Pharmaceutical Research, Vol. 10, No. 4, 1993

Review

Strategies for Indirect Computer-Aided Drug Design

Gilda H. Loew,[1,2] Hugo O. Villar,[1] and Ibon Alkorta[1]

This review is intended to describe some of the methods and procedures used for computer-aided drug design when the structure of the macromolecular target is unknown, as is the case for CNS active drugs. Strategies and methods used in computer-aided design of drugs in such instances must be "indirect," i.e., focusing on the characterization of the ligands themselves. This situation is different from one in which the three-dimensional structure of the macromolecular target for a drug is known, for example, for drugs that are enzyme inhibitors, allowing "direct" characterization of ligand–receptor interactions. Two qualitatively different "indirect" approaches are described here. One, called 2D-QSAR, is briefly reviewed. It is based on delineating regression relationships between a specified biological end point and properties of the compounds eliciting it. The other, based on pharmacophore development, constitutes the main part of this review. Several levels of pharmacophore development are described, which differ in the extent to which they encompass fundamental molecular properties that are determinants of receptor recognition and activation. The strengths and limitations of each procedure are discussed and illustrated by examples. Two methods for obtaining model receptor structures are then briefly described. Both rely on the prior success of the indirect methods in obtaining ligand properties that modulate receptor recognition and activation. These emerging capabilities have the potential to bridge the gap between indirect and direct methods of drug design, since, if successful, the design process can continue in a direct mode using explicit characterization of drug–receptor interactions. Strategies for hypothesis validation and use of hypothesis for drug design and discovery are also briefly reviewed. The final sections of this review describe specific computational tools such as molecular mechanics and quantum mechanical methods used to characterize and identify relevant molecular properties and indicate some areas for future development of computational chemistry methods that could increase its effectiveness in the design of novel drugs.

KEY WORDS: drug design; pharmacophore development; QSAR; molecular mechanics; quantum mechanics.

INTRODUCTION

Computers have become powerful tools in all areas of scientific research. Pharmaceutical and medicinal chemistry are no exception to the rule. They have profited from the use of the methods of theoretical chemistry to understand the structure and mechanism of action of biological systems, as well as to design new compounds that can be used to further this understanding or be investigated as potentially useful therapeutic agents.

The appropriate strategy to use in the design of novel drugs depends on the available knowledge about the structure of the macromolecular target. A "direct" strategy can be used if the three-dimensional structure of the binding sites is known, allowing explicit characterization of ligand–receptor interactions, for example, for the design of drugs that are enzyme inhibitors, since there are many enzymes with known structures (1–3). Such knowledge can be achieved either from appropriate experimental techniques, such as X-ray crystallography or NMR, or from homology modeling that uses theoretical tools to deduce the three-dimensional (3D) structure of a protein given structural data for a highly homologous one (4). If the 3D structure of the macromolecule is not known, then the clues for the design of new ligands for it are more "indirect" and are based on the analysis of the molecular properties of compounds known to have some interaction with it, resulting in diverse pharmacological activities.

In this review, we focus on strategies appropriate for the design of ligands when the 3D structure of the biological target is not known as is the case for CNS active drugs. Two qualitatively different "indirect" approaches are described here. One, called 2D-QSAR, is briefly reviewed. It is based on delineating regression relationships between a specified biological end point and properties of the compounds eliciting it. This relationship is then used to predict the activity of untested compounds at the same end point.

The main part of this review focuses on approaches that use initial hypothesis development as the basis for drug design. Several such strategies are described that differ in the extent to which they encompass fundamental molecular properties that are determinants of receptor recognition and activation.

[1] Molecular Research Institute, 845 Page Mill Road, Palo Alto, California 94304.

[2] To whom correspondence should be addressed.

0724-8741/93/0400-0475$07.00/0

The overall strategy involved in indirect design of ligands is summarized in Fig. 1. In any such approach, initial pharmacological data for hypothesis development must be obtained for a set of ligands for the system for which novel compounds are to be designed. The initial data should be homogeneous, i.e., obtained using uniform protocols and, ideally, from a single source. Otherwise, the data could have differences that may be misleading. If the data set contains only binding data for antagonists, then the hypothesis can include only molecular requirements for recognition. If the data set contains binding data only for active compounds, then the hypothesis will encompass molecular determinants for both recognition and activation but will not be able to distinguish between them. Only if the data set contains agonists and antagonists, identified as such by an activation end point, and that have different affinities for the receptor, will it be possible for the hypothesis to encompass separate determinants of recognition and activation. In parallel with the experimental effort, the techniques of computational chemistry should be used to calculate molecular properties of the same compounds. From an analysis of the relationship between the molecular properties and the pharmacological profile for each compound, a hypothesis is developed, which, in turn, permits the selection or design of novel ligands for the macromolecular target.

The validity of the initial hypothesis of the mechanism by which the compounds elicit their effect can be verified by acquisition or synthesis of the compounds selected or designed and their subsequent pharmacological evaluation. If the compounds tested have the predicted pharmacological profile, then they could be novel probes of mechanism or clinically useful drugs. Alternatively, if the compound does not have the profile expected, then the results can still be used to refine working hypothesis.

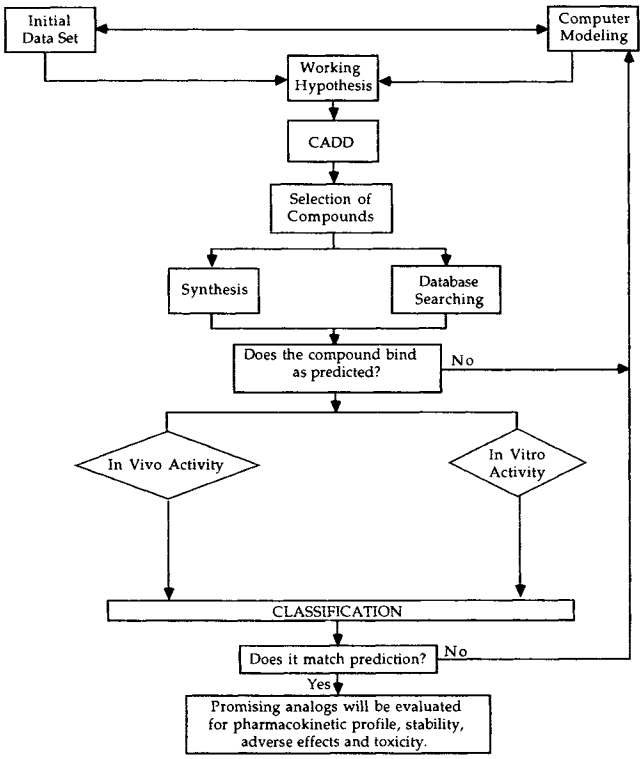

Fig. 1. Interdisciplinary approach to rational computer-aided drug design (CADD).

In the following section of this Review, different approaches to hypothesis development are described, and their strengths and limitations are discussed and illustrated by examples. In the next section, strategies for hypothesis validation and use of initial hypothesis for drug design and discovery are briefly reviewed.

In the fourth section, two strategies for obtaining model receptor structures are briefly described. Both rely on the prior success of the indirect methods in obtaining ligand properties that modulate receptor recognition and activation. These emerging capabilities have the potential to bridge the gap between indirect and direct methods of drug design, since, if successful, the design process can continue in a direct mode using explicit characterization of drug–receptor interactions.

The next part of this Review describes several specific computational tools, specifically, molecular mechanical, and quantum mechanical methods used to characterize and identify relevant molecular properties. Finally, we indicate some areas for improvement of computational chemistry methods that could increase their effectiveness when applied to design of novel compound.

This Review has three aims: (i) to serve as an introductory guide to how a working hypothesis can be developed if the 3D structure of the biological target is not known, (ii) to describe the tools required for such a purpose, and (iii) to indicate how to proceed to the design of drugs once the initial hypothesis has been formulated. The literature referenced is by no means exhaustive and it is provided with the main purpose of providing introductory material to the field.

