Similarity in Organic Chemistry: A Summary of the Beilstein Institute Conference

Stephen R. Heller

U.S. Department of Agriculture, Agricultural Research Service, NPS, Beltsville, Maryland 20705

This presentation is a summary of the Beilstein Bolzano meeting on Similarity in Chemistry.

As chemists increase their use of computer systems, a number of conferences have been established to meet the needs of the community. The biannual International Conference on Computers in Chemical Research and Education (IC-CCRE) held its 10th conference in Israel 1992.[1,2] A more recently established conference is the Beilstein Bozen (Bolzano) Meeting. This is also a biannual meeting and the third one in the series was held in May 1992. The papers from the 1992 Beilstein Bozen meeting are published in this issue of the *Journal of Chemical Information and Computer Sciences*.

Bozen conferences have previously been held on the Estimation of Physical Properties of Organic Compounds[3] and on Computer Reaction Management in Organic Chemistry.[4] The 1992 Beilstein Bozen Conference, attended by some 60 scientists from Europe, the United States, and Japan, took as its theme Similarity in Organic Chemistry. This paper summarizes the 31 papers and poster talks presented at the meeting, 23 of which are printed elsewhere in this issue.

Similarity in chemistry is a very venerable notion. The periodic table itself was developed from the observed similarity in the behavior and properties of various elements, and it stands as perhaps the most widely known and visible example of similarity in chemistry. The last 20 years have seen a great deal of work on exact matching of structures and substructures, but now similarity, or inexact ("fuzzy") structure matching, has recently been given a new examination. A number of workers have been providing this area of science with a more rigorous mathematical and conceptual foundation. Prominent among these efforts is the recent book by Johnson and Maggiora[5] on the subject. In fact many of the presentations that were made at Bozen in 1992 and which are printed in this issue of the *Journal of Chemical Information and Computer Sciences* represent continuing studies by workers who in 1988 and 1989 wrote chapters for Johnson and Maggiora's book.

A literature search of Chemical Abstracts for the term "similarity" produced the following results:

CA period	no. of hits
1967–1971	233
1972–1976	368
1977–1981	519
1982–1986	897
1987–mid-1992	1693

This is not of course a rigorous measure of the importance of this subject area, but it does suggest that there is increasing interest on the part of the scientific community in the concept of similarity.

The areas of chemistry which were covered in the 1992 Beilstein Bozen meeting were quite varied and extensive. Rouvray provided an excellent introduction to and review of the overall area of similarity.[6] A number of papers described different various methods for assessment or measurement of similarity. These included "quantum similarity", a method developed by Carbó and based upon quantum-mechanical principles,[7] and geometrical similarity or shape analysis approaches described by Venkatachalam of Molecular Simulations Inc. and by Mezey.[8] The importance of chemical similarity in reactions was the focus of a number of papers. Lawson[9] describes the concept of similarity between reactions and its use in information classification. Ponec[10] discusses the use of reaction similarity to rationalize various mechanistic features of pericyclic reactions, and Fontain[11] reports on the application of genetic algorithms that can determine constitutional similarity. A program, Beppe, which predicts the likely products of specific reactions was described by Sello;[12] Grethe of Molecular Design Ltd. discussed the similarity searching capabilities of the MACCS-II software; Mark Johnson[13] reported on the use of "site environments" in substrates to distinguish between xenobiotic amines susceptible to N-oxidation and those which are instead demethylated; and Gladkova (Moscow) described a new retrieval language based upon bond counts in molecules and showed that it is useful in inter alia reaction indexing. Synthesis design and planning can be influenced positively by considerations of similarity. Thus, Gasteiger[14] showed how quantitation of the electronic effects at reaction sites can be used to determine the similarity between reactions, and Hanessian[15] described the use of his program Chiron to identify the similar architectural features common to different molecules and so to predict biogenetic pathways. The application of similarity in structure and substructure searching was dealt with by a number of speakers. Willett[16] described the use of atom mapping as a means of three-dimensional similarity searching, Fisanick[17] reported on the work of the CAS group on the use of similarity searching to find "fuzzy" matches in a database, and Peter Johnson[18] discussed his multiple largest common subgraph routine, which finds the largest substructure common to several different molecules. Takahashi[19] covered the use of automatic similarity determination in QSAR; Razinger of the Kidrič Institute in Ljubljana described programs for the exhaustive generation of the stereoisomers corresponding to a given structural formula; and Marsili[20] of the University of L'Aquila in Italy provided some detail on his program ANALOGS, which generates sets of compounds related to a lead compound and useful in QSAR studies. Perry[21] reported on the use of molecular similarity in 3D searching, and Barnard[22] described his work on the clustering of chemical structures on the basis of 2D similarity measures. The use of similarity in QSAR was discussed by Judson,[23] and Allan[24] described a novel momentum-space approach to the similarity of electron densities and hence of molecules.

