Journal of Medicinal Chemistry

Volume 28, Number 9 September 1985

Perspective

Computer-Assisted Drug Design

A. J. Hopfinger

Department of Medicinal Chemistry, Searle Research and Development, Skokie, Illinois 60077. *Received November 27, 1984*

In March of 1981 this journal published a Perspective written by Yvonne Martin of Abbott Laboratories regarding the applications of quantitative structure-activity relationship (QSAR) analysis in medicinal chemistry.[1] The timeliness of Dr. Martin's article reflected the dominant role of QSAR methodology, as opposed to other computational approaches, in the design of new chemical entities. The leading references in conformational analysis at the time were cited, but the emphasis of the Perspective was clearly on practicing QSAR analysis in medicinal chemistry.

However, much has happened in the last 4 years to both change and dramatically extend the ways in which computers are being used in medicinal chemistry research. This Perspective attempts to overview and highlight what has happened these past 4 years regarding the evolution of computer-based tools employed in the design of molecules. Four major topics are discussed. A summary of each topic is as follows:

1. First, a set of definitions are suggested to describe individual components of this new emerging discipline growing out of the computer revolution. These definitions are deemed necessary in order to establish a common working vocabulary.

2. The past 4 years are next reviewed in terms of new approaches, as well as changes in existing methods, of applying the computer in medicinal chemistry. This discussion constitutes the major part of the Perspective.

3. An attempt is made to establish and apply some measures to both the impact and success of medicinal chemical design studies aided by the computer.

4. Last, there is a discussion of what might be expected in the immediate future (the next 3 years) regarding technical advances and new medicinal chemical applications of doing research and design on the computer.

This introductory section concludes with a series of references to augment the discussions of the four topics mentioned above. The selected references are targeted for that segment of the readership interested in computer applications to medicinal chemistry, but who are not necessarily well versed in this field.

General Background

1. Franke, R. "Theoretical Drug Design Methods"; Akademie-Verlag: Berlin, 1984.

2. Gund, P.; Andose, J. D.; Rhodes, J. B.; and Smith, G. M. *Science (Washington, D.C.)* **1980**, *No. 208*, 1425–1431.

3. Richards, W. G. "Quantum Pharmacology"; Buttersworths: Boston, 1977.

4. Hopfinger, A. J. "Intermolecular Interactions and Biomolecular Organization"; Wiley-Interscience: New York, 1977.

Applications of Methods and Approaches

1. Kuchar, M. Ed. "QSAR in the Design of Bioactive Compounds"; J. R. Prous International Press: Barcelona, 1984.

2. Topliss, J., Ed. "Quantitative Structure-Activity Relationships of Drugs"; Academic Press: New York, 1983.

3. Weinstein, H., and Green, J. P., Eds. "Quantum Chemistry in the Biomedical Sciences"; New York Academy of Sciences: New York, 1981; p 367.

4. Olson, E. C., Christoffersen, R. E., Eds. *ACS Monogr.* **1979**, *No. 112.*

Molecular Graphics

1. Max, N. L. *J. Mol. Graphics* **1984**, *2*, 8–13.

2. Langridge, R.; Martin, T. E.; Kuntz, I. D.; Connolly, M. L. *Science (Washington, D.C.)* **1981**, *No. 211*, 661–666.

It is also worth pointing out that the journal *Quantitative Structure-Activity Relationships* in *Pharmacology, Chemistry* and *Biology*, edited by J. K. Seydel and F. Darvas and published by Verlag Chemie, reports original works on applications and methods as well as a yearly reference listing of QSAR studies reported in the literature. The *Journal of Molecular Graphics*, edited by A. J. Morrffew and published by Butterworths Scientific Ltd., is devoted to the applications of computer graphics to chemistry and physics and currently has a major bias toward pharmaceutical applications.

Definitions

In any new and rapidly expanding field of science and technology many new terms arise that have somewhat different and/or multiple meanings. Generally, these terms grow in focus and consistency as the new discipline evolves and matures. Application of computer-based methodology in medicinal chemical research is no exception to such growing pains. Consequently, it is appropriate at the outset of this Perspective to set down some definitions that can be used in a self-consistent fashion throughout the balance of the text.

(1) Martin, Y. C. *J. Med. Chem.* 1981, *24*, 229–237.

0022-2623/85/1828-1133$01.50/0

Molecular modeling describes the generation, manipulation, and/or representation of three-dimensional structures of molecules and associated physicochemical properties.

Molecular modeling is a term not restricted in usage to drug design studies but is applicable to design research in all the chemical sciences.

Computer-assisted (or aided) drug design (CADD) includes any application of computer-based procedures for purposes of establishing criteria to relate activity to structure.

