Mathematics as a Basis for Chemistry

G. W. A. Milne

Laboratory of Medicinal Chemistry, NCI, NIH, Bethesda, Maryland 20892

Received December 18, 1996[⊗]

Mathematicians are increasingly publishing papers in which mathematics is applied to chemical problems. Examination of some of these papers reveals that while they contain genuine contributions to chemistry they tend to avoid the more interesting and difficult problems. The "forward" problem of estimation of physical properties from a compound's structure, for example, has seen many publications, a proportion of which have been successful. The "reverse" problem, prediction of the structures which possess given properties, is only rarely examined and never by mathematicians. This is unfortunate because these problems are mathematically complex and mathematicians could make significant contributions by bringing their skills to bear on such questions.

INTRODUCTION

From its traditional position as the "central pure science" mathematics is regarded by many as the true basis of other, more applied sciences, such as chemistry. Whether or not it is the progenitor of such sciences, however, is questionable. Most of the applied sciences grew up in their own right, taking their cues not from mathematics but from experiment, observation, classification, and the other familiar traditional techniques.

Since the 1950s however, the open space between mathematics and chemistry has been largely eliminated, and today, mathematics is infiltrating chemistry—and other sciences—at all levels, particularly at the foundations. Much of the credit for this must go to mathematicians, who have had the courage and energy to mount excursions to foreign disciplines where, with less knowledge and authority, they were almost certain to receive a chilly reception. Nevertheless, they have persevered, and because of this, "mathematical chemistry", once thought to be almost an oxymoron, is now tolerated, even well regarded as a subspecialty, and is the principal occupation of most of the speakers and audience at this Symposium honoring Milan Randić, an outstanding member of this group.

But what has mathematical chemistry done for chemistry? And for mathematics? Has either discipline benefitted from this alliance? Has chemistry benefitted more from other alliances? These are all interesting questions which are explored in this paper.

MATHEMATICS AND CHEMISTRY

During the 19th century, several distinguished mathematicians such as Cayley and Pólya studied specific problems in chemistry, such as calculating the precise number of isomers corresponding to a particular molecular formula, the "isomer enumeration" problem. Such pioneers, whose papers are cited almost *de rigor*, made valid contributions, but their efforts, when compared to work begun in the 1960s and still continuing, while ground-breaking, were minuscule. This was due partly to the extreme disconnect between different scientific disciplines at that time and also to the lack of availability of computers.

In the latter half of the 20th century, both of these difficulties have been largely eliminated. Mathematicians working on chemical problems are numerous, they publish widely and are beginning to be noticed, and their work is watched by the chemistry community. There have been some easy victories. Statistics for example has been readily adopted by chemists, some of whom think they invented it. Chemistry publications now make frequent and skilled use of methods like pattern recognition, linear regression, and principal component analysis, all statistical techniques, and it is fair to say that this branch of mathematics has long been firmly integrated into chemistry. Similarly, quantum mechanics, while arguably a branch of physics, nonetheless requires an understanding of and facility with differential calculus but has become a standard methodology in the chemists' armamentarium.
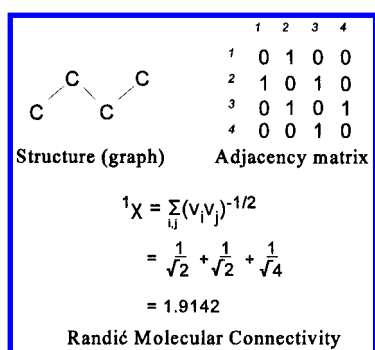
GRAPH THEORY

As more advanced mathematical techniques have been applied to chemistry, success has been harder to achieve. Graph theory, the mathematics of shapes such as chemical structure diagrams, a subject of interest to many attending this Symposium, is just such a case. Some of the early papers from the 1960s that used graph theory did so implicitly, even unconsciously. It is often said, by me as well as many others, that the well-known structure searching algorithms rest upon a graph theoretical foundation. This may be so, but the observation is one of hindsight. The chemists who wrote those programs were working from a knowledge of chemistry rather than graph theory. They unconsciously established some rules and used them to design the search algorithms. Only now are their rules recognized as graph theoretical in nature. The use of graph theory in isomer enumeration in the Dendral project, which was pursued at Stanford during the 1960s and 1970s, was more mathematical in nature;[1] topology was treated as the basis for the effort, and there was certainly some awareness among the Stanford group of the problem of the Königsberg bridges, a famous problem in mathematical topology.[2]