HYPOTHESIS DEVELOPMENT

2-D QSAR

A typical 2D-QSAR procedure assumes that, under certain conditions, the relationship between a biological end point and the molecular properties that determine it can be described in terms of a linear-free energy equation, for any congeneric set of drugs (5–7). A typical equation has the form

$$\log A = ch\, f_{hydr}(X_h) + ce\, f_{elec}(X_e) + cst\, f_{st}(X_s) + \text{constant}$$

where A is related to either the receptor binding affinity or a specific biological activity, and each of the terms is a congener property that can affect either receptor recognition or activation. Typically, hydrophobic (f_{hydr}), electronic (f_{elec}), and steric (f_{st}) properties of the ligands are used. Each term is a function of the corresponding parameter, X, which may have a linear or quadratic representations.

Although a large number of parameters have been used in connection with this approach, the most widely used set is (i) the octanol water partition coefficient for a hydrophobic term, (ii) Taft E_s quantities for steric effects, (iii) Hammett constants to describe electronic effects, and (iv) the molar refractivity to account for dispersion forces. Since these are all empirical parameters, this 2D-QSAR or "Hansch" ap-

proach has been simple to use and there are abundant examples of its application (5–7). Regression analysis is used for a set of compounds with known values of log A and known values of independent variable until values of the coefficients of each term are obtained that provide the best fit to the data.

In many cases, the parameters required for a meaningful application of the Hansch approach are not available. In such cases, variations have been formulated that resort to the use of parameters derived directly from the molecular topology. The simplest of these, called a Free–Wilson approach, assumes that structural fragments make additive contributions to a given biological activity (8–10). If the contribution of each fragment can be assigned, then the biological activity for all compounds representing different combination of the fragments can be estimated by simple addition of each of their contributions. Thus, the Free–Wilson approach introduces the concept of chemical structural moieties, at least in its minimum expression. In fact, the Free–Wilson method can be used with only a few substituents in many positions or several different substituents in a few positions, which is not possible with the Hansch approach. Variations of both of these approaches have been proposed, including more explicit treatment of the topology of the structure (11–13).

The 2D-QSAR procedures can be used only to predict a numerical value of the same property for which the regression analysis is performed and for compounds closely related to the original set. They cannot, in general, be used to deduce determinants of recognition or activation or to distinguish agonists from antagonists.

Pharmacophore Development (14,15)

The two-dimensional QSAR methods do not include information about the 3D structure of the compounds considered. Since the binding cavities are spatially constrained regions with specific steric requirements, this missing information is vital to the ability of ligands to bind to a receptor. A more complete and mechanistically relevant approach should indicate both the nature of the key moieties and their spatial relationship in the ligand, i.e., the "pharmacophore" required for both receptor recognition and activation. These molecular properties comprise both steric and electronic compounds and need not be the same since among compounds that bind to a given receptor are those that do and do not activate it, i.e., agonists and antagonists.

The term "pharmacophore" has many different meanings, depending on (i) the type of biological end points used to develop them, (ii) the properties incorporated into them, and (iii) the strategies used to obtain these properties. For example, to develop a pharmacophore for receptor recognition, binding data for antagonists only or binding data for both agonists and antagonists must be used. In both cases, common properties that are identified and characterized for high-affinity compounds, that are absent for low- or no-affinity compounds, relate directly to recognition. Only when both agonists and antagonists are included in the data set for hypothesis development will it be possible to also obtain a pharmacophore that includes requirements for activation and is hence different in some respects for agonists and antagonists. If, however, the pharmacological data being used for hypothesis development involve binding data for agonists only or are based only on agonist activity end points, then the resulting pharmacophore will have combined properties needed for both recognition and activation and a clear distinction between them will not be possible.

Pharmacophores can also be divided according to the different sets of molecular properties used to define them. On this basis, the following "hierarchy of pharmacophores" can be delineated. Type A pharmacophores are based primarily on conformational similarities, including (1) those that consider maximum spatial overlaps only and (2) those that incorporate implicit assumptions of key regions required for recognition and activation and require these to overlap in comparing structural similarities. Type B Pharmacophores emphasize common electronic properties explicitly identified as determinants of recognition and activation and are defined by the spatial arrangement of these moiety. This type of pharmacophore can also be further divided into those that (1) use the atoms themselves to represent the properties found to be important and define the common spatial relationships among them or (2) use the properties themselves instead of explicit atom types and establish the crucial steric requirements among them. The most complete pharmacophore would incorporate the requirements of this interactive pharmacophore and better-defined dynamic requirements.

Strategies for Development of Type A Pharmacophores

To a certain extent, the biologically active conformation of a given set of ligands may be obtained from a systematic comparison of the molecular geometries of these compounds interacting with the same receptor, using implicit assumptions about the elements considered to be essential for recognition. If the set contains both agonists and antagonists and at least one essentially rigid analogue that can recognize the receptor with a high affinity, then the task is much simplified. In such a case, the characterization of the bioactive form results from the identification of the conformation of each ligand that maintains a similar spatial arrangement of the elements considered to be essential for recognition found in the right analogue. In other cases, when all analogues show some degree of flexibility, then the bioactive conformation must be identified by an interactive procedure that attempts to determine which conformations of the ligands permit a unique arrangement of the elements thought to be important for recognition. In addition, the bioactive conformer should encompass a spatial arrangement of the key elements that other structurally similar, but low-affinity analogues cannot achieve. In the event that a unique 3D arrangement is found, it is selected as the bioactive conformation(s), i.e., the recognition pharmacophore for each ligand. When more than one such arrangement is possible, new analogues can be added to the data set, with the hope that some of the candidate conformers can be eliminated, either because they are not accessible in a new high-affinity analogue or because they are possible low-energy conformers for a compound that is known not to bind to the receptor.

An example of this procedure is given by our search for the candidate pharmacophore for a series of δ-selective opioid peptides (16). In this study, a set of 12 related analogues

was selected, 9 of them known to recognize the receptor and 3 with a low affinity. An extensive conformational search procedure, described below, was carried out for each cyclic peptide in the series. After the procedure was completed, the results for each of the 12 compounds were compared. The number of conformers found was very large, and only conformations that were within 5 kcal/mol of the lowest-energy form for each peptide were considered. The use of an energy cutoff can be rationalized as eliminating structures that are not accessible at physiologically relevant temperatures.

In addition to the extensive comparison of conformers for these analogues, an important ingredient in this pharmacophore development was the implicit assumption of key regions required for peptide recognition of opioid receptors. These were selected based on SAR studies that identified the *N*-tyramine region and a second aromatic ring Phe^4 as crucial moieties. Therefore, the goal was to identify a conformation of high-affinity ligands with a similar spatial arrangement of these two regions and to verify that no such conformation existed for the three low-affinity ones. To this end, three criteria were selected: (i) the distance between the center of the Tyr^1 and the Phe^4 aromatic rings, (ii) the distance between each of the two rings and the terminal-amine N atom, and (iii) the overall RMS of the cyclicized portion of the peptide. The third criteria represent a general steric requirement for receptor recognition because it ensures that when Tyr^1 and the Phe^4 are properly positioned, the remainder of the structure will not occupy significantly different areas of a hypothetical receptor cavity. The result of the comparison was that, one and only one conformation fulfilled all of the above criteria.

Once a conformer was identified, graphical tools were used to verify it. All graphical packages for molecular modeling include the capability of overlapping atoms of different molecules, at least in the approximation of keeping the molecules rigid. The graphics allows quick visualization of the differences in the structures in both the critical and other less critical regions of two superimposed ligands. The most effective means of choosing the most relevant overlaps between pairs of molecules is a combination of direct insight obtained from such graphical presentation and use of mathematical criteria such as the root mean square of the deviation.

For the nine high-affinity analogues, the Phe^4 and the Tyr^1 occupied identical positions and the cyclized skeleton "spacer" between these was in approximately the same region. For the three low-affinity analogues, the same relative orientations of the Phe^4 and the Tyr^1 were not possible. We were able to understand the fundamental reasons why those orientations of the Phe^4 and the Tyr^1 were not allowed in the low-affinity analogues.