Randić[25] discussed the use of an orthogonalized extended list of descriptors as a measure of similarity, but apart from this one paper there was less discussion of the presentation of similarity measures than one would have wished, although in

SUMMARY OF BEILSTEIN WORKSHOPS

J. Chem. Inf. Comput. Sci., Vol. 32, No. 6, 1992 **579**

3 days it is hard to completely cover such a broad subject matter.

In the area of structure searching, the value in similarity searching is that one is able to do a "loose" or "fuzzy" search, as opposed to the usual "tight" search, and we saw that having a good deal of slack in a search can lead to some interesting search results. One very important area of structure similarity searching which was not addressed was Markush searching.[26] The related area of homology searching in libraries of DNA base pairs was also not covered.

One session of the meeting was devoted to neural networks. This session began with a paper by Maggiora,[27] who gave an excellent introduction to the use of neutral networks in some chemical problems, as well as the use of artificial neural nets to model highly complex systems in the area of structure–activity research. Other papers in this session included one by Clark of the University of Erlangen on the use in neural networks of data developed by semiempirical MO theory in comparison with physically observed data, and a description by Herges[28] of the use of neural nets for determination of aromatic substitution patterns from infrared spectra. Finally, Kvasnička[29] spoke on the application of neural nets to the prediction of ^{13}C NMR shifts and metaproduct distribution of nitration in a series of monoaromatic compounds.

To paraphrase the old proverb: "Similarity is in the eyes of the beholder". As indicated by the varied subjects that were touched upon during this workshop, there are many ways to view similarity. There is little question that many new developments in applied organic chemistry, and especially pharmaceutical chemistry, could come from similarity-based approaches. If one knows enough about what the actual process or problem to be solved is (e.g., design of molecules that interact with a particular receptor site), an appropriate measure of similarity can be chosen which allows one to work on solving the problem most quickly and efficiently. One can easily envision using an appropriate similarity measure to find bioisosteres as new drug alternatives which would not violate an existing patent.

It is a paradox that more precise tools such as high-quality and high-resolution instruments and computers often lead us to more imprecision in science, particularly in the area of biological systems. This is becoming increasingly apparent to the public everyday. Perhaps the recent emphasis on the concept of similarity will lead us to better understand this fuzzy universe we inhabit. Then, again, perhaps the current fad will slow as people take a hard look at what this approach really does for their work and what practical results emerge from it. The Beilstein Institute has done us all a service by organizing this workshop, and it will be interesting to review this area in some five years time to see if the promise of similarity in chemistry has been fulfilled.

REFERENCES AND NOTES

(1) For further information concerning the 10th ICCCRE, please contact the Conference Chair, Prof. Y. Wolman, Department of Organic Chemistry, Hebrew University of Jerusalem, Jerusalem, 91904 Israel. Telephone: 972-2-585-282. FAX: 660-346; 666-804. Internet: CHESKEA@HUJIVMS.

(2) Papers presented at the 10th ICCCRE will be published in the March/April 1993 issue of the *Journal of Chemical Information and Computer Sciences*.