Hansch analysis[2] and satellite approaches fall under the umbrella of CADD by this definition.

Molecular graphics refers to the use of the technology of *computer graphics* for the representation and manipulation of molecular structure and properties.

Molecular graphics is the most powerful means of communicating information on chemical structure between man and machine. It is not surprising that in some quarters molecular graphics has come to mean drug design. However, molecular graphics is a CADD tool.

There are two types of computer graphics technologies. One is *vector* graphics which consists of drawing lines (vectors) between prescribed points on the surface of a cathode ray tube. Since the prescribed points can be set very close to one another, it is possible to generate all types of high-resolution curves including numbers and letters. Color vector graphics can be used to differentiate molecular structures through color coding.

Raster graphics is the other computer technology used to represent molecular structure. In this technology the surface of the cathode ray tube is divided into a matrix. Each element of the matrix, called a pixel, can be controlled from software. By requesting different pixels to be "painted" in different colors, or shaded, it is possible to generate space-filled graphical images like those seen on TV. The spatial resolution of this type or graphics representation increases as the number of pixels increases per unit area of the CRT tube.

Solid objects can be portrayed by raster graphics. Vector graphics is normally limited to outlining the surfaces of a solid. However, raster graphics can be used to emulate vector drawings. The quality of the raster-based vector emulation depends upon the size of the pixels. As the pixel size becomes smaller and image resolution increases, the raster representation of lines and curves becomes better.

In general, vectors can be drawn faster than corresponding sets of pixels can be turned on/off. The net result is that the time needed to draw a vector-based "picture" is normally less than that for a raster representation. The drawing time is referred to as the refresh time. If the refresh time is sufficiently small ($<[^1/_{30}]$ s), the human mind cannot differentiate between adjacent refresh cycles. Overall, an object will appear to move and/or change forms in "real time" at these short refresh times. Real-time computer graphics, when applied to molecular representation, allows "instantaneous" alterations to orientation, translation, and/or conformation of one or more molecules. Since refresh time is faster for vector graphics, it has preferentially been used in real-time applications. However, advances in the raster technology should ultimately make real-time raster graphics a practical reality.

Computational chemistry is the quantitative modeling of chemical behavior on a computer by the formalisms of theoretical chemistry.

Computational chemistry methods are essential for providing the most relevant estimations of molecular properties in three-dimensional drug design studies.

Conformation is the three-dimensional structure of a molecule as specified by atomic coordinates of the composite atoms.

Conformational analysis is a method of computational chemistry that allows a calculated relative energy to be associated with each conformation of a molecule.

Conformational energy can be calculated by *molecular mechanics* where the molecule is thought of as set of balls (atoms) held together by springs (bond lengths and angles).[3] Alternatively, or supplementarily, energies can be computed by *quantum mechanics* where molecular structure is represented by electrons flowing about nuclei fixed in space for each conformer state.[4]

Interestingly, many molecular orbital theory methods,[4] which are approximate quantum-mechanical representations of molecular structure and energetics, have been available for as long as 20 years. Molecular mechanics formalisms could also have been used in medicinal chemistry design applications for perhaps as long as the last 15 years. There are at least four interrelated reasons why these approaches to conformational analysis and drug design have not been used to the extent they are now being employed: (1) the high cost and low speed of the past generations of computers; (2) difficulty in using the available computer programs; (3) lack of molecular parameters needed to attack "real problems" in medicinal chemistry; (4) Lack of confidence by synthetic medicinal chemists in CADD.

Those conformations of a molecule that are low in energy are the most likely to be adopted. There are two ways of searching for thermodynamically stable molecular states, that is, doing a conformational analysis. One is to systematically vary each of the degrees of freedom (bond lengths, bond angles, torsional rotations) and to calculate the corresponding conformational energy. This *scanning* approach allows an investigator to approximately locate all stable states for a few degrees of structural freedom. The second means of seeking stable structures is to *minimize* the energy as a function of the degrees of freedom. An energy minimization can be carried out for a large number of degrees of structural freedom and locates stable minima to high precision. However, the minimization strategy does not necessarily locate all stable states.

Clearly, the strengths and shortcomings of scanning and minimizing complement one another. Consequently, these two approaches are often used in tandem in order to optimize the efficiency of identifying thermodynamically stable molecular structures.

Intermolecular energy calculations seeking stable molecular assemblies represent one of the cutting edges of current research in computational chemistry and drug design. No concise term has yet arisen to label intermolecular energy calculations.

Molecular shape refers to the conformation and some measure of the size, for example the volume encompassed by the van de Waals atomic volumes, of a molecule.

