CHAPTER THREE

Molecular Modeling in Drug Design

GARLAND R. MARSHALL
Washington University
Center for Computational Biology
St. Louis, Missouri

DENISE D. BEUSEN
Tripos, Inc.
St. Louis, Missouri

Contents

1 Introduction, 78
2 Background and Methods, 79
2.1 Molecular Mechanics, 79
2.1.1 Force Fields, 79
2.1.2 Electrostatics, 81
2.1.2.1 The Dielectric Problem and Solvation, 83
2.1.2.2 The "Hydrophobic" Effect, 85
2.1.2.3 Polarizability, 85
2.1.3 The Potential Surface, 85
2.1.3.1 Optimization, 86
2.1.3.2 Potential Smoothing, 86
2.1.3.3 Genetic Algorithm, 87
2.1.4 Systematic Search and Conformational Analysis, 89
2.1.4.1 Rigid Geometry Approximation, 89
2.1.4.2 Combinatorial Nature of the Problem, 89
2.1.4.3 Pruning the Combinatorial Tree, 90
2.1.4.4 Rigid Body Rotations, 90
2.1.4.5 The Concept and Exploitation of Rings, 91
2.1.4.6 Conformational Clustering and Families, 92
2.1.4.7 Conformational Analysis, 93
2.1.4.8 Other Implementations of Systematic Search, 94
2.1.5 Statistical Mechanics Foundation, 94
2.1.6 Molecular Dynamics, 95
2.1.6.1 Integration, 95
2.1.6.2 Temperature, 96

Burger's Medicinal Chemistry and Drug Discovery
Sixth Edition, Volume 1: Drug Discovery
Edited by Donald J. Abraham
ISBN 0-471-27090-3 © 2003 John Wiley & Sons, Inc.

2.1.6.3 Pressure and Volume, 96
2.1.7 Monte Carlo Simulations, 96
2.1.8 Thermodynamic Cycle Integration, 99
2.1.9 Non-Boltzmann Sampling, 100
2.2 Quantum Mechanics: Applications in Molecular Mechanics, 100
2.2.1 Parameterization of Charge, 101
2.2.1.1 Atom-Centered Point Charges, 101
2.2.1.2 Methods to Reproduce the Molecular Electrostatic Potential (MEP), 102
2.2.2 Parameter Derivation for Force Fields, 102
2.2.3 Modeling Chemical Reactions and Design of Transition-State Inhibitors, 103
3 Known Receptors, 103
3.1 Definition of Site, 103
3.2 Characterization of Site, 105
3.2.1 Volume and Shape, 105
3.2.2 Hydrogen-Bonding and Other Group Binding Sites, 107
3.2.3 Electrostatic and Hydrophobic Fields, 108
3.3 Design of Ligands, 110
3.3.1 Visually Assisted Design, 110
3.3.2 Three-Dimensional Databases, 110
3.3.3 De Novo Design, 113
3.3.4 Docking, 113
3.3.4.1 Docking Methods, 113
3.3.4.2 Scoring Functions, 115
3.3.4.3 Search for the Correct Binding Mode, 116
3.4 Calculation of Affinity, 118
3.4.1 Components of Binding Affinity, 118
3.4.2 Binding Energetics and Comparisons, 120
3.4.3 Atom-Pair Interaction Potentials, 120
3.4.4 Simulations and the Thermodynamic Cycle, 120
3.4.5 Multiple Binding Modes, 121
3.5 Protein Structure Prediction, 122
3.5.1 Homology Modeling, 123
3.5.2 Inverse Folding and Threading, 123
3.5.3 Contact Matrix, 125
4 Unknown Receptors, 127
4.1 Pharmacophore versus Binding-Site Models, 127
4.1.1 Pharmacophore Models, 127
4.1.2 Binding-Site Models, 130
4.1.3 Molecular Extensions, 130
4.1.4 Activity versus Affinity, 131
4.2 Searching for Similarity, 135
4.2.1 Simple Comparisons, 135
4.2.2 Visualization of Molecular Properties, 137
4.3 Molecular Comparisons, 138
4.3.1 Volume Mapping, 139
4.3.2 Field Effects, 140
4.3.3 Directionality, 140
4.3.4 Locus Maps, 140
4.3.5 Vector Maps and Conformational Mimicry, 140
4.4 Finding the Common Pattern, 142
4.4.1 Constrained Minimization, 143
4.4.2 Systematic Search and the Active Analog Approach, 144
4.4.3 Strategic Reductions of Computational Complexity, 145
4.4.4 Alternative Approaches, 146
4.4.5 Receptor Mapping, 148
4.4.6 Model Receptor Sites, 149
4.4.7 Assessment of Model Predictability, 151
5 Conclusions, 153
6 Acknowledgments, 155

1 INTRODUCTION

By historical imperative, the role of molecular modeling in drug design has been divided into two separate paradigms, one centered on the structure-activity problem that attempts to rationalize biological activity in the absence of detailed, three-dimensional structural information about the receptor, and the other focused on understanding the interactions seen in receptor-ligand complexes and using the known three-dimensional structure of the therapeutic target to design novel drugs. The rapid increase in relevant structural information, attributed to advances in molecular biology to generate the target proteins in adequate quantities for study, and the equally impressive gains in NMR (1–9) and crystallography (10, 11) to provide three-dimensional structures as well as identify leads, have stimulated the need for design tools and the molecular modeling community is rapidly evolving useful approaches. The more common problem, however, is one in which the receptor can only be inferred from pharmacological studies and little, if any, structural information is

available to guide in modeling. Nevertheless, useful information to guide the design and synthesis of potential novel therapeutics can be developed from an analysis of structure-activity data in the three-dimensional framework provided by current molecular modeling techniques. Although most of the techniques and approaches described have broader application than shown, the examples chosen should be sufficient to illustrate their use. A number of reviews (12–18) of computer-aided drug design have relevant sections covering portions of this chapter with different perspectives and are recommended for a more complete overview.

2 BACKGROUND AND METHODS

2.1 Molecular Mechanics

Molecular mechanics (19) treats a molecule as a collection of atoms whose interactions can be described by Newtonian mechanics. Because the mass of the nuclei is much greater than the mass of the electrons, one can separate (the Born-Oppenheimer approximation) the Schrodinger equation into a product of two functions: one for electrons, one for nuclei. For the purposes of molecular mechanics, the electronic function, initially developed to interpret spectroscopic data, is ignored; that is, the charge distribution is assumed to remain constant during changes in the position of the nuclei. Because molecular mechanics is based on classical physics, it cannot provide information about the electronic properties of molecules under study that are generally assumed fixed during the parameterization of the force field with experimental data.

A few words about the basics of molecular mechanics (19, 20) may provide the elements of understanding for what follows. This is not meant to be comprehensive, but rather a simple overview, to remind the reader of a few crucial points. For a comprehensive overview of molecular modeling, the reader is referred to the excellent text by Leach (21). The interactions between atoms are divided into bonded and nonbonded classes. Nonbonded forces between atoms are based on an attractive interaction that has a firm theoretical basis and varies as the inverse of the 6th power of the distance between the atoms. It is balanced by a repulsion between the electronic clouds as the atoms come close and this interaction has been represented empirically by a variety of functional forms: exponential, 12th power, or 9th power of the distance between the atoms. The coefficients for these two interactions are parameterized for atom types, usually by element, so that the minimum of the combined functions corresponds to the sum of the experimental van der Waals radii for the two atoms.

In addition, bonded atoms are considered as a special case, with a "spring constant" determining the energy of deformation from experimental bond lengths. Atoms directly bonded to the same atom (one-three interactions) are eliminated from the van der Waals list and have a special energetic term relating the deviation from an ideal bond angle. Atoms having a one-four interaction define a torsional relation that is usually parameterized based on the types of the four connected atoms defining the torsion angle. The numerous combinations of atom types require an enormous number of parameters to be determined from either theoretical (quantum mechanics) and/or experimental data. Simplified force fields in which the torsional parameters depend only on the atoms at the end of a bond have been developed, to give approximate geometries for further refinement by quantum mechanics.

2.1.1 Force Fields. The basic assumption underlying molecular mechanics is that classical physical concepts can be used to represent the forces between atoms. In other words, one can approximate the potential energy surface by the summation of a set of equations representing pairwise and multibody interactions. These equations represent forces between atoms related to bonded and nonbonded interactions. Pairwise interactions are often represented by a harmonic potential [$\frac{1}{2}K_b(b - b_0)^2$] that obeys Hooke's law (derived for a spring) for bonded atoms, restoring the bond distance to an equilibrium value b_0, and a van der Waals potential [$C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^{6}$] for nonbonded atoms. Similarly, distortion from an equilibrium valence angle (θ_0) describing the angle between three bonded atoms sharing a common atom is also penalized [$\frac{1}{2}K_\theta(\theta -$

$\theta_0)^2$]. A third class of interaction dependent on the dihedral angle ϕ between four bonded atoms is the torsional potential $\{K_\phi[1 + \cos(\phi - \delta)]\}$ used to account for orbital delocalization and to compensate for other deficiencies in the force field. A harmonic term $[\frac{1}{2}K_\xi(\xi - \xi_0)^2]$ is often introduced for dihedral angles 5 that are relatively fixed, such as those in aromatic rings. Coulomb's law $[q_i q_j/(4\pi\varepsilon_0\varepsilon_r r_{ij})]$ is the simplest approach to the contribution of electrostatics to the potential V:

$$V = \sum \frac{1}{2} K_b(b - b_0)^2 + \sum \frac{1}{2} K_\theta(\theta - \theta_0)^2 + \sum \frac{1}{2} K_\xi(\xi - \xi_0)^2 + \sum K_\phi[1 + \cos(\phi - \delta)] + \sum [C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^{6}] + \sum q_i q_j/(4\pi\varepsilon_0\varepsilon_r r_{ij}).$$

A central issue is the number of different atom types that are used in a particular force field. There is always a compromise between increasing the number to allow for the inclusion of more environmental effects (i.e., local electronic interactions) vs. the increase in the number of parameters to be determined to adequately represent a new atom type. In general, the more subtypes of atoms (how many different kinds of nitrogen, for example), the less likely that the parameters for a particular application will be available in the force field. The extreme, of course, would be a special atom type for each kind of atomic environment in which the parameters were chosen, so that the calculated properties of each molecule would simply reproduce the experimental observations. One major assumption, therefore, is that the force constants (parameters) and equilibrium values of the equations are functions of a limited number of atom types and can be transferred from one molecular environment to another. This assumption holds reasonably well where one may be primarily interested in geometric issues, but is not so valid in molecular spectroscopy. This had led to the introduction of additional equations, the so-called "cross-terms" which allow additional parameters to account for correlations between bond lengths and bond angles $[K_{b\theta}(b - b_0)(\theta - \theta_0)]$, dihedral angles and bond angles, and so forth. Because of the lack of adequate parameterization of the more complex force fields that are usually specialized to one kind of molecule (e.g., proteins or nucleic acids), more simplified force fields have gained some popularity because of their general applicability, despite limited accuracy.

Examples are the Tripos force field (22), the COSMIC force field (23), and that of White and Bovill (24), which uses only two atom types, those at the end of the bond to parameterize the torsional potential rather than the four types of the atoms used to define the torsional angle. One has only to consider the number of combinations of 20 atom subtypes taken four at time (160,000) versus two at a time (400) to understand the explosion of parameters that occurs with increased atom subtypes. The simplifying assumption in parameterization of the torsional potential reduces to some extent the quality of the results (25), but allows the use of the simplified force fields (22) in many situations where other force fields would lack appropriate parameters. The situation can become complicated, however. For example, the amide bond is normally represented by one set of parameters, whether the configuration is *cis* or *trans.* Experiment$ data are quite compelling that the electronic state is different between the two configurations, and different parameter sets should be used for accurate results (Fig. 3.1). Only AMBER/OPLS currently distinguishes between these two conformational states (26). Certainly, the limited parameterization of simplified force fields would not allow accurate prediction of spectra that is more reflective of the dynamic behavior of the molecule.

Accurate estimates of energy may require accurate representation of the dynamics of molecules and justify derivation of the larger number of parameters. The new version (27) of the Allinger force field, MM3, has the objective of reproducing spectral data more accurately than MM2. Much of the chemistry remains to be incorporated into appropriate force fields. Only recently have adequate modifications been made to the force fields developed for organic molecules to include some metals (28–31). Carlsson (32, 33) recently de-

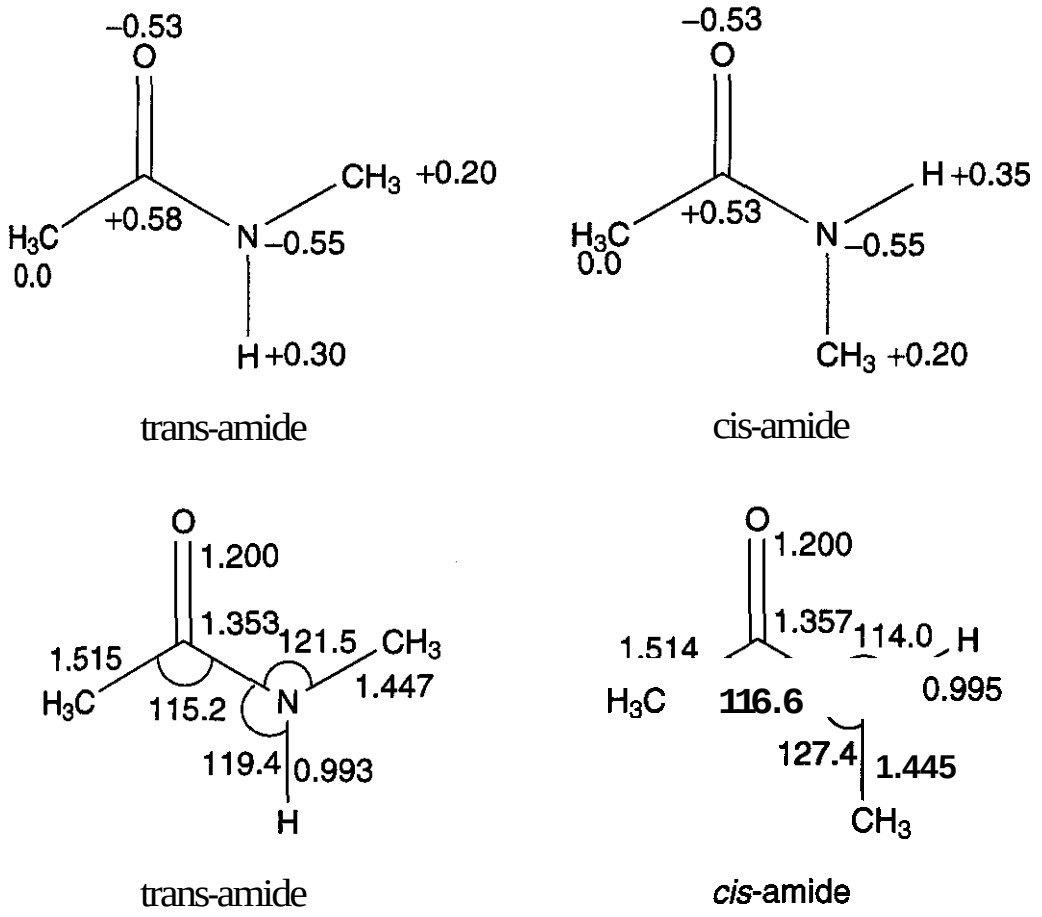


Figure 3.1. Differences in OPLS charge distribution (top) between cis- and trans-isomers of amide bond and geometries (bottom) as calculated by *ab* initio parameterization (26).

veloped a functional form that allows electronic d-orbitals of metals to be reasonably represented within molecular mechanics.

Because different force fields may use different mathematical representations of the forces between atoms and the details of their parameterization will in general differ also, it is unwise to use parameters derived for one force field to replace missing parameters in another. One often hears of a "balanced" parameter set that reproduces well the phenomena under consideration, but which is inadequate for other applications. A comparison by Burkert and Allinger (19) shows the different van der Waals (VDW) potentials used in several of the popular force fields, and the situation has not improved significantly in the intervening years. Because of other differences in parameters and functional forms of the equations used in the rest of the individual force fields, these quite different approaches to the VDW potential give excellent results when used in the correct combination. Indiscriminant combination of one part of a force field with another derived independently would lead to considerable divergence in the calculated results from those obtained by experimental observation.

The most extreme difference between force fields arises in the method by which the hydrogen bond is included. Because atoms involved in a hydrogen bond are often closer than the sum of their VDW radii, they must be handled in a special manner. Several force fields have special functional forms with angular dependency that not only have special VDW parameters, to ensure that the close approach of the atoms involved is calculated correctly, but that the angular distribution observed for hydrogen bonds is also reproduced. Hagler et al. (34) used an amide hydrogen with a zero VDW radius for hydrogen bonding and a slightly greater nitrogen radius to give a correct amide hydrogen bond distance. The charges on the atoms involved (including the amide hydrogen) are adjusted to give an appropriate balance of VDW repulsion and dipole attraction. Clearly, the method for handling the electrostatic interaction is an integral part of each force field and cannot be modified independently.

2.1.2 Electrostatics. The most difficult aspect of molecular mechanics is electrostatics (35–38). In most force fields, the electronic distribution surrounding each atom is treated as a monopole with a simple coulombic term for the interaction. The effect of the surrounding medium is generally treated with a continuum model by use of a dielectric constant. More

detailed approaches with distributed multipole representations of the electron distribution (39, 40) and/or efforts to deal with dielectric inhomogeneity through solution of the Poisson equation are clear improvements and have become routine in many studies. Other difficulties arise in dealing with macromolecular systems, given that the electrostatic interaction is long ranged ($1/r$) and the interactions cannot be arbitrarily terminated with distance. Electrostatic interactions range from those operating only at very short distances that are nonspecific (dispersive interactions, r^{-6} dependency) to those operating at very long distances with a high degree of specificity (charge-charge interactions, r^{-1} dependency).

Dispersive Interactions (r^{-6}). These are attributed to interaction of induced dipoles within the electron clouds as molecules come in proximity and are responsible for the attractive part of the nonbonded van der Waals interaction.

Dipole–Dipole Interactions (r^{-3}). Because of the nonsymmetrical distribution of electrons between atoms of different size and electronegativity, bonds have associated permanent dipoles. The interaction energy between two of these dipoles depends on their relative orientation. This is basically the interaction underlying the phenomenon of the hydrogen bond. Although some force field authors use a special hydrogen bonding potential with an orientation dependency, simple partial charge representations combined with appropriate VDW parameters can reproduce the effect as well (34).

Charge–Dipole Interactions (r^{-2}). A charge interacting with a permanent dipole can be handled simply by considering the charge interacting with the two charges at the poles of the dipole. Alternatively, if the distance between the poles of the dipole is small compared with that between the centers of the ion and the dipole, then the potential energy Φ can be approximated as

$$\Phi = e\mu \cos \Theta / r^2$$

where e is the charge of ion, μ is the dipole moment, Θ is the angle between the vector connecting the center of the dipole with charge and dipole orientation, and r is the distance between the center of the ion and the center of the dipole.

Charge-Charge Interactions (r^{-1}). The energy of interaction between two charges q_1 and q_2 is given by Coulomb's law:

$$E = \frac{q_1 q_2}{4\pi\varepsilon r_{12}}$$

where r_{12} is the distance separating charges and ε is the dielectric constant of the medium.