(3) *Proceedings of the Beilstein Worshop on the Estimation of Physical Data for Organic Compounds*, May 1988; Jochum, C., Hicks, M. G., Sunkel, J., Eds.; Springer-Verlag: New York, 1988.

(4) *J. Chem. Inf. Comp. Sci.* **1990**, *30*, 351–520.

(5) *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990.

(6) Rouvray, D. H. Definition and Role of Similarity Concepts in the Chemical and Physical Sciences. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(7) Carbó, R.; Calabuig, B. Quantum Similarity Measures, Molecular Cloud Description, and Structure–Properties Relationships. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(8) Mezey, P. G. Shape-Similarity Measures for Molecular Bodies: A Three-Dimensional Topological Approach to Quantitative Shape–Activity Relationships. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(9) Lawson, A. J. Organic Reaction Similarity in Information Processing. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(10) Ponec, R.; Strnad, M. Similarity Ideas in the Theory of Pericyclic Reactivity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(11) Fontain, E. Application of Genetic Algorithms in the Field of Constitutional Similarity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(12) Sello, G. Reaction Prediction: The Suggestions of the Beppe Program. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(13) Gifford, E.; Johnson, M. A.; Kaiser, D. G.; Tsai, C. Visualizable Relative Occurrences in Metabolic Transformations of Xenobiotics Using Structure–Activity Maps. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(14) Gasteiger, J.; Ihlenfeldt, W.-D.; Fick, R.; Rose, J. R. Similarity Concepts for the Planning of Organic Reactions and Syntheses. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(15) Hanessian, S.; Botta, M.; Larouche, B.; Boyaroglu, A. Computer-Assisted Perception of Similarity Using the Chiron Program: A Powerful Tool for the Analysis and Prediction of Biogenetic Patterns. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(16) Artymiuk, P. J.; Bath, P. A.; Grindley, H. M.; Pepperrell, C. A.; Poirrette, A. R.; Rice, D. W.; Thorner, D. A.; Wild, D. J.; Willett, P.; Allen, F. H.; Taylor, R. Similarity Searching in Databases of Three-Dimensional Molecules and Macromolecules. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(17) Fisanick, W.; Cross, K. P.; Rusinko, A., III. Similarity Searching on CAS Registry Substances. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue. 1. Global Molecular Property and Generic Atom Triangle Geometric Searching.

(18) Bayada, D.; Simpson, R. W.; Johnson, A. P.; Laurenço, C. Algorithm for the Multiple Common Subgraph Problem. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(19) Takahashi, Y.; Sukekawa, M.; Sasaki, S.-I. Automatic Identification of Molecular Similarity Using Reduced Graph Representation of Chemical Structure. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(20) Marsili, M.; Saller, H. ANALOGS: A Computer Programs for the Design of Multivariate Sets of Analogs. *J. Chem. Inf. Comput. Sci.* **1993**, in press.

(21) Perry, N. C.; van Geerestein, V. J. Database Searching on the Basis of Three-Dimensional Molecular Similarity Using the SPERM Program. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(22) Barnard, J. M.; Downs, G. M. Clustering of Chemical Structures on the Basis of Two-Dimensional Similarity Measures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(23) Judson, P. N. Relating Structural Similarity Searching to Biological Activity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(24) Allan, N. L.; Cooper, D. L. Momentum Space Approach to Molecular Similarity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(25) Randić, M. Similarity Based Upon Extended Basis Descriptors. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(26) See papers in Vol. 31, No. 1 (1991) of the *Journal of Chemical Information and Computer Sciences*.

(27) Maggiora, G. M.; Elrod, D. W.; Trenary, R. G. Computational Neural Nets as Model-Free Mapping Devices. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(28) Weigel, U.-M.; Herges, R. Automatic Interpretation of Infrared Spectra: Recognition of Aromatic Substitution Patterns Using Neural Networks. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.

(29) Kvasnička, V.; Sklenák, Š.; Pospíchal, J. An Application of Recurrent Neural Networks in Chemistry: Prediction and Classification of ^{13}C NMR Chemical Shifts in a Series of Monosubstituted Benzenes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, in this issue.