Relative similarity and/or differences in molecular shape among molecules is difficult to quantify. Molecular graphics has come to the rescue and provides a visual means of manipulating and comparing two, or more, molecular structures. Corresponding numerical fit procedures

(2) Tute, M. S. *Drug Res.* **1971**, *6*, 1–74.

(3) Boyd, D. B.; Lipkowitz, K. B. *J. Chem. Educ.* **1982**, *59*, 269–274.

(4) Segal, G. A., Ed. "Methods of Electronic Structure Calculations", Parts A and B; Plenum Press: New York, 1977.

have been designed to complement the graphical capabilities. The fit procedures have, for example, been used to minimize distances between specified atom pairs, maximize common overlap steric overlap volumes, and/or maximize potential energy fields.

1981–Now

A. QSAR. The last 4 years probably has not seen a decline in the use of QSAR methodology in medicinal chemistry research. However, there is the preception of a decline, owing to the explosive growth in the use of molecular modeling and molecular graphics in drug design. Research in traditional QSAR methodology continues with emphasis on derivation of new activity descriptors and development of statistical methods to relate structure to activity. Wold and Dunn,[5] for example, have investigated the criteria necessary for the meaningful application of several multivariate methods to establish QSARs. Charton[6] has explored the dependence of transport parameters on chemical structure with the goal of using transport parameters as potential activity correlates.

The octanol–water partition coefficient remains the premier activity correlate in both experimental and theoretical investigations. Environmental and toxicological QSAR-based studies have, in particular, found relative lipophilicity, as measured by partition coefficient, useful.[7] Progress has been made in the measurement and estimation of highly lipophilic materials like the polycyclic aromatic hydrocarbons.[8]

Graph theory and molecular connectivity are still being developed as both alternatives and complements to traditional QSAR methods. The concept of comparability graphics has been applied to the problem of hierarchial ordering of isomeric structures.[9] Kier and Hall remain active in extending and applying their formalism of molecular connectivity to QSAR problems. Recently, they have discussed the $^3\chi_p$ index in terms of its structural information content and measure of molecular flexibility.[10]

Two new trends in QSAR are emerging. First, it is increasingly being recognized that the correlation descriptors used to establish a QSAR need not be restricted to traditional indices like π and σ. Any measured or calculated molecular property can be included as part of a QSAR analysis. The statistical significance of fit between the measures of biological response and the molecular property descriptors ultimately dictates the selection of the preferred descriptors.

The generalization of activity correlates was first accomplished with the introduction of indicator variables by Hansch.[11] However, the intrinsic generality of such an action was not fully recognized. There now are several published QSARs that mix molecular descriptors from a wide range of computational chemistry methods and/or experimental measurements. Our laboratory has been successful in using common pairwise overlap steric volumes calculated from low-energy conformations of molecules and a variety of other molecular descriptors to generate significant QSARs. In the initial development of this formalism, the pairwise common overlap steric volumes and relative lipophilicities of meta and para substituents were used to generate a QSAR for a set of substituted phenyltriazines that inhibit dihydrofolate reductases.[12]

The other new emerging facet in QSAR methodology is to complement the QSAR analysis with a qualitative molecular graphics based ligand–receptor binding study. The goal is to be able to explain the terms in the QSAR through the favorable group interactions identified in the graphical display of the binding geometry. Obviously, the molecular geometry of the receptor must be known, usually from X-ray crystallography, in order to perform the molecular graphics study.

The laboratories of Hansch and Langridge have joined forces on several occasions to carry out combined traditional QSAR and molecular graphics studies. One such study[13] focused upon the binding of phenyl-substituted triazines to dihydrofolate reductases (DHFRs). Molecular graphics modeling was able to qualitatively account for a steric term that was necessary in one of the QSAR equations.

B. Molecular Modeling and Molecular Graphics. Molecular graphics is an integral part of most molecular modeling operations. Hence, the two topics need to be discussed together. Molecular modeling can be broken down into four subunits—structure building, structure analysis, structure comparison, and structure prediction.

The last 4 years has seen great progress in the ease of generating initial three-dimensional structures on a computer. There are two approaches to the generation of structures. The first is to input a description of the atom types and bonding scheme (connection table) for a molecule. This can be done from a terminal keyboard or the more popular and easier graphical route employing a "mouse", lite pen, or similar input control device. The connection table representation of the molecule is translated into a rough three-dimensional structure by a model-builder program. The rough three-dimensional structure can be refined by carrying out conformational energy calculations. A second method to build a three-dimensional molecular structure is to maintain a library of three-dimensional structure fragments. Appropriate structure fragments can be joined together graphically or by keyboard commands to form the molecule of interest. The resulting three-dimensional model is, again, refined by conformational energy calculations.

