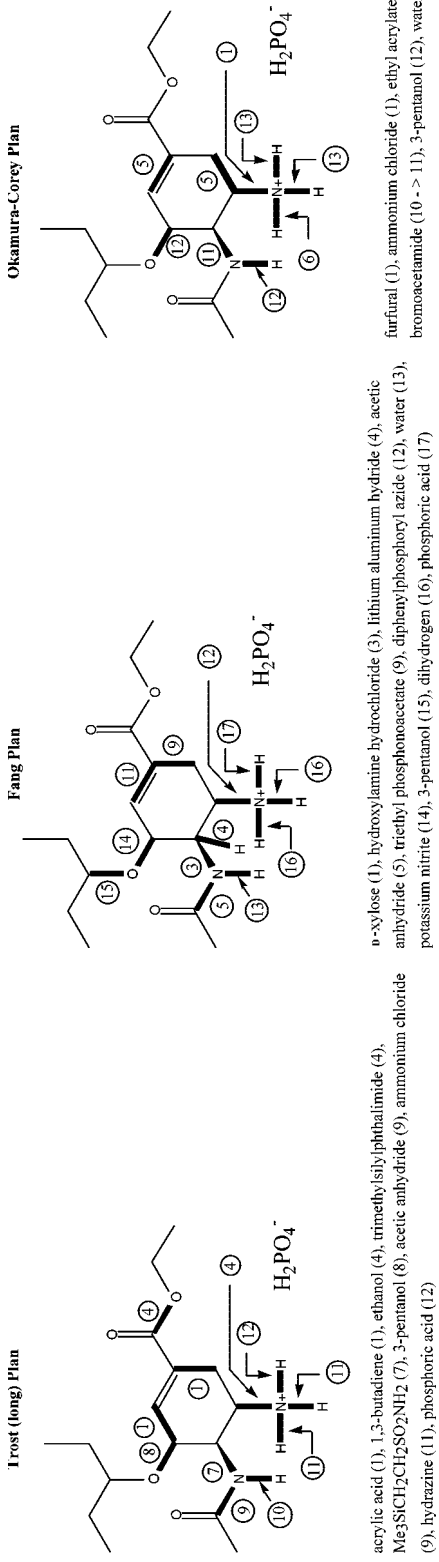
(g) The Okamura-Corey and Roche G4 plans have f_{match} equal to 0.80.

Scheme 3 depicts the most frequent target bonds made over the 15 plans examined. It is clear that almost all of the plans involve installation of the 3-pentyloxy, *N*-acetyl, and amino groups around a six-membered ring core starting material.

Figure 6 shows the target bond-forming reaction profiles. The Roche G3, Roche G2, Corey, Trost (long), Roche G1, Fang, Gilead, Okamura, Kann, and Shibasaki G2 plans achieve their respective target bonds in the late stages. The Fukuyama plan shows target bond-making activity in the middle stages, whereas the Shibasaki G3 and Roche G5 target bond distributions are bimodal with activity in the early and late stages, and middle and late stages, respectively. The Roche G4 and Shibasaki G1 plans show the most even distribution of target bond activity over the course of their respective plans. The gaps in these profiles may be directly correlated with the use of sacrificial reagents as noted in the synthesis bottlenecks shown in Table 3. The Shibasaki G3 and Roche G5 plans have the highest number of target bonds made per reaction step at 1.27 and 1.36, respectively. Graphs A and B of Figure 7 show correlations between molecular weight fraction of sacrificial reagents, *f*(sac), and AE and between *f*(sac) and number of target bonds per reaction step, *B*/*M*. Both plots show a general trend that as *f*(sac) decreases both AE and *B*/*M* tend to increase.

Figure 8 shows the hypsicity or oxidation level profiles and corresponding HI index values. The isohypsic condition is nearly met for the Roche G3, Roche G2, and Roche G1 plans. The Corey and Roche G4 plans show a monotonic increase in hypsicity and the Fukuyama does so approximately. The Fang plan has the highest positive HI, whereas the Shibasaki G1, Okamura, Roche G4, Fukuyama, Trost (long), and Corey plans have negative HI values, with the Fukuyama being the most negative.

The Gilead plan involves an additive reduction with dimethylsulfide borane complex (step 6) and additive catalytic hydrogenation with loss of nitrogen gas (step 12). The Roche G1 plan involves an additive reduction with dimethylsulfide borane complex (step 6) and an additive catalytic reduction using ethanolamine with loss of a protecting allyl group on the C5-amino group (step 12). The Roche G3 plan involves an additive reduction with triethylsilane (step 5) and an additive catalytic reduction with loss of two allyl protecting groups on the C5-amino group using *N*,*N*-dimethylbarbituric acid (step 13). The Roche G2 plan involves an additive reduction with dimethylsulfide borane complex (step 6) and an additive catalytic reduction with loss of two allyl protecting groups on the C5-amino group using *N*,*N*-dimethylbarbituric acid (step 14). The Roche G5 plan involves an additive catalytic asymmetric hydrogenation (step 5) and an additive catalytic hydrogenation with loss of nitrogen gas (step 11).



acrylic acid (1), 1,3-butadiene (1), ethanol (4), trimethylsilylphthalimide (4), $Me_3SiCH_2CH_2SO_2NH_2$ (7), 3-pentanol (8), acetic anhydride (9), ammonium chloride (9), hydrazine (11), phosphoric acid (12)

D-xylose (1), hydroxylamine hydrochloride (3), lithium aluminum hydride (4), acetic anhydride (5), triethyl phosphonoacetate (9), diphenylphosphoryl azide (12), water (13), potassium nitrite (14), 3-pentanol (15), dihydrogen (16), phosphoric acid (17)

furfural (1), ammonium chloride (1), ethyl acrylate (5), 1/4 sodium borohydride (6), N-bromoacetamide (10 - > 11), 3-pentanol (12), water (13), phosphoric acid (13)

Figure 5 Continued.