This study provides an example of the steps needed to characterize one type of very flexible pharmacophore, which heavily emphasizes steric similarity while implicitly implicating specific regions in receptor interactions. One caveat about this pharmacophore is that it could contain requirements for both receptor recognition and activation since affinities of only agonists were used in its development.

Another procedure that emphasizes structural overlap combined with implicitly assumed key atomic regions is called the active analogue approach (17). This approach also assumes that the presence of key regions are necessary but not sufficient for recognition or activation. An additional explicit steric element is added, namely, that the ligand should fit into the available volume of the binding site of the receptor and not overlap with an area that is occupied by the receptor itself. Strategies used to define this allowed volume of the ligand consider, in principle, all its accessible conformations and not just the minimum energy one. Thus conformational searches are combined with pairwise comparisons of active ligands. The matching of the specific key moieties and of the receptor-excluded volume are evaluated simultaneously for all conformations making this procedure very computationally intensive. This procedure can be simplified if a rigid analogue is used in the procedure. Important properties of this analogue are inferred rather than explored since the type of calculated properties that could be important is a computationally intensive exercise itself. Thus, while the method can be extremely powerful, in practice, the computational demands make it applicable only with restrictions. Not only are rigid analogues generally used, but important properties are inferred rather than explored.

Strategies for Development of a Type B Pharmacophore

The two types of Type A pharmacophores just described are based largely on conformational similarity but include assumptions as to which types of moieties or atoms should be regarded as essential for drug–receptor interactions. Then the spatial arrangement of these components is compared for the molecules under consideration and related to the biological activity until, eventually, a pharmacophore emerges. The assumptions made regarding the importance of the chosen components can change as the model evolves. The pharmacophore is expressed as a set of types of key atoms and the relative distance and overlaps among them.

The primary use of conformational similarities and the prior identification of atom centers important for drug receptor interaction have a number of limitations. One limitation of this class of pharmacophores is that there is ample room for subjective decisions because it is based on observer-defined choices of key atoms and moieties. In addition, it is possible that both recognition and activation are driven by factors other than maximum steric similarities such as key electronic properties. Finally, many of these properties are not centered on the atoms themselves but on the force fields they generate. Thus, an improved approach would be first to identify explicitly molecular properties important for recognition and activation and then to use them to develop a pharmacophore. In this approach, the molecular properties are the descriptors used to identify key moieties. Typically, the molecular electrostatic potential created by a molecule at a specific point in the surrounding space is one such property (18–20). A characteristic pattern of this property is then assumed to be required for recognition. However, other properties, in addition to the molecular electrostatic potentials, are related to ligand–receptor interactions and should be examined for their importance. The focus, then, changes from the type A pharmacophore in two ways: (i) physical and electronic properties are used as key determinants, and (ii) pharmacophores are identified by the spatial arrangement of these properties rather than of atoms that are independent of them.

The development of a Type B or interaction pharmacophore is based on explicit identification of molecular properties that are determinants of receptor recognition and interaction rather than a prior assumption of them. In this approach, a series of candidate properties is calculated and compared to experimental end points in order to explicitly select the important modulators of both recognition and activation. It can be illustrated by studies made in our laboratory for ligands of the benzodiazepine/$GABA_A$ receptor (21).

The first step in these studies was the generation of consistent experimental data on which to build the hypothesis. To this end, the *in vitro* binding affinity and *in vivo* anticonvulsant profile for a series of 15 compounds from five chemical classes were determined. In a parallel effort, theoretical studies were made for the same compound starting with a conformational analysis followed by a calculation of a series of physical and electronic properties that could be indicators of specific types of ligand–receptor interaction. The total and regional partition coefficients, a property directly related to the ability to participate in hydrophobic interactions with the receptor, were evaluated using an atom-based parametrization that permitted the characterization of hydrophobic centers. Among the electronic properties calculated were the nature and energy of the highest occupied and lowest unoccupied molecular orbital and atomic polarizabilities, related to the ability of the molecule as a whole, and of regions in particular, to be involved in an electron transfer interaction with the macromolecule as either donor or acceptor. In addition, the ability of different proton-accepting center of the ligands to interact with a complimentary proton-donating center of the receptor was computed, as the difference between the heat of formation of the neutral species and the protonated ligand. This property was computed for all possible centers, without any presumption regarding the most favorable center. To some extent, the heats of protonation replaced the molecular electrostatic potential in defining regions for interaction with charged centers.

In the next step, using the computed properties for each conformer of each ligand, distances between moieties with certain properties, rather than between particular atoms or group of atoms, were calculated and compared. This distance table provided the initial basis for the definition of the pharmacophore. The tables were searched for patterns common to all ligands, regardless of their activity profile. Two strong proton-accepting centers were found to be located 3.5 Å apart for all ligands, including agonists and antagonists even for inverse agonists compounds with significant structural differences. This common feature was therefore defined as a requirement for recognition and a common component in the pharmacophore for all types of compounds. In addition, the position of the most lipophilic center, as determined using a single geometric parameter, the angle from the center of the lipophilic area to the closest proton acceptor to the more distant proton acceptor, appeared to discriminate among agonists and antagonists/inverse agonists. Such a definition allowed the pharmacophore to be described in terms of a specific 3D relationship between probable types of interactions with the binding site cavity instead of in terms of specific atoms or structural elements. It is hence readily generalizable to structurally diverse compounds. It must also be noted that, because of the properties that are computed, the interaction pharmacophore is the only one that provides clues regarding the mechanism of interaction and/or activation of the receptor.

The Complete Pharmacophore

The major deficiency of the interaction pharmacophore is that general steric requirements, beyond those determined by the specific relationship of the recognition and activation regions, are not readily determined. The active analogue approach could, in principle, supply this missing steric component if it is used together with strategies that first identify the specific molecular properties required for recognition and activation, as was done for the BDZ ligands, rather than assuming such properties. Thus, combinations of these two approaches, the development of an interaction pharmacophore followed by active analogue comparisons, while computationally and labor intensive, could, in principle, lead to the most complete pharmacophore.

Undoubtedly, as computational resources grow, the pharmacophore based on the active analogue approach will be able to include some of the properties computed to develop an interaction pharmacophore.

3D-QSAR

A three-dimensional QSAR, called comparative molecular field analysis (COMFA), assumes that steric and electrostatic forces determine the nature of the ligand–receptor interactions (22,23). It uses the molecular electrostatic potential as the key property in the development of a pharmacophore without any evidence that it is a key modulator of either recognition or activation for the systems under consideration. Since this property depends on the shape of the compound, the method requires that a set of molecular overlaps be deduced by an independent method, such as the active analogue approach, thus satisfying the steric requirement. Around a set of overlapped molecules, a cubic grid is placed within which the electric field that each molecular would exert upon a probe atom placed at each lattice point is calculated. The value of the MEP at each of those points in the grid is then used in a linear regression equation. To extract a stable QSAR from this severely overdetermined system requires the use of a special mathematical tool, called the partial lease-squares (PLS) method. The PLS statistics permits the determination of a linear expression (3D-QSAR) which has the minimal set of lattice points that reproduces the measured activity of the set of compounds used. As in any linear regression procedure, once the equations have been deduced, they can be tested for their predictive utility. Cross-validation is normally used for this purpose. In simple terms, cross-validation is done by omitting compounds from the set used to develop the original equation and predicting the activities of these excluded compounds using it. This analysis is repeated with a randomly chosen subset of compounds excluded and used to test the resulting equation. The position in space of the most poorly predicted compounds is modified, the field is recomputed, and the PLS equations are rederived. The procedure is terminated when a satisfactory value of the cross-validation parameters is achieved. The

overall validity of the 3D-QSAR relies heavily on the cross-validation procedure.