These two model-building methods have often been combined to minimize the time and effort required for the three-dimensional assembly of virtually any molecule.

Structure analysis encompasses all computational-based approaches to estimate molecular properties. Intramolecular conformational analysis is central to most structure analyses. Unfortunately, there is considerable diversity among investigators doing conformational analyses regarding molecular energy representation, conformational search strategy, and interpretation of conformational data. Part of this diversity can be attributed to both the uniqueness of each drug design problem and the lack of "rules" associated with a new methodology. However, the

(5) Wold, S.; Dunn, W. J., III *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 6–13.
(6) Charton, M. In "QSAR in the Design of Bioactive Compounds"; Kuchar, M., Ed.; J. R. Prous International Press: Barcelona, 1984, pp 25–40.
(7) Leo, A. J. In "Environmental Health Chemistry"; McKinney, J. D., Ed.; Ann Arbor Science Press: Ann Arbor, MI, 1981; pp 323–336.
(8) Mallon, B. J.; Harrison, F. L. *Bull. Environ. Contam. Toxicol.* **1984**, *32*, 316–322.
(9) Bonchev, D.; Mekenyan, O. *J. Chem. Soc., Paraday Trans. 2*, **1984**, *80*, 695–712.
(10) Kier, L. B.; Hall, L. H. *Quant. Struct.-Act. Relat. Pharmacol. Chem. Biol.* **1983**, *2*, 215–227.
(11) Hansch, C.; Silipo, C.; Steller, E. E. *J. Pharm. Sci.* **1975**, *64*, 1186–1194.
(12) Hopfinger, A. J. *J. Am. Chem. Soc.* **1980**, *102*, 7196–7206.
(13) Hansch, C.; Hathaway, B. A.; Guo, Z.; Selassie, C. D.; Dietrich, S. W.; Blaney, J. M.; Langridge, R.; Volz, K. W.; Kaufman, B. T. *J. Med. Chem.* **1984**, *27*, 129–143.

diversity is also tied to the search for the "active conformation"—that conformation of a molecule critical to the expression of biological activity. The credence given to the concept of an "active conformation" and the perceived chances of discovering it from the ensemble of sterically allowed conformations will significantly impact the approach taken in a conformational analysis.

The identification of stable intramolecular conformers is an important consideration in the calculation of electronic properties of molecules using quantum mechanics. Atomic charge density, molecular orbital energy levels, dipole moment, and other electronic properties are dependent upon molecular conformation. The most realistic estimations of these properties are affiliated with low-energy conformers. We have found that the representation of electronic properties using molecular graphics is an effective means of increasing the user-friendly nature of drug design software to synthetic medicinal chemists.

Two physicochemical properties are enjoying growing attention as descriptors in CADD due to both increasing computer power and molecular graphics. The molecular potential energy field generated by a molecule in a particular conformation is increasingly being used to map receptor sites and/or compare molecules. Special molecular graphics representations, including color coding, have been devised to portray potential energy fields.[14] Numerical measures of potential field similarity and difference have also been used in QSAR construction.[15]

Solvent-accessible surface area[16] is another new molecular property gaining attention in structure analysis. The solvent-accessible surface area will likely be explored as a conformation-dependent substitute for lipophilicity (log P) and/or solubility.

Structure comparison largely involves the manipulation and superposition of molecules in various conformer states. Visual structure comparisons are most readily achieved by using real-time molecular graphics. Consequently, most visual structure comparisons have involved stick (vector) representations of molecules. As mentioned above, potential energy field comparisons of molecules are also being performed both visually and numerically.

The most straightforward means of numerically comparing two molecules involves computing distances between key pairs of atoms belonging to each molecule. Optimization methods are being used to maximize molecular shape similarity as a function of superposition criteria and conformation.

Distance geometry[17] is an evolving methodology that can be used to structurally compare a set of molecules and thereby derive a QSAR explicity dependent upon molecular shape. An added bonus from a distance geometry analysis is the postulation of the key interaction sites of the receptor. Molecular shape analysis[12] is another procedure for the quantitative comparison of the conformations of a set of molecules and the corresponding derivation of a shape-dependent QSAR.

Structure prediction is used here to describe two CADD components. One is the molecular graphics based design of new active analogues by constructing the three-dimensional shape of a "new" compound to mimic the shape of the known active compound frozen in its "active conformation". In essence, this process is a combination of structure building and structure comparison. Very often the new compounds are more restricted in conformational freedom, for example, by ring fusion(s), than the active compound being used as the template. Consequently, this strategy of CADD is often referred to as the restricted-analogue approach.