Table 4. **Summary of strategies to install four substituents on the six-membered ring of oseltamivir**

plan	alkoxy group at C3	amino group at C5	acetylamino group at C4	carboxyethyl group at C1
Roche G1	- 3-pentanone - quinic acid controls stereochemistry - swapping of diol protecting group on quinic acid	- allylamine - stereochemistry controlled via 1,2-walk across ring via aziridation	- acetic anhydride - direct S_N2 attack of allylamine via aziridine ring opening	- quinic acid (C=O) - ethanol (OEt)
Roche G2	- 3-pentanone - quinic acid controls stereochemistry - swapping of diol protecting group on quinic acid	- diallylamine - direct S_N2 attack via aziridine ring opening	- acetic anhydride - *tert*-butylamine - stereochemistry controlled via 1,2-walk across ring via aziridation	- quinic acid (C=O) - ethanol (OEt)
Roche G3	- 3-pentanone - shikimic acid controls stereochemistry - swapping of diol protecting group on quinic acid	- diallylamine - direct S_N2 attack via aziridine ring opening	- acetic anhydride - *tert*-butylamine - stereochemistry controlled via 1,2-walk across ring via aziridation	- shikimic acid (C=O) - ethanol (OEt)
Roche G4	- 3-pentanol[a]	- allylamine - direct S_N2 attack via displacement of mesylate group	- acetic anhydride - diphenylphosphorylazide via endo aziridation	- ethyl acrylate
Roche G5	- 3-pentanol - stereochemistry controlled via asymmetric hydrogenation	- diphenylphosphorylazide - Curtius rearrangement	- acetic anhydride - sodium azide - direct S_N2 attack via displacement of triflate group	- carbon monoxide (C=O) - ethanol (OEt)
Gilead	- 3-pentanone - quinic acid controls stereochemistry - swapping of diol protecting group on quinic acid	- sodium azide - direct S_N2 attack via aziridine ring opening	- acetic anhydride - sodium azide - stereochemistry controlled via 1,2-walk across ring via aziridation	- quinic acid (C=O) - ethanol (OEt)
Shibasaki G1	- 3-pentanol[a]	- trimethylsilylazide - direct attack - stereochemistry controlled via Shibasaki chiral auxiliary	- acetic anhydride - sodium azide- stereochemistry controlled via 1,2-walk across ring via aziridation	- trimethylsilylcyanide (C=O) - water (C=O) - ethanol (OEt)
Shibasaki G2	- 3-pentanol[a]	- sodium azide- stereochemistry controlled via 1,2-walk across ring via aziridation	- thioacetic acid - trimethylsilylazide- stereochemistry controlled via Shibasaki chiral auxiliary	- lithium cyanide (C=O) - water (C=O) - ethanol (OEt)
Shibasaki G3	- 3-pentanol[a]	- trimethylsilylazide - Curtius rearrangement	- acetic anhydride - trimethylsilylazide - Curtius rearrangement	- trimethylsilylcyanide (C=O) - water (C=O) - ethanol (OEt)
Corey	- 3-pentanol[a]	- ammonia- 1,3-walk across ring via lactam	- *N*-bromoacetamide	- trifluoroethylvinylketone(C=O) - ethanol (OEt)
Trost	- 3-pentanol[a]	- trimethylsilylphthalimide - direct S_N2 attack via lactone ring opening using Trost ligand to control stereochemistry	- acetic anhydride - $Me_3SiCH_2CH_2SO_2NH_2$- stereochemistry controlled via Du Bois chiral catalyst	- acrylic acid (C=O) - ethanol (OEt)
Kann	- 3-pentanol[a]	- *tert*-butylcarbamate - direct attack - stereochemistry controlled via $Fe(CO)_3$ group	- acetic anhydride - sodium azide - stereochemistry controlled via 1,2-walk across ring via aziridation	- $BrCH_2$-CH=CH-COOEt
Fukuyama	- 3-pentanol[a]	- NH_3 - Curtius rearrangement	- acetic anhydride - pyridine ring	- pyridine ring (C=O) - $NaIO_4$ oxidation (C=O) - ethanol (OEt)
Okamura	- 3-pentanol[a]	- ammonium chloride via pyridone synthesis from furfural	- *N*-bromoacetamide	- ethyl acrylate
Fang	- 3-pentanol via trichloromethylimidate intermediate - stereochemistry controlled by epimerization via sequential reaction with triflic anhydride and potassium nitrite	- diphenylphosphorylazide - direct S_N2 attack via displacement of OH group under Mitsunobu conditions	- acetic anhydride - hydroxylamine hydrochloride direct attack - stereochemistry controlled via asymmetric reduction of oxime intermediate	- $(EtO)_2(P=O)CH_2COOEt$ via Horner−Wittig−Emmons reaction

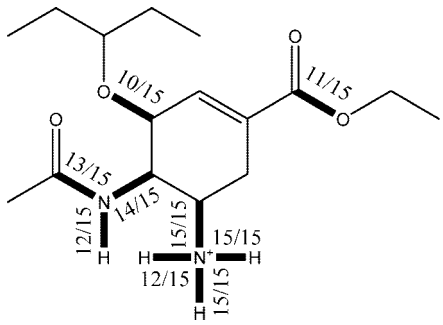
[a] Stereochemistry at C3 controlled via aziridation.

