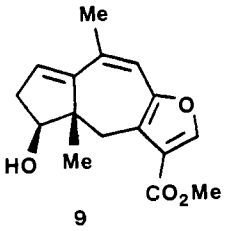
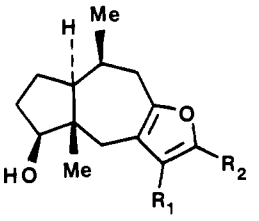
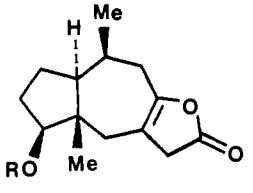


9

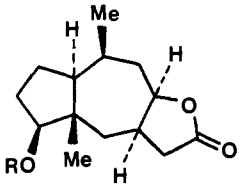


10a, R_1 = COOMe; R_2 = H
b, R_1 = COOH; R_2 = H
c, R_1 = R_2 = H
d, R_1 = H; R_2 = $SiMe_3$

Saponification of **10a** produces the carboxylic acid **10b** (mp 152–153 °C), and this is decarboxylated with copper in refluxing quinoline to give **10c**. Treatment of a THF solution of **10c** with *n*-butyllithium (2 equiv) and excess trimethylsilyl chloride followed by aqueous acid workup gives the (trimethylsilyl)furan **10d** in 82% overall yield from **10a**. The conversion of **10d** to enol lactone **11a**



11a, R = H
b, R = Ac

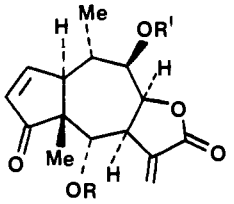


12a, R = H
b, R = Ac

is accomplished by the peracetic acid oxidation method of Kuwajima and Urabe.[10] Acetylation of **11a** with acetic anhydride/pyridine/4-(dimethylamino)pyridine gives **11b**, identical with previously reported **11b**[1,4] in all respects. It is important to note that in contrast to other syntheses that involve **11b**, this oxidative method of enol lactone elaboration gives pure **11a** (and, therefore, pure **11b**) uncontaminated by the isomeric α,β butenolide.

Hydrogenation of **11b** with rhodium on alumina in ethyl acetate at 60 psi[4] gives crystalline **12b** (mp 109–110 °C, lit. mp 110–111 °C).[4] In a more direct formal total synthesis of confertin, enol lactone **11a** is hydrogenated ($Rh{\cdot}Al_2O_3$ in ethyl acetate) to give lactone alcohol **12a** (white foam) in 89% isolated yield, identical with previously reported **12a**[1,4] in all respects.

As a result of this work, we have established the furan **4** based annelation approach to the pseudoguaianolide sesquiterpene lactones. We intend to apply this chemistry to the synthesis of other pseudoguaianes, with the fastigilins (**13**) representing our



13

ultimate goal. The correct placement of the C(9) oxygen atom and the potentially reactive methylene group at C(4) in the key tricyclic dienedione **8** should provide adequate functionality for synthesis of **13**.

Acknowledgment. This work was supported by the National Cancer Institute (CA 25787). We thank Professor R. H. Schlessinger for providing us with 100- and 400-MHz ^1H NMR spectra of our synthetic **12b**.

(10) Kuwajima, I.; Urabe, H. *Tetrahedron Lett.* **1981**, *22*, 5191.

Registry No. 1, 60426-81-5; **2**, 83096-96-2; **3**, 83096-97-3; **4**, 83096-98-4; **5a**, 56522-18-0; **5b**, 83096-99-5; **6**, 83097-00-1; **7**, 83097-01-2; **8**, 83097-02-3; **9**, 83097-03-4; **10a**, 83097-04-5; **10b**, 83097-05-6; **10c**, 83097-06-7; **10d**, 83097-07-8; **11a**, 83097-08-9; **11b**, 72341-84-5; **12a**, 72341-86-7; **12b**, 72341-85-6; 2-methyl-1,3-cyclopentanedione, 765-69-5.

Supplementary Material Available: Listing of spectra data for all new compounds prepared in this work (4 pages). Ordering information is given on any current masthead page.