The other part of structure prediction concerns the statistical methods used in CADD to derive QSARs. In some cases the large number of compounds in the data base under consideration, especially in toxicology studies, has necessitated the application of the statistical methods inherent to pattern-recognition theory.[18] The studies of Enslein[19] are exemplary of approaches taken on large data bases. Conversely, certain types of structure analyses lead to so many possible correlation descriptors that some means to prune down the number of descriptors is needed. The approach of Wise and colleagues to comparing molecular shapes is an example of a method that generates a huge number of spatial descriptors necessitating the use of data reduction techniques like principal-component analysis.[20] This aspect of structure prediction using descriptors derived from molecular modeling merges with the methods of traditional QSAR.

As mentioned in Definitions, intermolecular modeling is an active area of research in CADD. The ability to compute the binding constant of a ligand to a receptor macromolecule using intermolecular modeling methods would be a powerful tool for medicinal chemists. However, to reiterate, intermolecular modeling is limited to those cases where the geometry of the receptor is known. This information comes almost exclusively from X-ray crystallography although advances in NMR offer additional hope. At present the crystal structures of far too few enzymes have been determined to make intermolecular modeling a practical tool in pharmaceutical product research. Further, the capacity to meaningfully compute binding constants appears intimately dependent upon solvation/desolvation processes taking place during ligand binding, as well as the exasperatingly large numbers of geometric degrees of freedom inherent to the ligand–macromolecule coplex. These are difficult problems to solve in computational chemistry, but not insurmontable. All in all, there is a growing need for X-ray protein crystallographers and computational chemists to join forces to make intermolecular modeling a practical tool for medicinal chemists.

The major methods of CADD are shown in Figure 1. The flow chart representation suggests a hierarchial ranking of the order, from easiest to most difficult, in which they might be applied in a CADD study. Essentially, Figure 1 summarizes the CADD methods presented and discussed in this Perspective.

Measures of Success

There are a variety of measures that can be used to evaluate the success of CADD in pharmaceutical research. An attempt has been made in this section to associate some data with some of the more obvious measures. The examples cited, to be sure, constitute an incomplete set. My advance apology for any obvious omissions. The criteria used in selecting the examples are either documented

(14) Quarendon, P.; Naylor, C. B.; Richards, W. G. *J. Molec. Graphics* **1984**, *2*, 4–7.
(15) Hopfinger, A. J. *J. Med. Chem.* **1983**, *26*, 990–996.
(16) Connolly, M. *Science (Washington, D.C.)* **1983**, *No. 221*, 709–714.
(17) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1982**, *25*, 892–899.
(18) Wold, S.; Dunn, W. J., III; Hellberg, S. In "Drug Design: Fact or Fantasy"; Jolles, G., Woolridge, K. R. H., Eds.; Academic Press: New York, 1984; pp 95–117.
(19) Enslein, K. "HDI Toxicology Newsletter"; Health Designs Inc.; Rochester, NY, 1984; No. 3.
(20) Wise, M. In "Proceedings of the Fifth European Symposium on QSAR"; Seydel, J. K. Ed., Verlag-Chemie: Deerfield Beach, FL, in press.

Table I. Computer-Assisted Drug Design Success Stories

organization	therapeutic area	computer-assisted drug design method	status: development	status: clinic
Syntex	cardiovascular β-blocker	Hansch anal.		X
	antiinflammatory	Hansch anal.	X	
Searle	antiarrythmic	molec shape anal.	*a*	*a*
	peptide analgesic	conformnl anal.	X	
Smith-Kline Beckman	antiallergic pyraneamine	Hansch anal.	X	
Merck	antihypertension renin inhib	molec shape matching	X	
SRI	nonaddicting opiate analgesic	comput chem	*a*	*a*
FMC	pyrethroid insecticide[b]	Hansch anal.	X	

[a] Clinical testing is imminent. [b] Success stories from agriculture chemistry are considered valid entries to this table.