The Fang plan involves a subtractive oxidation with pyridinium dichromate (step 3), a stereoselective additive reduction with lithium aluminum hydride (step 4), an additive catalytic hydrogenation (step 10), and an additive catalytic hydrogenation with loss of nitrogen gas (step 16). The Fukuyama plan involves an additive reduction with sodium borohydride (step 1), additive oxidation with sodium chlorite (step 3), additive catalytic hydrogenation with loss of carbon dioxide and toluene (step 4), an additive oxidation with sodium periodate (step 5), and an additive catalytic reduction with loss of carbon dioxide and an allyl protecting group on the C5-amino group using *N*,*N*-dimethylbarbituric acid (step 13). The Kann plan involves additive oxidation with diiron nonacarbonyl to protect a cyclic diene (step 3), subtractive oxidation with trityl hexafluorophosphate via hydride transfer (step 4), subtractive oxidation with hydrogen peroxide to deprotect an iron tricarbonyl chiral discriminating group (step 9), and an additive oxidation with *m*-chloroperbenzoic acid (step 10). The Shibasaki G1 plan involves an additive oxidation with *m*-chloroperbenzoic acid (step 1), sequential additive oxidation with selenium dioxide and subtractive oxidation with Dess-Martin periodinane (step 7), and additive reduction with lithium tri-*tert*-butoxyaluminum hydride (step 9). The Shibasaki G2 plan involves an additive oxidation with *m*-chloroperbenzoic acid (step 1) and a subtractive oxidation with Dess-Martin periodinane (step 10). The Shibasaki G3 plan involves an additive stereoselective reduction with lithium tri-t-butoxyaluminum hydride (step 7). The Trost plan involves an additive oxidation with *m*-chloroperbenzoic

Table 5. **Pairwise matching frequencies of target bonds for various plans to oseltamivir phosphate**

plan	Corey	Fang G1	Fukuyama	Gilead	Kann	Okamura-Corey	Roche G1	Roche G2	Roche G3	Roche G4	Roche G5	Shibasaki G1	Shibasaki G2	Shibasaki G3	Trost-long
Corey		0.571	0.615	0.538	0.583	0.583	0.538	0.462	0.538	0.462	0.389	0.571	0.615	0.500	**0.909**
Fang G1			0.533	0.571	0.615	0.615	0.571	0.500	0.571	0.615	0.500	0.500	0.533	0.444	0.643
Fukuyama				0.500	0.538	0.429	0.400	0.429	0.500	0.429	0.368	0.643	0.692	0.563	0.692
Gilead					0.462	0.462	**1.000**	**0.900**	**0.100**	0.462	0.563	0.571	0.615	0.500	0.615
Kann						0.500	0.462	0.500	0.462	0.636	0.333	0.500	0.538	0.533	0.667
Okamura-Corey							0.462	0.385	0.462	**0.800**	0.333	0.500	0.538	0.533	0.538
Roche G1								**0.900**	**1.000**	0.462	0.563	0.571	0.615	0.500	0.615
Roche G2									0.900	0.500	0.500	0.500	0.538	0.438	0.538
Roche G3										0.462	0.563	0.571	0.615	0.500	0.615
Roche G4											0.333	0.500	0.538	0.533	0.538
Roche G5												0.588	0.529	0.526	0.444
Shibasaki G1													**0.917**	**0.857**	0.643
Shibasaki G2														**0.786**	0.692
Shibasaki G3															0.563
Trost-long															

Scheme 3. **Target bond-making frequencies for all 15 plans to oseltamivir phosphate**



acid (step 6). The Okamura−Corey plan involves an additive reduction with sodium borohydride (step 6), a subtractive oxidation with sodium periodate (step 8), and an additive reduction with sodium borohydride (step 9). The Corey and Roche G4 plans do not involve formal redox reactions.

Figure 9shows the molecular weight first-moment profiles for all 15 plans. The following plans show that intermediates are made along the way, having molecular weights exceeding the target molecular weight of 410 g/mol: Roche G2, Roche G3, Roche G4, Roche G5, Corey, Trost (long), Fang, Fukuyama, and Kann. For these plans, the sacrificial or protecting groups that are responsible are itemized below:

(a) Roche G2 plan: diallyl protecting groups in step 10 and hydrochloride salt formation in step 12;

(b) Roche G3 plan: diallyl protecting group in step 9 and hydrochloride salt formation in step 11;

(c) Roche G4 plan: methanesulfonyl group in step 5;

(d) Roche G5 plan: triflate group in step 9;

(e) Corey plan: bromo group in step 8;

(f) Trost plan: phenylsulfide group in step 5 and 2-trimethylsilylethanesulfonate group in 7;

(g) Fang plan: triflate group in step 8 and diethylphosphonate group in step 9;

(h) Fukuyama plan: methanesulfonyl group in step 7 and carboxyallyl group in step 8;

(i) Kann plan: triphenylphosphino group in step 1, iron tricarbonyl and hexafluorophosphate groups in steps 4 and 7, 2(*S*)-phenyl-cyclohexyloxy resolving group in step 5, and methanesulfonyl group in step 12.

