A New and Simple Approach to Chemical Complexity. Application to the Synthesis of Natural Products

René Barone* and Michel Chanon

Laboratoire AM3, Case 561, Faculte des Sciences de St. Jérôme, 13397 Marseille Cedex 20, France

Received October 27, 2000

A new approach to describe the complexity of chemical structures is proposed. This index is simple and can be easily programmed. It derives from Whitlock's index and, despite its empirical character, provides results paralleling those obtained with the mathematical Bertz index. It has been applied to follow the strategic progress of some natural product syntheses.

INTRODUCTION

Some years ago Bertz used information theory to propose a mathematical approach for calculating the complexity of chemical structures.[1] This approach was used to follow the complexity of intermediates during the synthesis of complex structures.[2] It has been adapted and programmed by J. B. Hendrickson.[3]

Recently, Whitlock[4] proposed an intuitive (metric) approach for molecular complexity. This formula was then again applied to several syntheses. In this approach complexity is the sum of four terms: 4 times the number of rings, 2 times the number of nonaromatic unsaturations, 1 times the number of heteratoms, and 2 times the number of chiral centers.

Despite its success, this approach presents some limitations since substituents are not taken into account and since each ring contributes the same amount, 4. We decided to ameliorate this approach by improving these two points. A new index has been developed and applied to the syntheses of some natural products.

DESCRIPTION OF THE NEW MOLECULAR COMPLEXITY INDEX

This new approach, as Whitlock's approach, is purely intuitive and empirical. In a first approximation we decided to neglect the chiral term, since it was omitted in Bertz's formula.

First, we adopted the following values for carbon atoms: $CH_3 = 1$, $CH_2 = 2$, $CH = 3$, $C = 4$. These values represent the number of neighbors for each carbon atom. These values, however, gave the same results for the isomers of an alkane family: all isomers of pentane have the same value of 8. So, instead of using a linear progression, we decided to use a geometrical progression, 1, 2, 4, and 8, and to have larger values for complexity, we multiplied these values by 3, which provides 3 (CH_3), 6 (CH_2), 12 (CH), and 24 (C). Since CH_2 can be found in a chain, $-CH_2-$, but also in a methylene group, $=CH_2$, the use of this approach allows us not to count unsaturations. These values are also used for heteroatoms: $-O-$ and $=O$ contribute 6 units to the overall complexity.

These values were used not only for the substituents, but for all atoms of the structure, even the cyclic atoms.

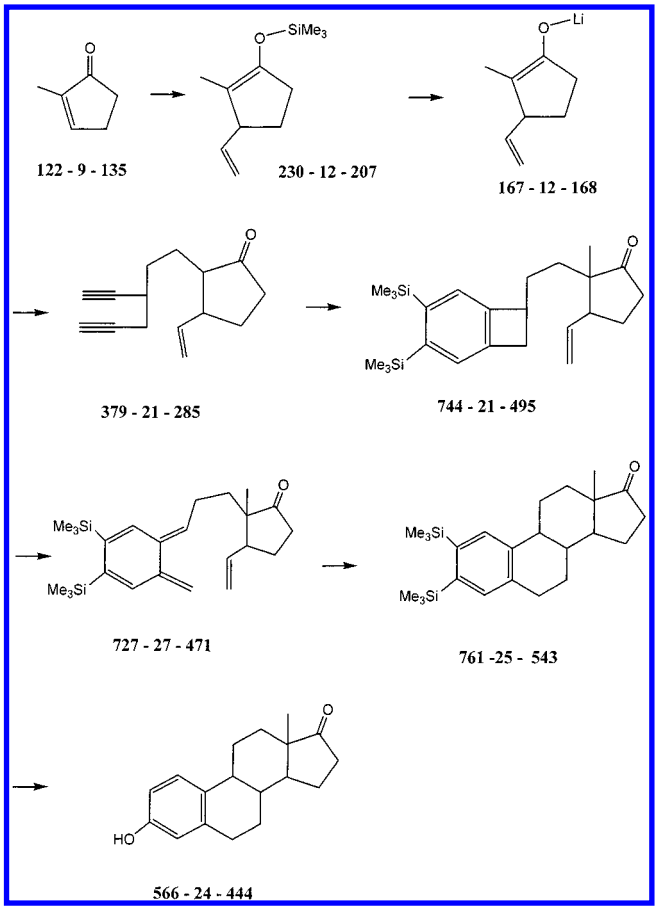

Figure 1. Volhardt's synthesis of estrone with Bertz's, Whitlock's, and our complexity indexes.

