The role of structure-based ligand design and molecular modelling in drug discovery

• J.P. Tollenaere

Introduction

Nowadays, new and better medicines result more often than not from a concerted multidisciplinary approach towards treating a medical problem or curing a disease. Modern drug research is a multidisciplinary endeavour, in which medicinal chemists, pharmacologists, molecular biologists, biochemists, pharmacists, computer scientists, theoreticians, clinicians, *etc.* work together towards the common goal of finding and developing an improved or new medicine.

Modern pharmaceutical research, however, faces fundamental obstacles bearing on yet ill-understood complexities of living matter. Scientific and technological developments in a wide range of fields, such as molecular biology, physiology, chemistry, physics, information science and computer technology, have facilitated the unravelling of these complexities. With our better understanding of normal and diseased states, our desire to intervene pharmacologically in disease processes has grown.

The need for the continual development of new drugs hardly needs to be pointed out, in view of the global health and disease situation. Painfully apparent is the need for drugs that will effectively halt the spread of the human immunodeficiency virus. Other scourges such as tuberculosis, malaria, and various parasitic plagues continue to afflict many millions of people. With increased life expectancy, diseases such as Alzheimer's and other central nervous system malfunctions will increasingly require the development of new drugs.

Traditionally, the drug discovery process relied almost exclusively on the synthesis and subsequent pharmacological testing of many thousands of chemicals. Although almost from the very beginning of contemporary medicinal chemistry the idea was accepted that there was a relationship between chemical properties and biological activity, the synthesis of compounds was to a large extent dictated by what was synthetically feasible. Pharmacological testing was based on relatively simple animal models. The entire process of drug discovery was not as multidisciplinary and integrated as it is today. Yet, it is an undeniable fact that the traditional approach has led to the discovery of most prototype structures and to the development of most drugs in use today.

In this overview the most prominent developments during the last decade or so in computer-assisted structure-based ligand design are discussed, with special emphasis on searching three-dimensional databases with a pharmacophore as a query. The structure-based chemical diversity of chemical libraries and combinatorial chemistry are discussed as key developments in the ever increasing armoury of tools available to those involved in the discovery and development of new or improved drugs.

Tollenaere JP. The role of structure-based ligand design and molecular modelling in drug discovery. Pharm World Sci 1996;18(2): 56-62.

J.P. Tollenaere: Janssen Research Foundation, 2340 Beerse, Belgium, and Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Utrecht University, P.O. Box 80082, 3508 TB Utrecht, The Netherlands.

Keywords
Computer graphics
Databases
Drug design
Molecular conformation
Nuclear magnetic resonance
Receptors
Structure–activity relationship
X-ray diffraction

Abstract
Structure-based ligand design is a technique that is used in the initial stages of a drug development programme. The role of various computational methods in the characterization of the chemical properties and behaviour of molecular systems is discussed. The determination of the three-dimensional properties of small molecules and macromolecular receptor structures is a core activity in the efforts towards a better understanding of structure–activity relationships.

Accepted September 1995

Basic assumption

The basic assumption when dealing with molecules that elicit a biological response, BR, is that there is a relation between the structure, S, of the molecule and the physicochemical properties of the molecules that are ultimately responsible for the biological activity. Succinctly stated:

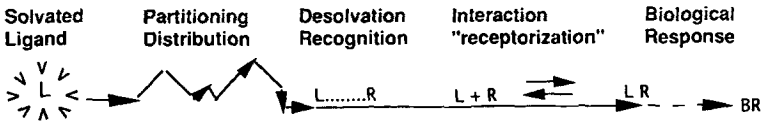
$$BR = f(S)$$

where S may be any or a combination of physical properties [1]. Properties of S include:

- lipophilicity (log P) and solubility;
- ionization constant (pK);
- shape (volume, conformation, configuration, solvent-accessible surface area) and steric factors;
- atomic charges and dipole moment (μ);
- molar refraction and polarizability;
- electronic structure, energy levels (*e.g.*, highest occupied molecular orbital, lowest unoccupied molecular orbital), reactivity indices, hydrogen bonds, rotational barriers, entropic and hydrophobic effects.