The rest of the plans do not show any overshoots above the target product molecular weight.

As noted in Shibasaki's excellent review, "although oseltamivir is a relatively small molecule, developing a practical synthesis that can satisfy worldwide demand in an environmentally friendly and safe way is quite challenging."[25a] The key constraints to consider in the target structure are that the three consecutive substituents at C3, C4, and C5 are electronegative in nature and that they alternate in orientation at the stereogenic centers. Consequently, these three consecutive carbon atoms making up the six-membered ring will likely originate from electrophilic fragments unless their electronic states are changed along the way via redox reactions. The important challenge is to solve the problem of introducing these groups sequentially via epimerizations and redox changes with the least number of steps since these kinds of reactions not only add extra steps to the overall plan but also they produce significant byproducts. The internal carbon−carbon double bond in the ring is strongly suggestive of a Diels−Alder strategy if ring formation is chosen as a strategy. This observation was exploited in several of the plans discussed.

One observation that can be made from the Gilead and Roche plans is that one step could possibly be saved if the 3-pentyl group is used directly instead of the 2-propyl group to protect the *cis*-diol of quinic acid or shikimic acid. Scheme 4 shows a proposed plan using shikimic acid as a starting material. As noted before, starting the plan from shikimic acid instead of quinic acid saves the step of dehydration to install the internal carbon−carbon double bond. Such a change under the assumption that the reaction yield performance for the 3-pentyl protecting group step is the same as that for the 2-propyl group at 93% would translate into a linear plan of 12 steps and 18 input materials with the following kernel material efficiency metrics for overall reaction yield, atom economy, kernel reaction mass efficiency, and *E*-total: 39.4%, 21.6%, 12.0%, and 222, respectively. Values for the synthesis efficiency metrics *f*(sac), HI, and *B*/*M* are 41.6%, +0.31, and 0.83, respectively. When these results are compared with metrics for the best-performing Roche G3 plan given in Tables 1 and 2, we see that they show marginal improvements, but they are improvements nonetheless.

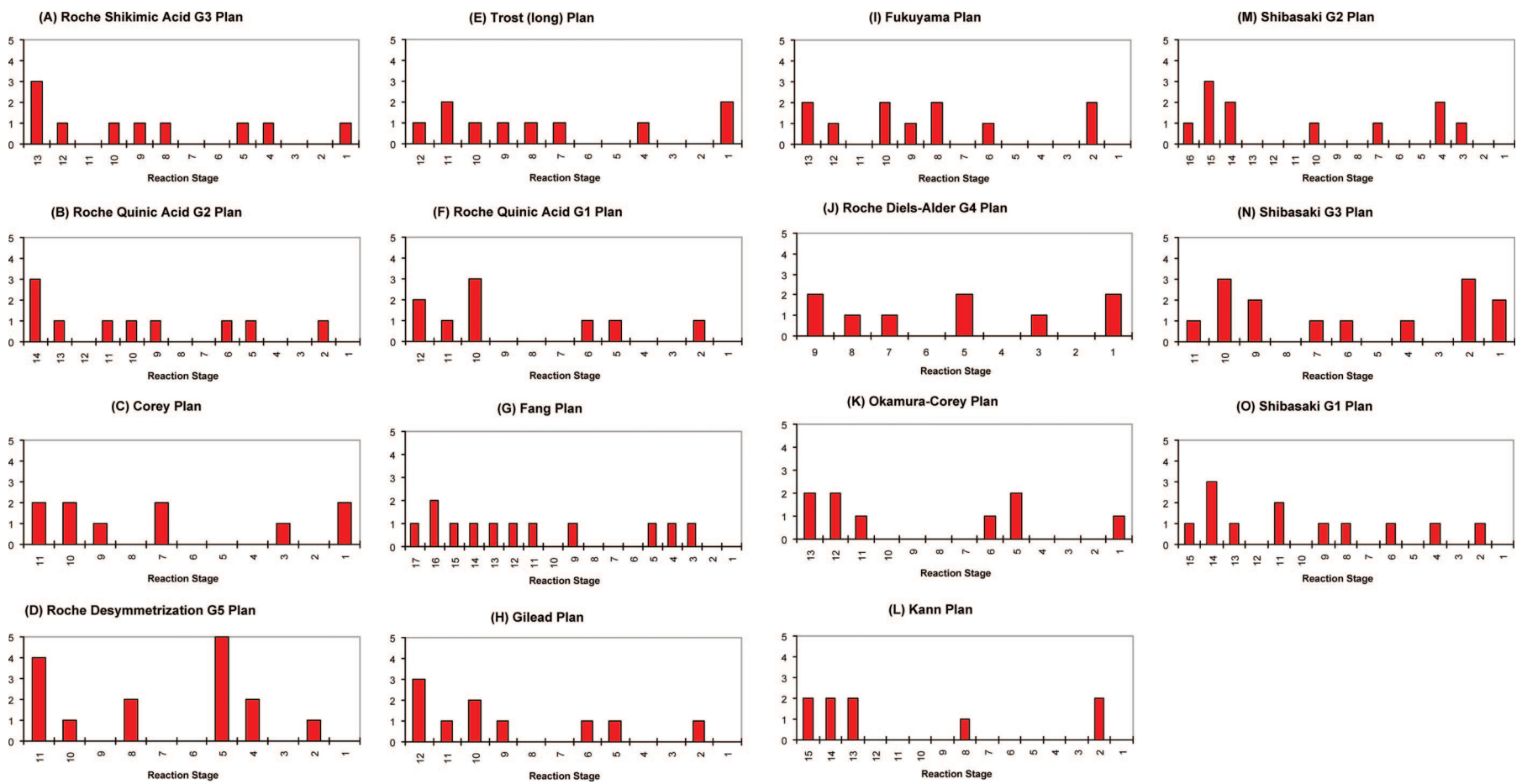


Figure 6. **Target bond-forming reaction profiles for various plans to oseltamivir phosphate. Ordinate is number of target bonds made.**