For the nature of atoms we count 3 for a carbon and 6 for any heteroatom.

The ring term should be proportional to the number of atoms of each ring, i.e.

$$\text{RingTerm} = \sum_{i=1}^{i=\text{nbring}} \text{constant*size(i)}$$

in which nbring is the number of rings in the structure and size(i) is the number of atoms in ring number i.

Or in Basic programming language

```
for i = 1 to nbring:  RingTerm = RingTerm +
                      constant*size(i):  next i.
```

* To whom correspondence should be addressed. E-mail: rene.barone@am3.u-3mrs.fr. Phone: (33) (0)4 91 28 84 23. Fax: (33) (0)4 91 28 82 34.

10.1021/ci000145p CCC: $20.00
Published on Web 02/06/2001

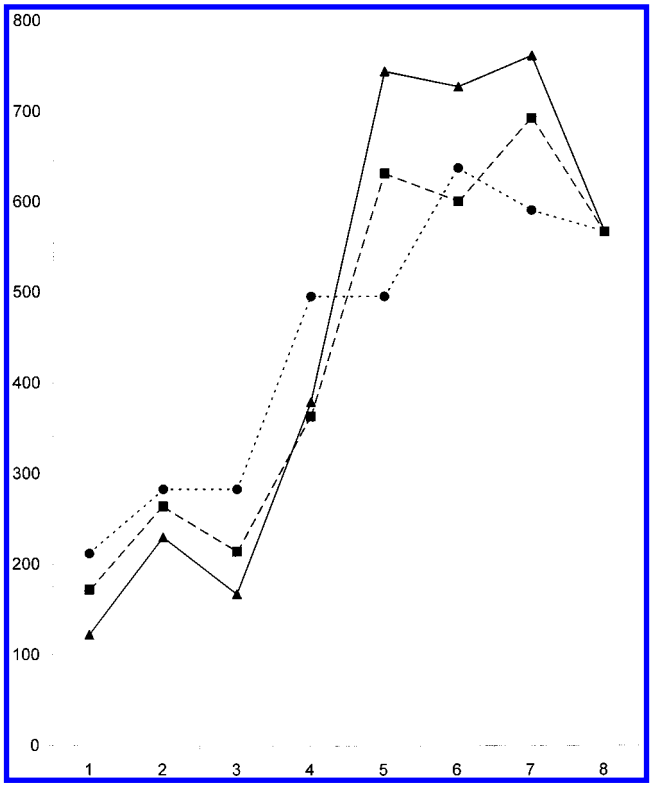

Figure 2. Complexities for estrone synthesis from Bertz's (triangles), Whitlock's (circles), and our (squares) indexes.