This list, though by no means exhaustive, contains a number of properties, many of which can only be obtained by computational means, for example, electronic structure, highest occupied molecular orbital, lowest occupied molecular orbital, and reactivity indices. Bulk properties such as pK and solubility are only accessible by experiment while log P can be experimentally determined if the compound is physically available or has to be calculated if the prospective compound has yet to be synthesized. Methods for calculating log P, being an additive constitutive property, are based on the addition of fragment values together with factors, taking into account any fragment interaction [2-3]. The identification of those properties that are possibly relevant or are correlated with the biological activity is not easy and often depends on trial-and-error, particularly when the underlying reaction mechanism is not known. Therefore, it is of interest to envisage the hypothetical sequence of events during the trajectory of a drug or ligand, L, from its injection into the biological system till it elicits a biological effect (Fig. 1).

At the stage of partitioning and distribution, one might envisage a drug or ligand molecule crossing several cell membranes and lipid barriers, and that properties such as partitioning behaviour, as reflected by the log P parameter, may satisfactorily describe the underlying phenomena. At the desolvation and recognition phase, physical quantities and concepts, such as hydrogen bond formation and breaking, and entropic and hydrophobic effects, may be invoked to rationalize biological data. At the drug–receptor interaction stage, physical properties such as conformation, configuration, dipolar effects, the highest occupied molecular orbital or the lowest occupied molecular orbital may effectively govern the interaction. In practice, however, there are no sharp divisions between the various stages. Thus, the appearance of log P in a regression equation may in fact reflect a genuine partitioning effect governing biological activity, as might be expected in *in vivo* testing, but may equally well model desolvation processes prior to binding to the hydrophobic surface of a protein. A strong correlation between log P and the biological response can also be interpreted in terms of a complete engulfment of the ligand in a hydrophobic space of the macromolecule or receptor [4].



▲ Figure 1

Trajectory of a ligand, L, as from the injection into the biological system till its interaction with the receptor, R, ultimately leading to the biological response, BR.

The conformation of a ligand or drug is of utmost importance in all quantitative structure–activity relationships, except in the traditional Hansch-type quantitative structure–activity relationships, where only topological or two-dimensional information of a congeneric series of molecules is needed [1]. In fact, many if not all physicochemical properties associated with structure, S, depend on the spatial arrangement of the atoms in the molecule. Therefore, conformational aspects are core concepts in any modelling experiment. A good understanding of the limitations and applicability of the techniques used to determine the conformation of molecular structures is therefore of crucial importance in structure-based ligand design.

Conformational analysis

At the forefront of any molecular modelling experiment is the question as to what shape(s) or conformation(s) a molecule can adopt. Molecules can be observed in three aggregation states, *i.e.*, the solid or crystalline, the dissolved, and the gaseous state [5].

Solid state

Provided suitable crystals are available, X-ray diffraction experiments on small molecules lead to the precise location of each atom of the molecule within the crystal lattice [6]. Although the conformation observed in the crystal possibly pertains to a minimum energy conformation, there is no *a priori* reason for this conformation to be biologically relevant. That is, though experimentally determined, this conformation is not necessarily the conformation recognized by the receptor or the conformation required for a productive drug–receptor interaction. Molecules such as polypeptides and proteins need special techniques to grow suitable crystals. Even

with suitable crystals, the X-ray structure determination of proteins becomes more difficult the larger the protein. Nevertheless, with the increasing availability of modern X-ray diffraction equipment, the pace of successful protein structure determination is picking up momentum [7], reminiscent of that of the small molecules in the Sixties and Seventies.

Dissolved state

Molecules in solution form the second aggregation state. The method of choice for determining a molecule's conformation in solution is nuclear magnetic resonance (NMR). Apart from small molecules for which NMR techniques have been used for decades, NMR spectroscopy over the last decade has become an important tool for the structure determination of polypeptides and small proteins [8]. In cases where no suitable crystals can be obtained, NMR is the only method available for the determination of the structure of peptides and proteins up to approximately 30,000 Da. Once more, although drug molecules in physiological conditions are in the dissolved state, the solvents used in an NMR experiment, such as chloroform or dimethylsulfoxide, must be considered to be poor mimics of physiological environmental conditions.