These two areas are now quite advanced. Modern structure searching algorithms are clearly written by chemists and computer scientists who are conversant with chemical graph theory, and isomer enumeration, having progressed

to problems like calculation of the numbers of benzene rings in polycyclic aromatic hydrocarbons and enjoying consistent support (for some reason) from Cancer funds, is established as a branch of mathematical chemistry. But these are highly specialized research areas. They are interesting and important, but their details are of little interest to most chemists who regard them, if at all, as a means to an end.

So has graph theory made any really significant inroads into chemistry? The subject on which most papers have been published in chemical journals is surely numeric matrices that represent chemical structures and topological indices derived from these matrices as shown below, and it is here that an answer to this question might be sought. Representation of a chemical structure as a graph and reduction of the graph to various matrices is very familiar to many chemists and this ground will not be covered again. Any of the different matrices can be manipulated so as to produce a single number, usually referred to as a local vertex invariant[3] or a topological index,[4] and much effort has been expended

$$\begin{array}{cccc} & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 3 & 0 & 1 & 0 & 1 \\ 4 & 0 & 0 & 1 & 0 \end{array}$$

Structure (graph) Adjacency matrix

$$^1\chi = \sum_{i,j} (v_i v_j)^{-1/2}$$

$$= \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{4}}$$

$$= 1.9142$$

Randić Molecular Connectivity

in attempts to understand the properties of such indices. Two questions that seem to be particularly interesting are is there a unique 1:1 relationship between a structure and its topological index and is the topological index related to the physical or chemical properties of a chemical.

Some 200 topological indices, first, second, and third generation,[3] have been devised and published,[5] and it seems increasingly likely, though hard to prove, that some of these uniquely represent a single structure. Such a topological index would become a unique identifier, and this would have significant impact throughout the whole area of chemical algorithms. The physical properties question is more difficult to resolve. Different topological indices say different things about a molecule. Some reflect the molecule's degree of branching, others its overall shape, or bulk. Those like Randić's molecular connectivity index,[6] which reflects branching, do indeed track with and can be regressed onto boiling point. Others reflect better some other property such as magnetic susceptibility,[7] and, in general, the answer is provided you choose the right topological index and try to relate it to the appropriate property, then indeed some relationship may emerge. Nobody, however, has even claimed to be able to predict melting point in this way, for example, because the relationship between structure and melting point is complex and obscure.

Prediction of properties is a very useful and commercially valuable capability, and graph theoretical work will continue if only for this reason. But it should be recognized as a part of a larger task, *viz.* design of a structure with specific physical properties. To do this it will be necessary to circumnavigate a central inconvenient fact, that generation
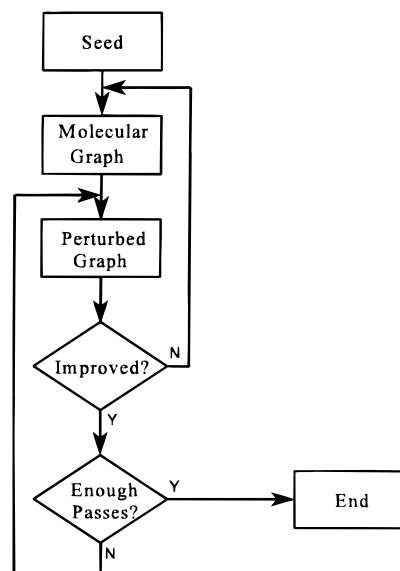

Figure 1. Simulated annealing algorithm. Simulated Annealing of Molecular Graphs. Kvasnička, Pospíchal, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 516−526.

of a topological index from a structure is a one-way process. It is not known how to convert a topological index back to the structure it represents; it is unlikely that it is even possible in a direct sense, and it is a challenge which seems largely to have been ignored by the mathematicians working on graph theory.