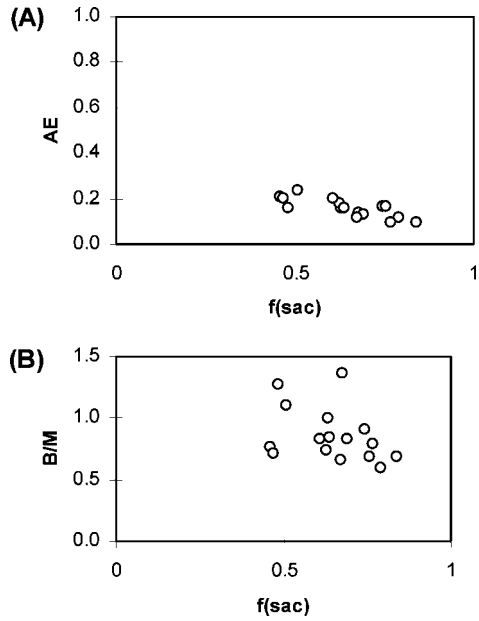


Figure 7. **(A) Plot of AE vs *f*(sac). (B) Plot of *B/M* vs *f*(sac).**

Since this scenario hinges on the performance of the assumed reaction, this could of course change dramatically in the negative direction if it were not to work as anticipated.

Note Added in Proof

A recently published route[26] has appeared which begins from shikimic acid and goes to the third structure shown in Scheme 4 in one step, thus validating further predictions for optimizing the synthesis of oseltamivir phosphate as described here. This modification improves the Roche G3 route so that the following new parameters are obtained:

(a) 11 linear steps (reduction of 2 steps);
(b) 18 input materials (reduction of 1 input material);
(c) atom economy = 20.2% (lowered by 4%);
(d) overall yield = 43.8% (raised by 12%);
(e) kernel RME = 11.3% (lowered by 2%);
(f) *E*-kernel = 7.9 (raised by 3%);
(g) *E*-excess = 15.7 (lowered by 36%);
(h) *E*-aux = 179.9 (lowered by 9.4%);
(i) *E*-total = 203.5 (lowered by 12%);
(j) *f*(sac) = 0.476 (raised by 4%); (k) HI = + 0.33; and (l) *B*/*M* = 0.91 (raised by 18%).

Fang and coworkers[27] have continued to improve their synthesis of oseltamivir phosphate in their recent report that describes a second-generation azide route using diphenylphosphorylazide and a third-generation azide-free route involving introduction of the amino group in the ring using tributylammonium cyanate followed by rearrangement to an isocyanate intermediate. Both new routes are linear and start by stereospecific enzymatic oxidation of bromobenzene to a catechol intermediate using *Pseudomonas putida.* For the azide Fang G2 route the following parameters are obtained in comparison with the Fang G1 route:

(a) 12 reaction stages (reduction of 5 stages);
(b) 12 reaction steps (reduction of 6 steps);
(c) 20 input materials (reduction of 15 input materials);
(d) atom economy = 14.7% (raised by 23%);
(e) overall yield = 26.5% (raised by 98%);
(f) kernel RME = 7.4 (raised by 139%);
(g) *E*-kernel = 12.6 (lowered by 59%);
(h) *E*-excess = 59.1 (lowered by 78%);
(i) *E*-aux = 1656.7 (lowered by 27%);
(j) *E*-total = 1728.3 (lowered by 33%);
(k) *f*(sac) = 0.438 (lowered by 35%);
(l) HI = +1.39;
(m) *B*/*M* = 0.75 (raised by 12%).

For the azide-free Fang G3 route the following parameters are obtained in comparison with the Fang G1 route:

(a) 12 reaction stages (reduction of 5 stages);
(b) 12 reaction steps (reduction of 6 steps);
(c) 21 input materials (reduction of 14 input materials);
(d) atom economy = 17.3% (raised by 44%);
(e) overall yield = 22.5% (raised by 68%);
(f) kernel RME = 7% (raised by 126%);
(g) *E*-kernel = 13.3 (lowered by 57%);
(h) *E*-excess = 104.3 (lowered by 62%);
(i) *E*-aux = 1858.6 (lowered by 18%);
(j) *E*-total = 1976.2 (lowered by 23%);
(k) *f*(sac) = 0.544 (lowered by 19%);
(l) HI = +1.39;
(m) *B*/*M* = 0.75 (raised by 12%).

As can be easily seen from these results the new Fang routes show the most dramatic improvements made by any group and thus are very competitive with the leading industrial routes in terms of material efficiency.

4. Conclusions

The present work has achieved the following points:

(a) A powerful algorithm has been introduced which allows facile computation of key kernel and global material efficiency metrics for any kind of synthesis plan regardless of complexity.