Gaseous state

Molecules can also be studied in a third aggregation state, namely, the gaseous or isolated state. Apart from microwave spectroscopic techniques suitable for the analysis of small molecules, the method of choice for studying the conformational aspects of drug molecules is computation, either by quantum chemical or molecular mechanical procedures. Calculations for the isolated state or *in vacuo* calculations are usually conducted in the complete absence of any environmental effects of the medium surrounding a molecule and therefore their biological relevance may be questioned.

Despite the fundamental objections that can be raised against any of these approaches, the combination of the three methods yields a complete picture of the conformational profile of a molecule. In the case of small organic molecules, this knowledge may be used to propose a pharmacophore (*i.e.,* the spatial disposition of atoms or groups of atoms that is required for the recognition of and interaction with a receptor) that can be either used as a lead for the synthesis of new compounds or used to construct a hypothetical receptor model. Pharmacophoric models are now also used as queries for searching three-dimensional databases, enabling the identification of structures that match the pharmacophoric pattern.

In the case of peptides and proteins, the combination of X-ray diffraction and NMR spectroscopy may provide meaningful comparisons between crystal and solution structures. Either the structural similarities or differences in both aggregation states may say something about the dynamic behaviour, salt and temperature effects of peptides and proteins [9].

Multiple minimum energy problem

As a molecule becomes conformationally more flexible, one is rapidly faced with the problem of finding the minimum energy conformation. The magnitude of the problem can be illustrated by consideration of a molecule with, for example, six rotatable bonds. If it is assumed that each bond can rotate a full circle in steps of 10°, one faces the combinatorial problem of calculating N_{conf} conformations:

$$N_{conf} = (360/10)^n = 36^6 \approx 2.1 \bullet 10^9.$$

At a speed of 100 conformation evaluations per second, this brute force approach means asking for some 250 days of CPU (central processing unit) time from your local computer department. It is clear that the brute force approach using a traditional sequential-architecture computer is only feasible for molecules with four to five rotatable bonds and that fundamentally different methodologies are required to approach the minimum energy conformation problem of polypeptides and proteins that have hundreds to many thousands of rotatable bonds.

Instead of using the systematic search in torsional angle space, also called the non-adiabatic grid search, one may resort to various energy minimization techniques [10 11]. All standard energy minimization methods, such as steepest descents, conjugate gradients, and Newton–Raphson, always proceed downhill. Thus, the energy and the corresponding structure obtained after the minimization are strongly dependent on the starting structure and do not necessarily represent the minimum energy conformation. Sampling the conformational hypersurface by random generation of starting structures followed by static energy minimization may lead to the identification of the minimum energy conformation. Molecular dynamics [11], simulated annealing [12], and Monte Carlo methods [13] are suitable for the generation of random starting structures.

Computational techniques

Broadly speaking, there are two fundamental classes of computational methodologies that can be used to simulate chemical behaviour, namely, the quantum chemical and the molecular mechanical approaches, each of which has its advantages and disadvantages. In general, quantum chemical calculations, even at the semi-empirical level of approximation, are not practically feasible for molecular systems containing more than 200 to 300 atoms. For systems such as polypeptides or proteins that have several thousands of atoms, molecular mechanics methodologies are invariably used.

Quantum chemical calculations

For relatively small molecular systems, quantum chemical calculations ranging from various semi-empirical approximations up to the highest quality *ab initio* methods are used. The choice of a particular approximation depends on the system at hand. Because even quantum chemical methodologies have their strong and weak points, only experience with the various methods can tell which method is to be used for a particular problem. It can be stated, however, that semi-empirical calculations prove to be of major value in the fairly accurate description of relatively small molecular systems and in predicting chemical behaviour in a series of compounds [14].