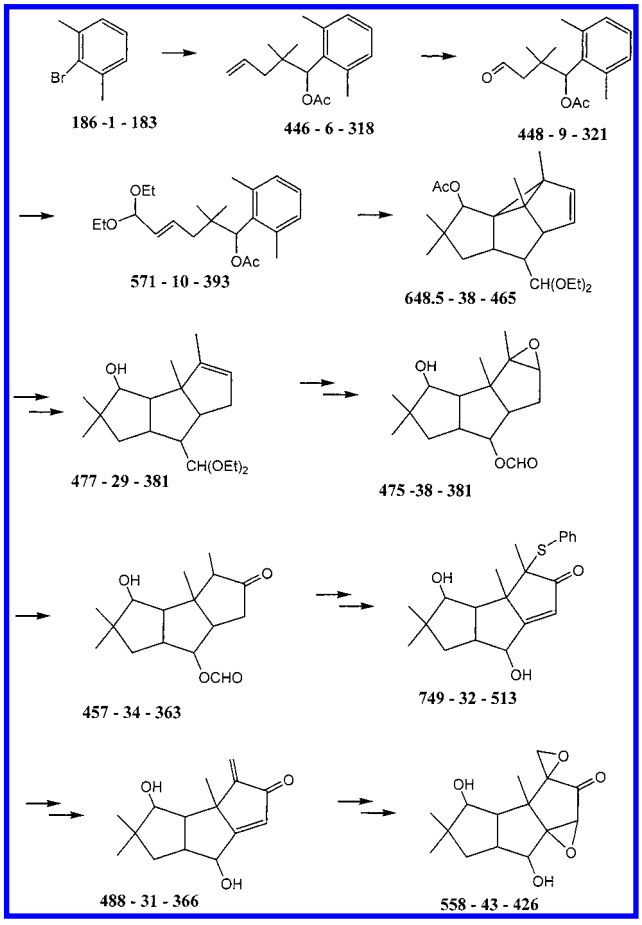

Figure 3. Wender's synthesis of coriolin with Bertz's, Whitlock's, and our complexity indexes.

After some attempts the constant was set to 6. For cyclobutane the ring term in complexity is 6*4 = 24, for cyclopentane the value is 30, and so on. This allows the contribution of each ring to be modulated.

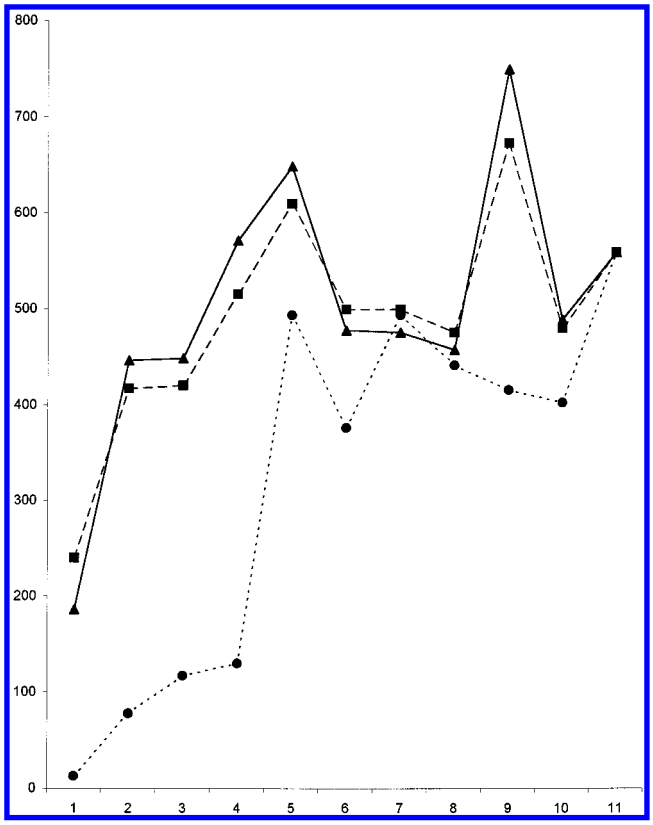

Figure 4. Complexities for coriolin synthesis from Bertz's (triangles), Whitlock's (circles), and our (squares) indexes.

Finally the program to compute the new value of complexity is simply the following:

(1) Search of rings: this subroutine returns the number of rings and the size of each ring. (The number of rings is the smallest set of smallest rings.[5])

(2) Calculate the connectivity (cnt) of each atom, that is, the sum of all the bonds (hydrogens are omitted).

Then

```
for i = 1 to nbring                'number of rings
 cpxty = cpxty + size(i)*6
next i

for i = 1 to nbatoms    'number of atoms in the structure
  if cnt(i) = 4 then cpxty = cpxty + 24
  if cnt(i) = 3 then cpxty = cpxty + 12
  if cnt(i) = 2 then cpxty = cpxty + 6
  if cnt(i) = 1 then cpxty =  cpxty + 3
next i

for i = 1 to nbatoms
 if naturatom(i) = "C" then cpxty = cpxty + 3 else cpxty = cxpty + 6
next i
```

APPLICATION OF COMPLEXITY TO SOME SYNTHESES

Having this new index in hand, we applied it to some syntheses involving two types of cyclic structures: linear, such as estrone[6] and coriolin,[7] and bridged, such as longifolene[8,9] and daucene.[10] These syntheses are depicted in Figures 1, 3, 5, 7, and 9.