THE SIMULATED ANNEALING APPROACH

Kvasnička and Pospíchal[8] have made some steps toward this goal by developing programs which can modify a chemical graph so as to modify the ^{13}C chemical shift of a specific carbon described in the graph in some desired direction. This process, which is known as *simulated annealing*, has also been used by Jurs[9] to develop chemical structures with LC_{50} (lethal concentration, a biochemical property) values in a given range. In Kvasnička's approach, which is summarized in Figure 1, a randomly selected molecular graph, represented in the Read linear notation,[10] was perturbed to produce another molecular graph in which, starting at a randomly selected point, all the code is replaced by a randomly generated code. Acceptance or rejection of the new molecular graph is controlled by the Metropolis criterion,[11] and each accepted structure is submitted to a new perturbation. The utility of the annealing approach was demonstrated by generation of 6 of the 128 possible[12] isomeric C_8-C_{10} alkanes which contain a carbon with a ^{13}C NMR chemical shift between 22.68 and 23.89 ppm. This is a very modest result, but it is important as a trailblazer: if this sort of approach blossoms into a generally useful method, it could catch the attention of chemists and chemical engineers alike, because it accomplishes a very elusive goal, the identification of unknown structures that possess specific properties.

INVERSE QUANTITATIVE STRUCTURE−ACTIVITY AND −PROPERTY RELATIONSHIPS

At the same time that chemical graph theory was under development, two different approaches, quantitative structure−activity relationships (QSAR) and the closely related quan-

titative structure−property relationships (QSPR) were being developed. There is a huge literature associated with both of these, but, in brief, the methods both depend upon some sort of a regression which establishes a mathematical relationship between a molecular descriptor and either the (biological) activity or a (physical) property. The method, at its heart, is simply a form of statistical analysis. It is desirable to keep the number of descriptors to a minimum, and so selection of descriptors is an important step in the process. The end product of a QSAR study is a regression equation which relates the property (biological activity or physical property) to the structure descriptors that were used. Thus a relationship between the activity on the one hand and the Hammett σ constant, the octanol/water partition coefficient, usually expressed as its logarithm (log$P$), and a hydrophobic substitution π constant, for example, may be obtained and will allow subsequent estimation of the activity of a new structure.

Just as with the calculation of topological indices, however, this is a one-way process. There is no simple way to proceed from a QSAR equation to an optimal structure. Zefirov and co-workers[13] have recognized this and devised a method to use a regression equation relating the Kier and Hall "κ" indices,[14] $^{0}\kappa$ (symmetry), $^{1}\kappa$ (size), $^{2}\kappa$ (shape), and $^{3}\kappa$ (centrality), with the heat of evaporation (ΔH_e) to obtain optimal structures corresponding to a specific heat of evaporation. The regression equation is

$$\Delta H_e = 3.971\kappa_1 + 1.285\kappa_2 - 0.253\kappa_3 + 2.683$$

for which $r = 0.996$, $s = 0.48$, and $n = 69$. Solution of this equation for all C_6 alkanes with ΔH_e between 28 and 30 kJ/mol gave only 2-methylpentane, 3-methylpentane, and 2,3-dimethylbutane whose measured heats of evaporation, respectively, are 29.86, 30.27, and 29.12 kJ/mol. In a somewhat similar approach, Kier and Hall[15] have developed a means of taking a QSAR equation covering alcohols, esters, and ketones and generating structures of compounds predicted by the equation. The method uses a regression equation cast in terms of the Randić[6] $^{1}\chi$ and $^{2}\chi$ indices derived from the structure and, for a given value of the property, calculates the allowable values of $^{1}\chi$ and $^{2}\chi$. From these values the permissible path counts (*i.e.,* the number of bonds in the shortest path between specified atoms) and then the degree counts (number of attached non-hydrogens) are calculated. Finally the connection matrices are rebuilt from the path and degree counts. The difficult step is generation of the path counts from the $^{1}\chi$ and $^{2}\chi$ values, because specific $^{1}\chi$ and $^{2}\chi$ values do not have a 1:1 correspondence to specific path counts. This one-to-many relationship produces numerous solutions, about half of which have the correct value of the property in question.