There are several drawbacks to the COMFA method (22). First, the results are very dependent on the initial molecular overlaps chosen. Second, despite the use of PLS methods, the system of equations remains inherently undetermined. Almost certainly, other QSAR equations equally consistent with any set of compounds could be found. Third, the COMFA analysis will fail when a few molecules, all very dissimilar from the rest, are included in the set, because of the impossibility of predicting the behavior of the dissimilar molecule from the others. In this situation, the PLS procedure can derive equations for which there is a much higher risk of chance correlation. Finally, in common with all QSAR procedures, it can be used only to predict numerical values of the same biological activity that is used to develop the equations, and not to characterize general molecular determinants of recognition and activation.

Strategies for Obtaining 3D Models of Receptors

The indirect methods described in the previous sections to identify and characterize molecular determinants of receptor recognition and activation are required because of the lack of a three-dimensional structure for the target macromolecule, usually a protein. Once these determinants are characterized, they can, however, be used in two different ways to begin to develop a 3D model for the receptor to which they bind. These emerging capabilities have the potential to bridge the gap between indirect and direct methods of drug design since, if successful, the design process can continue in a direct mode using explicit characterization of drug–receptor interactions.

One such procedure is to search for a surrogate protein with a known structure that permits the explicit docking of ligands (24). Once the spatial relationship between elements required for recognition and those required for activation are determined, a search can be made for a protein with sites complementary to this "interaction pharmacophore" in the protein structure database. If such a protein model is found, then it can be used to characterize explicitly the receptor–ligand interactions.

In addition to the search for heuristic receptor models, a new type of receptor model building is emerging. These are the models for the seven-transmembrane helical segments of G-coupled receptors built based on the structure of bacteriorhodopsin (25,26). Bacteriorhodopsin is a membrane-bound protein with seven transmembrane helical regions, now thought to be a common feature of all G-coupled receptors. Hence, efforts are being undertaken to build models for this portion of the receptors via homology modeling. Unfortunately, the homology between bacteriorhodopsin and any G-coupled protein, including the other opsins, is extremely limited even in the seven-helical segments, and therefore, it is not a simple task. Moreover, in addition to the extensive problems of modeling a protein from another that does not have any significant homology, the problem is compounded by uncertainties in the organization of the helices themselves. Although difficult, construction of even approximate 3D structures for these transmembrane helices is an effort that is worth undertaking. Helpful additional information is emerging, such as candidate ligand binding sites and the effects of point mutation on ligand recognition. Again, detailed knowledge of the ligand interaction pharmacophore will be helpful in this approach. As the models develop, they can be refined based on the molecular biology data, and they can provide valuable insights into receptor–ligand interaction, as well as guidance for future experimental work in the design of mutants.

HYPOTHESIS VALIDATION AND DESIGN OF CANDIDATE DRUGS

In the preceding sections, the different approaches to determining the relationship between molecular properties and drug activity were described. Here, we briefly mention two different approaches to the next step, validation of the hypothesis developed and use of them for drug design.

In the classical 2D-QSAR mode, the validity of the regression equations obtained can be tested by using them to predict the specific biological end point under consideration for a series of compounds with known activity. Then the possibility that novel compounds might have this activity can be directly tested by using the same regression equation. The only additional effort required is that of evaluating each of the properties used as independent variables in that equation. These compounds can then be acquired or synthesized and tested for that end point. The appeal of this method is its simplicity and the ease with which the screening of new candidate compounds can be done. The disadvantages are that (i) it predicts a value only for a single given end point, i.e., a receptor affinity of biochemical or behavioral activity; (ii) it is best used with closely related congeners since explicit 3D criteria are not included; (iii) the reasons for failures are not clear—For example, if an end point is activity, is the failure due to lack of recognition or activation?—and failures cannot be used to refine the equation; and (iv) it cannot be used to distinguish qualitatively different behavior, i.e., agonism from antagonism.

The other approaches to hypothesis or pharmacophore development discussed all have a 3D component of some kind. If these pharmacophores can be described by a set of geometric parameters, i.e., distances, angles, and torsion angles among key moieties (static) or properties (interactive) in the pharmacophore, then two approaches can be used to design new analogues that, at the same time, can test these hypothetical requirements for receptor recognition and of activation, if separate pharmacophores have already been developed for agonists and antagonists.

The first is the traditional approach of suggesting new analogues for synthesis which are variations of the families used to develop the hypothesis. For example, if requirements for recognition of a given receptor subtype are deduced, variations of known analogues can be suggested that would confer or increase receptor affinity at that subtype. If molecular determinants of activation of a given receptor have been developed, then these can be used to propose compounds for synthesis in the same chemical families that will have qualitatively different, i.e., agonist, antagonist, or even inverse agonist, activity at a given end point. Synthesis and testing of proposed congeners both would provide hypothesis validation or refinement and also could lead to a

promising new drug. The main drawbacks of this route are that promising compounds can be very difficult to synthesize and they are usually structurally related to the known chemical families used for hypothesis development.

The recent development of both 2D and, especially 3D, relational data bases and strategies to search them with user-provided criteria has provided a potentially efficient alternative to immediate synthesis for hypothesis validation and drug discovery (27–29). A 2D chemical data base is defined by its capability of storing and retrieving information based on a 2D chemical structure, i.e., its connectivity for all entries. The 2D information ensures that only one representation is associated with each entry. The connectivity table then constitutes the identifier (1D type) for a given compound. The 1D type is then use to recover chemical and pharmacological information stores as data types. Three-dimensional structural information can be included as an additional data type in a 2D data base, thus generating a 3D data base.

The 3D databases can be searched (30,31) to retrieve all compounds with user-provided stereochemical criteria that define the spatial relationship between all key moieties or properties deduced from pharmacophore development as essential modulators of receptor recognition or activation. Depending on the level at which this pharmacophore was developed, these criteria would involve different combinations of steric, electronic, and hydrophobic properties and be useful for either receptor recognition or activation or both. The 3D criteria can also be relaxed to corresponding 2D-level input. In this case, types of atoms, as well as the group that separates them (the spacers), have to be specified. At the 2D level, a chemical data base could be used to retrieve all compounds that have these characteristics.

Searching of a 3D data base is a better tool for the purpose of identifying novel leads, since the criteria provided need not involve connectivity or the specification of any specific atoms, functional groups, or other moieties. Thus, this procedure could retrieve a list of structurally and chemically diverse compounds that satisfy the criteria. Among them would be compounds known to have the property for which the criteria were developed but were not included in the data set for hypothesis development. The retrieval of such compounds then constitutes hypothesis validation. In addition, this search can retrieve structurally diverse compounds not known to be ligands for the particular receptor system under study and, hence, provide novel leads for new drugs. The 3D data bases and 3D searching capabilities have been extensively described in the literature.

Many of the compounds that are retrieved based on an initial hypothesis, when tested, will not express the desired profile. The compounds that fail to express the pharmacology expected can, however, be used to refine the working hypothesis in the sense shown in Fig. 1, hopefully leading to more stringent search criteria that would result in fewer candidate structures and, ultimately, to promising new therapeutic agents.

PROPERTIES AND METHODS (32)

Conformational Analysis

In order to derive any pharmacophore or perform 3D-QSAR, it is necessary to have knowledge of the three-dimensional structures accessible to the drug molecule so that these can be compared in order to deduce which, among these structures, is the form that is recognized by the receptor, i.e., its bioactive form. The ideal set of compounds on which to base the development of a pharmacophore for a given receptor would be rigid agonists and antagonists, since such drugs interact with the receptor in a clearly defined and unique form, making the task of deducing common properties required for both recognition and activation easier. However, ligands of most receptor families vary enormously in their conformational flexibility, and the task of obtaining their conformational profile is progressively more challenging the more flexible the ligand. Another complication for flexible ligands is that they do not necessarily interact with a macromolecular target either in their lowest-energy form, in their crystal structure, or in their most abundant structure in solution.