(b) An expression for the global *E*-factor applicable to any kind of plan has been derived which shows the individual contributors to waste production during the course of an entire synthesis plan.

(c) Microsoft Excel spreadsheets have been introduced for radial pentagon analyses applicable to assessing individual reaction performances and for linear and convergent plans applicable to assessing overall plan performance.

(d) The above-mentioned spreadsheets allow complete automation in computation and proofreading of synthesis plans and have several built-in check calculation features to ensure that correct results are obtained.

(e)Target bond-forming maps are introduced as powerful tools to understand and measure synthesis strategy efficiency.

(f) A number of visual aids have been introduced which assist chemists in making informed decisions regarding both material and synthesis strategy and also allow deeper critiques of synthesis plans.

(g) All of the above have been successfully demonstrated on the test compound oseltamivir phosphate.

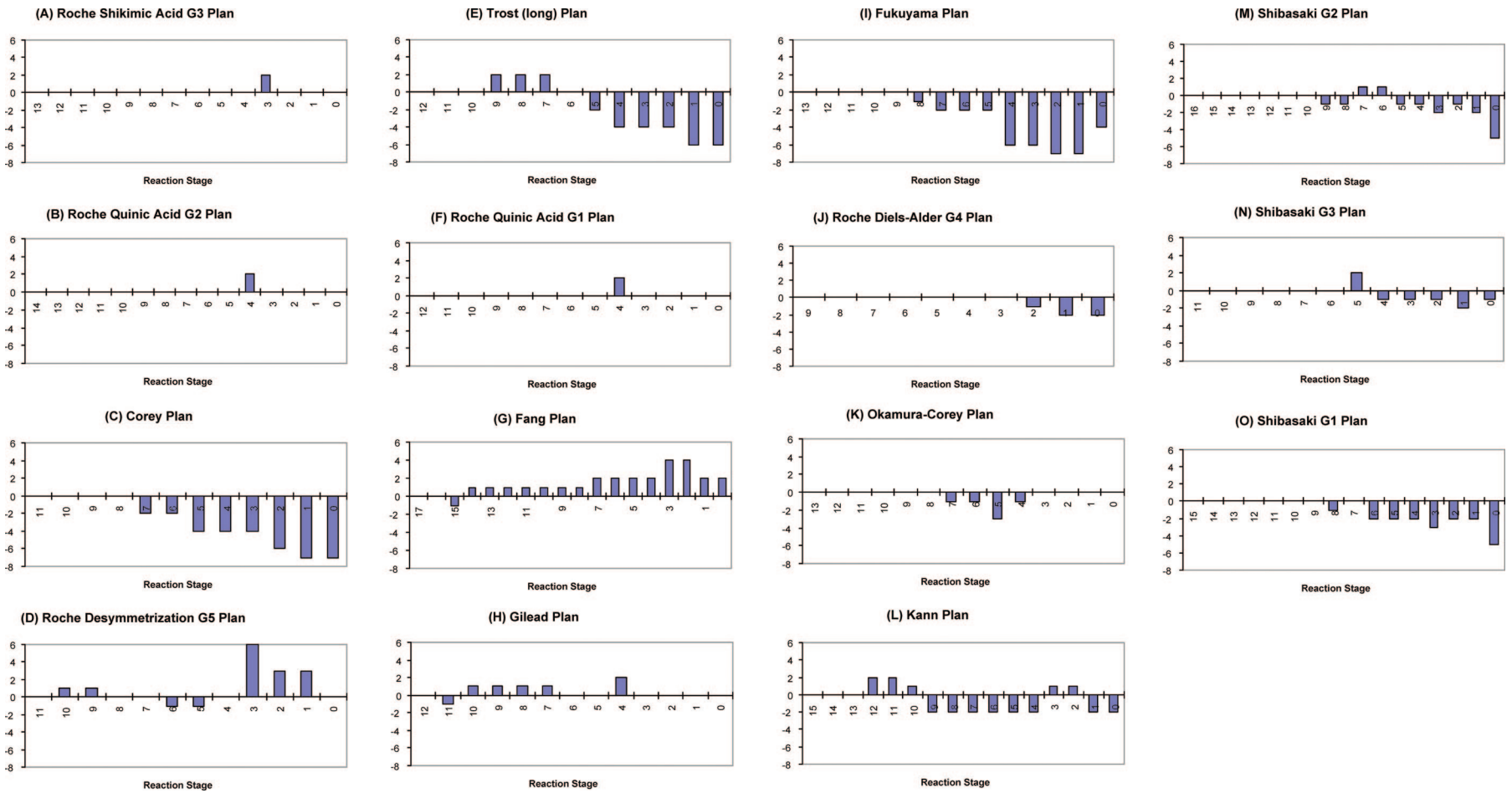


Figure 8. **Hypsicity profiles for various plans to oseltamivir phosphate: (A) HI = +0.14; (B) HI = +0.13; (C) HI = −3; (D) HI = +1; (E) HI = −1.54; (F) HI = +0.54; (G) HI = +1.44; (H) HI = +0.39; (I) HI = −2.64; (J) HI = −0.5; (K) HI = −0.43; (L) HI = −0.56; (M) HI = −0.71; (N) HI = −0.33; and (O) HI = −1.19.**