Papers such as those by Kvasnička,[8] Zefirov,[13] and Kier and Hall[15] are important, because they show that although difficult, the "inverse problem" is not insoluble. The problems that have been undertaken so far are fairly simple, and the results have not been unambiguous but have shown that the problem can be resolved. It is undeniably useful for chemists to be able to predict properties and activity from a structure, whether by means of topological indices or some form of a QSAR, but the ability to propose a structure with specific properties is a much more substantial goal, which is being sought by other means.
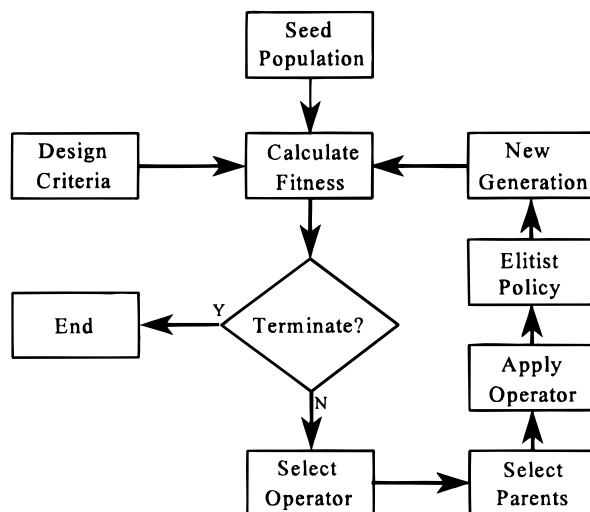

Figure 2. Genetic algorithm for polymer design. From: Venkatasubramanian, Chan, and Caruthers, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 188−195.

CHEMISTRY AND BIOLOGY

In the last decade, chemists have demonstrated an interesting tendency to examine other disciplines in search of ideas as to how to tackle some of the problems that they face. One area that has been very fruitful in this regard is biology. Organic chemists are reasonably familiar with many areas of biology, and one phenomenon which has always been intellectually attractive is that of natural selection. A system in which millions of attempts to reach a goal can be made, with the most successful surviving, has great appeal, because chemistry whose natural statistics are reflected in Avogadro's number, 6.02×10^{23} molecules/mol, lends itself to such an approach.

Genetic algorithms depend heavily upon rules of selection and survival, and work in this area is currently very fashionable. Between 1987 and 1992, just five papers on genetic algorithms in chemistry were published, but since 1993 there have been 210 such papers. The general structure of a genetic algorithm requires a starting point, a means of perturbation, a goal, and, most important, a scoring mechanism to determine whether the perturbation advanced the problem toward the goal or not. In an effort to design polymers with given physical properties, Venkatasubramanian and his group at Purdue[16] developed the genetic algorithm shown in Figure 2. Here, starting from some seed population of typical polymers, a fitness is calculated, genetic operators are selected, and then two "parents" are selected to produce an offspring. This step, controlled by the genetic operators, produces a new structure by crossover or mutation or addition from a working set of new groups, and this new structure is examined for fitness. Fitness is essentially the inverse of the similarity of the structure's estimated properties to the required properties, a normalized number which approaches 1 as the goal is neared. New structures which are more "fit" than their parents are allowed to proceed; those that are less fit do not. An elitist policy allows control by external factors such as chemical stability and environmental acceptability.

This algorithm works remarkably well. Given a specific target polymer with established physical properties (density, glass transition temperature, coefficient of thermal expansion, specific heat, and bulk modulus) it produces different
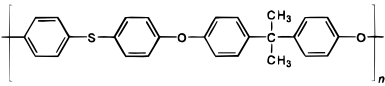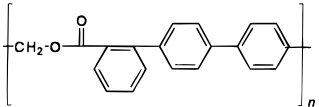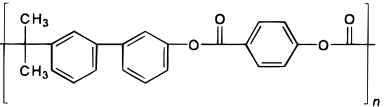
	Density	Tg	Thermal Expansion	Specific Heat	Bulk Modulus
Target compound	1.19	406.83	2.90×10^{-4}	1073.96	5.39×10^{9}

Two best solutions			% Error		
Fitness 0.999	-0.1	0.6	0.1	0.08	-0.04
Fitness 0.995	0.4	-1.0	0.02	1.8	-0.9

Figure 3. Results from polymer design by genetic algorithm.

polymeric structures whose (estimated) properties are within 0.6% of the target properties as is shown in Figure 3. This is typically accomplished within a few hundred generations, which often, but not always, succeed in generating the target structure along with other highly fit candidates. Evolution is a complex phenomenon, but algorithms like this appear to have a firm grip on its major tenets, and as the details are tuned, they presumably will perform even more effectively.