Molecular mechanics

The purely empirical approach or the so-called force field calculation is based on a purely mechanical model of a molecule, which is considered to be an assembly of point masses connected by springs, and susceptible to classical motions such as bond stretching, bond angle bending, torsional motions, and non-bonded Van der Waals and electrostatic interactions [15]. Owing to the simplicity of the mathematical expression (Eq. 1 and 2), molecular mechanics is particularly well-suited for the calculation of the internal energy, V, of a molecular system comprising several thousands of atoms.

$$V = \Sigma V_{bonds} + \Sigma V_{angles} + \Sigma V_{torsion} + \Sigma V_{vdw} + \Sigma V_{coul} \qquad \text{(Eq. 1)}$$

$$V = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{tors} \frac{1}{2} K_\phi [1 + \cos(n\phi - \gamma)] + \sum_{i<j} (b_{ij}/r_{ij}^{12} - a_{ij}/r_{ij}^{6}) + \frac{1}{\varepsilon} \sum_{i<j} q_i q_j / r_{ij} \qquad \text{(Eq. 2)}$$

The molecular mechanics approach is widely used for the elucidation of the conformational aspects of both small drug molecules and macromolecular structures. Because of their empirical nature, molecular mechanics calculations are often hampered by the lack of suitable parameters K_r, K_θ, K_f, a_{ij}, and b_{ij}, which are often not available for particular molecular fragments [16]. Because of its classical nature, molecular mechanics is totally inadequate to model bond breaking and bond formation processes.

The potential energy of a molecular system consisting of interacting particles should, in principle, be treated by using quantum mechanical methods. As the size of biological systems, and thus the number of particles, is so large, quantum mechanical methods are in practice not feasible for the description of peptides or proteins. However, semi-empirical quantum mechanical methods are suitable for relatively small molecules.

The reliability of empirical potential energy or molecular mechanics calculations depends on the energy terms included in the total energy function and the numerical values of the parameters. As molecular mechanics calculations can routinely be done by virtually all modelling software packages, it is of interest to have a closer look at the applicability and limitations of force fields.

The advantage of being conceptually simple, because valence force fields are expressed in terms of bond stretching, bond angles, torsion angles, *etc.*, is at the same time a weak point. Over the years, several force fields have evolved and gradually migrated from academic to commercial environments. By now, every commercially available molecular modelling software vendor has its own force field in different stages of development. In fact, companies that started offering small molecule modelling capability have acquired and incorporated an academic force field, in order to do macromolecule modelling.

Similarly, modelling software originally intended for macromolecules gradually saw the incorporation of force fields capable of treating small organic molecules. Not only have existing force fields different ranges of applicability, but they also have different mathematical functions for the expression of the potential energy. Force fields may, for example, differ in how non-bonded van der Waals interactions are treated; they may contain special functions to treat the interaction between the hydrogen atom and the receptor atom in a hydrogen bond in the form of a (10-12) function including functions reflecting the A–H...B angle dependence of a hydrogen bond whereas other force fields treat hydrogen bonds as purely Coulombic. Some are purely diagonal while others may include off-diagonal cross terms. The way electrostatic interactions are treated may vary widely, ranging from simply omitting electrostatics to the use of point charges, bond dipoles, atomic polarizability, and higher order atomic multipole moments [17-20]. Adding to the despair of the user, software vendors acknowledge the less-than-optimal quality of their force fields and therefore, more often than not, each new release of a modelling package contains some changes (not always fully documented) in the force field used. The effort to try to improve or alter things, laudable as it is in principle, has, in practice, the effect that customers have to invest much time in bug hunting and in evaluating the new 'improvements'. Furthermore, research projects lasting considerably longer than the software release cycle time may not always benefit from the latest release, because results of the computations may differ if the force field is altered from one release to the next.

In general, where it is even obvious that successive releases of one and the same software package are not always transferable, force fields from different origins are not transferable. Force fields that are accurate over a limited range of compounds are not necessarily accurate for a somewhat different set of compounds. As long as force fields continue to contain parameters that are not truly transferable, the situation will prevail in which the ultimate choice of a force field is to a large extent a software-vendor-driven process.

In conclusion, due to their importance in the simulation of the chemical behaviour of small molecules and macromolecular systems, the accuracy and the applicability of the routinely used force fields are and should be issues of constant concern for the theoretical medicinal chemist using them.