The first step in deducing the bioactive form in which flexible ligands, such as peptides, recognize a given receptor is to perform a thorough search of conformational space of each analogue in the set chosen for hypothesis development. This analysis of conformational space is required to determine the low-energy structures accessible to each compound and can be performed only using the techniques of computational chemistry. The bioactive form should be closely related to one of these conformations. While X-ray crystal structures or NMR solution conformations are helpful in validating the results of conformation search procedures used, they cannot, in general, provide this crucial information. Because of the crucial role of computational chemistry in this first step of pharmacophore development, a description of conformational search strategies is described, with emphasis on recently developed algorithms that have proven to be effective for flexible ligands such as peptides.

Recently, a number of novel approaches have been described in the literature (33–47), focusing mainly on obtaining the global or lowest energy minimum for a flexible ligand that has multiple minima with varying relative energies (Fig. 2). While this information is necessary to characterize the conformational profile, it is not sufficient, since, as already stated, the bioactive form need not be the global minimum of a given ligand, but only one that is relatively energy accessible from it. Thus, other strategies, including those developed in our laboratory, have been designed to identify both the global minimum and many other local minima.

The most straightforward approach to the identification of multiple minima is to perform nested rotations involving systematic rotation of all flexible torsion angles. In principle, this approach guarantees that all the minima will be found, provided that a small enough step is used to increment each torsion angle. In practice, however, this procedure rapidly becomes impractical, as the number of torsion angles to be scanned increases, since the number or conformations to be generated varies with the power of the number of variables. Therefore, other methods, such as random sampling or Monte Carlo methods, have been used and details of these can be found in the recent literature.

We wish to describe here an alternate strategy we have recently developed using molecular dynamics procedures. This use of molecular dynamics (48–50) in general is based

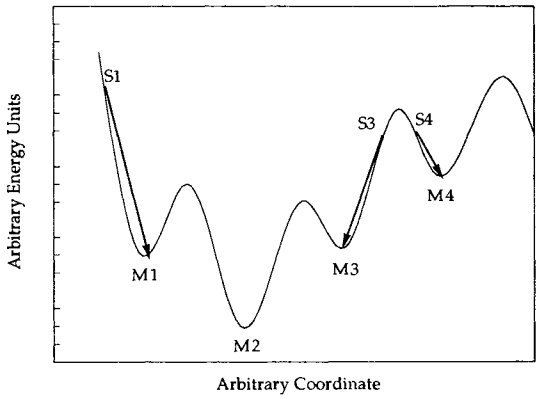

Fig. 2. Potential energy of an arbitrary molecular system as a function of an internal degree of freedom. Given a starting point on the surface such as S1, many algorithms permit the closest downhill minimum to be found. However, no method can find M2 starting from M1 or any other point in its domain such as S1. Similarly, starting from S3, only M3, and starting from S4, only M4, can be found. These procedures are called energy minimization algorithms.

on the idea that increasing the energy of the molecule by assigning each atom a random velocity can be done in such a way that the energy content of the molecule allows it to escape from the energy minima in which it is trapped, obeying Newton's laws. From Fig. 3, it should be easy to understand why it can be used to locate many minima in conformational space, when combined with energy minimization techniques.

The variation of the molecular dynamics procedures outlined in Fig. 2b that we have developed has been applied both to linear peptides, specifically the endogenous opioid peptide met-enkephalin (51), and to a series of cyclic peptides with a high specificity for one of the opiate receptor subtypes, the δ receptor (16). These two systems were chosen because of their usefulness, each in a different way, in validating the procedures used. The conformational behavior of met-enkephalin has been the subject of extensive investigations using diverse theoretical methods. Thus, we wished to test the power of our strategy by comparing it with the results from other procedures. The structures of the cyclic peptides have been studied by 2D-NMR and comparisons of low-energy structures from our search strategies with those deduced from NMR presented another opportunity to validate them.

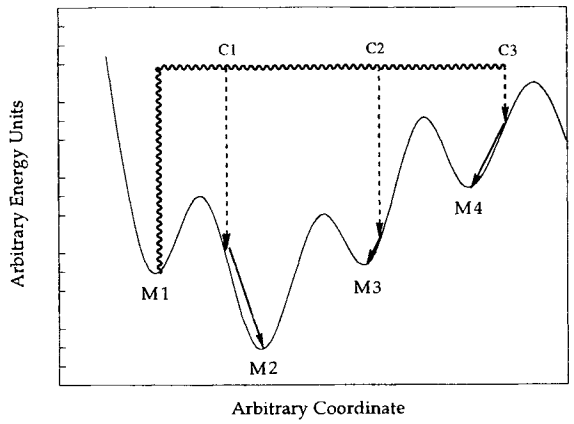

Fig. 3. In molecular dynamics, we can, starting from a low-energy point on the potential surface, provide energy to the molecule until it reaches a certain value dependent on the temperature of equilibration chosen and, then, let the system of atoms follow Newton's law in a classical manner. Along the trajectory, we can collect the coordinates (C1, C2, C3) of that point and use them as starting points in an energy minimization procedure.

The strategy investigated consisted of iterative cycles of high- and low-temperature molecular dynamics interspersed with energy minimizations. The rationale for combining high- and low-temperature molecular dynamics is illustrated in Fig. 4. The high-temperature simulations allowed the sampling of many local minima. The low-temperature simulations allowed the refinement of the structures.

To initiate this procedure, for the linear peptide, initial conformations were generated by systematic nested rotations of the backbone angles. The lowest-energy structures found by this procedure were then subjected to high-temperature molecular dynamics simulations at 900 K for 75 psec., during which the coordinates of 300 points were stored and energy minimized. The lowest-energy structure found by these means were then heated to 300 K. Along this trajectory, structures were also collected and energy minimized, completing the first cycle. The next cycle was begun by subjecting this new set to high-temperature molecular dynamics. The procedure was continued alternating high- and low-temperature molecular dynamics interspersed with structure collection and minimization, until no new low-energy structures were found.

To determine if the high- and low-temperature molecular dynamics procedure led to a reliable set of conformations for met-enkephalin, we compared the multiple minima found by us and those reported for this peptide using other methods for conformational searching. Our technique found all

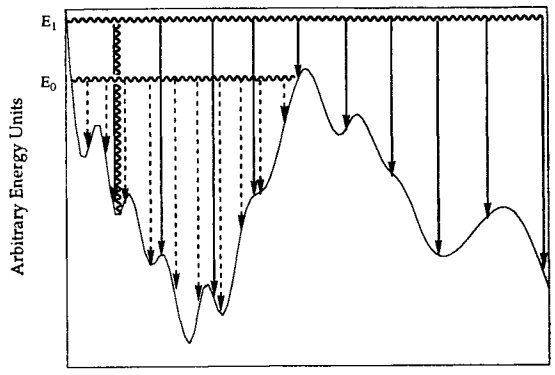

Fig. 4. Potential energy surfaces are considerably more complex than represented in Figs. 2 and 3. Indeed, they are N-dimensional, and therefore, a bidimensional cut will have additional saddle and other stationary points. At a high temperature, the energy of the molecule is also higher, which allows it to overcome numerous barriers, but since its motions are faster, it can significantly lose part of the detail of the potential surface. Ten points collected at equal intervals at a high temperature cover a larger portion of the conformational space than an equal number of points at a lower temperature during a similar interval. At a lower temperature, the system is confined to a smaller portion of the space determined by the energy barriers it cannot overcome, providing more refined details.

families of conformational domains that had been reported in the literature and some that had not been previously reported. This result reinforced our confidence in the search strategy developed as an effective means of scanning the conformational space of short linear peptides.

The study of the cyclic opioid peptide, DPDPE, using a search strategy similar to the one outlined above for metenkephalin, allowed the further validation of this procedure. For DPDPE, there is abundant experimental structural information derived from NMR data. Application of the above method, using environmental conditions similar to those used in the NMR studies, yielded a lowest-energy conformer that had all the structural characteristics determined from the NMR studies. Among the consistent characteristics found were the following: (i) all the interproton distances determined by NOE experiments were satisfied by this conformer within the experimental error; (ii) the populations of the rotamers for the side chains were in agreement with the highest-quality NMR data; (iii) the most buried NH proton in the predicted structure was also the one that, in the NMR studies, had a frequency with the smallest temperature dependence, an indication that it was the only one not exposed to the solvent; and (iv) a short distance between the disulfide bond and an aromatic ring was found both in the predicted structure and in the NMR conformer. In summary, the lowest energy conformer deduced "*de novo*" by our search strategy, without use of any experimental information, was able to reproduce all the characteristics that were experimentally observed for the conformation found in solution. While any one of these characteristics in the 2D-NMR spectra could perhaps be found in other conformers, the fact that it satisfies all experimental data increases our confidence in the methodology used.