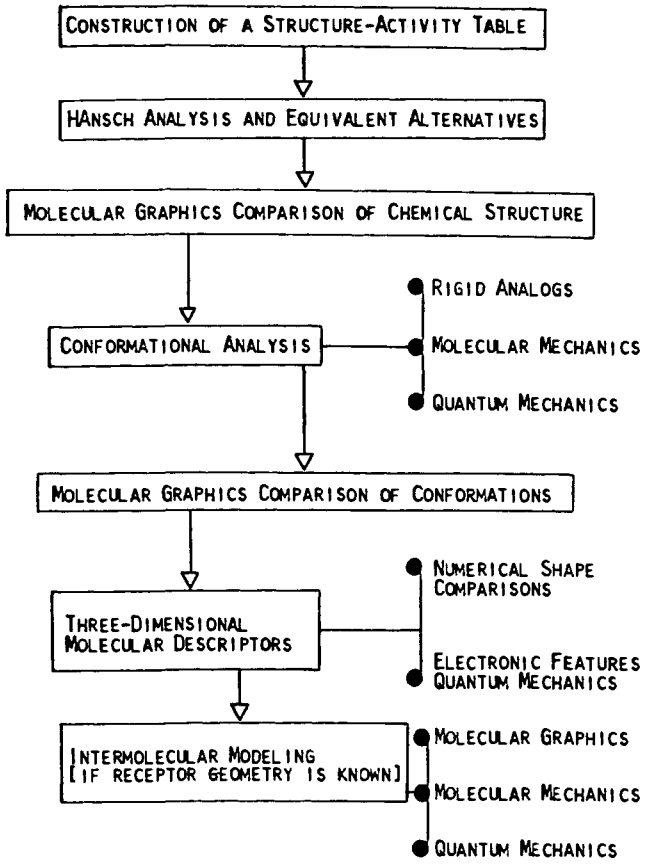


Figure 1.

publication, private discussions with key scientists on successful projects, and/or personal involvement.

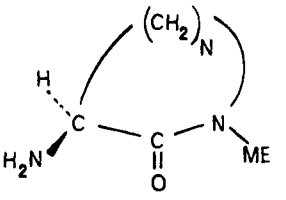
The questions most often asked of industrial applications of CADD is how many drug products have been computer designed. This is not a fair question to ask since CADD is applied only to the research phase of the preclinical component of the drug discovery process. All the pitfalls inherent to preclinical development and clinical testing that "kill" a drug candidate are beyond the control of a CADD analysis. Hence, a more accurate question to ask is how many compounds coming from projects using CADD have gone into the development phase.

This author contacted most major U.S. pharmaceutical companies to learn of successful design stories. Nearly all the companies interviewed had examples of successes but often could not reveal details because the work was quite recent and the findings still proprietary. Table I contains what must be viewed as a partial list of success stories in which a compound has been designed with some computational assistance and has reached at least the preclinical development phase. The CADD method used to conceive each compound is also stated. The predominance of Hansch analysis[2] in Table I as the method of drug design probably reflects its longer use as a drug design tool when compared to other methods.

Fujita recently summarized the successful applications of QSAR analysis in Japan.[21] One of the examples cited is the design of AM-715, a quinolonecarboxylic acid antibacterial agent, at Kyorin Pharmaceutical Co. This compound has been recently marketed by Merck in Italy. As such, it may be the first CADD product.

Another measure of the successful application of computer-assisted molecular design involves predictions that have been proven correct. This measure relates more to the scientific fidelity of the methodology than to the "bottom line" of enhancing scientific productivity (producing new drugs). There have been several novel predictions that have been substantiated.

At Merck the prediction was made, and experimentally validated, that the inhibition potency of a set of aminolactam compounds (I) against the angiotensin-converting enzyme (ACE) would be maximum at $N = 4$.[22] This prediction was based upon comparisons of the conformations of I to that of MK-422, a potent Merck ACE inhibitor.



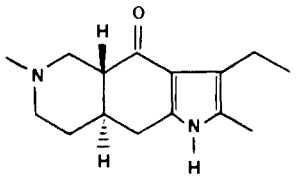
I

Scientists at Roche were able to successfully design bicyclic mimetics of captropril, another potent ACE inhibitor developed by Squibb, using a combination of X-ray crystallography, nuclear magnetic spectroscopy, and molecular graphics methods.[23]

Cohen carried out a conformational comparison of six dopamine receptor ligands.[24] This led him to propose that

(21) Fujita, T. In "Drug Design: Fact or Fantasy"; Jolles, G., Woolridge, K. R. J. H., Eds.; Academic Press: New York, 1984; pp 19–33.
(22) Thorsett, E. D.; Harris, E. E.; Aster, S. D.; Peterson, E. R.; Tristram, E. W.; Snyder, J. P.; Springer, J.; Patchett, A. A.; Proceedings of the 8th American Peptide Symposium, Tucson, AZ, May 22–27, 1983; Abstr. 6–1.
(23) Hassall, C. H.; Kiohn, A.; Moody, C. J.; Thomas, W. A. *FEBS Lett.* **1982**, *147*, 175–179.

a set of pyrroloisoquinolines (II) should have the required

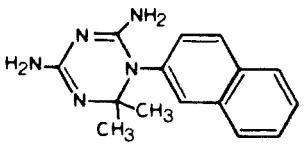


II

conformation and functional groups for biological activity. It was also possible to predict that only the $4\alpha(R),8\alpha(R)$ configuration (the trans fusion of the two six-membered rings as indicated in II) of the chiral derivatives would have the expected biological properties. This was fully confirmed. Studies with avoidance blockade tests showed that the antipsychotic-like activity of the (−)-II isomer is 7 times more potent than that of molindone and comparable to that of haloperidol.