It is interesting to compare Kvasnička's simulated annealing and Venkatasubramanian's genetic algorithm. While there is much similarity between them, the major difference is in the propagation step. Annealing is a stepwise building process in which all earlier modifications are retained so long as they improved the candidate. In the genetic algorithm, mechanisms such as crossovers allows massive deletion of information. As a result, the genetic algorithm explores a much wider search space than the annealing algorithm which tends to proceed more conservatively. Otherwise, there is little difference between the two techniques. Both seek unswervingly to generate species that will best meet the established criteria, and both succeed or fail on the basis of the fitness calculation.

A different computer technique which is also modeled upon a biological process is the neural network. This is an algorithm which makes decisions in a way which is similar to that in which the brain is thought to operate.[17] Some tasks, such as driving a car, are performed very efficiently by humans, apparently because the brain is capable of processing in parallel all the data that go into the calculations that support such activities. The artificial neural network[17] seeks to capitalize upon this design by providing a network of neurons, arranged in layers, which all contribute in some measure to complex decisions. In the typical neural net shown in Figure 4, every connection between two neurons is associated with a weight, a positive or negative real number which multiplies the signal from the preceding neuron. Each node accumulates and sums its various weighted inputs until some preset level is reached, at which point it fires and sends its signal to nodes in the next layer. The classification results achieved by the network are compared to known results, and then the weighting factors are corrected until the correct
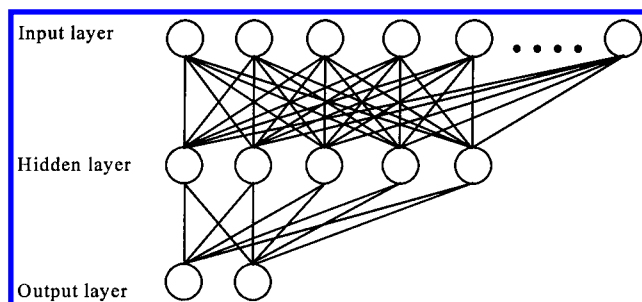

Figure 4. Three layer feedforward neural network. From: Elrod, Maggiora, and Trenary, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 477–484.

results are produced by the net. The correction process moves backwards through every node of the net and is thus called "back propagation". Artificial neural networks do indeed appear to operate much as the human brain does in that they can be trained with actual data and are capable of self-correcting behavior based upon additional experience that is gained in operation. Their application to problems in chemistry is increasingly common, and they can be regarded as examples of the chemist benefitting from a biological design.[18]

ACCEPTANCE OF GRAPH THEORY

Molecular structure and the structure diagram are both fundamental and unique to chemistry. The structure diagram is in effect a specialized language which must be understood by anyone working in chemistry. Mathematicians who have brought graph theory to bear on chemical problems began by replacing the structure diagram with a series of numeric matrices. These are not as easy for chemists to understand as are structure diagrams, but they are immediately manageable in the mathematical sense. As has been shown, mathematical manipulation of such matrices can lead to useful and valuable results.

In spite of this, acceptance by chemists of these methods is minimal, and this can be traced to two causes. First, general applications of graph theory have not been adequately pursued, and second, the "inverse" problems referred to above have been largely avoided by the graph theory school.
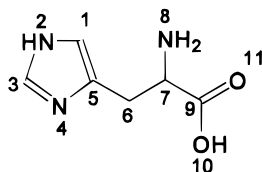
Atom	Symbol	E-State Value		E-State Symbol	Sum Index Value
1	aaCH	1.628			
2	aaNH	2.714		SdO	10.263
3	aaCH	1.490		SsOH	8.418
4	aaN	3.840		SsNH$_2$	5.257
5	aasC	0.666		SaaN	3.840
6	ssCH$_2$	0.263		SaaNH	2.714
7	sssCH	-0.863		SaaCH	3.118
8	sNH$_2$	5.257		SaasC	0.666
9	dssC	-1.006		SsssCH	-0.863
10	sOH	8.418		SdssC	-1.006
11	dO	10.263			

Figure 5. Computed E-state values for atoms in a molecule. From: Kier, Hall, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039−1045.