Molecular dynamics

Instead of the static picture one gets from the computations thus far discussed, the atoms of a molecule are actually in constant motion. These motions around an equilibrium position, or even larger fluctuations involving the movement of side chains of the amino acids of proteins can be simulated by the technique of molecular dynamics. Molecular dynamics simulations are based on the knowledge of the energy of a system (Eq. 1) as a function of the atomic coordinates [21-22]. The force acting on each atom, F_i, is related to the first derivative of the potential energy, V, with respect to the atom position. Solving Newton's equation by using this force leads to the motion of the atoms as a function of time:

$$F_i = -\partial V_i / \partial x_i, \; F_i = m_i a_i$$

Molecular dynamics simulations are frequently used to examine the possible conformational domains of small molecules and macromolecular systems and may lead to the assessment of the conformational flexibility of a molecule. Molecular dynamics calculations are one of the strategies that can be used to generate low-energy conformations.

Databases of molecular structure

Molecular structure databases combined with computer-assisted molecular modelling are of vital importance in ligand design. As already mentioned, X-ray crystallography is one of the main sources of information bearing on the three-dimensional structure of small molecules and macromolecules. It is clear that molecular modelling without the knowledge gained from X-ray crystallography would have to rely solely on theoretical models of molecular structure. There are two X-ray crystallographic databases that are used by those interested in structure-based ligand design.

The Cambridge Structural Database, which contains the X-ray structure coordinates of small organic and organometallic compounds (120,481 entries in the April 1994 release), is the prime source of experimentally determined information regarding the three-dimensional characteristics of small organic molecules [6]. In many instances retrieving structures from the Cambridge Structural Database will give high-quality models of structures, which can subsequently be modified into the desired structures by use of molecular modelling techniques. Often X-ray structures are used as input structure for theoretical calculations.

The Brookhaven Protein Data Bank contains the coordinates of protein and nucleic acid structures [7]. The Protein Data Bank (2,327 entries in the January 1994 release) offers a rich source of information about the tertiary structure of proteins. Detailed analysis of the data may often give a better understanding of specific molecular characteristics of ligand–protein interactions and interaction sites.

Pharmaceutical companies typically maintain databases of the compounds that have been synthesized over the years. These databases, containing often several hundreds of thousands of compounds, store two-dimensional information of chemical structures. With the advent of software that can generate three-dimensional structures from two-dimensional information, a new wealth of three-dimensional information becomes available for three-dimensional searches and for identifying pharmacophoric patterns of functional groups or atoms important for recognition of and interaction with a receptor leading to a given biological response. Three-dimensional searches, or data mining, are becoming an important tool in lead generation [23 24]. In particular, three-dimensional searching seems to be particularly useful in 'reviving' the older structures of a corporate database. In fact, it often happens that, in a given current research project, medicinal chemists tend to concentrate on the structures they are currently dealing with, and their analogues, while forgetting that older structures that have not been evaluated in the pharmacological tests of today could equally well be candidate compounds for the current research project.

In general, databases containing either experimentally derived structures, such as the Cambridge Structural Database and the Brookhaven Protein Data Bank, or calculated structures are indispensable tools for theoretical or computational medicinal chemists. Fast and easy access to three-dimensional structural information is of crucial importance for molecular modelling. It offers a wealth of information regarding intermolecular and intramolecular architecture and may provide suitable template structures that can then be used for further *in computro* modification.

The new field of combinatorial chemistry, which offers the possibility to synthesize many thousands of compounds simultaneously by combining structural elements from a set of building blocks (reactants), is emerging as an important new technology for drug discovery [25-29]. Given the possibility of high-throughput screening of this multitude of molecules, combinatorial chemistry has the potential to revolutionalize the process of lead finding. Recently, the combination of combinatorial chemistry with multidimensional NMR techniques and biochemical methods led to the identification of two classes of ligands for SH-3 domains (small receptor areas consisting of about 60 amino acids of some proteins which are involved in signal transduction processes) [30]. In a certain sense, although on a vastly larger scale, the combinatorial chemistry approach is reminiscent of traditional drug discovery based on screening strategies. Yet, the potential of the combinatorial chemical approach will be maximized by combining it with appropriate computational chemistry methods in making rational chemical libraries.