Convergence, Molecular Complexity, and Synthetic Analysis

Steven H. Bertz

Bell Laboratories, Murray Hill, New Jersey 07974

Received January 8, 1982

Consideration of molecular complexity[1] results in a clearer understanding of convergence and leads to the more general principle of minimization of molecular complexity for synthetic planning. Convergence is one of the most valuable heuristics organic chemists have to aid in planning the sophisticated syntheses that are the hallmark of modern organic chemistry. In the 2 decades since Velluz introduced the "convergent" synthesis,[2] "doubly convergent"[3] and "triply convergent"[4] syntheses have been devised. If Fuchs' definition of triple convergence is adopted for a process in which three components are brought together in one step, then an ordinary convergent synthesis in which two components are brought together at some stage would be termed "doubly convergent". But then the "doubly convergent" synthesis of Carrupt and Vogel[3] should be called "doubly, doubly convergent", as two components are joined in two different steps. Add to these the definitions[5] of "partially convergent", "fully convergent", and "perfect convergency", and one sees that a simple, universal measure of the degree of probable efficiency in a synthetic plan is needed in order to provide an operational definition of this important concept as well as a numerical basis for the comparison of alternative synthetic routes, a plethora of which can be generated by computer.[6-11]

As summarized by Hendrickson,[5] the *qualitative* basis for the economy of a convergent synthesis is "the idea that when a reaction is carried out on an intermediate, it usually involves only one or two of the synthons that make up the intermediate so that the other, uninvolved synthons comprising the intermediate are subjected to needless waste from yield loss in the reaction." Most treatments of convergence assume equal yields for all reactions, which leads to the conclusion that the convergent synthesis is *always* more efficient than the corresponding linear one. Similarly, Hendrickson's proposed index of convergency, L_k, the sum of all the path lengths (number of steps) from all the starting materials

(1) Bertz, S. H. *J. Am. Chem. Soc.* **1981**, *103*, 3599. *J. Chem. Soc., Chem. Commun.* **1981**, 818. Merrifield, R. E.; Simmons, H. E. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 1329.
(2) Velluz, L.; Valls, J.; Mathieu, J. *Angew. Chem., Int. Ed. Engl.* **1967**, *6*, 778. Velluz, L.; Valls, J.; Nominé, G. *Ibid.* **1965**, *4*, 181.
(3) Carrupt, P.-A.; Vogel, P. *Tetrahedron Lett.* **1979**, 4533.
(4) Donaldson, R. E.; Fuchs, P. L. *J. Am. Chem. Soc.* **1981**, *103*, 2108.
(5) Hendrickson, J. B. *J. Am. Chem. Soc.* **1977**, *99*, 5439.
(6) Corey, E. J. *Q. Rev. Chem. Soc.* **1971**, *25*, 455. Corey, E. J.; Wipke, W. T. *Science (Washington, D.C.)* **1969**, *166*, 178.
(7) Gelernter, H.; Sridharan, N. S.; Hart, A. J.; Yen, S.-C.; Fowler, F. W.; Shue, H.-J. *Top. Curr. Chem.* **1973**, *41*, 113.
(8) Jochum, C.; Gasteiger, J.; Ugi, I. *Angew. Chem., Int. Ed. Engl.* **1980**, *19*, 495.
(9) Hendrickson, J. B. *Top. Curr. Chem.* **1976**, *62*, 49; *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 129.
(10) Salatin, T. D.; Jorgensen, W. L. *J. Org. Chem.* **1980**, *45*, 2043.
(11) Bersohn, M.; Esack, A. *Chem. Rev.* **1976**, *76*, 269 and refs. therein.