For each structure we calculated the complexity with Bertz's, Whitlock's, and our indexes. The values obtained are indicated below each structure in this order. The

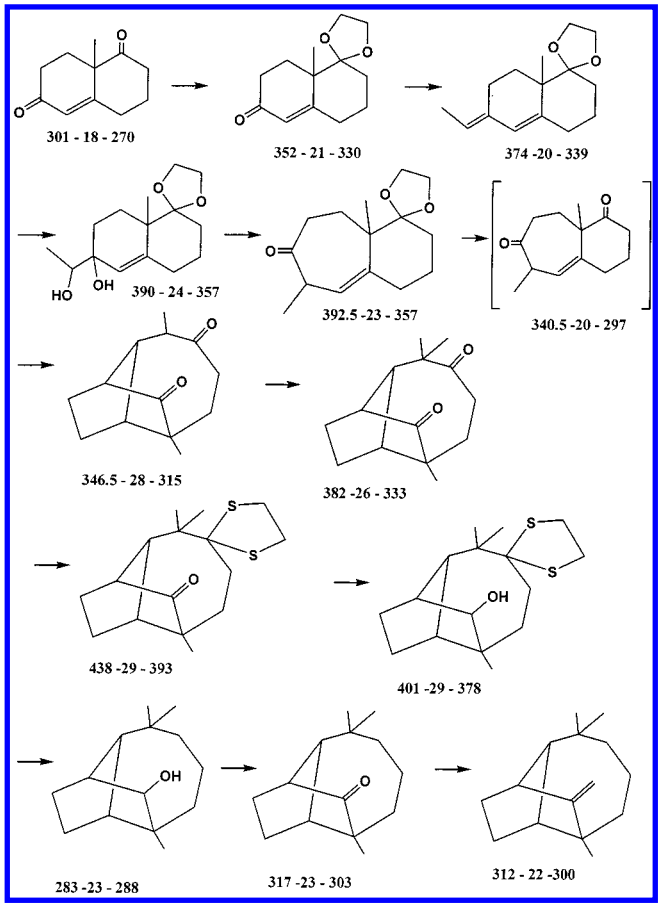

Figure 5. Corey's synthesis of longifolene with complexities from Bertz's, Whitlock's and our indexes.

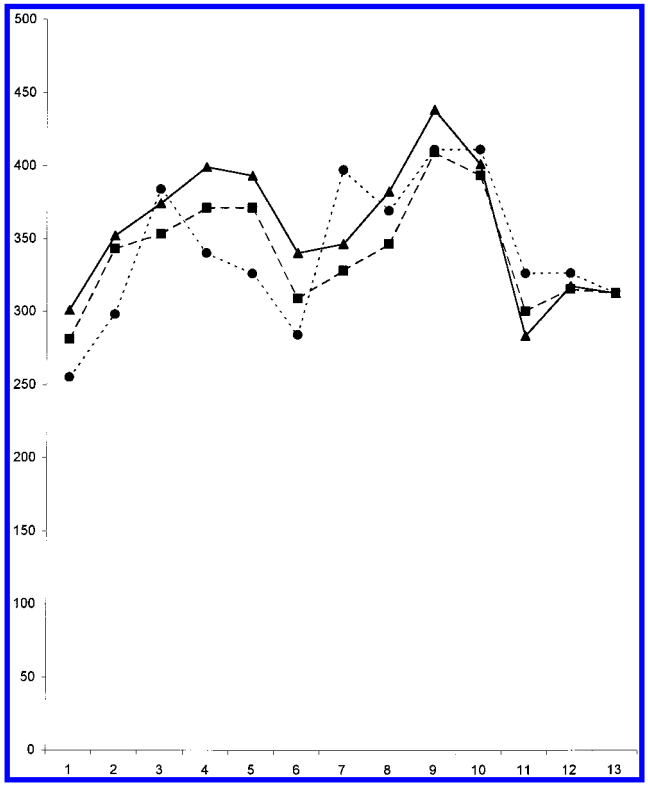

Figure 6. Complexities for Corey's longifolene synthesis from Bertz's (triangles), Whitlock's (circles), and our (squares) indexes.