To evaluate atom-atom interactions using Coulomb's law, the concept of net atomic charge is invoked. This amounts to representing charge as a point, a monopole, and is an artificial construct. Nevertheless, this is the common method. Recent improvements in calculating an appropriate set of point charges, to accurately reproduce the molecular electrostatic potential derived by quantum calculations, have been reported (41).

In an effort to increase the quality of electrostatic representations, dipole and higher multipole moments have been used. There are advantages in these more accurate representations, with a relatively small computational increase attributed to the reductions in distances over which the higher moments have to be summed, although they do require additional effort in the derivation of the parameters for the higher moments themselves. A good example is the distributed multipole model of electrostatics derived for peptides. A review by Williams (42) discusses the problems of deriving a distributed multipole expansion of charge representation that accurately reproduces the molecular electrostatic potential derived from quantum calculations. Comparisons were made between atomic multipoles, bond dipole, and restricted bond dipole models. Williams finds that a model for the electrostatic potential based on bond dipoles supplemented with monopoles (for ions) and atomic dipoles (for lone pairs) is most useful. Dipole-dipole energy converges much faster than monopole-monopole energy. Molecular charge at any desired position in a molecule is not a physically measurable quantity; one can only calculate a delocalized electron probabil-

ity distribution from quantum theory. Clearly, the more complex the representation, the more accurately one can approximate the quantum mechanical results, and the more realistic should be the results obtained. One complexity of electrostatics is the long distances over which interactions occur. Appropriate means of truncating the long-range forces to maintain the accuracy of simulations are necessary (43–45) and progress in better approximations has been reported (46). The difficulties with cutoff schemes were demonstrated (47, 48) by significant variations in the behavior of a 17-residue helical peptide simulated with explicit waters, using various electrostatic schemes and by studies (49) of a pentapeptide in aqueous ionic solution (50). In both cases, the Ewald approximation in which periodicity is assumed (which allows summation over much longer distances) gave superior results (47–49).

2.1.2.1 The Dielectric Problem and Solvation. Although methods of localizing charge just described may give reasonable results, the use of Coulomb's law with a dielectric constant, a scaling factor related to the polarizability of the medium between the charges, is clearly of concern. The dielectric at the molecular level is neither homogeneous nor continuous, nor even well defined, and thus violates the basic assumption of Coulomb's law. Although the use of a low, uniform dielectric is more nearly correct in dynamical simulations where all solute and solvent atoms are explicitly included, a variety of comparisons of experimental data with the results of calculation by use of a simplified solvent model have led to the realization that much better approaches are needed. Initial efforts (51) led to the proposal of a variable dielectric ($1/R$ or $1/4R$). More recently, the use of approaches that model the inhomogeneity of the dielectric at the interface between the solute and solvent by use of the Poisson-Boltzman equation have shown considerable promise (52, 53). An alternative approach that uses the mirror charge approximation has been described by Schaefer and Froemmel (54). Excellent reviews (35–38) of the electrostatic problem have appeared, to which the reader is referred.

Much effort has been given to simple continuum models of solvation to explain the origin of solvent effects on conformational equilibria and reaction rates. The current status of such efforts, as well as simulations to rationalize solvation effects, has been reviewed by Richards et al. (55). There are two general approaches to the continuum models. The first is reaction field theory (Bell, Kirkwood, Onsager) that follows the classical treatment of Debye-Huckel. The solvent is considered in terms of charge distribution, polarizability, and dielectric constant. The solvation energy is determined simply by considering the solute as a point dipole that interacts with the induced charge distribution in the solvent (Onsager reaction field). An extension by Sinangolou in the 1960s partitioned solvation energy into cavity formation, solvent-solute interaction, and the "free volume" of the solute. The logical extension of this approach is scaled-particle theory (56), in which the free energy of formation of a hard-sphere cavity of diameter a2 in a hard-sphere solvent of diameter a and number density p is scaled to the exact solution for small cavity sizes. Alternatively, the virtual charge approach used a system of effective and virtual charges interacting in the gas phase. The Hamiltonian of the system is modified to include an imaginary particle, a "solvaton" with an opposite charge for each of the solute atoms and solved by the SCF procedure. These continuum models have met with limited success (trends and relative effects of solvation can be predicted), although highly specific molecular interactions, such as those involving hydrogen-bonding groups, cannot be accommodated.

In the equation for calculating affinity of a drug for a receptor, the ligand is solvated either by the receptor or by the solvent. This competition means that accurate determination of the free energy of solvation is important in understanding differences in affinities. Solvation free energy (G_{sol}) can be approximated by three terms: G, the formation of a cavity in the solvent to hold the solute; G_{vdW} and G_{pol}, the interaction between solute and solvent divided between VDW and electrostatic forces, respectively:

$$G_{sol} = G_{cav} + G_{vdW} + G_{pol}.$$

There are four theoretical approaches to the problem:

1. Scaled Particle Theory (56)

   The essence of the scaled particle theory is that formation of a cavity in a fluid requires work. The theory for hard spheres has been well developed from statistical mechanics, and the work, $W(R, \rho)$, can be calculated as follows:

$$W(R,\rho)/kT = -\ln(1-y) + (3y/1-y)R + [(3y/1-y) + 9/2*(y/1-y)^2]R^2 + yPR^3/rkT$$

   where y = $\pi\rho\sigma_1^3/6$, R = σ_2/σ_1, σ_2 is the diameter of the hard-sphere solute, σ_1 is the diameter of the solvent, and p is the number density of the fluid (N/V).

   Because this theory includes no interaction between solvent and solute (i.e., only G_{cav} is calculated), effective volumes for nonspherical compounds with interactive groups are normally calibrated from experiment. This is one way to deal with the energy of interaction between solvent and solute. For further discussion, see Pollack (57).

2. Charge Image (or Virtual Charge) Method (54) (Method for G_{pol} calculation)

   This model replaces the solute-continuum model with one in which a system of charges derived from the solute and virtual charges in the adjacent space interact in the gas phase. A set of mirror charges reflected at the dielectric boundary are created and used in the calculation of the electrostatics.

3. Boundary Element Method (58) (Method for G_{pol} calculation)

   In this approximation, the system is modeled by calculating the appropriate surface charges at the dielectric boundary. This is similar to fitting charges at atomic centers to reproduce the molecular electrostatic potential. For a quantum-mechanical equivalent, Tomasi et al. (59) introduced a charge distribution on the surface of a cavity of realistic shape to introduce the solvation term in the Hamiltonian of the solute. The charge distribution on the surface of the cavity depends on the solute's electric field, which is affected in turn by polarization from the cavity's surface. An iterative QM procedure is used to obtain the perturbation term. Cramer and Trular have developed AMSOL to include a solvent approximation in calculations of molecular systems. The approach has been calibrated by comparison of theoretical and experimental solvation free energies for numerous molecular species (60).

4. Poisson-Boltzmann Equation (53) (Method for G_{pol} calculation)

   Generalization of the Debye-Huckel theory leads directly to the Poisson-Boltzmann equation that describes the electrostatic potential of a field of charges with dielectric discontinuities. This equation has been solved analytically for spherical and elliptical cavities, but must be solved by finite-difference methods on a grid for more complicated systems. One exciting advance in this area is the development of an approximate equation for the reaction field acting on a macromolecular solute, attributed to the surrounding water and ions (61). By combining these equations with conventional molecular dynamics, solvation free energies were obtained similar to those with explicit solvent molecules, at little computational cost over vacuum simulations. This implies that a more nearly correct solution to the electrostatics problem might minimize the solvation problem. Other approaches to evaluations of G_{sol} have recently appeared in the literature. Still et al. (62) estimated $G_{cav} + G_{vdW}$ by the solvent-accessible surface area times 7.2 cal/mol/Å^2. G_{pol} is estimated from the generalized Born equation. Effective solvation terms have been added (63, 64) to molecular mechanics force fields to improve molecular dynamics simulations without the cost of modeling explicit solvent. Zauhar (65) combined the polarization-charge technique with molecular mechanics to effectively minimize a tripeptide in solvent.

One final refinement may be necessary in some situations: the inclusion of electric polar-

izability, for example, by inclusion of induced dipoles, or distributed polarizability (66) in the electrostatic representation of the model. Kuwajima and Warshel (67) recently examined the effects of this refinement in modeling crystal structures of polymorphs of ice. Such models including polarizability have been previously shown useful for predicting the properties of crystalline polymorphs of polymers by Sorensen et al. (68). Caldwell et al. (69) included implicit nonadditive polarization energies in water-ion outcomes, resulting in improved accuracy. At the semiempirical level of quantum theory, Cramer and Truhlar (70–73) added solvation and solvent effects on polarizability to AM1, with impressive agreement between experimental and calculated solvation energies (60). Rauhut et al. (74) also introduced an arbitrarily shaped cavity model by use of standard AM1 theory.

2.1.2.2 The "Hydrophobic" Effect. Water has been the nemesis of solvation modeling because of its rather unique thermodynamic properties, as reviewed by Frank (75) and Stillinger (76). The biochemical literature discusses at length "hydrophobic effects" (77). This effect is not "hydrophobic" at all because the enthalpic interaction of nonpolar solutes with water is favorable. This, however, is counterbalanced by an unfavorable entropic interaction that is interpreted as an induced structuring of the water by the nonpolar solute. Water interacts less well with the nonpolar solute than it does with itself because of the lack of hydrogen-bonding groups on the solute. This creates an interface similar to the air-water interface, with a resulting surface tension attributed to the organization of the hydrogen-bonded patterns available. This is the so-called iceberg formation around nonpolar solutes in water, first suggested by Frank and Evans. Studies by both molecular dynamics (78–80) and Monte Carlo simulations (81) support this interpretation (76), although there is still considerable controversy in interpretation of experimental data (82).

2.1.2.3 Polarizability. The traditional approaches in molecular mechanics have excluded the effects of charge on induced dipoles and multibody effects. This approximation becomes a serious limitation when dealing with charged systems and molecules like water that are highly polar. A recent paper (83) from the Kollman group described nonadditive many-body potential models to calculate ion solvation in polarizable water with good agreement with experimental observation. It was necessary to include a three-body potential (ion-water-water) in the molecular dynamics simulation of the ionic solution to obtain quantitative agreement with solvation enthalpies and coordination numbers. Inclusion of a bond-dipole model with polarizability in molecular dynamics simulations has given excellent agreement in predicting physical properties of polymers by Sorensen et al. (68).

A novel approach based on the concept of charge equilibration has been suggested by Rappe and Goddard (84) that allows the inclusion of polarizabilities in molecular dynamics calculations.

2.1.3 The Potential Surface. The set of equations that describe the sum of interactions between the ensemble of atoms under consideration is an analytical representation of the Born-Oppenheimer surface, which describes the energy of the molecule as a function of the atomic positions. Many important properties of the molecule can be derived by evaluation of this function and its derivatives. For example, setting the value of the first derivative to zero and solving for the coordinates of the atoms leads one to minima, maxima, and saddlepoints. Evaluation of the sign of the second derivative can determine which of the above have been found. It is a straightforward procedure to calculate the vibrational frequencies from the force constants by evaluation of the eigenvalues of the secular determinant (the mass-weighted matrix; see textbook on vibrational spectroscopy). Gradient methods for the location of energy minima and transition states are an essential part of any molecular modeling package. It is essential to remember, however, that minimization is an iterative method of geometrical optimization that is dependent on starting geometry, unless the potential surface contains only one minimum (a condition not found for any system of sufficient complexity to be of real interest).

The ability to locate both minima and transition points enables one to determine the minimum energy reaction path between any

two minima. In the case of flexible molecules, these minima could correspond to conformers and the reaction path would correspond to the most likely reaction coordinate. One could estimate the rate of transition by determination of the height of the transition states (the activation energy) between the minima. Elbers (85) developed a new protocol for the location of minima and transition states and applied it to the determination of reaction paths for the conformational transition of a tetrapeptide (86). Huston and Marshall (87) used this approach to map the reaction coordinates of the a- to 3_{10}-helical transition in model peptides.

Despite the limitations that curtail exact quantitative applications, molecular mechanics can provide three-dimensional insight as the geometric relations between molecules are adequately represented. Electrical field potentials can be calculated and compared to give a qualitative basis for rationalizing differences in activity. Molecular modeling and its graphical representation allow the medicinal chemist to explore the three-dimensional aspects of molecular recognition and to generate hypotheses that lead to design and synthesis of new ligands. The more accurate the representation of the potential surface of the molecular system under investigation, the more likely that the modeling studies will provide qualitatively correct solutions.

2.1.3.1 Optimization. The search for the optimal solution to a complex problem is common to many areas in science and engineering and does not have a general solution. Numerous approaches to this problem, which is generally referred to as optimization, have been used in chemistry: most commonly, distance geometry, molecular dynamics, stochastic methods such as Monte Carlo sampling, and systematic, or grid, search. Most rely on minimization, often combined with a stochastic search. Minimization algorithms have been thoroughly characterized with regard to their convergence properties, but, in general only locate the closest local minima to the starting geometry of the system. A stochastic approach to starting geometries can be combined with minimization to find a subset of minima in the hope that the global minimal is contained within the subset and can readily be identified by its potential value compared with that of the other minima.

2.1.3.2 Potential Smoothing. One approach to global optimization that has shown promise is potential smoothing (88). This approach uses a mathematical transformation to smooth the multidimensional potential energy surface of a molecule, reducing the high frequency complexity of the surface and making it much easier to search for minimum energy conformations. This concept was first used to deform the conformational potential energy surface in the diffusion equation method (DEM) of Piela and coworkers (89). Search procedures will not confront multiple local minima on the deformed surface. If the procedure is reversed iteratively, then one can trace the path back into a region that lies near the global minimum of the undeformed potential surface. Ponder et al. (88, 90) improved the procedure for tracing back from one partially deformed surface to the next by including a local search procedure to limit detection of false minima.

One of the best known benchmark problems for conformational search involves the determination of the low energy conformations of the highly flexible cycloheptadecane (91, 92). This system continues to serve as a test for newly developed search methods (93). Although not a particularly large molecule, this system is a challenge because of its flexibility and the close energy spacing of the lower lying minima. Extensive analysis through a variety of search methods has located exactly 263 minima within 3.0 kcal/mole of the purported global minimum. The potential smoothing search (PSS) (88) was dramatically effective at locating many of the lowest energy structures for cycloheptadecane. Although the global minimum for cycloheptadecane was not located, the second lowest energy structure was located and differed by only 0.01 kcal/mole. Based on its MM2 vibrational frequencies, the global minimum is entropically disfavored relative to all of the minima located by the smoothing procedure. The PSS method was also applied to obtain the minimum energy conformation of the TM helix dimer of glycophorin A (GpA) (94), previously solved by solution NMR spectroscopy (95).

2.1.3.3 Genetic Algorithm. Another approach to global optimization is the genetic algorithm. This approach is based on biological evolution and is analogous to natural selection (96–98). In applications to computational chemistry, evolution on the computer has been shown to be an efficient approach to global optimization, although because of sampling issues, there is no guarantee that the global optimum has been found in any particular application (99).

2.1.3.3.1 Characteristics of the Genetic Algorithm. In analogy to natural selection, the parameters to be optimized are encoded in a bit string and strung together in a "chromosome." Each chromosome in the population represents a particular genotype or solution to the problem under consideration (i.e., a specific set of values for the parameters that determine the configuration of the system under study). The values of the parameters have to be decoded for the "fitness" of a particular genotype to be evaluated. Once the fitness of each chromosome in the population has been evaluated, then the more "fit" members are allowed to reproduce, mutate, or cross over with other members of the parent population to generate a new daughter population. This process is repeated until the fitness of the population converges, or until the available computer cycles are consumed.

2.1.3.3.2 Example of Conformational Analysis. The simplifying assumption of rigid geometry is used to reduce the computational complexity of the model problem of conformational analysis. The elimination of variables is rationalized based on the high energy cost associated with bond length distortions and the ability to accommodate bond angle deformations by a reduced set of van der Waals radii. To represent the conformation of a molecule, one needs only to specify the values of the torsional angles associated with rotatable bonds. One can assign a set number N of bits, 6 for example, to represent 2^N values for the torsional angles. Each set of 6 bits can be considered a "gene" and crossover allowed only at gene boundaries, if desired. Thus, the conformation of a molecule can be encoded as a set of torsional genes. The actual coordinates of the molecule corresponding to each genotype must be generated for the fitness function F, in this case internal energy, to be numerically evaluated by molecular mechanics. Each chromosome in the population is evaluated for its internal energy and a subset of the more fit selected for reproduction. The degree of limitation on reproductive fitness is analogous to the selective pressure brought to bear on a population (i.e., selection of the fittest). This is a parameter that can be varied in most GA programs and one must balance selective pressure against maintaining some variation in the population for evolution to occur (to avoid being trapped in a local minimum). The set of chromosomes to be reproduced can be based on some arbitrary criteria (the top 50%, all those with fitness at least half that of the most fit chromosome detected, or the fitness scaled in some way and chromosomes reproduced in proportion to their scaled fitness.

Given a subset of chromosomes to reproduce, several operations analogous to evolution are invoked. First is mutation, where a certain number of randomly selected bits are mutated from 0 to 1 or vice versa in the daughter chromosome. This would allow for changes in the settings of one or more torsional angles. A certain number of pairs of chromosomes are also selected for crossover and one or more locations between genes (if specified) are randomly selected and the two pieces derived from each parent chromosome swapped, to generate two or more novel chromosomes. This would allow for different subsets of conformations to be combined; this provides a mechanism for concerted changes or jumps over barriers to find minima that would be difficult to sample by mutation alone. This would appear to be the feature that provides the analogous behavior to simulated annealing in efficient searching of parameter space. In this case, however, the search is more directed by the selective pressure of increasing the "fitness" or facing elimination from the population. In other words, each new generation should have eliminated a significant portion of the less fit members of the previous generation and propagated those torsional values that generate good local conformational states.

2.1.3.3.3 Schema and the Building Block Hypothesis. Once a population of good local substates has been established, then crossover

can probe the combination of these subconformations that have positive interactions leading to more fit progeny. In the jargon of computer science, the subpattern of 1's and 0's giving a preferred subconformation would be a schema (or building block). According to the most accepted theory, the building block hypothesis, the genetic algorithm initially detects biases toward fitness in lower order (fewer identical bits) schemas and converges on this part of search space (the entire set of bit strings). By combining information from lower order schema through crossovers, biases in higher order schemas are detected and propagated.

The strong convergence property of the genetic algorithm is a major attraction. Given sufficient members of the population and sufficient evolutionary time (number of generations), then one can expect convergence if the fitness function is based on the optimal combination of locally optimized substructures. Some fitness functions are termed "deceptive," in that low order schemas are not present in higher order schemas and their propagation slows detection of the more fit higher order schemas. Another problem arises when the population size is too small or the selection factor too high. Then, the genetic algorithm can magnify a small sampling error and prematurely converge in a local optimum.

2.1.3.3.4 Mutations and Encoding. There are different ways to encode binary numbers by bit strings and these can have some influence on the impact of mutation. Traditional binary encoding requires that all bits be changed for some cases if the digital value is to be simply incremented. This causes erratic behavior near an optimum, with mutation and mutations in higher order bits having more effect than in lower order bits.