Computational Methods

Thus far, the basic strategies for indirect drug design using the techniques of computational chemistry, together with experimental pharmacology, have been discussed. Drug design is possible in these circumstances; although the 3D structure of the receptors may not be known, the types of molecular properties of the ligands determine ligand receptor interactions. Drug–receptor interactions depend on hydrophobic effects, dispersion forces, induction forces, electrostatic forces, ion-induced dipole interactions, ion-permanent dipole interactions, hydrogen bonds, ionic bonds, electrostatic repulsion, and steric hindrance (14). All these forces, except the hydrophobic and steric effects, require knowledge of the electronic structure of the molecule.

The fundamental laws describing the hydrophobic interactions are not fully understood yet. Estimates of the octanol/water partition coefficient or related hydrophobic indices of the ligand, derived using different parametrizations, are thought to be determinants of this type of interaction with receptors. These properties can be measured or calculated.

Steric properties, including conformational searches, can be done using either molecular or quantum mechanical methods, or a combination of both, while all electronic properties, upon which a description of the ability of the ligand to interact with a receptor of unknown structure is based, can be computed only using quantum mechanical methods. A description of the main features of these two types of methods is given below.

Molecular Mechanics (52–54)

The molecular mechanics, also called "empirical energy" or "force field" methods, are based on the idea that the forces to which atoms in a molecule are subjected can be described in classical terms. This is a pragmatic approach since it is known that these forces are determined by the principles of quantum mechanics. However, this approach has proven reliable when carried out with extreme care.

A generic empirical energy potential has the form

$$E_{\text{total}} = E_{\text{stretching}} + E_{\text{bending}} + E_{\text{torsional}} + E_{\text{electrostatic}} (+ E_{\text{hydrogen bond}}) + E_{\text{steric}}$$

In the model system, all atoms are represented as spheres, with van der Waals radii. The interactions between the bonding atoms are described by bond stretching, bending, and torsional potentials. Most molecular mechanics methods assume that $E_{\text{stretching}}$ and E_{bending} are governed by a Hook's law term, i.e., that the forces that describe the chemical bond are harmonic. However, there are major efforts under way to develop methods containing nonharmonic terms. These second-generation force fields also contain cross terms, which would be higher-order terms. The MM3 (55) and the CVFF-89 (56) force fields belong to this class. The torsional energy, assuming a rigid rotor model, is composed of three periodic terms. The nonbonding, E_{steric}, term is represented by a Lennard–Jones model potential. The electrostatic term most commonly used is Coulomb's law. However, in some potentials, such as MM2 (57) and MM3 (54), bond–dipole–bond–dipole interactions are used to calculate the electrostatic components. Some empirical energy expressions also contain a hydrogen bond term.

Using empirical energy expressions, the total energy of a molecule can be computed analytically. There are several methods that permit the minimization of the total energy as a function of the internal coordinates, which permits the localization of the resting states of the molecule or minimum energy structures.

Several force fields are currently in wide use, such as those embodied in the AMBER (58,59), CHARMm (60,61), CVFF (56), and TRIPOS (62) and the MM3 suite of programs for peptides, nucleic acids, and small molecules. In addition, some of the pioneering work on the representation of peptides, using molecular mechanics methods, was done using the ECEPP potential (63), which lacks the stretching and bending terms in it.

The ability of the approach to describe molecular structure rests on the set of parameters contained in each term. For instance, for each bond in the molecule, in the harmonic approximation, there are two parameters required to describe the stretching term: the spring constant and the spring length (bond length) in a resting state. The number of parameters required to obtain a reliable description of a medium sized molecule is large, and thus, parameter development for a molecular mechanics method is a major undertaking.

Because of the significant effort that developing one of these potentials represents, most of them started by focusing on application to a particular type of molecular systems. However, all these force fields contain approximately the same terms in the potential and, therefore, could potentially be applied to any system with an adequate parametrization.

Quantum Mechanical Methods

In contrast to the molecular mechanics approach, which is phenomenological, quantum mechanical techniques are directly derived from the physical principles that govern the molecular structure, by solution of the stationary Schrödinger equation in an approximate manner. Briefly, the techniques can be divided into *ab initio* and semiempirical methods (65). While *ab initio* methods do not resort to parametrization to solve this equation, semiempirical methods contain parameters that avoid the computation of some time-consuming integrals required in *ab initio* procedures. Moreover, the semiempirical techniques consider only the valence electrons, i.e., those in the outer atomic shells. The parameters used are far fewer and less intuitive than those used in molecular mechanics methods. Both methods provide a wave function from which all electronic properties can be computed as expectation values.

The appeal of quantum mechanical methods is that they can, in principle, be used to calculate the entire range of properties that are necessary to understand the characteristics of the ligand that allow recognition and activation of receptors. However, these techniques are computationally intensive and require significantly more expertise for the interpretation of the data than do molecular mechanics methods.

Among the semiempirical methods, those developed by Dewar, Stewart, and co-workers are the most popular (66). The different variations of these methods MNDO (67), AM1 (68), and MNDO-PM3 (69), and MNDO-H (70) indicate different treatments of the multicenter integrals that are retained.

Ab initio methods do not contain parameters and, in most cases, the simplest level can provide the information required for drug design. However, the description of the atomic orbitals that define the spatial distribution of the electrons have different degrees of accuracy and reliability. The *ab initio* packages HONDO (71,72) and GAUSSIAN (73) are the most popular. It should be stressed that all the *ab initio* packages, in principle, have the capability of giving exactly the same results for the same system and the differences among them are secondary, such as speed, a major concern with *ab initio* calculations, ability to calculate certain properties, or ease of input preparation.

CHALLENGES FOR THE FUTURE

Computer-aided drug design is not governed by the clear-cut rules of engineering, and hence, these methods do not produce a finished product by a fully prescribed procedure in the same sense that CAD can produce other goods. The limitations of the rational computer-aided drug design approach arise because of the complexity of the biological processes involved in drug action at the molecular level and the level of approximation that must be used in describing their molecular properties.

Perhaps the most important challenge on the computational side relates to the inability of the current methods to provide a description of the molecular properties beyond the intimate complex between the ligand and the receptor. In particular, two additional capabilities would help improve the insights obtained. One advantage would be more frequent inclusion of solvent effects in developing criteria for drug–receptor interactions (74–78). An example of how this missing insight could impact on drug design can be cited. It is possible that congeners have been identified that by all criteria used are a perfect match for the receptor. However, their high affinity for the solvent could prevent them from having any significant interaction with the receptor. A number of different theoretical techniques for including solvent effects are in current development. Solvent effects can be treated using explicit water interaction via computation of the free energy of solvation, using statistical methods such as Monte Carlo or sampling in a molecular dynamics trajectory. Alternately, the molecules could be embedded in a continuum dielectric and some implementation of the Born equation could be applied.

Another aspect that still remains to be symmetrically included is the computation of free energies as opposed to potential energies of interaction (79–82). While molecular methods emphasize the importance of the internal energy, the free energy of a process is the fundamental quantity that determines its feasibility. For direct methods, in which the 3D structure of the macromolecular target is known, free energies can be calculated using free energy perturbation methods. In indirect drug design, even when there is no such structure, computation of free energies can be useful in several stages of analysis, for example, it can be used as a more refined criteria for choosing the most favorable conformers of a flexible ligand and for calculating their interactions with solvent.