Table 1. Graph Theory Papers from *JCICS*

total papers (1995 *JCICS*) no.	papers dealing with graph theory		papers on graph theory of hydrocarbons	
	no.	%	no.	%
145	65	45	45	70

HETEROATOMS

With respect to the first of these problems, a pair of statistics is illustrative. A tabulation of papers published in 1995 in the *Journal of Chemical Information and Computer Sciences* is shown in Table 1. That 45% of all papers published during 1995 dealt with graph theory indicates the Journal's sympathetic reception of such papers, but that 70% of all graph theory papers make no attempt to move beyond hydrocarbons speaks to the acceptance problem.

Hydrocarbons are undoubtedly important, but, as a matter of perspective, a standard organic textbook[19] devotes only 27/951 (2.84%) of its pages to them. The richness of organic chemistry flows in large part from the roles played in organic structures by, particularly, nitrogen and oxygen. Further, while the geometry of chemical structures is a major determinant of properties, electrostatic properties are at least equally important. The elemental identity of an atom profoundly affects both of these, yet many of those active in chemical graph theory persist in the apparent belief that the hydrocarbon molecule has broad chemical relevance. This damages the credibility of graph theory in the eyes of chemists who know how different the chemical world is from this view.

An attempt to address this problem is being made by Kier and Hall with their "electrotopological state (E-state) indices" which retain the important electronic characteristics of atoms while adopting a graph theoretical approach to structure representation. The E-state index of an atom is defined as $S = I + \Delta I$ where I is the intrinsic state of the atom and ΔI is a perturbation term. I is given by

$$I = ((2/N)^2 \delta^v + 1)/\delta$$

where N is the principal quantum number of the valence shell of the atom and δ and δ^v are given by

$$\delta = \sigma - h$$

$$\delta^v = \sigma + \pi + n - h$$

(σ and π are the numbers of σ and π electrons, respectively, h is the number of hydrogens bonded to the atom, and n is the number of electron in lone pairs). Thus for an =O atom, $\delta^v = 6$ and $\delta = 1$ and so $I = 7$. Similarly, for a =CH$_2$ group, $I = 3$. The quantities $\delta^v + \delta$ and $\delta^v - \delta$ are related, respectively, to van der Waals volume and valence state electronegativity. The quantity ΔI represents the effect upon I of every other atom in the molecule. The E-state can be computed for every atom in a molecule, and the E-state of the molecule is simply the sum of these atomic E-states as can be seen in Figure 5. When compared to the binary digital so often used in graph theory, the atomic E-state is rich in information, because its value is determined by the elemental character, the electronic properties, and the environment of the atom. These E-states are thus very useful in regression activities, such as boiling points of members of a dataset containing 245 alcohols and alkanes.[20] The regression equation obtained in this way allowed estimation of boiling points with an average error of 5.9 °C—comparable to the traditional estimates conducted on alkanes alone. This demonstration that heteroatoms can indeed be handled by these methods is encouraging, and it is to be hoped that more work along these lines will be forthcoming.

THE INVERSE PROBLEM

The second problem is that insufficient attention has been paid to the "inverse" problems. As has been pointed out, the establishing of a regression equation relating physical properties to structural features such as topological indices

or Hammett constants is useful because such an equation allows the calculation of the physical properties for new molecules. A much more valuable program would be one which, given a set of physical properties, can generate structures which possess these properties. This problem has been addressed with some success by scientists such as Zefirov,[13] Kvasnička,[8] and Kier and Hall,[15] whose work has been discussed above.

Both these problems are difficult. They both are challenging at the mathematical level, but, in addition, they require a significant understanding of the behavior of atoms in molecules, *i.e.*, of the chemistry of the system. Resolution of such problems is essential if this mathematical treatment of chemical structures is to progress. Graph theory has had encouraging success with alkanes, where the chemical complexities are minimal. Continued manipulation of alkanes, with the generation of more matrices and topological indices, is doubtless attractive (and safe), but if such work addresses any real problems, they are for the most part problems in mathematics, not chemistry. The chemistry community has many real problems in the interpretation of chemical structure−property relationships and will support, even collaborate, with serious attempts from the graph theory school to resolve such problems. The mathematicians, on the other hand, must come to terms with the fact that some chemical insight will be necessary if there is to be progress in this area. The skills of both groups are essential, and, if they work together, there is an opportunity for significant progress in this lively area of theoretical chemistry.

REFERENCES AND NOTES

(1) Lederberg, J. Topological Mapping of Organic Molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1965**, *53,* 134−138.