Scientists of Burroughs Wellcome have been able to design potent in vitro DHFR inhibitors similar in structure to trimethoprim from the manipulation of X-ray structural data using molecular graphics.[25]

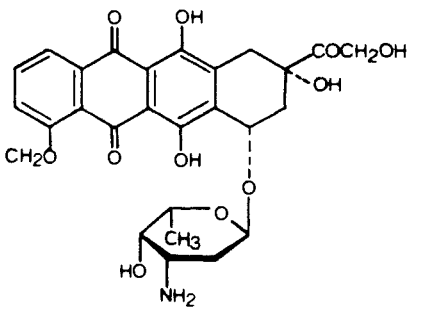
This author was fortunate enough to be involved in two correct predictions prior to the publication of Yvonne Martin's Perspective.[1] On Jan 15, 1980, as part of a seminar at Warner-Lambert Co., III was predicted to have log



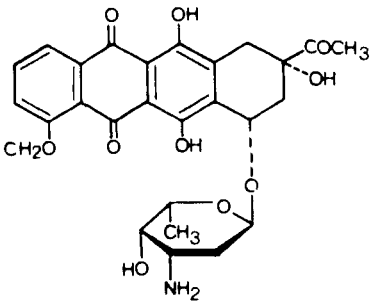
III

$(1/C) = 7.62$ where C is the molar concentration necessary for 50% inhibition of bovine liver DHFR based upon a molecular shape analysis QSAR study.[12] Dr. M. L. Black of Warner-Lambert noted that III had been synthesized for other purposed at Warner-Lambert. A sample of III, provided by Warner-Lambert, was subsequently tested in Corwin Hansch's laboratory at Ponoma College. Log $(1/C)$ was found to be 7.36!

The second successful prediction involves a novel intermolecular binding geometry.[26] The intercalation of anthracyclines like doxorubicin (IV) into DNA was assumed to be similar to that observed by X-ray crystallography for other known intercalators such as ethidum and proflavine. Succinctly, this intermolecular geometry is characterized by the long axis of the ring system of the drug being nearly parallel to the two long axes of the DNA hydrogen-bonded base pairs above and below the intercalated drug. Insertion of the drug takes place along the major groove of the DNA. Our calculations indicated the complete opposite for doxorubicin. That is, insertion of the anthracycline ring system proceeds along the minor groove, with the long axis of the ring inserting nearly perpendicular to the long axes of the adjacent DNA base pairs. The crystal structure of the complex of daunomycin (V), a doxorubicin congener, with a hexanucleotide dimer was subsequently solved by Quigley et al.[27] The crystal structure is virtually identical with the predicted geometry.



IV



V

Many investigators have wondered how patents might be handled for compounds synthesized and tested on the basis of CADD predictions. Interestingly, K. Yano of Yano-Miyazaki Research Laboratoty has taken this issue one step further by successfully obtaining U.S. Patent 4 432 732 for the invention titled "Method for Predicting Properties of a Chemical Compound"![28] The major demonstration of the claim in this patent is the design of artificial sweeteners. This apparent confidence in molecular design shown by the U.S. Patent Office can be viewed as a measure of success of CADD in product research.

The last measure for evaluation of the role of CADD in medicinal chemical research to be mentioned here is prediction accuracy. This author has not been able to obtain the data on prediction accuracy at any organization other than his own–Searle. Thus, only one "data point" is available, but it is worth discussing. A total of 83 compound predictions have been made at Searle with computer assistance over the last 3 years. Only 57 of these compounds were synthesized and tested due to a variety of reasons, the most common being unforeseen synthetic difficulties. A total of 24 compound predictions were accurate, yielding a correct prediction percentage of 42%.

A critic might argue that slightly better predictive accuracy could be achieved by flipping a coin! However, such a point of view assumes that each prediction depends upon a single variable. In reality, several structural properties must be concurrently considered, and synthetic constraints met, in making a compound prediction. A more realistic analogy, at least in terms of the actual number of variables (>10) involved in a "typical" prediction, would correspond to successfully selecting the faces of two dice thrown simultaneously. In this case the probability of a correct prediction is $^1/_{18}$, or about 6%.