complexities were then plotted in graphics (the curves have been scaled to obtain the same values for the targets). It is

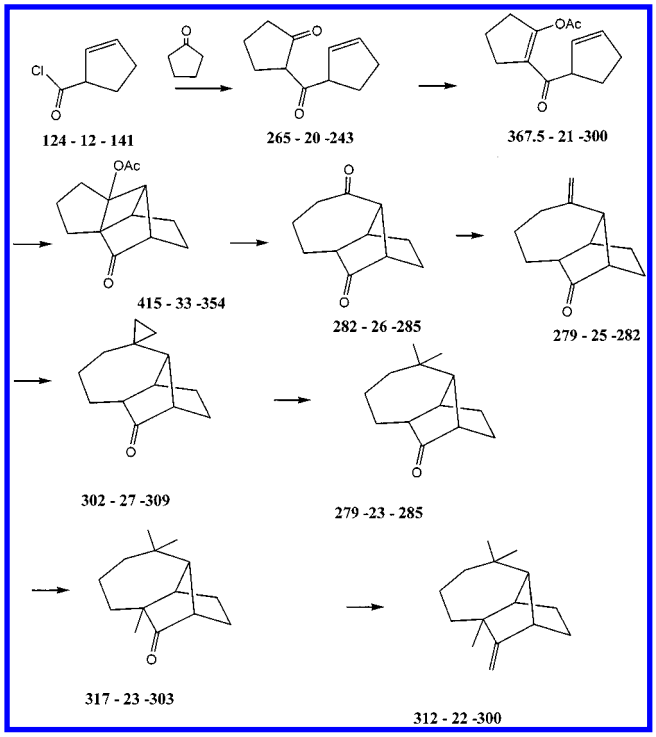

Figure 7. Oppolzer's synthesis of longifolene with Bertz's, Whitlock's, and our complexity indexes.

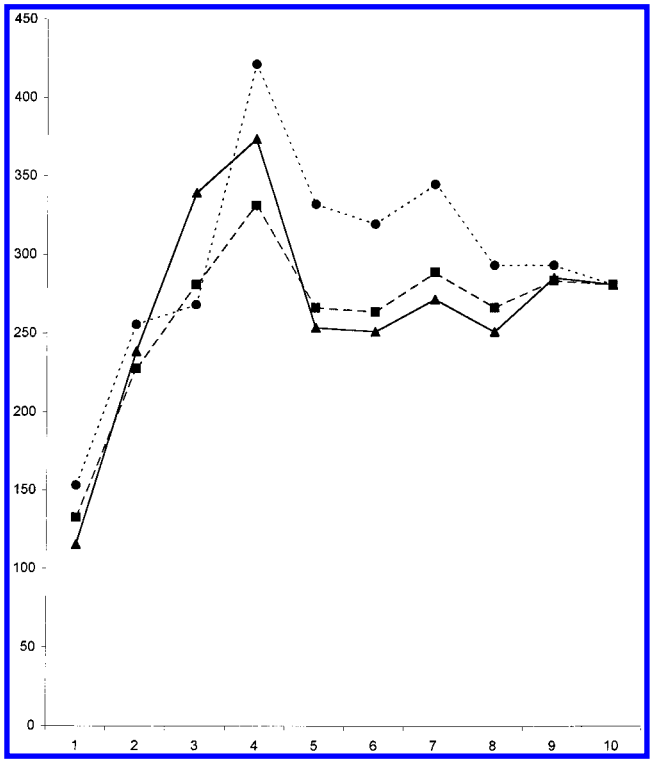

Figure 8. Complexities for Oppolzer's longifolene synthesis from Bertz's (triangles), Whitlock's (circles), and our (squares) indexes.