2.1.3.3.5 Crossovers and Encoding. In our example, we indicated that one might want to separate the bit string into genes corresponding to torsional angles because the gene has a coherent meaning in the context of the problem. If one restricts crossovers to the junctions between genes, then the coherence of the conformation of molecular fragments is preserved and one is more likely to make a successful crossover producing more fit offspring. There are methods such as random-key encoding (97) to generalize the process of crossovers without requiring customized crossover operators that are problem specific, although this is beyond the scope of this chapter.

2.1.3.3.6 Examples of Applications to Biochemical Problems. McGarrah and Judson (100) explored the impact of different parameters setting on the ability of the GA to explore the conformational space of cyclo(Gly_6). Each residue was represented by four angles, each with a string of four bits (1/16 of range). A selection fraction of 50% was used, which eliminated the lower half in fitness from reproduction. Population sizes of 10, 50, and 100 were tested. Each group was divided into four niche populations with communication between groups. Local minimization was performed for each chromosome before evaluation. They concluded that it was of little use to examine a population size of less than 100 members for the 24 variables examined. As soon as convergence in the average is detected in a population, it should be cross-fertilized from another niche or GA evolution should terminate. It is a clear example of a hybrid approach, in which GA does a rough search for minima and local minimization to find the closest local minimum.

Judson et al. (101) examined the use of a genetic algorithm to find low energy conformers of 72 small to medium organic molecules (1–12 rotatable bonds) whose crystal structures were known. They used the elitist strategy, in which the best individual from each generation is propagated without modification. A population size of 10 times the number of the nonring dihedral angles being varied was chosen. Each molecule was allowed to run for 10,000 energy evaluations, or until the population was bit converged. In a few cases, conformers with lower energies than those observed in the crystal structure were found. A comparison with CSEARCH in SYBYL (Tripos, Inc.) was made, but the differences in efficiencies found were not compelling. In only 9 of the 72 cases examined. did the GA find its best conformer had energy greater than the crystal structure, with the largest deviation being only 0.8 kcal/mol.

The GA approach has also been applied to the docking problem with dihydrofolate reductase, arabinose binding protein, and sialidase

(98).A typical run took minutes on a workstation and the predicted conformations agreed with those observed crystallographicallyin all cases. Meadows and Hajduk (102) used experimental constraints with a GA algorithm to dock biotin to stepavidin. Judson et al. (101) also reported docking of flexible molecules into the active sites of thermolysin, carboxypeptidase, and dihydrofolate reductase. In 9 of the 10 cases examined, the GA found conformations within 1.6 Å root-mean-square (rms) of the relaxed crystal conformation.

This approach has also been used in the PRO_LIGAND *de novo* design program (103) to optimize the structure of ligands for a binding site. Aset of candidate structures was generated and then crossover between molecular fragments used to optimize the predicted binding mode. This is similar to the SPLICE program of Ho and Marshall (104)that evolves ligands with more favorable interactions with a given site.

Payne and Glen (105) studied several different aspects of molecular recognition with genetic algorithms. Conformations and orientations were determined which best-fit constraints such as inter- or intramolecular distances, electrostatic surface potentials, or volume overlaps with up to 30 degrees of freedom.

2.1.4 Svstematic Search and Conformational Analysis. Because of the convoluted nature of the potential energy surface of molecules, minimization usually leads to the nearest local minimum (106,107) and not the global minimum. In addition, many problems in structure-activity studies require geometric solutions that may not be at the global minimum of the isolated molecule. To scan the potential surface with some surety of completeness, systematic, or grid, search procedures have been developed. To understand the strengths and limitations of this approach, some of the algorithmic details must be considered. These are discussed in depth in a review by Beusen et al. (108).

2.1.4.1 Rigid Geometry Approximation. A simplifying assumption that is usually invoked to reduce the computational complexity of the problem through elimination of variables is that of rigid geometry. The rationale is based on the high energy cost associated with bond-length distortions and the ability to accommodate bond-angle deformations by a reduced set of VDW radii. This approach is compatible with problems where one is most interested in eliminating conformations that are energetically unlikely (i.e., sterically disallowed) because of VDW interactions, which cannot be relieved by bond-angle deformation. A successful application requires that one calibrate an appropriate set of VDW radii for the particular application area. Iijima et al. (109) calibrated such a set (Fig. 3.2) for peptide application by comparison with experimental crystallographic data from proteins and peptides.

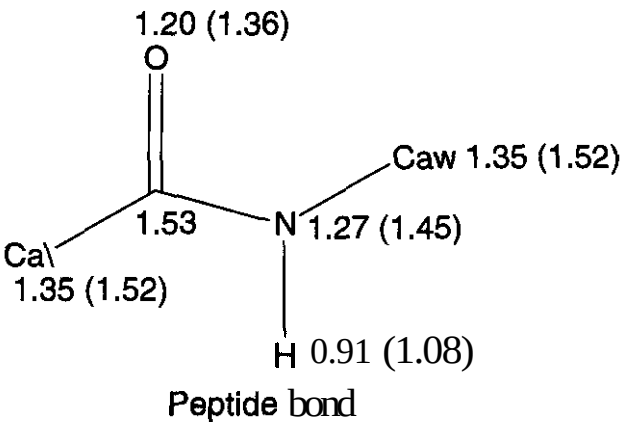


Figure 3.2. Calibrated set of van der Waals radii for peptide backbone for use with rigid geometry approximation (109). Usual radii shown in parentheses. Carbonyl carbon not modified.

2.1.4.2 Combinatorial Nature of the Problem. Using the rigid geometry assumption, one can analyze the combinatorial complexity of a simplified approach to the problem with some ease. Let us assume a molecule (Fig. 3.3) of N atoms with T torsional degrees of freedom (i.e., rotatable bonds). For each torsional degree of freedom T, explored at a given angular increment in degrees A, there are $360/A$ values to be examined for each T. This means that $(360/A)^T$ sets of angles, each describing a unique conformation, need to be examined for steric conflict. For each conformer, the starting geometry will have to be modified by applying the appropriate transformation matrices to different subsets of atoms to generate the coordinates of the conformation. For each conformation, $N(N-1)/2$ distance determinations will have to be calculated to a first approximation (this does not exclude bonded at-

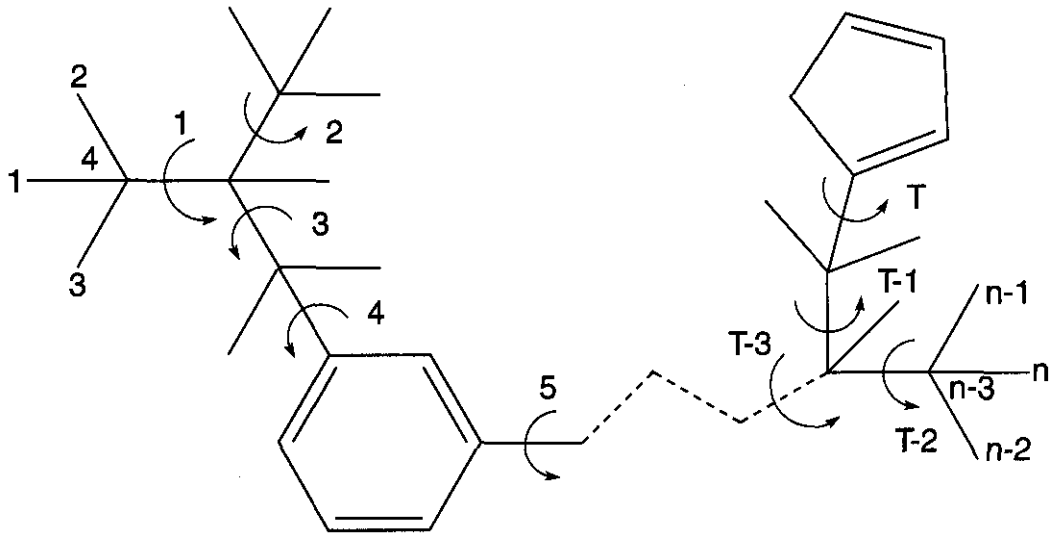


Figure 3.3. Schematic diagram of molecule with N atoms and T rotatable bonds.

oms and atoms bonded to the same atom from the check, which is necessary) and checked against the allowed sum of VDW radii for the two atoms involved. The number of VDW comparisons V is given by

$$V = (360/A)^T \times N(N-1)/2$$

It should be clear that the VDW comparisons are the rate-limiting step by their sheer number, and any algorithmic improvement that reduces the number of such checks or enhances the efficiency of performing such checks is of value.

2.1.4.3 Pruning the Combinatorial Tree. From this simplified analysis, a systematic search of other than the smallest molecules at a coarse increment would appear daunting. A hybrid approach with a coarse grid search followed by minimization has been successfully used to locate minima. There are a number of algorithmic improvements over the "brute force" approach that enhances the applicability of the systematic search itself. To understand these improvements, some concepts need to be defined. First is the concept (110) of aggregate, a set of atoms whose relative positions are invariant to rotation of the T rotational degrees of freedom. n-Butane is divided into aggregates as an illustration (Fig. 3.4).

In this simple example, the atoms in an aggregate are all either directly bonded or have a 1–3 relationship (i.e., are related by a bond angle). Because of the rigid geometry approximation, their relative positions are fixed. Atoms contained within the same aggregate do not, therefore, have to be included in the set of those that undergo VDW checks for each conformation. For linear molecules, there are $n - 1$ bonds and the number of 1–3 interactions depends on the valence of the atom. This simplification leads to a reduction of the number of VDW checks by the factor $N(N-1)/2$, which is multiplied by the number of conformations.

How can one reduce the number of conformations that have to be checked? Here the concept of construction becomes useful. One constructs the conformations in a stepwise fashion, starting with an initial aggregate and adding a second aggregate at a given torsional increment for the torsional variable T that is applied to the rotatable bond connecting the two. If any pair of atoms overlaps for that increment, then one can terminate the construction because no addition operation will relieve that steric overlap. In effect, one has truncated the combinatorial possibilities that would have included that subconformation; that is, one has pruned the combinatorial tree.

2.1.4.4 Rigid Body Rotations. If one constructs the molecule stepwise by the addition of aggregates, then one has two sets of atoms to consider. First are those in the partial mol-

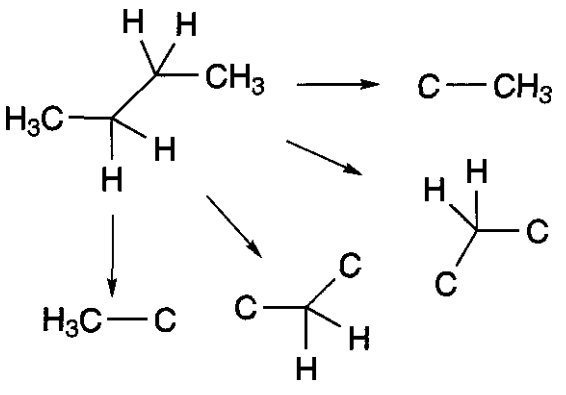


Figure 3.4. Decomposition of *n*-butane molecule into aggregates.

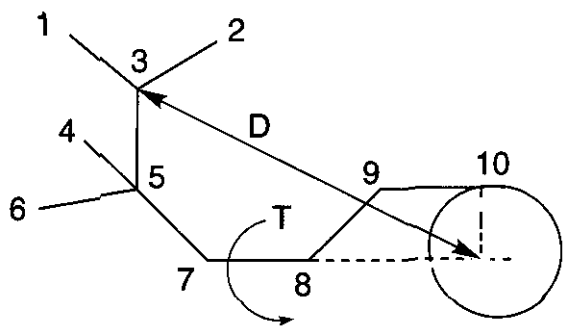


Figure 3.5. Distance between atoms (1–7) and atom 10 separated by a single rotatable bond T can be described with a transformation of the equation of a circle describing the locus of atom 10 as bond T is rotated. Notice that distance D between any atom (1–7) and center of circle of rotation of atom 10 that is on axis of rotation is fixed, regardless of value of T.

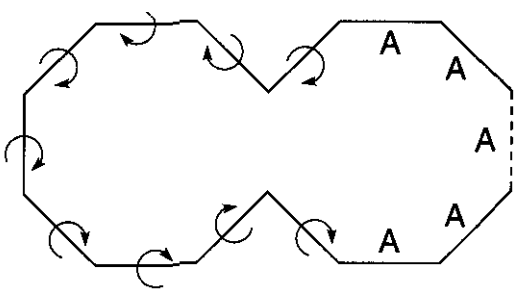


Figure 3.6. Scheme for combining systematic search with analytical solution for closure. Bonds indicated by arrows were systematically scanned, whereas those indicated by A were analytically determined. Dotted bond can represent either chemical bond or experimental distance determination (NOE, etc.).

ecule (set A), previously constructed, that have been found to be in a sterically allowed partial conformation. For each possible addition of the aggregate, the atoms of the aggregate (set B) must be checked against those in the partial molecule. If one uses the concept of a rigid body rotation, then one can describe the locus of possible positions of any atom in set B as a circle whose center lies on the axis of rotation T_i (the interconnectingbond) at a distance along the axis that can be calculated. The formula for a circle can be transformed to represent the possible distances between the atom b in set B and any atom *a* in set A, as shown in Fig. 3.5. An equation with scalar coefficientsthat describes the variable distance between two atoms as a function of a single torsional variable was derived (111), which has a discriminant whose evaluation can be used to determine whether atom *a* and atom b will:

- he in contact, despite changes in the value of the torsional rotation of the aggregate, which implies that the current partial conformation has to be discarded, given that there is no possible way to add the aggregate that is sterically allowed;
- never come in contact for any value of the torsional rotation, so that this pair of atoms can be removed from consideration regarding this aggregate; or
- come in contact for some values of the torsional rotation that can be calculated for that pair and that removes a segment of the torsional circle from consideration for other atom pairs. If all segments of the torsional circle are disallowed by combinations of the angular requirements of different atom pairs, then the partial conformation of the molecule is disallowed because further construction is not feasible. As a first approximation, this removes a degree of torsional freedom from the problem, reducing T to T – 1 torsional degrees of freedom. At a 10'' torsional scan, an approximate reduction in computational complexity of a factor of 36 results.

2.1.4.5 The Concept and Exploitation of, Rings. Realization that many of the relevant constraints in chemistry can be expressed as interatomic distances, VDW interactions, nuclear Overhauser effect constraints, and so forth allows use of the concept of a virtual ring in which the constraint forms the closure bond. Small rings up to six members can be solved analytically (112), so that one can search the torsional degrees of freedom associated with a constraint until only five remain and then solve the problem analytically (Fig. 3.6). The torsional angles for those degrees of freedom are no longer sampled on a grid, thus removing the problem of grid tyranny, in which valid conformations are missed by the choice of increment and starting conformation. This approach is then a hybrid because only part of the conformational space is searched with regular torsional increments. It is, however, much more efficient to solve a set of equations than search 5 torsional degrees of freedom.

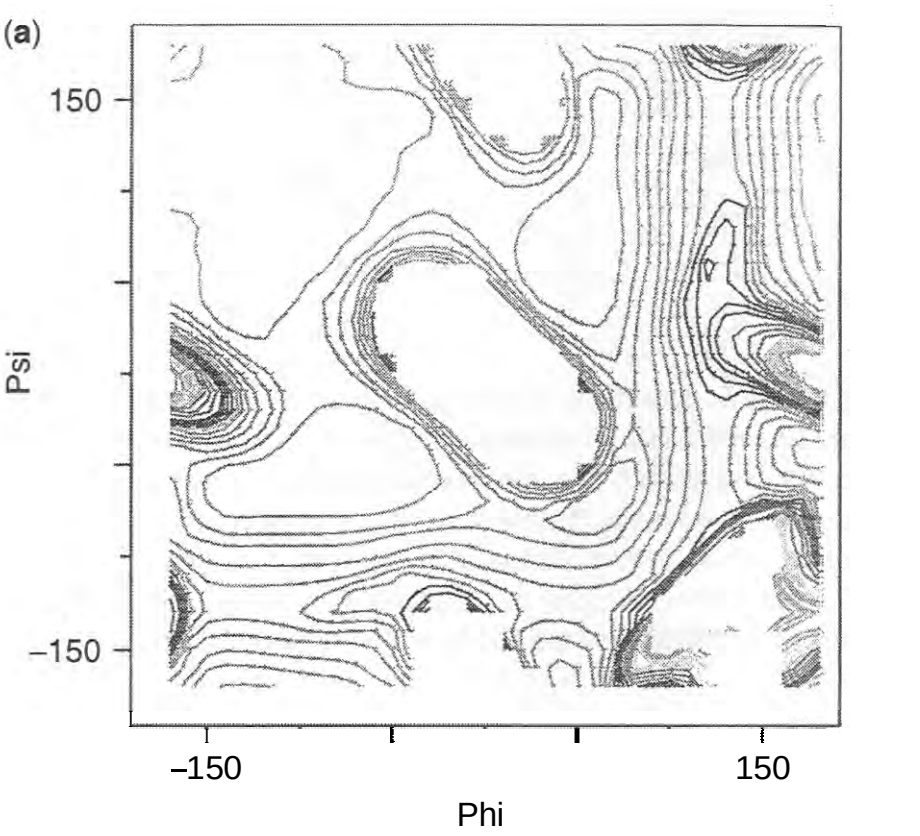


Figure 3.7. (a) Two-dimensional (Ramachandran) plot of energy vs. backbone torsional angles, @ and Ψ, for *N*-acetyl-valine-methylarnide. (b) Three-dimensional plot of energy vs. torsional angles, Φ, Ψ, and $\chi 1$, for N-acetyl-valine-methylamide.