Role of computational chemistry in drug discovery

Some topics have been presented related to structure-based ligand design. Admittedly, the number of topics discussed is by no means exhaustive and is mainly inspired by personal bias and daily practice over the last 25 years. Nevertheless, it should be clear that, due to the position of the theoretical medicinal chemist in the long process from the original concept to the point of making a drug available to physicians and patients, computational chemists are dealing in the vast majority of cases with ligands. From this point of view, 'computer-assisted drug design' is an utter misnomer. Also the expression 'rational drug design', which falsely may imply that drugs can be designed liked pieces of furniture and that the historically successful screening approach was irrational, has to be avoided.

Computer-assisted molecular modelling uses the methods and techniques of computational chemistry to describe and possibly predict physicochemical properties of molecules or ligands and to simulate their behaviour along their itinerary from their injection into the biological system towards recognition of and interaction with a macromolecular receptor molecule.

It is beyond any doubt that computational chemistry has made great strides in achieving satisfactory agreement between computed and experimentally determined properties of molecules. Even casual browsing through the current chemical literature should convince anyone that molecular properties

such as three-dimensional structure, energies, molecular interactions, and spectroscopic properties are amenable to successful computation. Applying the methods and techniques of computational chemistry to ligands or biologically active molecules does not warrant the notion that one is designing drugs – it simply means that one is calculating molecular properties that possibly lie at the basis of biological activity. As such, computational chemistry may be used to find useful correlations between chemical properties and biological activity, thereby providing a rationale why some compounds are biologically active and some are not. At this stage, theoretical medicinal chemists can influence the direction a research project may take.

A more direct contribution from theoretical medicinal chemists can be expected if the three-dimensional structure of the binding site of the receptor is known. In this case, a more detailed description of the energetics and the three-dimensional characteristics of the ligand–receptor complex can be computed. These types of computational experiments may result in a better understanding of how and why a ligand binds to a receptor, in terms of its structural characteristics, and may lead to suggestions of the synthesis of new analogues.

Frequently, the three-dimensional structure of the target is not known and then one can resort to homology modelling [31] to deduce the three-dimensional structure of a protein if its amino acid sequence is known and if structural data for a homologous protein are available. Other strategies include receptor mapping and *de novo* ligand design [32-34], whereby possible interaction sites are identified and molecules that are complementary to these interaction sites are searched for in three-dimensional databases.

If theoretical chemists really want to have a significant impact on the daily life of experimental chemists or biochemists, the quantity of interest is the free energy, ΔG, of a system. All chemical behaviour is determined by differences in ΔG between reactants and products in a reaction, or between reactants and the transition state. Thus, the free energy of binding $\Delta G = \Delta H - T\Delta S$ has contributions from the enthalpy, ΔH, and the entropy, ΔS. As ΔG is a state function that depends on the extent of phase or configuration space accessible to the molecular system, the computation of ΔG of a molecular system is virtually impossible. However, the relative ΔG of a molecular association between two structurally closely related ligands can be approached on the basis of the so-called thermodynamic cycle and free energy perturbation methods [35]. The implementation of free energy perturbation algorithms in molecular modelling software packages, combined with the ever-increasing development and availability of faster computer hardware, ushers in a new era in which entropic and solvent effects may be properly taken into account.

Conclusion

Some aspects of computational chemistry pertaining to structure-based ligand design have been briefly discussed. An attempt has been made to delineate the role of computational techniques that can be used to characterize the chemical properties and behaviour of molecular systems. As such, the use of computational methods only produces physical quantities and numbers. The connotation of (biologically active) ligand design comes in when these physical quantities are related and used to rationalize biological data. It should be stressed that the results of structure-based ligand design are not always clear-cut recipes useful for further action by organic synthetic chemists or pharmacologists as is often seen in computer-assisted manufacturing [36]. In fact, as stated in the Introduction, modern pharmaceutical research is limited by the complexities of the biological processes and by the same token also the theoretical chemist who wants to relate these biological processes to the physical properties of the ligands and receptors presumably responsible or involved in drug action. Therefore, it should be realized that because of these staggering complexities, awesome simplifications and approximations must be used. On the one hand, owing to the sheer size of macromolecular structures in general, their energetics and three-dimensional aspects can only be approached by molecular mechanics, thereby excluding, for example, the simulation of bond breaking and bond formation processes. On the other hand, the complexities of the processes playing a role in living matter, which pre-eminently is a dissipative system, force one to construct and study simplified models of biochemical reality.