Despite all the approximations made in its current applications, the techniques of computational chemistry have proven successful in helping to identify promising leads for novel families of compounds. The number of compounds that need to be synthesized or evaluated in preclinical pharmacological assessment can be greatly decreased using these procedures, compared to the totally empirical method of drug discovery. These techniques should be used in the multidisciplinary and iterative environment, illustrated in Fig. 1, to have the best chance of successful convergence to a clinically useful therapeutic agent.

REFERENCES

1. E. C. Meng, B. K. Shoichet, and I. D. Kuntz. Automated docking with grid based energy evaluation. *J. Comput. Chem.* 13:505–524 (1992).
2. E. F. Meyer. Molecular modelling: 1) playing the game when you do not know the rules, 2) how to learn the rules, and 3) some results. In H. van der Goot, G. Domany, L. Pallos, and H. Timerman (eds.), *Trends in Medicinal Chemistry '88*, Elsevier, Amsterdam, 1989, pp. 121–144.
3. H. J. Bohm. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comp.-Aided Mol. Design* 6:61–78 (1992).
4. W. R. Taylor and D. T. Jones. Templates, consensus patterns and motifs. *Curr. Opin. Struct. Biol.* 1:327–333 (1991).

5. C. Hansch and T. E. Klein. Molecular graphics and QSAR in the study of enzyme-ligand interactions. On the definition of bioreceptors. *Acc. Chem. Res.* 19:392–400 (1986).
6. T. Fujita. The extrathermodynamic approach to drug design. In C. Hansch, P. G. Sammes, J. B. Taylor, and C. A. Ramdsen (eds.), *Comprehensive Medicinal Chemistry. Vol. 4. Quantitative Drug Design,* Pergamon, 1990, pp. 497–560.
7. R. G. Gould (ed.). *Biological Correlations—The Hansch Approach,* Advances in Chemistry 114, American Chemical Society, 1972.
8. S. M. Free and J. N. Wilson. A mathematical contribution to structure-activity studies. *J. Med. Chem.* 7:395–399 (1964).
9. H. Kubinyi. The Free-Wilson method and its relationship to the extrathermodynamic approach. In C. Hansch, P. G. Sammes, J. B. Taylor, and C. A. Ramdsen (eds.), *Comprehensive Medicinal Chemistry. Vol. 4. Quantitative Drug Design,* Pergamon, 1990, pp. 589–643.
10. L. Goldberg (ed). *Structure-Activity Correlation as a Predictive Tool in Toxicology Fundamental, Methods and Applications,* Hemisphere, 1983.
11. L. H. Hall and L. B. Kier. The molecular connectivity chi indexes and kappa shape indexes in structure-property modeling. In K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry, Vol. 1,* VCH, 1991, pp. 367–422.
12. I. B. Bersuker and A. S. Dimoglio. The electron-topoogical approach to the QSAR problem. In K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry, Vol. 1,* VCH, 1991, pp. 423–479.
13. H. Weinstein, R. Osman, and J. P. Green. *The Molecular Basis of Structure Activity Relationships: Quantum Chemical Recognition Mechanisms in Drug Receptor Interactions,* In E. C. Olson and R. E. Christofferson (eds.), *Computer Assisted Drug Design,* ACS Symposium Series 112, American Chemical Society, 1979, pp. 161–187.
14. P. M. Dean. *Molecular Foundations of Drug-Receptor Interactions,* Cambridge University Press, Cambridge, 1987.
15. R. Franke. *Theoretical Drug Design Methods,* Akademie Verlag, 1984.
16. C. Chew, H. O. Villar, and G. H. Loew. Theoretical study of the flexibility and solution conformation of the cyclic opioid peptides [D-Pen2,D-Pen5]enkephalin and [D-Pen2,L-Pen5]enkephalin. *Mol. Pharmacol.* 39:502–510 (1991).
17. G. R. Marshall, C. D. Barry, H. E. Bosshard, R. A. Dammkoehler, and D. A. Dunn. The conformational parameter in drug dosing: The active analog approach. In E. C. Olson and R. E. Christoffersen (eds.), *Computer Assisted Drug Design,* ACS Symposium Series 112, American Chemical Society, 1979, pp. 205–226.
18. H. Weinstein, R. Osman, S. Topiol, and J. P. Green. Quantum chemical studies on molecular determinants for drug action. *Ann N.Y. Acad. Sci.* 367:434–448 (1981).
19. H. Weinstein, R. Osman, S. Topiol, J. P. Green, and S. Topiol. Electrostatic potentials as descriptors of molecular reactivity: The basis for some successful predictions of biological activity. In P. Politzer and D. G. Truhlar (eds.), *Chemical Applications of Atomic and Molecular Electrostatic Potentials,* Plenum, New York, 1980.
20. P. Carrupt, N. El Tayar, A. Karlen, and B. Testa. Molecular electrostatic potentials for characterizing drug-biosystem interactions. *Methods Enzymol.* 203:638–677 (1991).
21. H. O. Villar, E. T. Uyeno, L. Toll, W. Polgar, M. F. Davies, and G. H. Loew. Molecular determinants of benzodiazepine receptor affinities and anticonvulsant activities. *Mol. Pharmacol.* 38:589–600 (1989).
22. R. D. Cramer III, D. E. Patterson, and J. D. Bunce. Comparative molecular field analysis (COMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 110:5959–5967 (1988).
23. G. R. Marshall and R. D. Cramer III. Three dimensional structure activity relationships. *Trends Pharm. Sci.* 9:285–289 (1988).
24. H. Weinstein, M. N. Liebman, and C. A. Venanzi. Theoretical principles of drug action: The use of enzymes to model receptor recognition and activity. In A. Makriyannis (ed.), *New Methods in Drug Research, Vol. 1,* J. R. Prous, 1985, pp. 233–246.
25. J. Findlay and E. Eliopoulos. Three dimensional modelling of G-protein linked receptors. *Trends Pharm. Sci.* 11:492–499 (1990).
26. M. F. Hibert, S. Trump-Kallmeyer, A. Bruinvels, and J. Hoflack. Three-dimensional models of neurotransmitter G-binding protein-coupled receptors. *Mol. Pharmacol.* 40:8–15 (1991).
27. D. Weininger and J. L. Weininger, Chemical structures and computers. In C. Hansch, P. G. Sammes, J. B. Taylor, and C. A. Ramdsen (eds.), *Comprehensive Medicinal Chemistry. Vol. 4. Quantitative Drug Design,* Pergamon, 1990, pp. 59–82.
28. Y. C. Martin, E. B. Danaher, C. S. May, and D. Weininger. MENTHOR, a database system for the storage and retrieval of three-dimensional molecular structures and associated data searchable by substructural, biologic, physical or geometric properties. *J. Comp.-Aided Mol. Design* 2:15–29 (1988).
29. Y. C. Martin. Computer-assisted rational drug design. *Methods Enzymol.* 203:587–613 (1991).
30. J. H. Van Drie, D. Weininger, and Y. C. Martin. ALADDIN: An integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric and substructure searching of three-dimensional molecular structures. *J. Comp.-Aided Mol. Design* 3:225–251 (1989).
31. R. P. Sheridan, A. Rushinko III, R. Nilakantan, and R. Venkataraghavan. Searching for pharmacophores in large coordinate data bases and its use in drug design. *Proc. Natl. Acad. Sci. USA* 86:8165–8169 (1989).
32. N. C. Cohen, J. M. Blaney, C. Humblet, P. Gund, and D. C. Barry. Molecular modelling software and methods for medicinal chemistry. *J. Med. Chem.* 33:883–894 (1990).
33. A. R. Leach. A survey of methods for searching the conformational space of small and medium-sized molecules. In K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry, Vol. 2,* VCH, 1991, pp. 1–55.
34. H. A. Scheraga. Calculations of stable conformations of polypeptides, proteins and protein complexes. *Chem. Scripta* 29A:3–13 (1989).
35. H. A. Scheraga. Theoretical and experimental studies of conformations of polypeptides. *Chem. Rev.* 71:195–217 (1971).
36. M. Saunders. Searching for conformers of nine- to twelve-ring hydrocarbonds on the MM2 and MM3 energy surfaces: Stochastic search for interconversion pathways. *J. Comput. Chem.* 12:645–663 (1991).
37. S. R. Wilson, W. Cui, J. W. Moskowitz, and K. E. Schmidt. Application of simulated annealing to the conformational analysis of flexible molecules. *J. Comput. Chem.* 12:342–349 (1991).
38. P. Auffinger and G. Wipff. High temperature annealed molecular dynamics simulations as a tool for conformational sampling. Application to the bicyclic "222" cryptand. *J. Comput. Chem.* 11:19–31 (1990).
39. A. E. Howard and P. A. Kollman. An analysis of the current methodologies for conformational searching of complex molecules. *J. Med. Chem.* 31:1669–1675 (1988).
40. M. Saunders, K. N. Houk, Y. D. Wu, W. C. Still, L. Lipton, G. Chang, and W. Guida. Conformations of cycloheptadecane. A comparison of methods for conformational searching. *J.Am. Chem. Soc.* 112:1419–1427 (1990).
41. H. J. Bohm, G. Klebe, T. Lorenz, T. Meitzner, and L. Siggel. Different approaches to conformational analysis: A comparison of completeness, efficiency and reliability based on the study of a nine-membered lactam. *J. Comput. Chem.* 11:1021–1028 (1990).
42. T. D. J. Perkins and D. J. Barlow. RAMBLE: A conformational search program. *J. Mol. Graph.* 8:156–162 (1990).
43. B. von Freyberg and W. Braun. Efficient search for all low energy conformations of polypeptides by Monte Carlo methods. *J. Comput. Chem.* 12:1065–1075 (1991).
44. J. K. Shin and M. S. Jhon. High directional Monte Carlo procedure coupled with the temperature heating and annealing as a method to obtain the global energy minimum structure of polypeptides and proteins. *Biopolymers* 31:177–185 (1991).
45. S. D. Morley, D. Jackson, M. R. Saunders, and J. G. Vinter. DMC: A multifunctional hybrid dynamics/Monte Carlo simulation algorithm for the evaluation of conformational space. *J. Comput. Chem.* 13:693–703 (1992).