2.1.4.6 Conformational Clustering and families. In a congeneric series, the correspondence between torsional rotation variables is maintained as one compares molecules, and a direct comparison of the values allowed for one molecule with those allowed for another is meaningful. Two- (2D) or three-dimensional (3D) plots (Fig. 3.7) of torsional variables against energy often provide considerable insight into the difference in conformational flexibility between two molecules. Such a plot of the peptide backbone torsional angles Φ, Ψ is known as a Ramachandran plot. When more than three torsional variables become necessary to define the conformation of the molecule under consideration, then multiple plots become necessary to represent the variables. Unless special graphical functions are included in the software, then correlations between plots become difficult, given that each plot is a projection of a multidimensional space. One approach to this problem is to use cluster analysis programs to identify those values of the multidimensional variables that

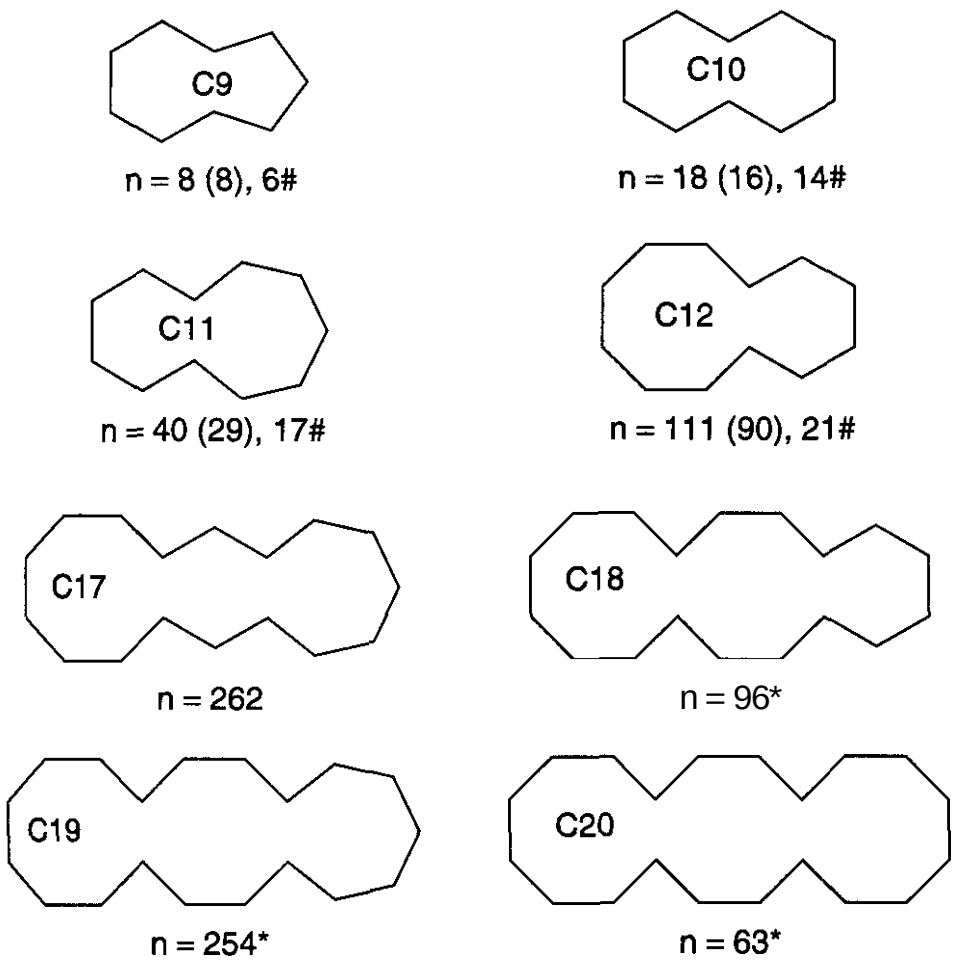


Figure 3.8. Cycloalkane rings and number of local minima found by various search strategies. n, number of conformers with MM2 (117); parentheses, number of conformers with MM3 (117); #, number of conformers within 25 kJ/mol of global minima [MM2 (92)]; *, number of minima found within 3 kcal/mol of global minima (115).

are adjacent in N-space. The clusters of conformers that result have been referred to as families. A member of a family is capable of being transformed into another conformer belonging to the same family without having to pass over an energy barrier; that is, the members of a family exist within the same energy valley.

Because of the combinatorial nature of systematic search, one is often faced with large numbers of conformers that have to be analyzed. For some problems, energetic considerations are appropriate and conformers can be clustered with the closest local minimum, providing to a first approximation an estimate of the entropy associated with each minima by the number of conformers associated, in that they can come from a grid search that approximates the volume of the potential well. A single conformer, perhaps the one of lowest energy, can be used with appropriately adjusted error limits in further analyses as representative of the family.

2.1.4.7 Conformational Analysis. Although interaction with a receptor will certainly perturb the conformational energy surface of a flexible ligand, high affinity would suggest that the ligand binds in a conformation that is not exceptionally different from one of its low energy minima. Mapping the energy surface of the ligand in isolation to determine the low energy minima will, at the very least, provide a set of candidate conformations for consideration, or as starting points for further analyses. The problem of finding the global minimum on a complicated potential surface is common to many areas, and lacks a general solution. Minimization procedures locate the closest local minimum depending on the starting conformation. Several strategies have developed to map the potential surface and locate minima. For an excellent overview of the different approaches, the reader is referred to the surveys by Leach (113) and by Burt and Greer (114). Stochastic methods such as Monte Carlo have been advocated (115) for conformational analysis and their usefulness demonstrated on carbocyclic ring systems (91, 115–121) (Fig. 3.8). Molecular dynamics can be used to explore the potential energy surface, often with simulated annealing to help overcome activation-energy barriers, but exploration is concentrated in local minima and duplication of the surface explored is controlled by Boltzmann's law. A systematic, or grid, search samples conformations in a regular fashion, at least in the parameter space (usually torsional space) that is incremented.

Comparisons of a variety of methods were made on cycloheptadecane by Saunders et al. (91) and it was concluded that the stochastic method was most efficient. In one of the few independent comparisons of the effectiveness of these procedures, Boehm et al. (122) studied the sampling properties on the model system caprylolactam, a nine-membered ring, and concluded that systematic search was both inefficient and ineffective at finding the minima found by the other methods when the number of conformers examined was limited.

2.1.4.8 Other Implementations of Systematic Search. Numerous other implementations of systematic, or grid, search programs exist in the literature and those with protein applications have been reviewed by Howard and Kollman (123), whereas those for small or medium sized molecules are included in the reviews by Burt and Greer (114) and by Leach (113). One of the more widely used programs in organic chemistry, MACROMODEL, has a search module (124) coupled to energy minimization for conformational analysis. MACROSEARCH has been developed by Beusen et al. (125) to generate the set of conformers consistent with experimental NMR data and used to determine the conformation of a 15-residue peptide antibiotic.

2.1.5 Statistical Mechanics Foundation (126). To understand the relationships between the simulation methods and the desired thermodynamic quantities, a short review of the major concepts of statistical mechanics may be in order. This is not meant to be comprehensive, but rather to remind the reader of the relevant ideas.

The set of configurations generated by the Monte Carlo simulation generates what J. Willard Gibbs would call an "ensemble," assuming that the number of molecules in the simulation was large and the number of configurations was also large. This ensures that the possible arrangements of molecules that are energetically reasonable have been adequately sampled. One is often interested in the statistical weight Wof a particular observable. For example, a particular conformation of a solute molecule, say, the staggered rotamer of ethane, could be compared with another conformer, the eclipsed rotamer, in a simulation with solvent. If more configurations of the surrounding solvent molecules of equivalent energy were available to the staggered than to the eclipsed, then the staggered would have a higher statistical weight. From the inscription on Boltzmann's tomb, we all recall that S = k ln W, where S is the entropy and k is Boltzmann's constant. Thus, we have a link between statistics and thermodynamics. W in this case would be the number of configurations associated with the particular conformation of ethane under consideration divided by the total number of configurations sampled. This would have to be weighted by their energy, of course, unless the distribution was already Boltzmann weighted, as happens when one uses the Metropolis algorithm (127).

Another way of stating this is that the probability P_i of a particular configuration N_i is proportional to its Boltzmann probability divided by the Boltzmann probability of all the other configurations or states:

$$P_i = \exp(-E_i/kT) \Big/ \sum_{i=1}^{N} \exp(-E_i/kT)$$

The denominator in this equation has been given a special name, partition *function*, often symbolized by Z, which is derived from the German *Zustandsumme* (sum over states). The successive terms in the partition function describe the partition of the configurations among the respectives states available. One can express the thermodynamic state functions of an ideal gas in terms of the molecular partition function Z as follows:

$$S = k \ln W = kN \ln Z/N + U/T + kN$$

where N is the number of molecules and U is the internal energy. From this and the assumption of an ideal gas $pV = NkT$, the Gibbs free energy $G = U - TS + pV$ leads to

$$G = -NkT \ln Z/N$$

and similarly, the Helmholtz free energy A = U – TS leads to the expression

$$A = -kT \ln Z^N/N!$$

all of which may be more familiar if expressed in terms of enthalpy, $H = U + pV$.

In summary, by simulating a relevant statistical sample of the possible arrangements of molecules when interacting, one can derive the macroscopic thermodynamic properties by statistical analysis of the results. In this case, one is deriving the partition function not by theoretical analysis of the quantum states available to the molecule, but through simulation. In other words, the average properties are valid if the Monte Carlo or molecular dynamics trajectories are ergodic, that is, constructed such that the Boltzman distribution law is in accord with the relative frequencies with which the different configurations are sampled. (An ergodic system is by definition one in which the time average of the system is the same as the ensemble average.) A basic concept in statistical mechanics is that the system will eventually sample all configurations, or microscopic states, consistent with the conditions (temperature, pressure, volume, other constraints) given sufficient time. In other words, a trajectory of sufficient length (in time) would sample configuration space.

***2.1.6* Molecular Dynamics *(37, 126,* 128).** Molecular dynamics is a deterministic process based on the simulation of molecular motion by solving Newton's equations of motion for each atom and incrementing the position and velocity of each atom by use of a small time increment. If a molecular mechanics force field of adequate parameterization is available for the molecular system of interest and the phenomenon under study occurs within the time scale of simulation, this technique offers an extremely powerful tool for dissecting the molecular nature of the phenomenon and the details of the forces contributing to the behavior of the system.

In this paradigm, atoms are essentially a collection of billiard balls, with classical mechanics determining their positions and velocities at any moment in time. As the position of one atom changes with respect to the others, the forces that it experiences also change. The forces on any particular atom can be calculated by evaluation of the energy of the system using the appropriate force field. From physics,

$$F = ma = -\delta V/\delta r = m\delta^2 r/\delta t^2$$

where F is the force on the atom, m is the mass of the atom, a is the acceleration, V is the potential energy function, and r represents the cartesian coordinates of the atom. Using the first derivative of the analytical expression for the force field allows the calculation of the force felt on any atom as a function of the position of the other atoms.

2.1.6.1 ***integration.*** In this simulation, we use numerical integration; that is, we choose a small time step (smaller than the period of fastest local motion in the system) such that our simulation moves atoms in sufficiently small increments, so that the position of surrounding atoms does not change significantly per incremental move. In general, this means that the time increment is on the order of 10^{-15} s (1 femtosecond). This reflects the need to adequately represent atomic vibrations that have a time scale of 10^{-15} to 10^{-11} s. For each picosecond of simulation. we need to do 1000 iterations of the simulation. For each iteration, the force on each atom must be evaluated and its next position calculated. For simulations involving molecules in solvent. sufficient solvent molecules must be included, so that the distance from any atom in the solute to the boundary of the solvent is larger than the decay of the intermolecular interaction between the solute and solvent molecules. This requires several hundred solvent molecules for even small solutes, and the computations to do a single iteration are sufficiently large that simulations of more than several hundred picoseconds for proteins with explicit solvent are still rare. Efforts to increase the time step and thus allow for longer simulations without sacrificing the accuracy of the methodology are under investigation. Combination of normal mode calculations with explicit numerical integration allows time steps up to 50 ps for model systems (129). A similar approach has been shown effective by Schlick and Olson (130) in modeling supercoiling of DNA.

Let us attempt a rough trajectory through molecular dynamics. We have a system of N atoms obeying classical Newtonian mechan-

ics. In such a system, we can represent the total energy E_{tot} as the sum of kinetic energy E_{kin} and potential energy V_{pot}:

$$E_{tot}(t) = E_{kin}(t) + V_{pot}(t)$$

where the potential energy is a function of the coordinates, $V_i = f(r_i)$ for atoms i to N and r_i represents cartesian coordinates of atom i; and the kinetic energy depends on the motion of the atoms:

$$E_{kin}(t) = \sum \frac{1}{2} M_i V_i^2(t)$$

where M_i is the mass of atom i and V_i is the velocity of atom i.

The energy undergoes constant redistribution because of the movements of the atoms, resulting in changes in their positions on the potential surface and in their velocities. At each iteration ($t \rightarrow t +$ At), an atom i moves to a new position $[r_i(t) \rightarrow r_i(t + \Delta t)]$, and it experiences a new set of forces. The basic assumption is that the time step At is sufficiently small that the position of atom i at $t +$ At *can be* linearly extrapolated from its velocity at time t and the acceleration resulting from the forces felt by atom i at time t. If At is long enough for the atoms surrounding atom i to change their position so that the forces felt by atom i will change during At, then the approximation is not valid and the simulation will deviate from that observed with a shorter At. After each atom is moved, the forces on the first atom based on the new positions of the other N – 1 atoms can be recalculated and a new iteration begun. Several algorithms exist for numerical integration. The ones by Verlet and Gear are in common use, with the one by Verlet being computationally more efficient *(126).*A variant of the Verlet algorithm in common use is called the leapfrog algorithm. The calculation of the velocity is done at $t - \Delta t/2$, whereas the calculation of the force occurs at t to derive the new velocity at $t = \Delta t/2$. In other words,

$$V_i(t + \Delta t/2) = V_i(t - \Delta t/2) + F_i(t)\Delta t/M_i.$$

The atomic position of atom i is calculated by adding the incremental change in position, $V_i(t + \Delta t/2)$. AT, to the original position $V_i(t)$. By staggering the evaluation of the velocity and force calculations by $\Delta t/2$, an improvement in the simulation performance is obtained.

2.1.6.2 Temperafure. For simulations that can be compared with experimental results, one must be able to control the temperature of the simulation. The temperature of a system is a function of the kinetic energy, $E_{kin}(t)$:

$$T(t) = E_{kin}(t)/\frac{3}{2}Nk$$

where k is Boltzmann's constant.

One can perform molecular dynamics simulations, at a constant temperature T_c, by scaling all atomic velocities $V_i(t)$ at each step by a factor t derived from

$$\delta T(t)/\delta t = [T, - T(t)]/t$$

where T_c is the desired temperature.

2.1.6.3 Pressure and Volume. Depending on the simulation that one desires to accomplish, either the pressure or volume must be maintained constant. Constant volume is the easiest to perform because the boundaries of the system are maintained with all molecules confined within those boundaries and the pressure allowed to change during the simulation.

2.1.7 Monte Carlo Simulations. The Monte Carlo method (126) is based on statistical mechanics and generates sufficient different configurations of a system by computer simulation to allow the desired structural, statistical, and thermodynamic properties to be calculated as a weighted average of these properties over these configurations. The average value $\langle X \rangle$ of the property X can be calculated by the following formula:

$$<X> = \sum_{i=1}^{N} X_i \exp(-E_i/kT)V'$$

$$\sum_{i=1}^{N} \exp(-E_i/kT)$$

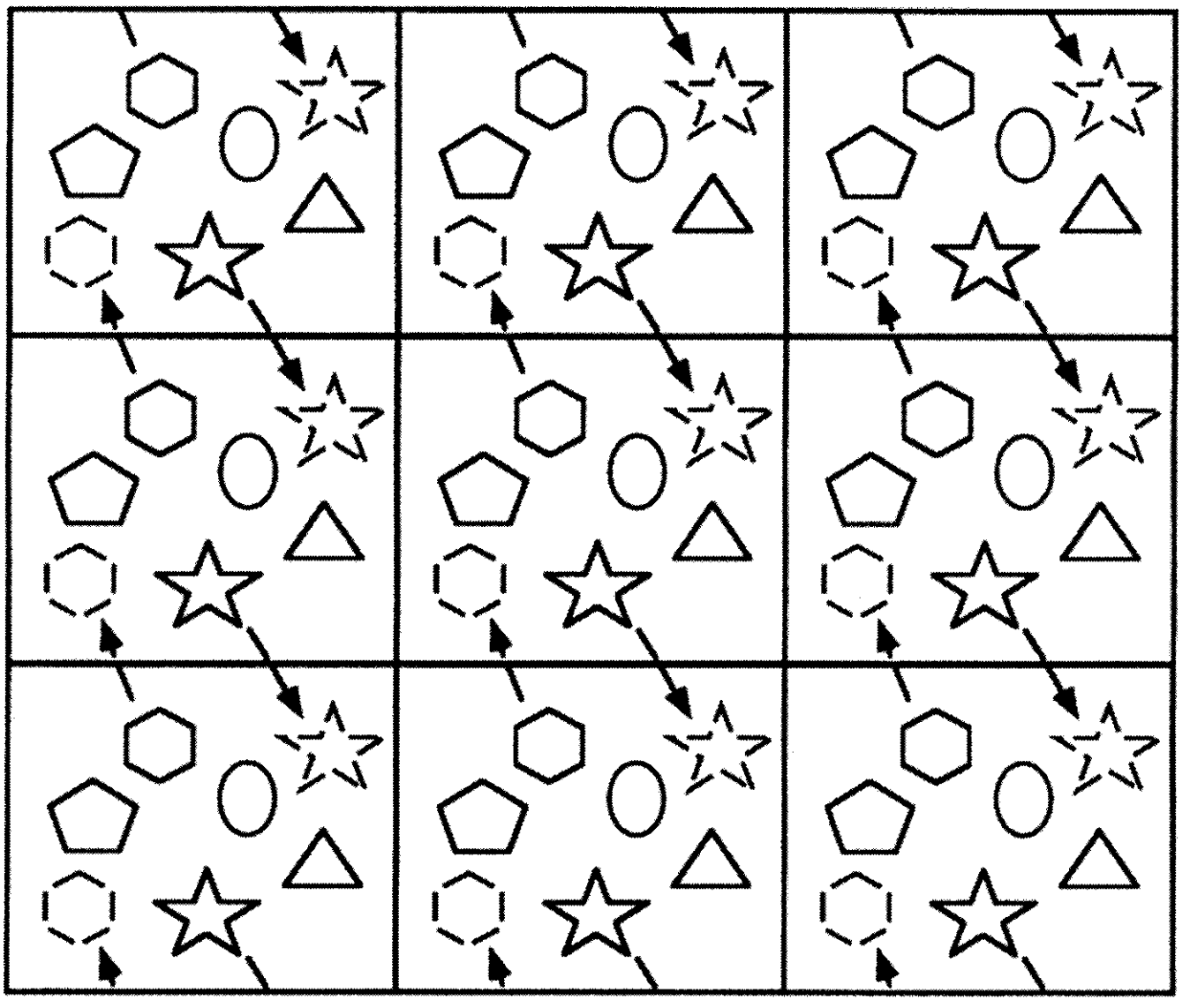


Figure 3.9. Schematic diagram of simulation with periodic boundary conditions in which adjacent cells are generated by simple translations of coordinates.

where N is the number of configurations, E_i is the energy of configuration i, k is Boltzmann's constant, and T is temperature.

If we have sufficiently sampled the possible arrangements of molecules in the simulation and have an accurate method to calculate their energy E, then the above formula will give a Boltzmann weighted average of the property X.

In practice, one must compromise the number of molecules in the simulation and/or the number of configurations calculated to conserve computer cycles. Two essential techniques that are utilized are periodic boundary conditions and sampling algorithms, which we discuss separately.

Although it is important to minimize the number of molecules in either Monte Carlo or molecular dynamics simulations for computational convenience, surface effects at the interface between the simulated solvent and the surrounding vacuum could seriously distort the results. To approximate an "infinite" liquid, one can surround the box of molecules by simple translations to generate periodic images. Each atom in the central box has a set of related molecules in the virtual boxes surrounding the central one (Fig. **3.9).** The energy calculations for pairwise interactions consider only the interaction of a molecule, or its "ghost," with any other molecule, but not both. In practice, this is accomplished by limiting pairwise interactions to distances less than one-half the length of the side of the box. Real concerns often arise regarding convergence of electrostatic terms because of the linear dependency on distance.

For any large nontrivial system, the total number of possible configurations is beyond comprehension. Consider a set of protons in a magnetic field: the magnetic moments can be either aligned with or opposed to the magnetic field. For only 50 protons, there are 2^{50} combinations, which is a large number. For a

small cyclic pentapeptide, there are potentially 36^{10} conformations if one considers a 10° scan of the torsional variables Φ, Ψ. Clearly, some of these are energetically unreasonable because the conformation requires overlap of two or more atoms in the structure. Monte Carlo simulations are successfully performed by sampling only a limited set of the energetically feasible conformations, say, 10^6 out of 10^{100} theoretical possibilities. The reason for this success is that the Monte Carlo schemes sample those states that are statistically most important. One could sample all states, calculate the energy of each, and then Boltzmann-weight its contribution to the average. Alternatively, one can ignore those states that are energetically high so that they contribute little, if any, weight to the average, and concentrate on those of low energy. In other words, we look only where there are reasonable answers energetically. This is called ***importance sampling,*** which is the key to the Monte Carlo procedure.