Nevertheless, if ligand–receptor recognition and interaction are considered to be necessary, but not sufficient steps for biological activity, the theoretical chemical approach to the problem, despite all its approximations, does make significant contributions to structure-based ligand design [37 38]. Experience shows that computer-assisted molecular modelling is an indispensable tool for displaying and manipulating molecular structures generated by experimental and/or theoretical techniques. In fact, the mere viewing and manipulation of three-dimensional models of small molecules on a computer graphics screen gives synthetic organic chemists a better understanding of their current molecules and quite frequently offer them some clues of what else they could synthesize. Likewise, it appears that biochemists benefit from 'seeing' and 'looking' into their target structure. In general, molecular modelling definitely stimulates the creativity of those involved in the study and analysis of biologically active ligands.

It should be pointed out, however, that whatever the degree of sophistication of the hardware and software used, structure-based ligand design is but a small step in the arduous and costly process from concept to a useful medicine.

References

1 Fujita T. The extrathermodynamic approach to drug design. In: Hansch C, Sammes PG, Taylor JB, Ramdsen CA, editors. Comprehensive medicinal chemistry. Vol. 4. Quantitative drug design. Oxford: Pergamon, 1990: 497-560.

2 Rekker R. The hydrophobic fragmental constant. Amsterdam: Elsevier Scientific Publishing, 1977.

3 Leo AJ. Calculating log Poct from structures. Chem Rev 1993;93:1281-1306.

4 Hansch C, Klein T, McClarin J, Langridge R, Cornell NW. A quantitative structure–activity relationship and molecular graphics analysis of hydrophobic effects in the interactions