interesting to see that the curves obtained with our formula are rather similar to the curves obtained with Bertz's formula (Figures 2, 4, 6, 8, and 10). The values are, obviously, different, but they follow the same general trend, increasing and decreasing simultaneously. The graphs obtained using Whitlock's approach are also in agreement with those obtained using Bertz's approach, except for structures in which substituents and/or aromatic rings are present since they are not taken into account in Whitlock's treatment. This

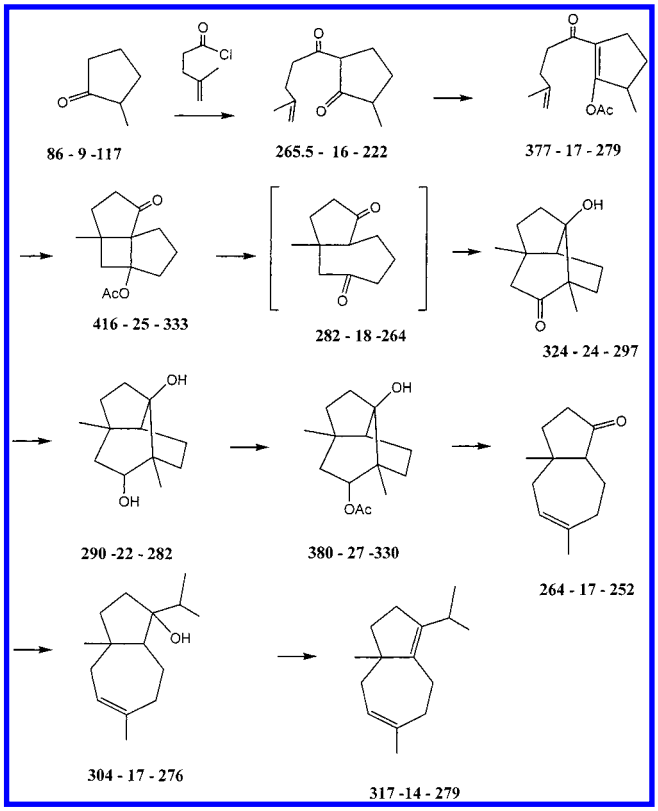

Figure 9. Seto's synthesis of daucene with complexities from Bertz's, Whitlock's, and our indexes.

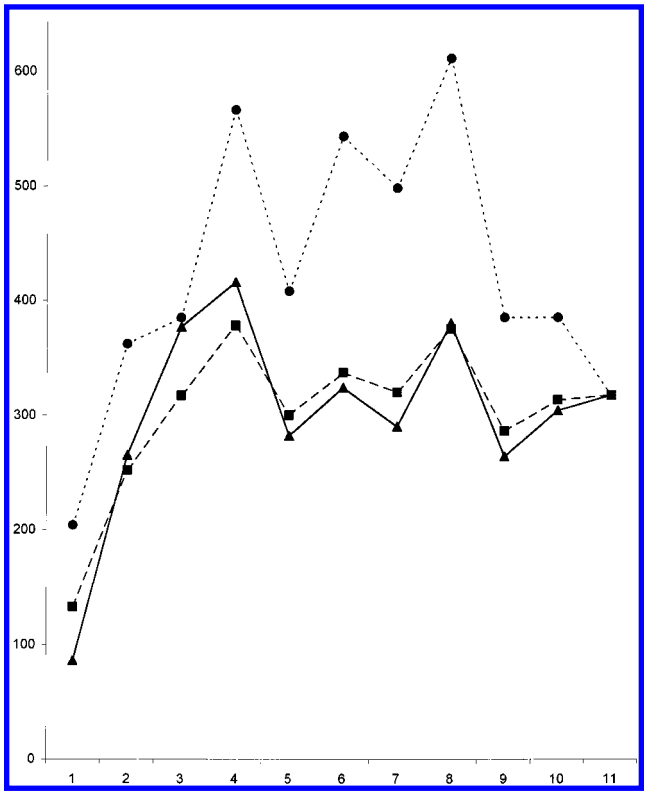

Figure 10. Complexities for daucene synthesis from Bertz's (triangles), Whitlock's (circles), and our (squares) indexes.

is the case for estrone (Figures 1 and 2) and even more for coriolin, where intermediates **1**−**4** have a very low complexity (Figures 3 and 4), or for intermediate **9**, where the phenyl group is omitted. Generally in Whitlock's approach the evolutions are often more uneven.