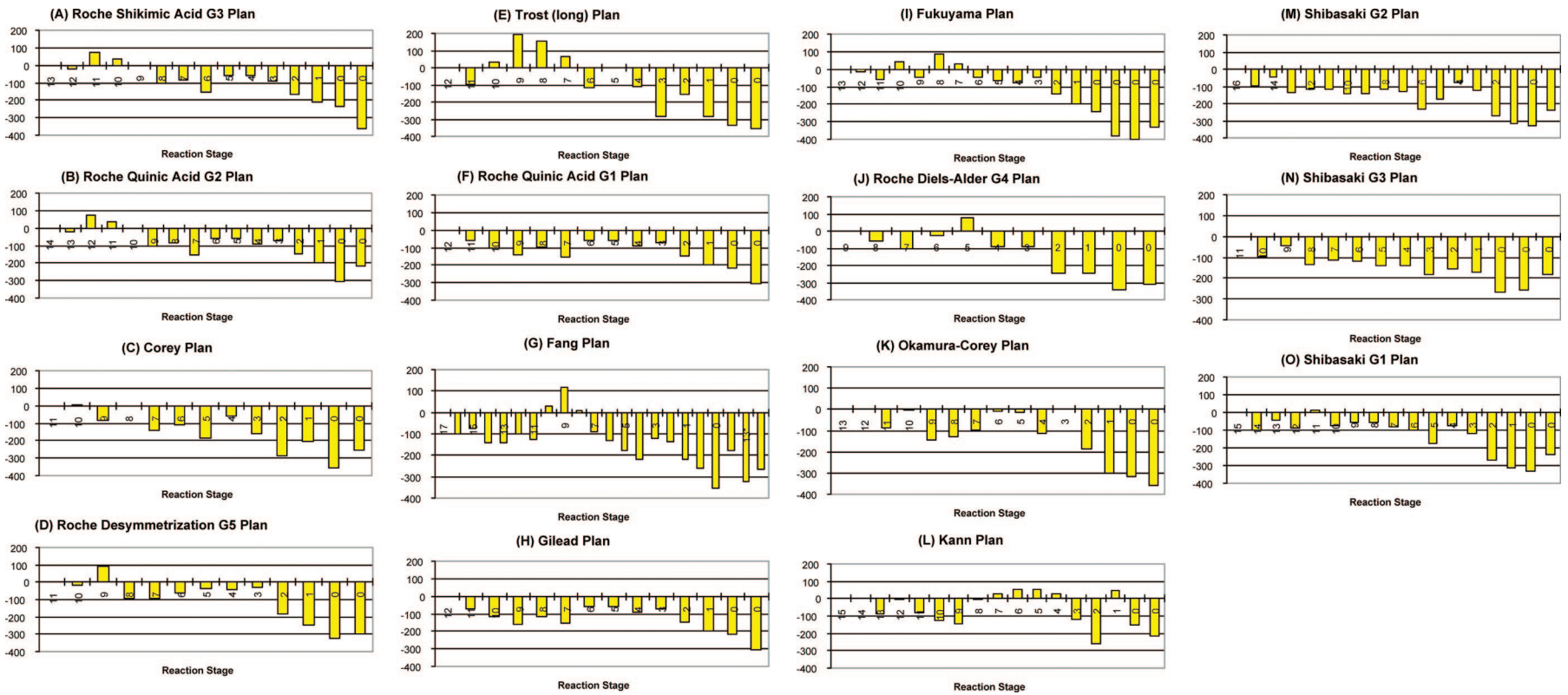
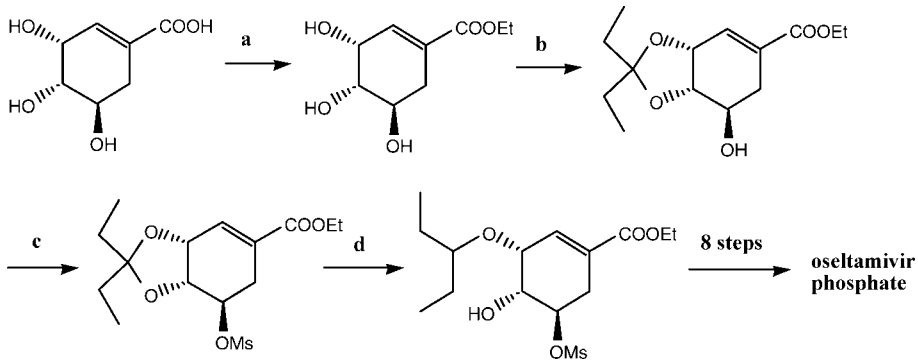


Figure 9. **Molecular weight first moment profiles for various plans to oseltamivir phosphate: (A) $\mu_1 = -102.68$; (B) $\mu_1 = -93.57$; (C) $\mu_1 = -153.68$; (D) $\mu_1 = -110.48$; (E) $\mu_1 = -99.92$; (F) $\mu_1 = -131.23$; (G) $\mu_1 = -167.02$; (H) $\mu_1 = -135.85$; (I) $\mu_1 = -133.48$; (J) $\mu_1 = -142.56$; (K) $\mu_1 = -124.76$; (L) $\mu_1 = -61.26$; (M) $\mu_1 = -163.86$; (N) $\mu_1 = -167.26$; (O) $\mu_1 = -132.10$ g/mol/stage.**

Scheme 4. **Proposed shortened Roche G3 plan**[a]



[a] Reagents: (a) $SOCl_2$, EtOH (93%); (b) 3,3-dimethoxypentane, $TsOH \cdot H_2O$ (cat.) (93%, assumed); (c) MsCl, Et_3N (93%); (d) Et_3SiH, $TiCl_4$ (cat.), then H_2O (86%).