46. I. Kolossvary and W. C. Guida. Compare conformer: A program for the rapid comparison of molecular conformers based on interatomic distances and torsional angles. *J. Chem. Inf. Comput. Sci.* 32:191–199 (1992).
47. Z. Li and H. A. Scheraga. Monte-Carlo minimization approach to the multiple minima problem in protein folding. *Proc. Natl. Acad. Sci.* 84:6611–6615 (1987).
48. W. F. van Gusteren and H. J. C. Beredsen. Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew. Chem.* 29:992–1023 (1990).
49. A. T. Brunger and M. Karplus. Molecular dynamics simulations with experimental restraints. *Acc. Chem. Res.* 24:54–61 (1991).
50. J. Andrew McCammon and S. C. Harvey. *Dynamics of Proteins and Nucleic Acids,* Cambridge University Press, Cambridge, 1987.
51. J. J. Perez, H. O. Villar, and G. H. Loew. Characterization of low-energy conformational domians for Met-enkephalin. *J. Comp.-Aided Mol. Design* 6:175–190 (1992).
52. G. L. Siebel and P. A. Kollman. Molecular mechanics and the modelling of drug structures. In C. Hansch, P. G. Sammes, J. B. Taylor, and C. A. Ramdsen (eds.), *Comprehensive Medicinal Chemistry. Vol. 4. Quantitative Drug Design,* Pergamon, 1990, pp. 125–138.
53. N. L. Allinger. Calculation of molecular structure and energy by force field methods. *Adv. Phys. Org. Chem.* 13:1 (1976).
54. J. P. Bowen and N. L. Allinger. Molecular mechanics: The art and science of parametrization. In K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry, Vol. 2,* VCH, 1991, pp. 81–97.
55. N. L. Allinger, Y. H. Yuh, and J. H. Lii. Molecular mechanics. The MM3 force field for hydrocarbonds. 1. *J. Am. Chem. Soc.* 111:8551 (1989).
56. U. Dinur and A. T. Hagler. New approaches to empirical force fields. In K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry, Vol. 2,* VCH, 1991, pp. 99–164.
57. U. Burkert and N. L. Allinger. Molecular Mechanics, ACS Monograph 177, American Chemical Society, 1982.
58. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Proferta Jr., and P. Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106:765–784 (1984).
59. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case. An all atom force field for simulation of proteins and nucleic acids. *J. Comp. Chem.* 2:230–252 (1986).
60. B. R. Brooks, E. R. Bruccoleri, E. R. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMm: A program for macromolecular energy minimization and dynamic calculations. *J. Comput. Chem.* 4:187–217 (1983).
61. F. Momany and R. Rone. Validation of the general purpose QUANTA3.2/CHARMm force field. *J. Comput. Chem.* 13:888–900 (1992).
62. M. Clark, R. D. Cramer III, and N. Van Opdenbosch. Validation of the general purpose tripos 5.2 force field. *J. Comput. Chem.* 10:982–1012 (1989).
63. F. A. Momany, R. F. McGuire, A. W. Burgess, and A. Scheraga. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, non-bonded interactions, hydrogen-bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.* 79:2361–2381 (1975).
64. J. J. P. Stewart. Semi-empirical molecular orbital methods. In K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry, Vol. 1,* VCH, 1990, pp. 45–81.
65. M. C. Zerner. Semi-empirical molecular orbital methods, 10. In K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry, Vol. 2,* VCH, 1991, pp. 313–365.
66. J. J. P. Stewart. MOPAC: A semi-empirical molecular orbital program. *J. Comp.-Aided Mol. Design* 4:1–105 (1990).
67. M. J. S. Dewar and W. Thiel. ground states of molecules. 38. The MNDO method: Approximations and parameters. *J. Am. Chem. Soc.* 99:4899–4906 (1977).
68. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* 107: 3902–3909 (1985).
69. J. J. P. Stewart. Optimization of parameters for semi-empirical methods. I. Method. *J. Comput. Chem.* 10:209–220 (1989).
70. A. Goldblum. Improvement of the hydrogen bonding correction to MNDO for calculations of biochemical interests. *J. Comp. Chem.* 8:835–839 (1987).
71. M. Dupuis, J. D. Watts, H. O. Villar, and G. J. B. Hurst. The general atomic and molecular electronic structure system HONDO: Version 7.0. *Comput. Phys. Comm.* 52:415–425 (1989).
72. M. Dupuis and S. Maluendes. HONDO: A general atomic and molecular electronic structure system. In E. Clementi (ed.), *Modern Techniques in Computational Chemistry: MOTECC-91,* ESCOM, 1991.
73. M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople. *Gaussian 92: Revision A.* Gaussian, Pittsburgh, PA, 1992.
74. P. R. Bergethon and E. R. Simons. *Biophysical Chemistry Molecules to Membranes,* Springer, 1990.
75. M. E. Davies and J. A. McCammon. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* 90:509–521 (1990).
76. S. E. Harvey. Treatment of electrostatic effects in macromolecular modelling. *Proteins* 5:78–92 (1989).
77. A. Warshel, F. Sussman, and G. King. Free energies of charges in solvated proteins: Microscopic calculations using a reversible charging process. *Biochemistry* 25:8368–8372 (1986).
78. M. K. Gilson and B. Honig. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins* 4:7–18 (1988).
79. P. A. Bash, U. C. Singh, R. Langridge, and P. A. Kolman. Free energy calculations by computer simulation. *Science* 236:564–568 (1987).
80. D. L. Beveridge and F. M. DiCapua. Free energy via molecular simulation: Application to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* 18:431–492 (1989).
81. A. G. Anderson and J. Herman. Micro-folding: conformational probability map for the alanine dipeptide in water from molecular dynamics simulations. Proteins 3:262–265 (1988).
82. D. J. Tobias and C. L. Brooks III. Molecular dynamics with internal coordinate constraints. *J. Chem. Phys.* 89:5115–5127 (1988).