0002-7863/82/1504-5801$01.25/0

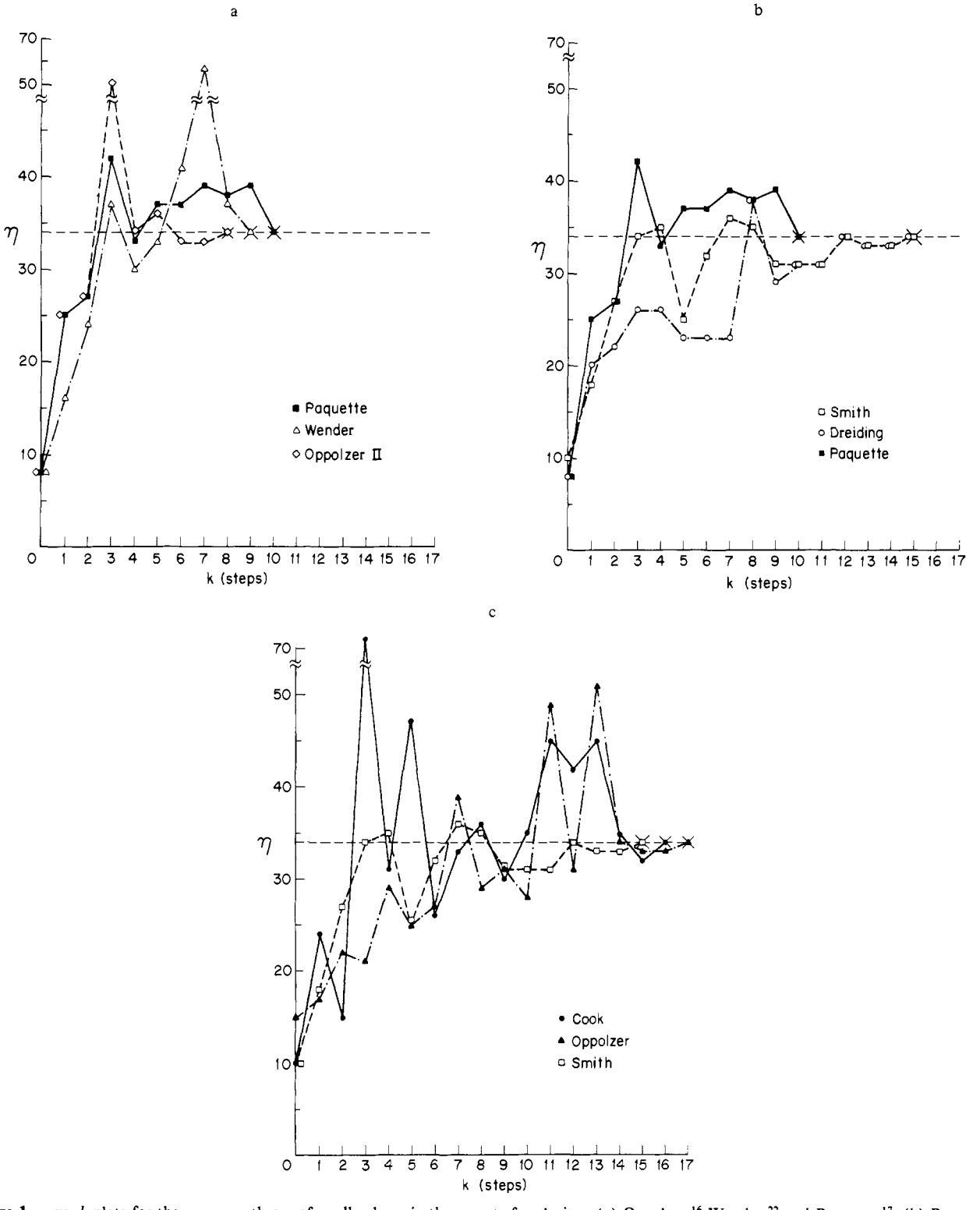


Figure 1. η vs. k plots for the seven syntheses of modhephene in three parts for clarity: (a) Oppolzer,[16] Wender,[22] and Paquette;[17] (b) Paquette,[17] Dreiding,[18] and Smith;[19] (c) Smith,[19] Oppolzer,[20] and Cook.[21] See also Table I.

to the target, is *always* lower for a convergent plan than for a linear one from the same starting materials. Nevertheless, Hendrickson[5] points the way to improvement as he continues, "Indeed the functionality present on the uninvolved synthons may contribute to yield loss through unwanted side reactions." In terms of molecular complexity, a more complex intermediate is *likely* to have more interfering functionality than a less complex one and therefore more opportunities for yield-decreasing side reactions. For a given state of the art, more steps will *probably* be required to make the more complicated molecule, and at a given level of complexity, more steps mean more chances for misadventure.