One aspect shared by Monte Carlo methods and molecular dynamics is the ability to cross barriers. In the case of Monte Carlo, barrier crossing occurs both by random selection of variables and by acceptance of higher energy states on occasion. Both methods require an equilibration period to eliminate bias associated with the starting configuration. When one considers randomly filling a box with molecules with arbitrary choices for position and orientation, it should be obvious that most examples would result in high energy, especially if the density of such a simulation is made to resemble that of a liquid in which adjacent molecules are often in VDW contact. High energy configurations contribute very little to the properties we are trying to evaluate because they are Boltzmann weighted. It is, therefore, extremely inefficient to randomly calculate configurations. One needs procedures, often referred to as importance sampling, that selectively calculate configurations that will be representative of allowed states. In fact, if one can guarantee that the energy of the configurations actually has a Boltzmann distribution, then one can simply average the properties. In practice, this has been accomplished by an algorithm suggested by Metropolis et al. **(127).** One essentially uses a Markov process in which the current configuration becomes the basis for generating the next.

1. A molecule in the current configuration is chosen at random and its degrees of freedom randomly varied by small increments.
2. The energy of the new configuration is evaluated and compared with that of the starting configuration.
3. If the new energy is lower, the new configuration is accepted and becomes the basis for the next random perturbation.
4. If the energy is higher, $E(\text{new}) > E(\text{old})$, then a random number between 0 and 1 is generated and compared with $\exp\{-[E(\text{new}) - E(\text{old})]\}/kT$. If the number is less, then the configuration is accepted and the process continues by generating a new configuration. If the number is greater, then the configuration is rejected and the process resumes with the old configuration.

In this way, configurations of lower energy are accepted and the system eventually "minimizes" to sample the higher populated lower energy configurations; at the same time, higher energy configurations are included but only in proportion to their Boltzmann distribution, which is clearly a function of temperature of the simulation. Because the configurations occur with a probability depending on their energy and proportional to the Boltzmann distribution, one can simply average thermodynamic properties over this distribution of configurations,

$$\langle X \rangle = 1/N \sum_{i=1}^{N} X_i$$

where the sum covers the N configurations generated. Because one often does not know an appropriate starting configuration, the initial part of the run may be used to "minimize," or equilibrate the system, and only

中国药科大学
图书馆藏书

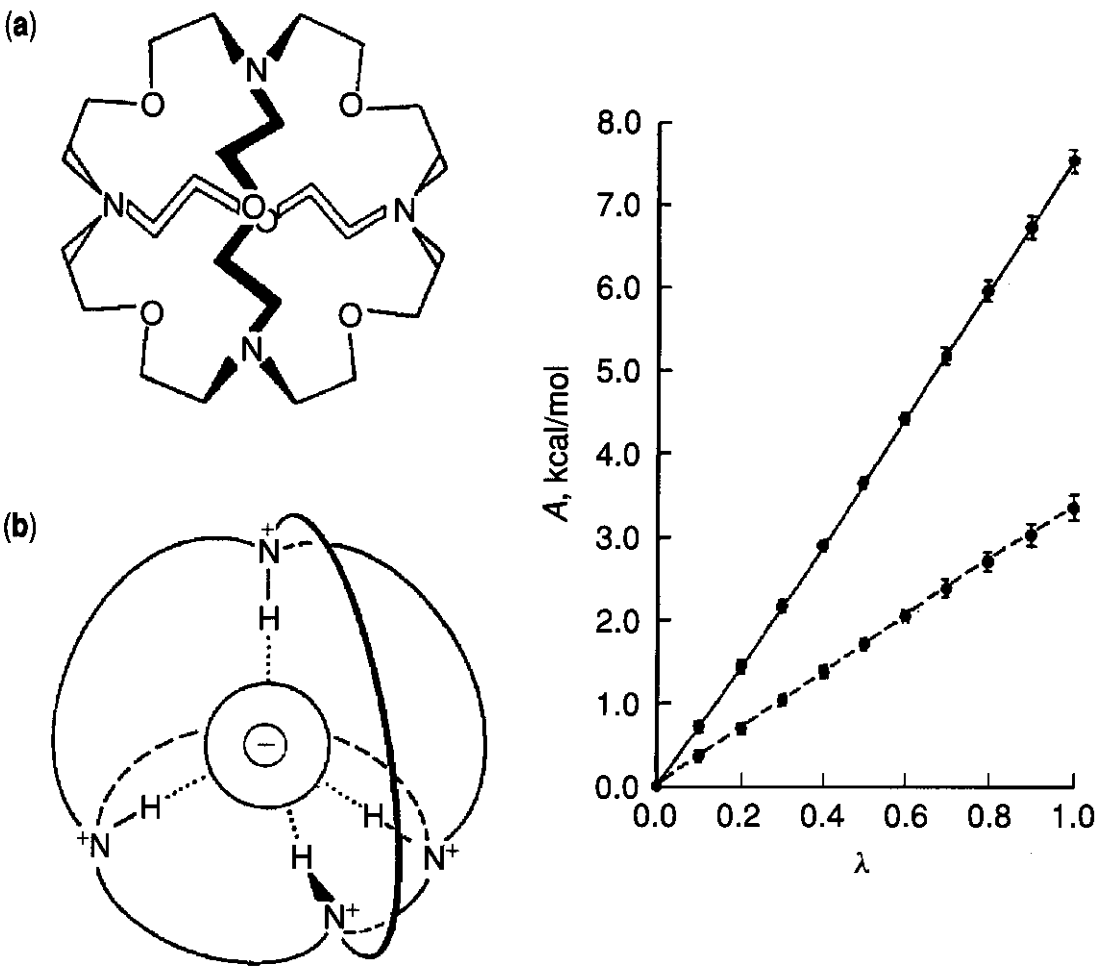


Figure 3.10. Estimation of difference in affinity ($\Delta\Delta G$) of the two anions Cl^- and Br^- for the cryptand SC24 [(a) structural formula; (b) schematic of complex formed with halide ion] as the parameters for Cl^- are slowly mutated into those for Br– in water (- - -) as well as in the complex (—). Used with permission (138).

the latter part of the simulation analyzed once the configurational energy has stabilized.

A useful application has combined Monte Carlo sampling with variable temperatures (simulated annealing) to encourage barrier crossing to optimize the docking of ligands into active sites. Random displacements of rigid body translation and rotation (6 degrees of freedom) and of internal torsional rotations in a substrate within the binding site cavity were performed with Metropolis sampling and a temperature program. This procedure reproduced the crystallographically observed structure of the complex for several test cases (131).

2.1.8 Thermodynamic Cycle Integration (132–134). Thermodynamic cycle integration is an approach that allows calculation of the free energy difference between two states. In this method, one takes advantage of the state-function nature of a thermodynamic cycle and eliminates the paths of the simulation with long time constants (e.g., formation of a complex requiring diffusion). As an example, the difference in affinity of two ligands (L and M) for the same enzyme or receptor R is described by the following thermodynamic cycle:

$$\begin{array}{ccc} R + L & \xrightarrow{\Delta A1} & RL \\ \Delta A3 \downarrow & & \Delta A4 \uparrow \\ R + M & \xrightarrow{\Delta A2} & RM \end{array}$$

Because the thermodynamic values of the two states do not depend on the path between the states, one can write the following equation:

$$\Delta\Delta A = \text{difference in affinity of L and M for R}$$
$$= \Delta A2 - \Delta A1 = \Delta A4 - \Delta A3$$

By simulating the mutation of L into M, paths A3 and A4, one can avoid the long simulation required for diffusion of the ligands, paths A1 and A2, into the receptor. One simply incrementally modifies the potential functions representing ligand L to those representing ligand M during the course of the simulation, making sure that the perturbations are introduced gradually and that the surrounding atoms have time to relax from the perturbation (Fig. 3.10). Either Monte Carlo (135) or molecular dynamics simulations can utilize this technique. Many interesting applications have

appeared in the literature (132,134,136,137). Their success appears directly related to sampling problems and minimal perturbation of the ligand to ensure equilibration.

2.1.9 Non-Boltzmann Sampling. There are equivalent molecular dynamics and Monte Carlo procedures that allow one to sample regions of configuration space that are not minima, transition states, for example. One can generate a Monte Carlo trajectory for a system E_V that has energetics similar to that of the Boltzmann system E„ with sampling in the region associated with a transition barrier by subtracting a potential V, to reduce the barrier:

$$E_0 = E_V - V.$$

Alternatively, one may want to obtain meaningful statistics for a rare event without oversampling the lower energy states. This can be accomplished by adding a potential W, which is zero for the the interesting class of configurations and very large for all others (Fig. 3.11):

$$E_0 = E_V + W$$

The details of these sampling procedures that allow one to focus on the aspect of the problem of interest are the subject of a review by Beveridge (133). Application of this approach to determining conformational transitions in model peptides (137,139,140) are exemplified in the work of Elber's group on helix-coil (85, 86, 141), the Brooks group on turn-coil (142–146), and Huston and Marshall and Smythe et al. (147,148) on helical transitions in peptides.

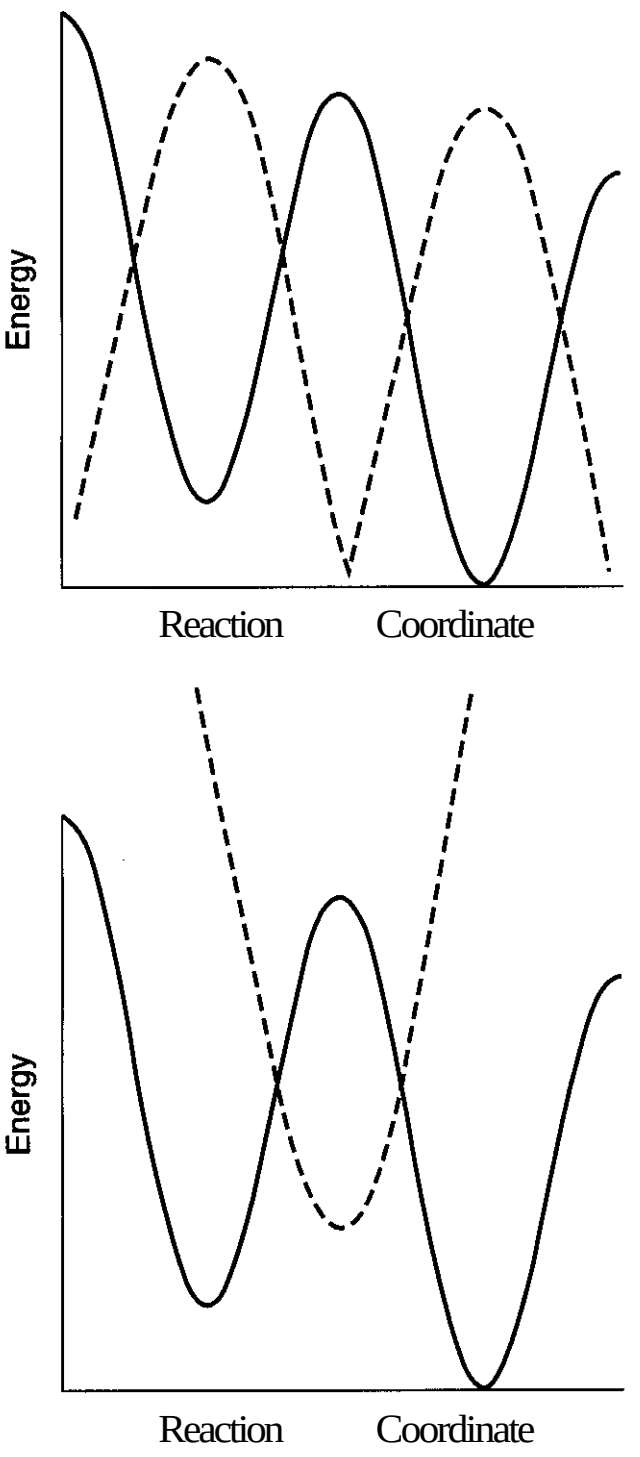


Figure 3.11. Schematic diagrams of methods for modifying the potential surface to allow adequate sampling during simulations.

2.2 Quantum Mechanics: Applications in Molecular Mechanics

Detailed discussion of quantum mechanics (149) is clearly beyond the scope of this review, and its applications to molecular mechanics and modeling will be briefly summarized. Molecular mechanics is based on the laws of classical physics and deals with electronic interactions by highly simplified approximations such as Coulomb's law. All forces operating in intermolecular interactions are essentially electronic in nature. Any effort to quantitate those forces requires detailed information about the nuclear positions and the electron distribution of the molecules involved. At considerable computational cost, quantum mechanics provides information about both nuclear position and electronic distribution. Molecular mechanics is built on the assumption that electronic interactions can be adequately accounted for by parameterization. Although most of the systems of interest in biology are too large for the direct application of quantum mechanics, quantum mechanics has at least three essential roles to play in drug design (149): (1) charge approximations, (2)

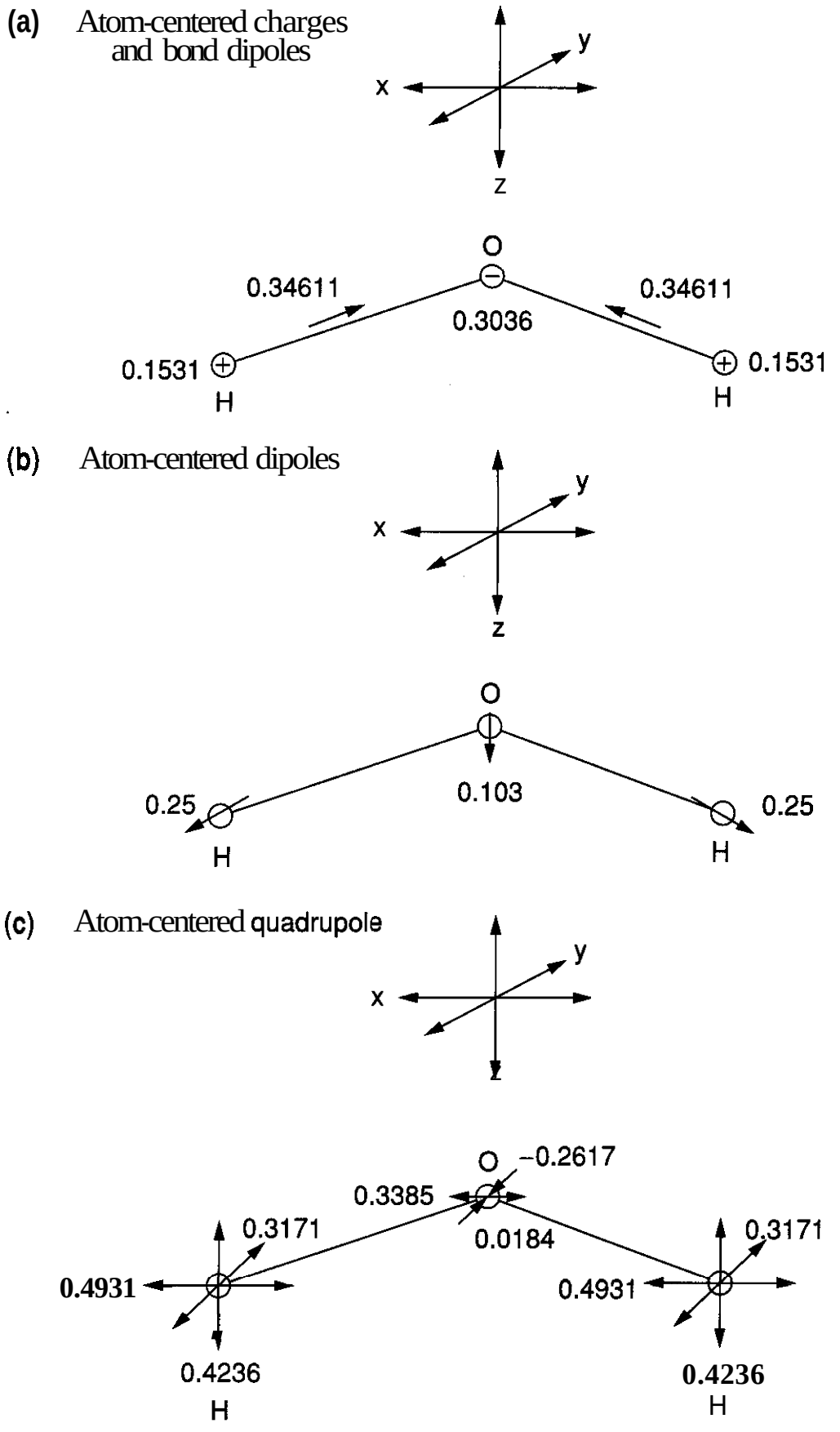


Figure 3.12. Different approaches to localization of charge used in electrostatic models. (a) Atom-centered monopole; (b) atom-centered dipole; and (c) atom-centered quadrapole.

characterization of molecular electrostatic potentials, and (3) parameter development for molecular mechanics.

2.2.1 Parameterization of Charge. Estimates of charges in molecular mechanics can be derived, in general, by application of one of the many different quantum chemical approaches, either *ab initio* or semiempirically. Quantum mechanical methods are available for calculating the electron probability distributions for all the electrons in a molecule and then partitioning those distributions to yield representations for the net atomic charges of atoms in the molecule, either as atom-centered charges or as more complex distributed multipole models (39, 42) (Fig. 3.12).

2.2.1.1 ***Atom-Centered Point Charges.*** In the Mulliken population analysis, all the one-center charge on an atom is assigned to that atom, whereas the two-center charge is divided equally between the two atoms in the

overlap (even if the electronegativities of the two atoms are quite dissimilar). The sum is the gross atomic population, and the net atomic charge is simply this plus the nuclear charge. The result is very sensitive to the basis set (the number of atomic orbitals) used. Despite poor fit of the molecular electrostatic potential derived with point charges to the *ab initio* electrostatic potential, or that derived from a distributed multipole analysis (150), widespread use continues because they do reflect chemical trends and are reportedly compatible with known electronegativities. In addition, this option is commonly available in software packages. Unfortunately, poor representation of the electric field surrounding the molecule results from use of atom-centered monopole models (42), even when more careful methods are used to distribute the charge.

2.2.1.2 Methods to Reproduce the Molecular Electrostatic Potential (MEP). The electrostatic potential surrounding the molecule that is created by the nuclear and electronic charge distribution of the molecule is a dominant feature in molecular recognition. Williams reviews (42) methods to calculate charge models to accurately represent the MEP as calculated by *ab initio* methods by use of large basis sets. The choice between models (monopole, dipole, quadrapole, bond dipole, etc., Fig. 3.12) depends on the accuracy with which one desires to reproduce the MEP. This desire has to be balanced by the increased complexity of the model and its resulting computational costs when implemented in molecular mechanics.

The first problem is to select points where the MEP is to be evaluated and eventually fitted, the position of the shell outside the VDW radii of the atoms in the molecule, and the spacing of grid points on that shell. Sampling too close to the nuclei gives rise to anomalies because the potential around nuclei is always positive. Singh and Kollman (151) report the use of four surfaces at 1.4, 1.6, 1.8, and 2.0 times the VDW radii, with a density of one to five points per Å^2. This paradigm was reported to give an adequate sampling to which the fitted charges were fairly insensitive, at least at the higher values. An improved procedure, the restrained electrostatic potential fit (RESP), was developed by Bayly et al. (41) to enhance transferability of the resulting point charges. Williams (42) derived a procedure to derive the best fit to a given MEP with a defined set of monopoles, dipoles, and so forth.