In order for these protocols to be implemented as standard practice by synthetic and process chemists in their ongoing work to optimize reactions and synthesis plans and for the principles of green chemistry to be widely accepted among these scientists, key recommendations need to be put forward. The most important of these is that a major overhaul in the reporting of synthesis procedures in the literature needs to be made. It is up to editors and reviewers of leading journals publishing original research in synthetic chemistry to clamp down on omissions, errors, or casual reporting of experimental write-ups that lack full disclosure of all aspects of material usage when authors make claims that significant advances in synthetic efficiency have been achieved. Supporting Information is unfortunately often not reviewed with the same care and rigor as accompanying manuscripts. If any claim of "greenness" or synthetic efficiency is made, it is imperative that such a claim be backed up with some kind of metrics analysis; otherwise that claim cannot be properly validated. Claims made simply on the basis of overall yield performance and step count are clearly insufficient.

The future success of applying green chemistry thinking to synthesis design and optimization will therefore depend on:

(a) making use of multicomponent and tandem one-pot reactions as central theme reactions in synthesis planning;

(b) disclosing all material and energy consumption parameters for each reaction in a given plan so that true assessments of global RME and *E*-factors can be made;

(c) reporting green metrics as proof of greenness and part of the standard protocol in reports of synthesis plans in the literature especially if plans are advertised as "green-er" than prior published plans;

(d) realizing that the concept of "greenness" is not absolute; rather it is comparative, and so it is more correct to state that a given plan is "green-er" than other competing plans; clearly, deeper understandings of strategy are achieved when several plans to a common target are studied in detail;

(e) accepting that optimization is a continuous ongoing iterative exercise, that ranking of plans is the inevitable consequence of metrics analysis, and that it absolutely does not degrade the value of any disclosed plan for it is only through the exploration of several routes that the true optimum will be found;

(f) accepting that optimization is a multivariable problem and that claims of achievements of optimization to a given target molecule are legitimate when the magnitudes of all metrics parameters in the set of plans considered are appropriately maximized and minimized and appear in the *same* plan;

(g) realizing that, when a new target molecule is desired to be made with no prior published guidance available, it is necessary to go through poor-performing plans before hitting on the "right" one;

(h) realizing that quantitative assessment of synthesis plans has an important role in deciding which may be good candidate plans to pursue;

(i) changing the well-worn paradigm of designing synthesis plans around a fixed type of reaction to one that uses material and synthetic efficiency as the uppermost constraints in the choice of the set of reactions ultimately selected for a plan from the entire pool of documented organic reactions known; and

(j) being especially careful in using the phrase "readily available" to describe starting materials chosen for a synthesis plan since apparent claims of greenness or synthetic efficiency may be artificially inflated.

Some caveats to bear in mind when carrying out green metrics analyses include:

(a) generalizations must be taken with caution about relative efficiencies of linear versus convergent routes, short plans versus longer ones, or stereoselective versus racemic with resolution, as there exist plenty of examples in the literature when counterintuitive results occur because gains made in one set of parameters are lost in others;

(b) the synthesis plans of specialized catalysts or solvents used, such as chiral catalysts, ligands, and ionic liquids, must also be worked out using separate synthesis tree diagrams with appropriate mole scaling factors as part of the overall assessment of material efficiency to a given target molecule;

(26) Carr, R.; Ciccone, F.; Gabel, R.; Guinn, M.; Johnston, D.; Mastriona, J.; Vandermeer, T.; Groaning, M. *Green Chem.* **2008**, *10*, 743.
(27) Shie, J. J.; Fang, J. M.; Wong, C. H. *Angew. Chem., Int. Ed.* **2008**, *47*, 5788.

(c) when using the radial pentagon analysis for single reactions, one must be aware of determining correct mole amounts for reagents that are reported in procedures such as solutions given in terms of weight percent, and properly assigning the role of each material used in the right place in the template spreadsheet;

(d) no claims of greenness can be made if only one plan exists for a given target molecule;

(e) claims of greenness cannot be made based on one criterion, such as the use of a "green" solvent for example; and

(f) once green metrics analyses are done on a set of literature procedures to a given target molecule and it is found in the ranking process that optimized parameters are scattered over a number of plans, it is imperative that the next disclosed plan should strive to demonstrate clear improvements over the currently reported highest ranked plan.

Acknowledgment

This paper is dedicated to the memory of Dr. Christopher R. Schmid, a man I did not know or meet, who was the first person to recognize the value of quantitative analysis to total synthesis design, and who took a risk in giving me an opportunity to disclose my work on green metrics in this journal, thereby opening a new door in my career as a chemist.

Supporting Information Available

PENTAGON, LINEAR-kernel, LINEAR-complete, CONVERGENT-kernel, and CONVERGENT-complete Microsoft Excel template files; detailed instructions and notes for spreadsheet usage; synthesis schemes, synthesis trees, and radial pentagons for all 15 plans to oseltamivir phosphate; explanatory note on errors in Auge analysis; Microsoft Excel files for radial pentagon and synthesis spreadsheet analyses for sildenafil synthesis. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review July 2, 2008.

OP800157Z