of inhibitors with alcohol dehydrogenase. J Med Chem 1986;29:615-20.
5 Tollenaere JP, Moereels H, Raymaekers LA. Structural aspects of the structure–activity relationships of neuroleptics: principles and methods. In: Ariëns EJ, editor. Drug design. Vol. 10. Oxford: Academic Press, 1980: 71-118.
6 Allen FH, Davies JE. File structures and search strategies for the Cambridge Structural Database. Crystallogr Comput 1988;2:271-89.
7 Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535-42.
8 Wuthrich K. The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination. Acc Chem Res 1989;22:36-44.
9 Lee GM, Chen C, Marschner TM, Andersen NH. Does the solid-state structure of endothelin-1 provide insights concerning the solution-state conformational equilibrium? FEBS Lett 1994;355:140-6.
10 Leach AR. A survey of methods for searching the conformational space of small and medium-sized molecules. In: Lipkowitz KB, Boyd DB, editors. Reviews in computational chemistry. Vol. 2. Weinheim: VCH Publishers, 1991:1-55.
11 Howard AE, Kollman PA. An analysis of the current methodologies for conformational searching of complex molecules. J Med Chem 1988;31:1669-75.
12 Wilson SR, Cui W, Moskowitz JM, Schmidt KE. Applications of simulated annealing to the conformational analysis of flexible molecules. J Comput Chem 1991;12:342-9.
13 Nayeem A, Villa J, Scheraga HA. A comparative study of simulated-annealing and Monte Carlo with minimization approaches to the minimum-energy structures of polypeptides: [Met]-enkephalin. J Comput Chem 1991;12:594-605.
14 Zerner MC. Semiempirical molecular orbital methods. In: Lipkowitz KB, Boyd DB, editors. Reviews in computational chemistry. Vol. 2. Weinheim: VCH Publishers, 1991:313-65.
15 Siebel GL, Kollman PA. Molecular mechanics and the modelling of drug structures. In: Hansch C, Sammes PG, Taylor JB, Ramdsen CA, editors. Comprehensive medicinal chemistry. Vol. 4. Quantitative drug design. Oxford: Pergamon, 1990: 125-38.
16 Bowen JP, Allinger NL. Molecular mechanics: The art and science of parametrization. In: Lipkowitz KB, Boyd DB, editors. Reviews in computational chemistry. Vol. 2. Weinheim: VCH Publishers, 1991:81-97.
17 Brooks BR, Bruccoleri ER, Olafson ER, States DJ, Swaminathan S, Karplus M. CHARMm: A program for macromolecular energy minimization and dynamic calculations. J Comput Chem 1983;4:187-217.
18 Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, et al. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984; 106:765-84.
19 Clark M, Cramer RD III, Van Opdenbosch N. Validation of the general purpose tripos 5.2 force field. J Comput Chem 1989;10:982-1012.
20 Dinur U, Hagler AT. New approaches to empirical force fields. In: Lipkowitz KB, Boyd DB, editors. Reviews in computational chemistry. Vol. 2. Weinheim: VCH Publishers, 1991:99-164.
21 van Gunsteren WF, Berendsen HJC. Computer simulation of molecular dynamics: methodology, applications and perspectives in chemistry. Angew Chem 1990;29:992-1023.
22 Brunger AT, Karplus M. Molecular dynamics simulations with experimental restraints. Acc Chem Res 1991;24:54-61.
23 Sheridan RP, Rushinko A III, Nilakantan R, Venkataraghavan R. Searching for pharmacophores in large coordinate data bases and its use in drug design. Proc Natl Acad Sci USA 1989;86:8165-9.
24 Martin YC, Bures MG, Willet P. Searching databases of three-dimensional structures. In: Lipkowitz KB, Boyd DB, editors. Reviews in computational chemistry. Vol 2. Weinheim: VCH Publishers, 1991:213-63.
25 Zuckermann RN, Kerr JM, Kent SBH, Moos WH. Efficient method for the preparation of peptoids [oligo(*N*-substituted glycines)] by submonomer solid-phase synthesis. J Am Chem Soc 1992;114:10646-7.
26 Simon RJ, Kania RS, Zuckermann RN, Huebner VD, Jewell DA, Banville S, et al. Peptoids. a modular approach to drug discovery. Proc Natl Acad Sci USA 1992;89:9367-71.
27 Zuckermann RN, Martin EJ, Spellmeyer DC, Stauber GB, Shoemaker KR, Kerr JM, et al. Discovery of nanomolar ligands for 7-transmembrane G-protein coupled receptors from a diverse (*N*-substituted) glycine peptoid library. J Med Chem 1994;37:2678-85.
28 Felder ER. The challenge of preparing and testing combinatorial compound libraries in the fast lane, at the front end of drug development. Chimia 1994;48:531-41.
29 Liskamp RMJ. Opportunities for new libraries: unnatural biopolymer and diversomers. Angew Chem Int Ed (Engl) 1994, 33:633-6.
30 Chen JK, Schreiber SL. Kombinatorische Synthese und mehrdimensionale NMR-Spektroskopie: ein Beitrag zum Verstandnis von Protein-Ligand-Wechselwirkungen. Angew Chem 1995;107:1041-58.
31 Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO J 1986;5:823-6.
32 Bohm H-J. The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. J Comput-Aided Mol Design 1992;6:61-78.
33 Böhm H-J. The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. J Comput-Aided Mol Design 1994;8:243-56.
34 Leach AR, Kilvington SR. Automated molecular design: a new fragment-joining algorithm. J Comput-Aided Mol Design 1994,8:283-98.
35 Kollman P. Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 1993;93:2395-2417.
36 Loew GH, Villar HO, Alkorta I. Strategies for indirect computer-aided drug design. Pharm Res 1993;10:475-86.
37 Lam PYS, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, et al. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. Science 1994;263:380-4.
38 Colman PM. Structure-based drug design. Curr Opin Struct Biol 1994;4:868-74.