Typically, fragments of molecules of interest are analyzed by *ab initio* techniques to generate their MEPs that are the reference for parameterization of charge. Besler et al. (152) reported fitting of atomic charges to the electrostatic potentials calculated by the semiempirical methods AM1 and MINDO. The MINDO charges derived by fitting the MEP can be linearly scaled to agree with results derived from *ab initio* calculations. Among the motivations for semiempirical methods are the facts that semiempirical methods using high quality basis sets often yield better results than *ab initio* techniques employing minimal basis sets, and the significant reduction in computational time in moving from *ab initio* to semiempirical calculations. Rauhut and Clark (153) used the AM1 wave function to develop a multicenter point-charge model in which each hybrid natural atomic orbital is represented by two charges located at the centroid of each lobe. Thus, up to nine charges (4 orbitals and 1 core charge) are used to represent heavy atoms. Results using this approach affirm the observations that distributed charges are more successful than atom-centered charges in reproducing intermolecular interactions (154, 155).

2.2.2 Parameter Derivation for Force Fields.

Because molecular mechanics is empirical, parameters are derived by iterative evaluation of computational results, such as molecular geometry (bond lengths, bond angles, dihedrals) and heats of formation, compared with experimental values (20). Lifson has coined the expression "consistent" for force fields in which structures, energies of formation, and vibrational spectra have all been used in parameterization by least-squares optimization. In the case of bond lengths, bond angles, and VDW parameters, crystallography has provided most of the essential experimental database. Major efforts (156) to derive general sets of parameters from quantum mechanical calculation have been made, especially for systems for which adequate experimental data are unavailable. Although quantum mechanics is certainly adequate for initial approxima-

tions of parameters and essential for charge approximations, a detailed analysis indicates that *in vacuo* calculations neglect many-body effects and can be misleading. A major effort by Hehre (personal communication) to derive parameters for water from extensive *ab initio* calculations with large basis sets failed even to give a parameter set that reproduced the radial distribution for bulk water. Parameters derived from relevant experimental data in condensed phase (especiallyif available in the solvent of theoretical interest) are generally more capable of accurately predicting results because the many-body effects are implicitly included in the parameterization. The basic assumption is that these "effective" two-body potentials implicitly incorporate many-body interaction energies.

Jorgensen has parameterized by fitting properties of bulk liquids to Monte Carlo simulations to give the AMBER/OPLS force field (26, 157, 158). Conceptually, one is attracted by the use of liquids and their observable properties as constraints during the derivation of a force field that is destined to study the properties of solvated molecules.

2.2.3 Modeling Chemical Reactions and Design of Transition-State Inhibitors. In cases, such as enzyme reactions, where chemical transformations occur, quantum chemical methods must be used to deal with electronic changes in hybridization and bond cleavage (159, 160). Hybrid applications (161–163) in which the reaction core is modeled quantum mechanically and the rest by molecular mechanics would appear a viable option. Alternatively, the geometry of the transition state has been modeled by molecular mechanics, with force constants derived from *ab initio* calculations that predict with amazing accuracy the relative selectivity of reactions. Andrews and coworkers (164) pioneered modeling of transition states (165) of enzymatic reactions to design transition-state inhibitors.

3 KNOWN RECEPTORS

A significant challenge is the design of novel ligands for therapeutic targets in which the three-dimensional structure has been determined by either X-ray crystallography or NMR (12, 13, 166). The availability of the coordinates of all the atoms of the target suggests use of modeling of the site and interaction with prospective ligands. Qualitative information can be discerned by simple examination of complexes by the use of molecular graphics and improvement of known ligands made by searching for accessory binding interactions through ligand modification. This approach was pioneered by groups at Wellcome Research Laboratories (167–169) in designing analogs of 2,3-diphosphorylglycerate (Fig. 3.13), to modulate oxygen binding to hemoglobin, and at Burroughs-Wellcome (170), to enhance affinity of dihydrofolate reductase (DHFR) antagonists. When used in an iterative fashion, novel compounds with improved affinity result (166, 171, 172). Quantification of interactions and design of novel ligands require application of molecular and statistical mechanics to quantify the enthalpy and entropy of binding. In other words, experimental measurements reflect free energies of binding and both enthalpic and entropic contributions must be estimated for prediction of affinities as part of the design process. When combined with combinatorial chemistry and high throughput screening, rapid identification of therapeutic candidates is feasible, as witnessed in the case of factor Xa antagonists (173) or TAR RNA inhibitors as possible HIV drugs (174).

3.1 Definition of Site

The availability of three-dimensional structural information on a potential therapeutic target does not guarantee identification of the site of action of the substrate, or inhibitor, unless the structure of a relevant complex has been determined. In fact, conformational changes often occur during binding of ligands to enzymes that are not reflected in the three-dimensional structure of the enzyme alone. Illustrative examples are the major conformational changes seen (175, 176) in HIV protease on binding the inhibitor MVT-101 (Fig. 3.14) and the changes in domain orientation observed (177) in the complex of an anti-HIV peptide antibody with the peptide. Until the two β-strand flaps have been folded in, to complete the active site of HIV protease, many of

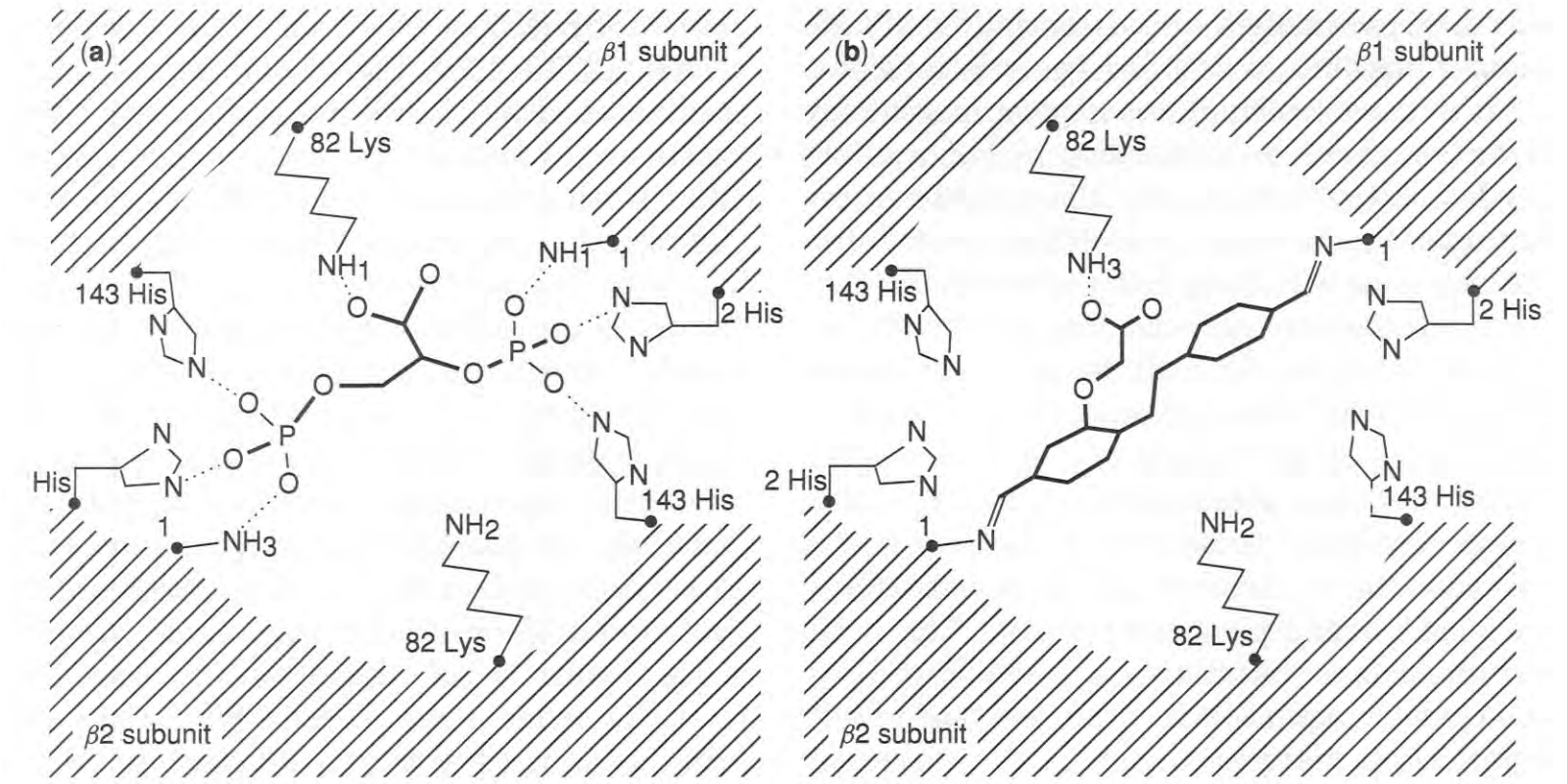


Figure 3.13. Diphosphoglycerate (a) and analogs (b-d) designed to optimize interactions bound in schematic model of hemoglobin. Used with permission (169).

the important interactions for recognition in this proteolytic system have not been defined. In other cases of therapeutic targets, allosteric sites are involved in regulation of binding and cannot clearly be discerned from the crystal structure available. Here NMR offers a highly complementary approach where transfer and isotope-edited NOEs as well as magic angle spinning NMR on solid samples can help identify those residues of the therapeutic target (Fig. 3.15) involved in receptor interaction (178–180).

One significant concern of structure-based design is the dynamics of the target itself. How stable is the active site to modifications in the ligand? Are there alternative potential binding sites that could compete for the ligand? The geometrical identity of serine protease catalytic residues, for example, argues that the specificity essential for biological utility ensures a relatively rigid three-dimensional arrangement of functionality in the active site that determines molecular recognition and

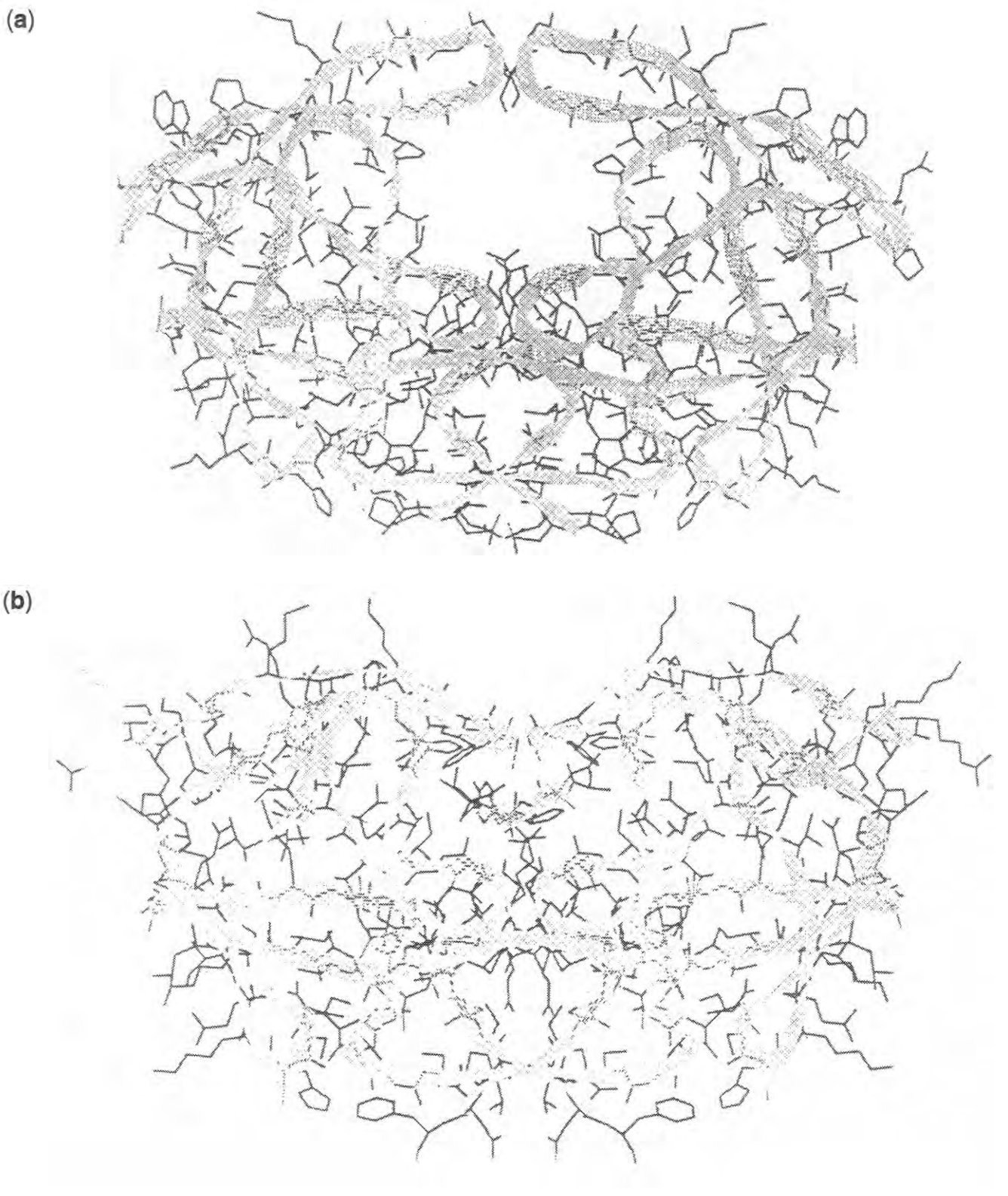


Figure 3.14. Ribbon diagram of HIV-1 protease in the absence of inhibitor (a) and when bound to the inhibitor MVT-101 (b). Diagrams based on crystal structures as reported by Miller et al. (175,176).

crimination. The active site has had no evolutionary pressure to optimize binding *per se*, but rather rates of interaction and discrimination among the limited repertoire of the biological milieu. One classic example (181) of difficulty in interpretation of binding as a result of ligand modification occurred when an analog designed to bind to a specific site on hemoglobin actually found a more appropriate site within the packed side chains of the protein molecule (Fig. **3.16).** This example emphasizes the importance of protein dynamics. Alternate conformations of the protein that are easily accessible at room temperature may be difficult to characterize experimentally because of relatively low abundance and/or lack of resolution of the experimental techniques used. Computationally, they are problematic as well because of the complexity of the energy surface for a macromolecule.

3.2 Characterization of Site

3.2.1 Volume and Shape. Most substrate-enzyme or receptor-ligand interactions occur within pockets, or cavities, buried within pro-

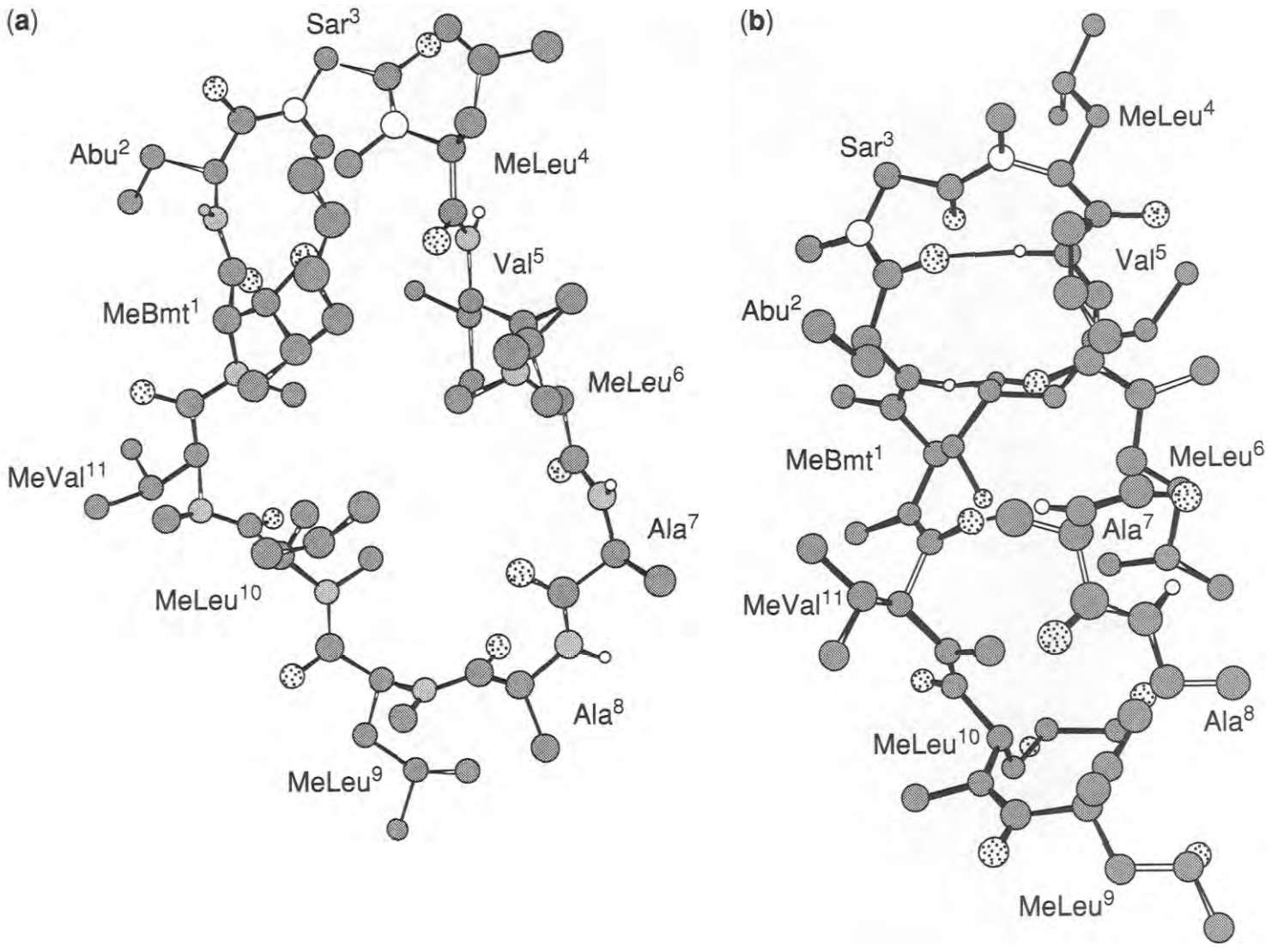


Figure 3.15. Bound conformation of cyclosporin (a) as determined by NMR compared with solution conformation (b) (178). Residues involved with interaction with cyclophilin are indicated on (a) in bold.

teins. Inside these invaginations, a *microenvironment* is established that favors desolvation and binding of the ligand, despite the entropic cost of fixing the relative geometries of the two molecules. Knowledge of the three-dimensional structure of such cavities can assist the study of binding interactions and the design of novel ligands as potential therapeutics. Several algorithms to find, display, and characterize cavity-like regions of proteins as potential binding sites have been developed. Kuntz et al. (13, 183) described a program, DOCK, to explore the steric complementarity between ligands and receptors of known three-dimensional structure. Using the molecular surface of a receptor, a volumetric representation of the chosen binding cavity is approximated by use of a set of spheres of various sizes that have been mathematically "packed" within it (Fig. 3.17). The set of distances between the centers of the spheres serves as a compact representation of the shape of the cavity. The use of the relative distance paradigm allows comparison without the need for orientation of one shape with respect to the other. Potential ligands are characterized in a similar fashion by generating a set of spheres that mimic the shape of the ligand. Matching the distance matrix of the cavity with that of a potential ligand provides an efficient screen for selection of complementary shapes. Voorintholt et al. (184) used three-dimensional lattices to calculate density maps of proteins. In these maps, lattice points were assigned as a function of the distance to the nearest atom. This technique is effective in delineating regions of low density where channels and cavities exist. Ho and Marshall (185) implemented a search function in CAVITY to allow the investigator to isolate a single cavity of interest by specifying a seed point. From this seed point, the algorithm systematically explored the entire volume of the cavity, following its borders and effectively filling every crevice within it; that

tions of parameters and essential for charge approximations, a detailed analysis indicates that *in vacuo* calculations neglect many-body effects and can be misleading. A major effort by Hehre (personal communication) to derive parameters for water from extensive *ab initio* calculations with large basis sets failed even to give a parameter set that reproduced the radial distribution for bulk water. Parameters derived from relevant experimental data in condensed phase (especiallyif available in the solvent of theoretical interest) are generally more capable of accurately predicting results because the many-body effects are implicitly included in the parameterization. The basic assumption is that these "effective" two-body potentials implicitly incorporate many-body interaction energies.