By changing to these statistical arguments, it is concluded that the synthetic route with the highest probability of success from a starting material of complexity C_0 to a target of complexity C_f is the one that minimizes the function

$$F = \int_0^{k_f} C(k)\, dk \qquad (1)$$

where the molecular complexity, C, is a function of the steps, k. This is simply the area under a plot of complexity vs. steps (e.g.,

Table I[a]

synthesis	$\Sigma\eta$	yield, %
Oppolzer II[16]	187	26.4
Wender[22]	235	8.2
Paquette[17]	265	6.3
Smith[19]	296	4.0 (6.1)[b]
Dreiding[18]	350	4.7
Cook[21]	508[c] (406)[d]	3.8[c] (9.4)[d]
(Oppolzer[20]	482	0.7)

[a] See supplementary material for data used to calculate values listed here. [b] Calculated by using Dreiding's last four yields (same intermediates). [c] Based on monocyclic starting material. [d] Based on tricyclic starting material.

see Figure 1). Since $C(k)$ is a polygonal function,[12] F can be expanded into the sum

$$F = \tfrac{1}{2}(C_0 + C_1)k_1 + \tfrac{1}{2}(C_1 + C_2)k_2 + \ldots + \tfrac{1}{2}(C_{f-1} + C_f)k_f = \sum_{i=1}^{f-1} C_i + \tfrac{1}{2}(C_0 + C_f) \quad (2)$$

in which all the k_i, the size of the individual steps, are taken to equal 1. *Ideally*,[13] the starting materials would be converted to the target in a single step, for which $F_0 = \tfrac{1}{2}(C_0 + C_f)$. Then the *excess complexity*, C_x, the net complexity above this ideal minimum, is the difference $F - F_0$, which is simply the sum of the complexities of the intermediates (eq 3). For a synthesis with

$$C_x = \sum_i C_i \quad (3)$$

more than one starting material and thus more than one branch in the synthesis tree, the summation is taken over all intermediates so that all the branches are treated in the same way.[14] The result contained in eq 3 is remarkably simple, considering the fact that we begin with a two-dimensional model considerably more sophisticated than previous one-dimensional treatments. Moreover, our result is completely general, as $C(k)$ can represent any measure of complexity, either empirically[6,7] or mathematically[1] derived. Note that it is possible for C_x to be lower for a linear route *if the intermediates are less complex*. The "principle of minimum chemical distance", by which the total number of bonds made and broken is minimized,[8] is successful in synthetic analysis because it tends to keep the complexities of the intermediates as close as possible to those of the starting material and the target, thus tending to minimize the excess complexity.

The use of plots of complexity vs. number of steps in the evaluation of alternative paths to a target is illustrated in Figure 1, which follows the topological evolution of a number of routes to modhephene.[16-22] The topological parameter η, the number of pairs of adjacent bonds, is used because it is the simplest general one thus far proposed.[1,15] Several things are obvious from these plots that might otherwise have eluded attention. For example, the intermediates in these routes reach the topological complexity of modhephene quickly—some in a few steps, all in the first half of the synthesis. The rest of the steps are devoted to changing functionality and fixing stereochemistry. At the present state of the art, the principal way in which stereochemistry contributes to excess complexity is by requiring more steps. Clearly, reactions need to be invented that give desired arrays of functionality directly and that set the desired stereochemistry at the same time. (Only Oppolzer and Bättig[16] do this exceptionally well: their ene reaction fixes the stereochemistry of the isolated methyl group at the same time it forms the third ring.) The choice of synthesis strategies and reaction conditions that avoid protecting groups also helps to minimize excess complexity. (The use of dichloroethylene as an acetylene equivalent in ref 19 requires a strong reducing agent for unmasking, which necessitates the protection of a ketone and its subsequent unmasking—all contributing excess complexity.)