(2) The problem of the Königsberg bridges and Euler's solution to it was profitably revisited by the Stanford group during their successful effort to write a ring-detection algorithm for use with matrices derived from graphs of organic structures. See: Buchanan, B. G.; Duffield, A. M.; Robertson, A. V. An Application of Artificial Intelligence to the Interpretation of Mass Spectra. In *Mass Spectrometry. Techniques and Applications*; Milne, G. W. A., Ed.; Wiley Interscience: New York, 1971; p 129.

(3) Balaban, A. T. Chemical Graphs: Looking Back and Glimpsing Ahead. *J. Chem. Inf. Comput. Sci.* **1995**, *35,* 339−350.

(4) Hosoya, H. Topological Index. A Proposed Quantity Characterizing the Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44,* 2332.

(5) Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structure*; Research Studies Press-Wiley: Chichester, 1993.

(6) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97,* 6609−6615.

(7) Schmalz, T. G.; Klein, D. J.; Sandleback, B. L. Chemical Graph-Theoretical Cluster Expansion and Diamagnetic Susceptibility. *J. Chem. Inf. Comput. Sci.* **1992**, *32,* 54−57.

(8) Kvasnička, V.; Pospíchal, J. Simulated Annealing Construction of Molecular Graphs with Required Properties. *J. Chem. Inf. Comput. Sci.* **1996**, *36,* 516−526.

(9) Sutter, J. M.; Dixon, S. L.; Jurs, P. C. Automated Descriptor Selection for Quantitative Structure-Activity Relationships Used Generalized Simulated Annealing. *J. Chem. Inf. Comput. Sci.* **1995**, *35,* 77−84.

(10) Read, R. C. The Enumeration of Acyclic Chemical Compounds. In *Chemical Application of Graph Theory*: Balaban, A. T., Ed.; Academic Press: London, 1976; pp 25−61.

(11) Metropolis, M.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations for Fast Computing Machines. *J. Chem. Phys.* **1953**, *21,* 1087−1092.

(12) Contreras, M. L.; Valdivia, R.; Rozas, R. Exhaustive Generation of Organic Isomers. 1. Acyclic Structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32,* 323−330.

(13) Skvortsova, M. I.; Baskin, I. I.; Slovokhotova, O. L.; Palyulin, P. A.' Zefirov, N. S. Inverse Problem in QSAR/QSPR Studies for the Case of Topological Indices Characterizing Molecular Shape (Kier Indices). *J. Chem. Inf. Comput. Sci.* **1993**, *33,* 630−634.

(14) Kier, L. B. Indexes of Molecular Shape from Chemical Graphs. *Med. Res. Rev.* **1987**, *7,* 417−440.

(15) Kier, L. B.; Hall, L. H. The generation of molecular structures from a graph-based QSAR equation. *Quant. Struct.-Act. Relat.* **1993, *12,*** 383−388.

(16) Venkatasubramanian, V.; Chan, K.; Caruthers, J. M. Evolutionary Design of Molecules with Desired Properties Using the Genetic Algorithm. *J. Chem. Inf. Comput. Sci.* **1995**, *35,* 188−195.

(17) An early and very clear explanation of the design and operation of a neural network has been provided by Elrod et al. See: Elrod, D.; Maggiora, G. M.; Trenary, R. G. Applications of Neural Networks in Chemistry. 1. Prediction of Electrophilic Aromatic Substitution Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30,* 477−484.

(18) The use of neural networks in chemistry has been pioneered by Gasteiger and Zupan, who have produced two landmark publications on the subject. See: Zupan, J.; Gasteiger, J. Neural Networks: A New Method for Solving Chemical Problems or Just a Passing Phase? *Anal. Chim. Acta* **1991**, *248,* 1−30. Zupan, J.; Gasteiger, J. *Neural Networks for Chemists*; VCH Publishers: New York, 1993.

(19) Allinger, N. L.; Cava, M. P.; DeJongh, D. C.; Johnson, C. R.; LeBel, N. A.; Stevens, C. L. *Organic Chemistry*; Worth Publishers, Inc.: New York, NY, 1971; p 951.

(20) Hall, L. H.; Kier, L. B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence Shell Information. *J. Chem. Inf. Comput. Sci.* **1995**, *35,* 1039−1045.