Jorgensen has parameterized by fitting properties of bulk liquids to Monte Carlo simulations to give the AMBER/OPLS force field (26, 157, 158). Conceptually, one is attracted by the use of liquids and their observable properties as constraints during the derivation of a force field that is destined to study the properties of solvated molecules.

2.2.3 Modeling Chemical Reactions and Design of Transition-State Inhibitors. In cases, such as enzyme reactions, where chemical transformations occur, quantum chemical methods must be used to deal with electronic changes in hybridization and bond cleavage (159, 160). Hybrid applications (161–163) in which the reaction core is modeled quantum mechanically and the rest by molecular mechanics would appear a viable option. Alternatively, the geometry of the transition state has been modeled by molecular mechanics, with force constants derived from *ab initio* calculations that predict with amazing accuracy the relative selectivity of reactions. Andrews and coworkers (164) pioneered modeling of transition states (165) of enzymatic reactions to design transition-state inhibitors.

3 KNOWN RECEPTORS

A significant challenge is the design of novel ligands for therapeutic targets in which the three-dimensional structure has been determined by either X-ray crystallography or NMR (12, 13, 166). The availability of the coordinates of all the atoms of the target suggests use of modeling of the site and interaction with prospective ligands. Qualitative information can be discerned by simple examination of complexes by the use of molecular graphics and improvement of known ligands made by searching for accessory binding interactions through ligand modification. This approach was pioneered by groups at Wellcome Research Laboratories (167–169) in designing analogs of 2,3-diphosphorylglycerate (Fig. 3.13), to modulate oxygen binding to hemoglobin, and at Burroughs-Wellcome (170), to enhance affinity of dihydrofolate reductase (DHFR) antagonists. When used in an iterative fashion, novel compounds with improved affinity result (166, 171, 172). Quantification of interactions and design of novel ligands require application of molecular and statistical mechanics to quantify the enthalpy and entropy of binding. In other words, experimental measurements reflect free energies of binding and both enthalpic and entropic contributions must be estimated for prediction of affinities as part of the design process. When combined with combinatorial chemistry and high throughput screening, rapid identification of therapeutic candidates is feasible, as witnessed in the case of factor Xa antagonists (173) or TAR RNA inhibitors as possible HIV drugs (174).

3.1 Definition of Site

The availability of three-dimensional structural information on a potential therapeutic target does not guarantee identification of the site of action of the substrate, or inhibitor, unless the structure of a relevant complex has been determined. In fact, conformational changes often occur during binding of ligands to enzymes that are not reflected in the three-dimensional structure of the enzyme alone. Illustrative examples are the major conformational changes seen (175,176) in HIV protease on binding the inhibitor MVT-101 (Fig. 3.14) and the changes in domain orientation observed (177) in the complex of an anti-HIV peptide antibody with the peptide. Until the two β-strand flaps have been folded in, to complete the active site of HIV protease, many of

overlap (even if the electronegativities of the two atoms are quite dissimilar). The sum is the gross atomic population, and the net atomic charge is simply this plus the nuclear charge. The result is very sensitive to the basis set (the number of atomic orbitals) used. Despite poor fit of the molecular electrostatic potential derived with point charges to the *ab initio* electrostatic potential, or that derived from a distributed multipole analysis (150), widespread use continues because they do reflect chemical trends and are reportedly compatible with known electronegativities. In addition, this option is commonly available in software packages. Unfortunately, poor representation of the electric field surrounding the molecule results from use of atom-centered monopole models (42), even when more careful methods are used to distribute the charge.

2.2.1.2 Methods to Reproduce the Molecular Electrostatic Potential (MEP). The electrostatic potential surrounding the molecule that is created by the nuclear and electronic charge distribution of the molecule is a dominant feature in molecular recognition. Williams reviews (42) methods to calculate charge models to accurately represent the MEP as calculated by *ab initio* methods by use of large basis sets. The choice between models (monopole, dipole, quadrapole, bond dipole, etc., Fig. 3.12) depends on the accuracy with which one desires to reproduce the MEP. This desire has to be balanced by the increased complexity of the model and its resulting computational costs when implemented in molecular mechanics.

The first problem is to select points where the MEP is to be evaluated and eventually fitted, the position of the shell outside the VDW radii of the atoms in the molecule, and the spacing of grid points on that shell. Sampling too close to the nuclei gives rise to anomalies because the potential around nuclei is always positive. Singh and Kollman (151) report the use of four surfaces at 1.4, 1.6, 1.8, and 2.0 times the VDW radii, with a density of one to five points per Å^2. This paradigm was reported to give an adequate sampling to which the fitted charges were fairly insensitive, at least at the higher values. An improved procedure, the restrained electrostatic potential fit (RESP), was developed by Bayly et al. (41) to enhance transferability of the resulting point charges. Williams (42) derived a procedure to derive the best fit to a given MEP with a defined set of monopoles, dipoles, and so forth.

Typically, fragments of molecules of interest are analyzed by *ab initio* techniques to generate their MEPs that are the reference for parameterization of charge. Besler et al. (152) reported fitting of atomic charges to the electrostatic potentials calculated by the semiempirical methods AM1 and MINDO. The MINDO charges derived by fitting the MEP can be linearly scaled to agree with results derived from *ab initio* calculations. Among the motivations for semiempirical methods are the facts that semiempirical methods using high quality basis sets often yield better results than *ab initio* techniques employing minimal basis sets, and the significant reduction in computational time in moving from *ab initio* to semiempirical calculations. Rauhut and Clark (153) used the AM1 wave function to develop a multicenter point-charge model in which each hybrid natural atomic orbital is represented by two charges located at the centroid of each lobe. Thus, up to nine charges (4 orbitals and 1 core charge) are used to represent heavy atoms. Results using this approach affirm the observations that distributed charges are more successful than atom-centered charges in reproducing intermolecular interactions (154, 155).

2.2.2 Parameter Derivation for Force Fields.

Because molecular mechanics is empirical, parameters are derived by iterative evaluation of computational results, such as molecular geometry (bond lengths, bond angles, dihedrals) and heats of formation, compared with experimental values (20). Lifson has coined the expression "consistent" for force fields in which structures, energies of formation, and vibrational spectra have all been used in parameterization by least-squares optimization. In the case of bond lengths, bond angles, and VDW parameters, crystallography has provided most of the essential experimental database. Major efforts (156) to derive general sets of parameters from quantum mechanical calculation have been made, especially for systems for which adequate experimental data are unavailable. Although quantum mechanics is certainly adequate for initial approxima-

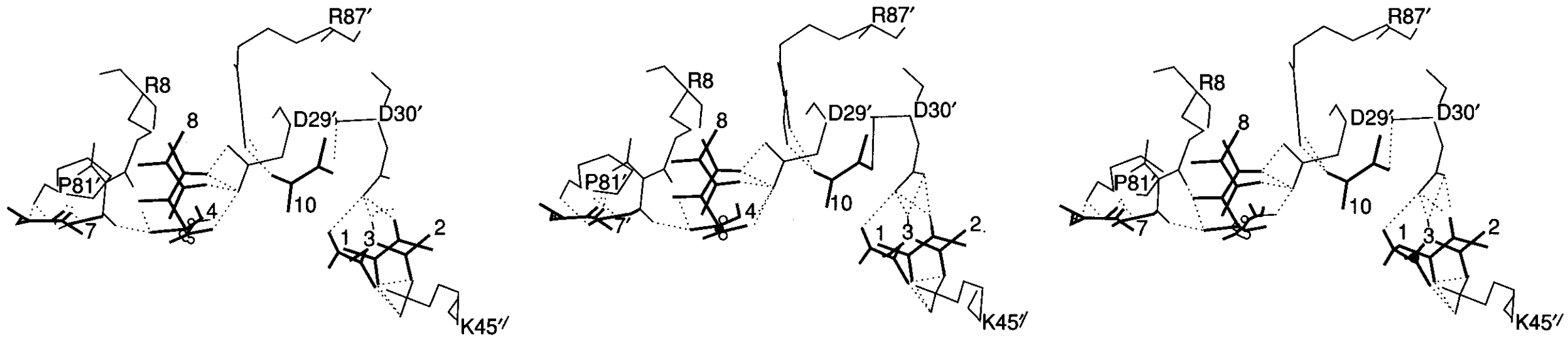


Figure 3.19. Stereoview of 10 minima of *N*-methyl-acetamide bound to open end (S4′) of the active site region of HIV protease by simulation with multiple fragments (189, 190). Left pair for cross-eyed and right pair for wall-eyed viewing. Used with permission.

and receptor (185). At every cavity-pocket interface point, the electrostatic potential of both the atoms forming the cavity and those of the binding ligand are calculated. A rough approximation of complementarity is computed by multiplying these potentials together. A favorable electrostatic interaction is produced when the electrostatic potentials are opposite in sign. Therefore, favorable interactions are indicated when the product of these values is a negative number. Likewise, unfavorable interactions are indicated when the product of these values is a positive number and the potential of the cavity and that of the binding ligand have the same sign. These products are then normalized, assigned a color, and displayed.

In a similar way, an estimate of the hydrophobic character of a segment of the surface can be quantitated and indicated through color coding. The ability to rapidly switch between these hydrophobic and electrostatic surface representations, to visually integrate the optimal complementarity between site and potential ligand to be designed, is helpful.

3.3 Design of Ligands

3.3.1 Visually Assisted Design. In the process of optimization of a lead, one needs to ascertain where modification is feasible. Although visualization of the excess space available in the active-site cavity by directly examining ligands is useful for locating selected regions where ligand modifications may be made, it is not well suited for fully characterizing the void that exists between the ligand and the receptor, the ligand-receptor gap region; information concerning the relative dimensions of free space is difficult to discern. To facilitate the display of this information, Ho and Marshall (185) developed another algorithm to color-code the cavity display by the ligand-receptor nearest atom gap distance. The actual VDW, surface-to-surface distance (not center to center) between the ligand and enzyme atoms is calculated. When the ligand-receptor distances have been calculated at all cavity-pocket interface lattice points, a user-defined color-coding scale is implemented to generate the displays. This highlights those areas that are less well packed and available for ligand modification.

3.3.2 Three-Dimensional Databases. Medicinal chemists have recognized the potential of searching three-dimensional chemical databases to aid in the process of designing drugs for known, or hypothetical, receptor sites. Several databases are well known, such as the Cambridge Crystallographic Database (194) (CSD). The crystal coordinates of proteins and other large macromolecules are deposited into the Brookhaven Protein Databank (195). The conformations present in crystallographic databases reflect low energy conformers that should be readily attainable in solution and in the receptor complex. The three-dimensional orientation of the key regions of the drug that are crucial for molecular recognition and binding are termed the pharmacophore. The investigator searches the three-dimensional database through a query for fragments that contain the pharmacophoric functional groups in the proper three-dimensional orientation. Using these fragments as "building blocks," completely novel structures may be constructed through assembly and pruning (196). Receptor sites are complex both in geometrical features and in their potential energy fields, and many diverse compounds can bind to the same protein by occupying various combinations of subsites. Noncrystallographic databases have been developed as well. One example is the three-dimensional database of structures from Chemical Abstracts generated through CONCORD (197–199) that contains over 700,000 entries. The use of such databases is most applicable when the binding of a particular ligand and its receptor is well understood in terms of functional group recognition, and a crystal structure of the complex is known (200). One approach to ligand design is to develop novel chemical architectures (i.e., scaffolds) that position the pharmacophoric groups, or their bioisosteres, in the correct three-dimensional arrangement.

Gund conceived the first prototypic program designed to search for molecules that match three-dimensional pharmacophoric patterns (201, 202). This program, MOLPAT, performed atom-by-atom searches to verify comparable interatomic distances between

pattern and candidate structures. Although rigorous, this approach was tedious and required optimization. Lesk (203) devised a method that used the geometric attributes of the query to screen potential candidates. Similarly, Jakes and Willett (204) proposed that screens based on interatomic distances and atom types could considerably augment search efficiency. Furthermore, Jakes et al. (205) showed that methods widely used in two-dimensional structure retrieval could be applied to three-dimensional searches, to remove the vast majority of compounds before more rigorous comparisons. This was validated in test searches against a subset of the CSD. This concept was furthered by Sheridan et al. (200), who included screens based on aromaticity, hybridization, connectivity, charge, position of lone pairs, and centers of mass of rings. To contain this wealth of information, an inverted bit map [the presence or absence of a feature is encoded as a 1 or 0 (bit) at a particular location in a "keyword"] was employed for highly efficient screening, hundreds of thousands of compounds in minutes.

Similar database searching methods have been incorporated into a number of current database searching systems. Programs such as CAVEAT (206), ALADDIN (Abbott) (207), 3DSEARCH (Lederle) (208), MACCS-3D (209), CHEM-X (210), UNITY (211), and others contain considerable functionality useful for such an approach. CAVEAT (206) is designed to assist a chemist in identifying cyclic structures that could serve as the foundation for novel compounds. In particular, it allows an investigator to rapidly search structural databases for compounds containing substituent bonds that satisfy a specific geometric relationship. ALADDIN (207), 3DSEARCH (208), MACCS-3D (209), and CHEM-X (210) are similar, in that geometric relationships between various user-defined atomic components can be used as a query to retrieve matching structures. Features have been included to allow the user to delineate molecular characteristics (atom type, bond angles, torsional constraints, etc.) to ensure the retrieval of relevant compounds. Additional constraints have been incorporated into 3DSEARCH (208) and ALADDIN (207), including the consideration of retrieved ligand-receptor volume complementarity. Furthermore, CHEM-X (210) performs a rule-based conformational search on each structure in the database to account for conformational flexibility. For a comprehensive review of three-dimensional chemical database searching, see Martin et al. (212,213).

Pharmaceutical companies have developed three-dimensional databases for their compound files to help prioritize candidates for screening (210, 214). An essential component in such a system is a method for assessing similarity (212,215). Because most compound databases were entered as two-dimensional structures, this has required conversion to a three-dimensional format. Programs have proved (197–199, 216) useful in generating plausible three-dimensional structures from the connectivity data, as reviewed by Sadowski and Gasteiger (217). Because of the inherent flexibility in most compounds, the use of a single conformation to represent the three-dimensional potential for interaction of a molecule is a clear limitation. Development of three-dimensional databases with a compact, coded representation of the conformational states available to each compound is a logical next step. Efficient use of such a database requires methods for evaluating three-dimensional similarities. In addition to identification of compounds that can present an appropriate three-dimensional pattern, compounds must also fit within the receptor cavity. Based on a shape-matching algorithm, Sheridan et al. (200) screened candidate compounds to select those whose volumes would fit within the combined volumes of known active compounds. Previously, this group used (218) the same algorithm to help identify potential ligands for papain and carbonic anhydrase, by screening compounds from the CSD. Screening of the active site of HIV protease identified (219) haloperidol (Fig. 3.20) as an inhibitor of the enzyme and provided a novel chemical lead for further investigation. Burt and Richards (220) introduced flexible fitting of molecules to a target structure, with assessment of molecular similarity as a means of dealing with the conformational problem.

The use of preliminary screens can eliminate the vast majority of compounds before more rigorous, and computationally demanding, pattern-matching comparisons (212,213).

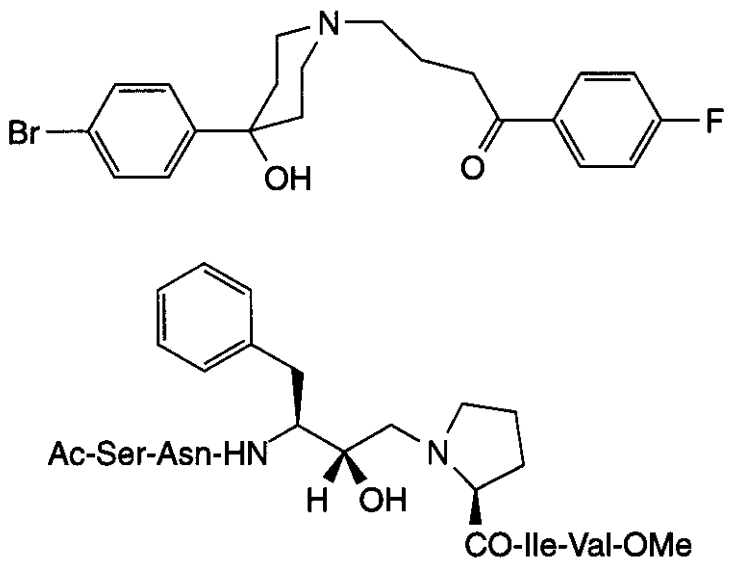


Figure 3.20. Structure of bromperidol (top) found by **DOCK program** when used on active site of HIV-1 **protease** (219) compared with structure of **JG-365** (bottom), a typical substrate-derived inhibitor.

This search strategy is indeed very quick and efficient; however, all retrieved compounds must contain every query component as defined in the preliminary screens. **As** the number and complexity of the query elements increase, one would anticipate fewer true hits, but a corresponding rise in the number of near-misses. If such near-misses could be recovered, effective ligands may simply arise from slight conformational modification to maximize receptor interactions. Furthermore, the retrieval and combinatorial assembly of numerous pharmacophore subcomponents would intuitively produce many more diverse structures than the quest for a single compound in the database incorporating the entire pharmacophore, that is, all requirements of the query. This suggests an approach that would retrieve compounds containing any combination of a minimum number of matching pharmacophoric elements.

Methods have been developed that employ this "divide-and-conquer" approach to ligand development. The active site is partitioned into subsites, each containing several **pharmacophoric** elements. Chemical fragments complementary to each **subsite** are then designed or retrieved from databases. Finally, fragments are linked to form aggregate ligands. The advantage of this approach is that ligand diversity can be tremendously augmented through the combinatorial assembly of numerous subcomponents. **DeJarlais** was perhaps the first to employ this philosophy in a novel application (220) of the program DOCK. This well-known program searches three-dimensional databases of ligands and determines potential binding modes of any that will fit within a target receptor (183). However, only a single, static conformation of each database structure is maintained, disregarding ligand flexibility. In **DeJarlais'** method, **conformational** flexibility was later introduced by dividing individual ligands into fragments overlapping at rotatable bonds. Each fragment was first docked separately into various receptor regions. Attempts were then made to reassemble the component parts into a legitimate structure. **A** current example of this approach is the program LUDI, written by Bohm (221,222). In this program, a receptor volume of interest is scanned to determine **subsites** where hydrogen bonding or hydrophobic contact can occur. Small complementary molecules are then chosen from a database and positioned within these **subsites** to optimize binding energy. The process concludes with the selection of various bridging fragments to link subsets of small molecules.