Those who support CADD would argue a 42% correct prediction rate is a major success, the point being that currently far more than two (50%) to three (33%) compounds have to be made, on the average, in order to

(24) Cohen, N. C. Presented at the North American Medicinal Chemistry Symposium, Toronto, CA, June 20–24 1982.
(25) Kuyper, L. F.; Roth, B.; Baccanai, D. P.; Perone, R.; Beddel, C. R.; Champnees, J. N.; Stammers, D. K.; Cann, J. G.; Norrington, F. E. A.; Baker, D. J.; Goodford, P. J. *J. Med. Chem.* 1982, *25*, 1122–1124.
(26) Nakata, Y.; Hopfinger, A. J. *Biochem. Biophys. Res. Commun.* 1980, *95*, 583–588.
(27) Quigley, G. J.; Wang, A. H.-J.; Ughetto, C.; Van der Marel, G.; Van Boom, J. H.; Rich, A. *Proc. Natl. Acad. Sci. USA* 1980, *77*, 7204–7208.
(28) Yano, K. U.S. Patent 4 432 732, 1984.

achieve some expected (predicted) result—potency, therapeutic index, etc. However, while only 83 compound predictions were made at Searle, we used all relevant structure–activity data within and outside Searle in designing these 83 compounds. Thus, the actual number of compounds considered in making the predictions is considerably more than 83. This should be reflected in any measure of predictive accuracy.

Summary and Future Trends

CADD is evolving into the application of an increasingly large variety of computer-based methods with the intermediate objective of identifying relationships between molecular structure and corresponding biological activity. The ultimate goal of CADD is to use these identified structure–activity relationships to predict compounds having predesired activity profiles.

The pharmaceutical industry has been first to apply computer-based molecular design methods. Most major drug companies in the U.S., Western Europe, and Japan have CADD programs as part of their research organizations. These groups have probably not yet been in place long enough to fully ascertain their actual impact on the drug discovery process. However, some preliminary data, in terms of successfully designed compounds and efficiency of lead optimization, have been reported in this Perspective.

The comments on the Searle drug design program probably raise the most controversy. CADD has not had the *net* impact on Searle so as to enhance the research phase of the preclinical drug discovery process to near 50% reliability. However, there is a clear increase in the efficiency of lead identification, and/or lead optimization, for most research projects in which CADD methods have been employed. This is likely the case at other organizations where CADD is being carefully used by experts as a tool in medicinal chemistry research.

In the immediate future, research in intermolecular modeling should intensify and breakthroughs can be expected in both theory and application. Certainly real-time intermolecular energy calculations should become commonplace. The potential of intermolecular modeling in designing novel ligands will, in turn, spur growth in the X-ray crystallography of biological macromolecules. X-ray crystallography groups can be expected to become commonplace in pharmaceutical research departments and complement CADD groups.

The extended usage of molecular spectroscopy, especially NMR and IR methods, in medicinal chemistry applications can be expected to benefit from CADD. Computational chemistry appears to be on the verge of reliably predicting spectral properties as a function of both chemical structure and conformation. This capability should make it possible to match calculated and observed spectral properties. Since the calculated spectral properties are directly related to conformation, it may be possible to reliably identify conformations in solution and/or conformations of bound ligands. The capacity to compute spectra as a function of conformation may go a long way to making NMR and IR spectroscopies equally useful to X-ray diffraction in the elucidation of three-dimensional structures of molecules.

The next few years should see the emergence of artificial intelligence as a means to make computers more friendly, and more forgiving. This should be reflected in CADD software so that more synthetic medicinal chemists will find it agreeable to do hands-on CADD. Along these lines, the fledging, but emerging, chemical sciences software industry should continue to grow and become increasingly competitive. However, it is extremely difficult to speculate whether or not groups doing software development within pharmaceutical companies will continue their activities as the corresponding software service industry grows. By analogy to the management information systems, MIS, field, the CADD software development groups within individual pharmaceutical companies would become caretaker users of commercial packages.

Suggestions have been made to this author that he give his opinions and/or viewpoints regarding (1) what the "best" way (methods) of doing CADD is, (2) whether practicing synthetic medicinal chemists should do hands-on CADD studies, (3) how CADD groups should be organized and staffed in pharmaceutical houses, and (4) what will be the long-term impact of most CADD being done in industry as opposed to academia.

These are important questions, but they do not necessarily have definitive and/or unique answers. Consequently, this author feels it is important to state the questions but leaves it to the reader to answer these questions as it relates to his/her situation.

Other disciplines within the chemical sciences are now beginning to pay attention to molecular modeling, computational chemistry, and molecular graphics. In particular, polymer scientists appear poised to use this new technology. Quite likely, CADD will become an extinct term being replaced by CAMD—computer-assisted molecular design—to reflect the general utility of this new and exciting technology.