The overall convergence of the results for these three approaches could be explained by the fact that there is an underlying similarity between them. The values that we selected for the complexity of atoms (3, 6, 12, and 24) are similar to the number of propane units for saturated carbons (0, 1, 3, and 6)[3] in Bertz's approach, and the chemical features counted by Whitlock (rings, unsaturations, chiral centers) are the main sources of propane substructures. We have, elsewhere, discussed the present limitations of the different treatments of complexity with respect to the evaluation of a synthetic strategy.[11] The present treatment underlines the similarities of different approaches; it does not, however, bring a definitive answer to the previously discussed flaws. Work is in progress in this direction.

CONCLUSION

We developed a new intuitive and empirical approach to calculate the complexity of chemical structures. It is an extension of Whitlock's work. This treatment was then applied to several syntheses of natural products. The results obtained converge with Bertz's approach. Since it is very simple to program, we plan to use it in our computer-aided organic synthesis program Holowin[12] and for our study about complexity vs similarity.[11]

ACKNOWLEDGMENT

We thank Professor J. B. Hendrickson for providing his program CPXCAL to calculate complexity from Bertz's approach and Professor F. Petra for helpful discussions.

REFERENCES AND NOTES

(1) Bertz, S. H. The First General Index of Molecular Complexity. *J. Am. Chem. Soc.* **1981**, *103*, 3599−3601.

(2) Bertz, S. H. Convergence, Molecular Complexity, and Synthetic Analysis. *J. Am. Chem. Soc.* **1982**, *104*, 5801−5803.

(3) Hendrickson, J. B.; Huang, P.; Toczko, A. G. Molecular Complexity: A Simplified Formula Adapted to Individual Atoms. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 63−67.

(4) Whitlock, H. W. On the Structure of Total Synthesis of Complex Natural Products. *J. Org. Chem.* **1998**, *63*, 7982−7989.

(5) Downs, G. M.; Gillett, V. J.; Holliday, J. D.; Lynch, M. F. Review of Ring Perception Algorithms for Chemical Graphs. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 172−187.

(6) Funk, R. L.; Vollhardt, K. P. C. A Cobalt-Catalyzed Steroid Synthesis. *J. Am. Chem. Soc.* **1977**, *99*, 5483−5484; The Cobalt Way to *dl*-Estrone, a Highly Regiospecific Functionalization of 2,3-Bis(trimethylsylyl)estratrien-17-one. *J. Am. Chem. Soc.* **1979**, *101*, 215−217.

(7) Wender, P. A.; Howbert, J. J. Synthetic Studies on Areneolefin Cycloadditions −6- Two Syntheses of (+)-Coriolin. *Tetrahedron Lett.* **1983**, *24*, 5325−5328.

(8) Corey, E. J.; Ohno, M.; Mitra, R. B.; Vatakencherry, P. A. Total Synthesis of Longifolene. *J. Am. Chem. Soc.* **1964**, *86*, 478−485.

(9) Oppolzer, W.; Godel, T. A New and Efficient Synthesis of (±)-Longifolene. *J. Am. Chem. Soc.* **1978**, *100*, 2583−2584.

(10) Seto, H.; Fujimoto, Y.; Tatsuno, T.; Yoshioka, H. SyntheticStudies on Carotane and Dolastane Type Terpenes: A New Entry to the Total Synthesis of (±)-Daucene. *Synth. Commun.* **1985**, *15*, 1217−1224.

(11) Chanon, M.; Barone, R.; Baralotto, C.; Julliard, M.; Hendrickson, J. B. Information Theory Description of Synthetic Strategies in the Polyquinane Series. The Holosynthon Concept. *Synthesis* **1998**, 1559−1583.

(12) Barberis, F.; Barone, R.; Chanon, M. Holowin: A Fast Way to Search for Tandem Reactions with Computer. Application to the Taxane Framework. *Tetrahedron* **1996**, *52*, 14625−14630.

CI000145P