These observations are made more quantitatively in Table I, which lists for each synthesis the sum over the intermediates of η as well as the overall yield.[23] For every pair of syntheses but two (note that there are 21 possible pairs), the order of *decreasing* excess topological complexity is the same as the order of *increasing* overall yield. A fair comparison cannot be made with Oppolzer's first synthesis[20] as it was not optimized as much as the others.[24] Furthermore, if Smith had used Dreiding's reaction conditions to carry out the last four steps (for which the intermediates are the same), his overall yield would have been 6.1%. After Oppolzer's first synthesis is dropped from consideration and this reasonable correction to Smith's yield is made, *the order of yields is predicted by the order of excess complexities for all six completed syntheses of modhephene*! However, considering Cook's starting material to be tricyclic (7) instead of monocyclic (cyclopentane-1,5-dione), $\eta = 406$ and the overall yield is 9.4%. This example also illustrates the semiquantitative nature of the theory, due to differences in amounts of effort spent perfecting different routes. Indeed, again emphasizing the probabilistic basis of the theory, it is not to be expected that such a statistical model will hold all of the time. Nonetheless, it does lead to important insights: there can be little doubt that topology is a major factor in these syntheses of modhephene. Of optimization Woodward[25] wrote "These developments measure the extent to which the arithmetic fiend, which besets all multi-stage synthetic activity, can be conquered by careful and intensive developmental work. . . . It is perhaps less widely appreciated . . . how many reactions can be coaxed into giving yields in the range of 90 to 100 per cent." When the general state of the art reaches this level, topology will be the controlling factor in the synthesis of complex molecules.

We propose that the topological complexity, as well as other sources of molecular complexity,[1] be examined *in addition to* the classical considerations[2,5,6,13] of synthetic efficiency when evaluating synthetic routes. Unlike the empirical indices,[6,7] which depend upon the state of the art, the indices of complexity grounded in graph theory and information theory provide a fixed frame of reference against which, ultimately, the state of the art can be measured. Whichever index is used, the C vs. k plot is a useful tool for visualizing the course of a synthesis and comparing alternative routes.

Acknowledgment. Professors J. M. Cook (University of Wisconsin—Milwaukee) and P. A. Wender (Stanford University) generously supplied details of their modhephene syntheses prior to publication. Professor J. B. Hendrickson checked the calculations and pointed out several errors, for which assistance I am most grateful.

Supplementary Material Available: Listings of the intermediates of all the routes along with their complexities and yields (the data used to calculate the values in Table I) (2 pages). Ordering information is given on any current masthead page.

(12) Protter, M. H.; Morrey, C. B. "Calculus for College Students", 2nd ed.; Addison-Wesley: Reading, MA, 1973; p 451.
(13) Turner, S. "The Design of Organic Synthesis"; Elsevier: Amsterdam, 1976; p 10.
(14) For a relevant discussion of the importance of symmetry and mathematical simplicity in science, see: Dirac, P. A. M. "Symmetries in Science"; Gruber, B., Millman, R. S., Eds.; Plenum Press: New York, 1980; p 1.
(15) For syntheses in which symmetry is not a major factor, we prefer to use plots of η vs. k rather than $C(\eta)$ cs. k, as η is much easier to calculate.
(16) Oppolzer, W.; Bättig, K. *Helv. Chim. Acta* **1981**, *64*, 2489.
(17) Schostarez, H.; Paquette, L. A. *Tetrahedron* **1981**, *37*, 4431.
(18) Karpf, M.; Dreiding, A. S. *Helv. Chim. Acta* **1981**, *64*, 1123.
(19) Smith, A. B.; Jerris, P. J. *J. Org. Chem.* **1982**, *47*, 1845.
(20) Oppolzer, W.; Marazza, F. *Helv. Chim. Acta* **1981**, *64*, 1575.
(21) Wrobel, J.; Takahashi, K.; Honkan, V.; Lannoye, G.; Cook, J. M.; Bertz, S. H., submitted for publication.
(22) Wender, P. A.; Dreyer, G. *J. Am. Chem. Soc.*, in press.
(23) While the routes of Figure 1 have all been taken back to simple monocyclic starting materials to make the plots more aesthetically appealing, the yields and complexities of Table I are based on the starting materials specified by the authors to avoid editorial bias. All begin with bicyclic ones except for Cook's route, which goes directly from monocyclic to tricyclic compounds, and Oppolzer's first route. All conventionally isolable intermediates are included, whether or not they were actually isolated (see supplementary material).
(24) Work on it was discontinued in favor of the second synthesis (ref 16): W. Oppolzer, private communication.
(25) Woodward, R. B. "Perspectives in Organic Chemistry"; Todd, A., Ed.; Interscience: New York, 1956; p 164.