Chau and Dean published a series of articles addressing whether small molecular fragments, with transferable properties, could be generated for further use in automated **site-directed** drug design (223–225). A **program** was developed to combinatorially generate all **three-, four-,** and five-atom fragments containing any geometrically allowed combination of H, C, N, O, F, and Cl. Aromatic fragments were produced as well. Searches of the Cambridge Structural Database (194) were performed to determine the most frequently occurring fragments. To utilize these fragments as components for ligand assembly, more data were necessary to better characterize them. They were analyzed, therefore, to statistically ascertain bond lengths from the CSD to provide some geometrical constraints for structure assembly. Finally, the transferability of atomic residual charges was studied by comparing charges generated for the atoms in each fragment with charges calculated for whole molecules containing the fragment.

Another approach, FOUNDATION (226), searches three-dimensional databases of chemical structures for a user-defined query

consisting of the coordinates of atoms and/or bonds. All possible structures that contain any combination of a user-specified minimum number of matching atoms and/or bonds are retrieved. Combinations of hits can be generated automatically by a companion program (104), SPLICE, which trims molecules found from the database to fit within the active site and then logically combines them by overlapping bonds to maximize their interactions with the site (Fig. 3.21). The addition of bridging fragments to those recovered from the database allows generation of many novel ligands for further evaluation.

3.3.3 De Novo Design. Design of novel chemical structures that are capable of interacting with a receptor of known structure uses methodology that is much more robust, given that the geometric foundations of molecular sciences are much firmer than the thermodynamic ones. Techniques for the design of novel structures to interact with a known receptor site are becoming more available and show promise (227–229). It has become quite evident that much of a molecule acts simply as a scaffold to align the appropriate groups in the three-dimensional arrangement that is crucial for molecular recognition. By understanding the pattern for a particular receptor, one can transcend a given chemical series by replacing one scaffold with another of geometric equivalence. This offers a logical way to dramatically change the side-effect profile of the drug as well as its physical and metabolic attributes. Various software tools are already under development to assist the chemist in this design objective. Lewis and Dean described their approaches to molecular templates in a series of papers (230, 231). An alternative approach, BRIDGE (Dammkoehler et al., unpublished), is based on geometric generation of possible cyclic compounds as scaffolds, given constraints derived from the types of chemistry the chemist is willing to consider. Nishibata and Itai (232, 233) published a Monte Carlo approach to generating novel structures that fit a receptor cavity. Pearlman and Murko (234) combined a similar approach with molecular dynamics with illustrative applications to HIV protease and FK506 binding protein. CAVEAT is a program developed by Bartlett to find cyclic scaffolds (207) by searching the CSD (195) for the correct vectorial arrangement of appended groups.

All of these approaches attempt to help the chemist discover novel compounds that will be recognized at a given receptor. Van Drie et al. (207) described a program, ALADDIN, for the design or recognition of compounds that meet geometric, steric, or substructural criteria, and Bures et al. (235) described its successful application to the discovery of novel auxin transport inhibitors. As our knowledge base of receptors grows, such tools will prove increasingly useful. The ability to transcend the chemical structure of lead compounds, while retaining the desired activity, should dramatically improve the ability to design away undesirable side effects. Bohm developed the program LUDI (221,222) to construct ligands for active sites with an empirical scoring function to evaluate their construction.

3.3.4 Docking. The search for the global minimum, or the complete set of low energy minima, on the free energy surface when two molecules come in contact is commonly referred to as the "docking" problem [(236); see also Leach (21)]. Any useful molecular docking program must be computationally efficient in determining the most favorable binding mode, sufficiently sensitive in its scoring function to discriminate between alternate binding modes and the correct mode, and robust enough to allow various ligand-receptor systems to be studied.

3.3.4.1 ***Docking Methods.*** In the case of two proteins of known structure that can be approximated as rigid bodies, there are 6 degrees of freedom, the relative position (x, y, and z coordinates), and relative orientation (roll, pitch, and yaw to use the aeronautical expressions) to be explored. Several very intelligent approaches to this problem have been developed. The first and most well known approach is the DOCK program (http://www.cmpharm.ucsf.edu/kuntz/dock.html) (183) that was developed to solve the ligand-receptor problem. This program uses abstract representations (a set of spheres) of the convex shape on the receptor to be filled and the concave ligand and matches them to generate plausible binding modes with complementary

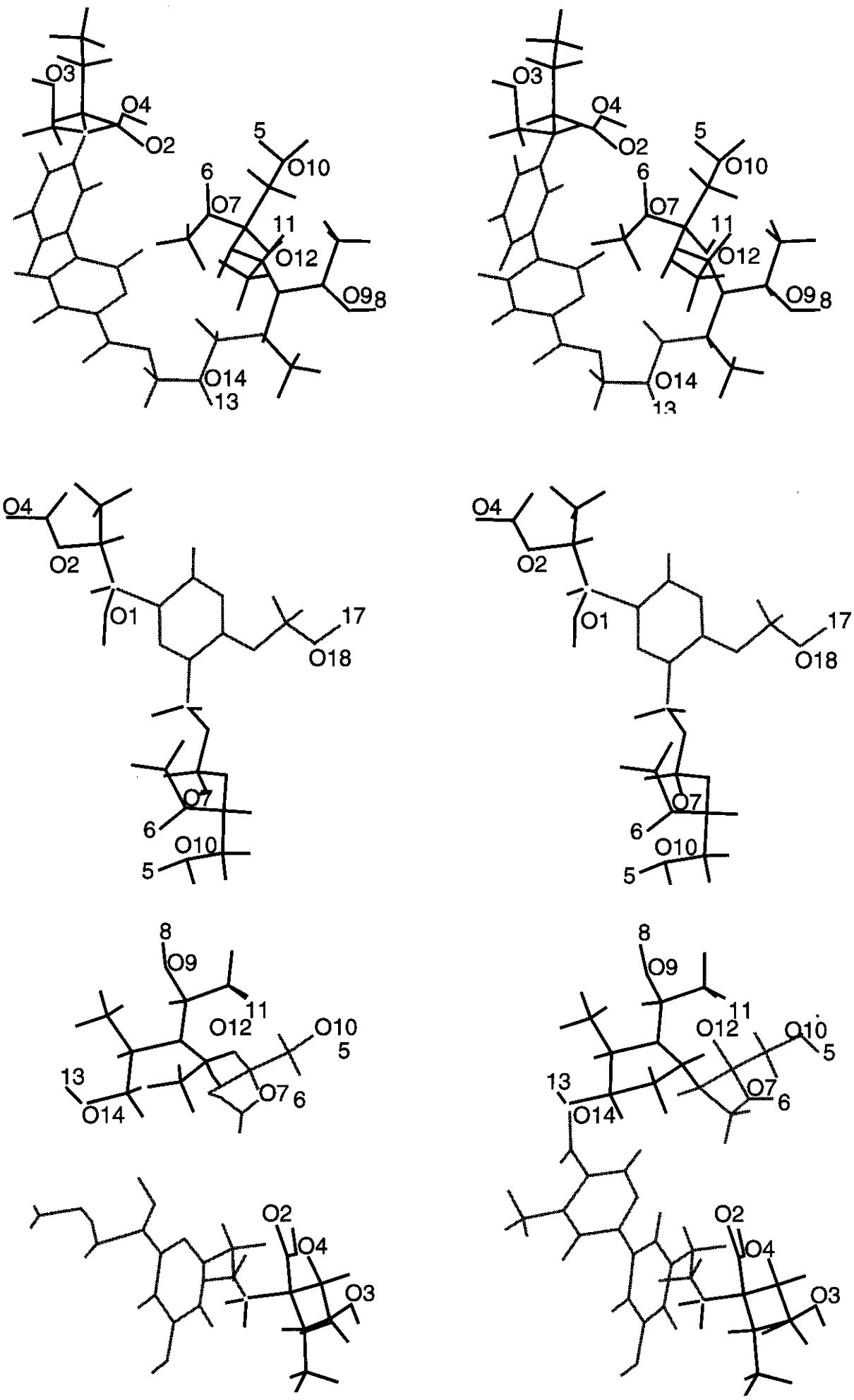


Figure 3.21. Combination by SPLICE (104) of fragments that bind to different subsites of NADP binding site of DHFR to generate a more optimal ligand.

surfaces. An example of the successful use of DOCK was the identification of 13 inhibitors of DHFR from P. *carinii* selected from the Fine Chemicals Directory. Of 40 compounds predicted to be active, these 13 showed IC_{50} values less than 150 micromolar. DOCK (13, 183) has been quite successful in finding noncongeneric molecules of the correct shape to interact with a receptor cavity (237–239). An overview of docking and scoring functions is available (240).

Another approach focusing on complementary surface maximization uses a grid representation of the surface in a series of slices. The slices from the target molecule are processed against the slices from the other molecules by use of a variant of the fast-Fourier transform (241–244) to identify those sections with the greatest complementarity. This approach has been incorporated and extended to electrostatic complementarity in FTDOCK (http://www.bmm.icnet.uk/ftdock/ftdock.html) by Gabb et al. (245). This approach is a relatively fast method for searching the 6 degrees of freedom and has reproduced the binding mode of several macromolecular complexes and is available in GRAMM (Global Range Molecular Matching, http://reco3.musc.edu/gramm/) that was judged the best when applied to identify the binding modes in a set of macromolecular complexes at the second (Fall, 1996) CASP evaluation of prediction methods.

Obviously, other degrees of freedom should be included to allow both molecules to undergo conformational changes (side-chain relaxation, at the very least, in the case of proteins). In many cases, the active site of the receptor is assumed to be rigid (rationalized on the basis of the specificity and affinity of the system) and a flexible ligand is docked. This limits the number of degrees of freedom to be explored. By simply generating a set of low energy conformers of the ligand and processing them sequentially with DOCK (220), one can sample on a low resolution scale; the flexible ligand problem can be addressed on the basis of shape complementarity.

FlexX is a program for flexibly docking ligands into binding sites, by use of an incremental construction algorithm that builds the ligands in the binding site (246). It starts by extracting a core fragment from the ligand. The algorithm is dependent on selection of an appropriate base fragment, requiring one that makes enough specific contacts with the protein that a definite preference for binding orientation can be determined. FlexX holds bond lengths and angles invariant, using the values of the input ligand. The core is used as the base to which low energy fragment conformers are added, with these conformers based on a statistical evaluation of fragments in the Cambridge Structural Database.

3.3.4.2 ***Scoring Functions*** (247–260). Three-dimensional qualitative structure-activity relationship (3D QSAR) approaches based on the use of training sets of structures with measured affinities are often used to generate a model with predictive powers. The limitation in such methodologies is the necessity for a robust training set of diverse chemical structures to encompass the domain of possible interactions with the therapeutic target. At the beginning of a project, or when three-dimensional information on a novel target first becomes available, such data on a diverse set of chemical ligands are usually not available. For this reason, one would like to capitalize as much as possible on the physical chemistry of the possible interactions between the ligand and its receptor when the structure of the receptor is available. Because of the need to prioritize synthesis in structure-based design efforts and prioritize compounds in combinatorial libraries for screening as well as predict the structure of protein complexes, an increased interest in scoring functions (i.e., empirical approaches to predict affinities) have emerged. Several early attempts and their reported predictive ability are cited next.

1. Bohm (221, 222) analyzed 45 protein-ligand complexes (affinity range = −9 to −76 kJ/mol) and found the following equation by multiple regression analysis:

$$\Delta G_{\text{binding (kJ/mol)}} = 5.4\Delta G_0 - 4.7\Delta G_{\text{hb}} - 8.3\Delta G_{\text{ionic}} - 0.17\Delta G_{\text{lipo}} + 1.4\Delta G_{\text{rot}}$$

$$r^2 = 0.76,\ S = 7.9,\ q^2 = 0.696,$$

$$S\ (\text{press}) = 9.3\ (2.2\ \text{kcal/mol})$$

2. Krystek et al. (261) analyzed 19 protein-ligand complexes in an update of the Novotny approach (262).

$$\Delta G_{\text{binding (kcal/mol)}} = 11 - 0.025\Delta G_{\text{CSA}} - \Delta G_{\text{EL}} + 0.6\,T_{\text{SC}}$$

$$r^2 = 0.69, S = 4.0$$

3. VALIDATE is a hybrid approach to predict the binding affinity of novel ligands for a receptor of known three-dimensional structure based on the calculation of several physicochemical properties of the ligand itself as well as a molecular mechanics analysis of the receptor-ligand complex (263). The properties of a diverse training set ($-\log K_i$, range = 2.47–14.00) of 51 crystalline complexes were analyzed by partial least squares (PLS) statistical methodology and neural network analysis to select a statistical model from a variety of parameters with the following properties:

$$r^2 = 0.81, S = 1.15, q^2 = 0.72,$$

$$\text{S (press)} = 1.29\ (1.75\ \text{kcal/mol})$$

The true measure of any model rests in its ability to predict the affinity of new compounds. This would include the prediction of unique ligands bound to receptors that exist in the base set as well as the affinities of unique ligand/receptor complexes. Three separate test sets were compiled for this purpose. The first set consisted of 14 inhibitors that were obtained from crystalline receptor/ligand complexes. Neither ligands nor their receptor classes were included in this training set. Included were 2 DHFR, 2 penicillipepsin, 3 carboxypeptidase, 2 alpha-thrombin, and 2 trypsinogen inhibitors as well as 3 DNA-binding molecules. Prediction of binding affinities gave a PLS predictive $r^2 = 0.786$, with an absolute average error of 0.693 log units. The second test set consisted of 13 HIV protease inhibitors whose initial conformation and alignment were derived from the CoMFA analysis done by Waller et al. (264). The selection of the inhibitors was based on maintaining a good range of activity as well as using several inhibitors from the published test set. The PLS predictive r^2 value was 0.565, with an absolute average error of 0.694. The predictive r^2 value is considerably lower than that of the first test set, although this is attributed to the smaller range and distribution of activity in this set. The absolute average error is almost identical.

Although shape complementarity is an important consideration and shows correlation with the energy of interaction, it does not consider the electrostatics of the system (the relative positioning of hydrogen-bond donors and acceptors, etc.). More sophisticated energetic functions are often used to refine the candidate binding modes found by DOCK, or in the docking process itself. The assumption of rigid geometry for the receptor allows a preprocessing of the energetic contribution of the receptor to each grid point of a lattice constructed within the active site cavity (131, 265, 266). This allows a simple estimation of the energy of interaction of each atom in the ligand by finding the energy of the lattice points that are closest followed by interpolation. By increasing the efficiency of the scoring function, more candidate binding modes can be evaluated and, thus, one resembling the global minimum is more likely to be found. This assumes that. the scoring function used is sufficiently accurate to discriminate between the correct binding mode and others, and the problem is simply one of sampling. Most scoring functions used, however, deal almost essentially with the enthalpy of binding and ignore the entropy of binding. It should not be surprising, therefore, that the agreement between the predicted binding modes and those observed experimentally are not always perfect. As one is attempting to discriminate between alternate binding modes of the same complex, difficulties in estimating entropy and desolvation are minimal because many of the terms (solvation and entropy of isolated ligand and receptor) in the comparison cancel.

3.3.4.3 Search for the Correct Binding Mode (267–283). Just as there are many different approaches to the global minimization problem, most, if not all, have been applied to the docking problem. These include molecular dynamics, Monte Carlo sampling, systematic

search (284), the genetic algorithm (101, 102, 105, 285, 286), and straight derivative optimization with multiple starting geometries. A combination of MD/MC has been shown (287, 288) to be a fairly efficient method for determining the free energy surface in smaller host-guest systems (289). The combination of molecular dynamics to locally sample with Monte Carlo that allows for conformational transitions provides adequate sampling if sufficient computational resources are available.

Wasserman and Hodge (290) used molecular dynamics to dock thermolysin inhibitors to an approximate model of the enzyme, with flexibility in the active site (38 of 314 residues) and ligand and with the rest of the enzyme represented by a grid approximation. A solvation model was used to compensate for desolvation in complex formation. To get 22 of 25 runs to orient the hydroxamate function correctly, the hydroxamate oxygens of the starting conformation were initialized within 4 Å of the zinc. If they were allowed to vary to 8 Å, then only 3 of 24 runs placed the ligand correctly. Obviously, there is a serious sampling problem.

Desmet et al. (291) used a truncated (dead-end elimination) search procedure to bind flexible peptides to the MHC I receptor. The translation/rotational space covered 6636 relative orientations and each nonglycine/proline residue of the peptide had 47 main-chain conformers. Side chains had threefold rotations about their chi angles and 28 side chains of the receptor were allowed to rotate. Seventy-four low energy structures were obtained with an average rmsd of 1 Å. The lowest energy structure had an rmsd of 0.56 Å. Peptides up to 20 residues were docked with this procedure.

King et al. (292) used an empirical binding free-energy function when docking MVT-101 to HIV protease. Forty-nine translation/rotations were examined with the Ponder/Richard rotamer library. Only a limited number of rotamers for each amino acid were examined: Thr(2), Ile(3), Nle(3), Nle(3), Gln(6), and Arg(5). According to the authors, 2.24×10^{10} discrete states were examined. Sixty-four low energy structures with an average rmsd of 1.36 Å were found. If the CHARMM potential was used with the same protocol, then the average rmsd was increased to 1.68 Å.

The genetic algorithm has been used by several groups (101,102,105,285,286,293) to optimize the scoring function used. Encoding of the conformation of the ligand by torsional degrees of freedom and generating increasingly more fit sets of progeny by mutation and crossover have proved to be an effective search strategy. In one example (285), a Gray-coded binary string was used for the three translations, three rotations, and bond rotations that specified the binding mode, and a two-point crossover operator was used in the GA algorithm. In the four examples of complexes with known crystal structures, the results of rigid-body docking with a straightforward application of the GA were not encouraging, in that the correct binding mode was identified in only two of the four test cases. Restraining the GA to search subdomains (different binding hypotheses) in a systematic manner corrected this problem. Only the ligand was allowed flexibility and the GA procedure was repeated. Several binding modes similar to that seen in the experimental complex were found in each example, but ones with the lowest energy did not necessarily have the lowest rms from the experimental, pointing out deficiencies in the AMBER-like scoring function used.

Generally, no single scoring function can accurately predict the binding affinities for all types of ligands with all types of receptors. Consensus scoring (294, 295) is the simultaneous use of multiple different scoring functions to make virtual screening more predictive. CScore (Tripos, Inc.) is a consensus-scoring program that integrates several well-known scoring functions from the scientific literature. Each individual scoring function is used to predict the affinity of ligands in candidate complexes. CScore also creates a consensus column, containing integers that range from 0 to the total number of scoring functions. Each complex whose score exceeds the threshold for a particular function adds 1 to the value of the consensus; configurations below the threshold contribute a zero. Consensus columns can also be calculated from any combination of externally supplied indicators, so that key aspects of binding (e.g., the presence of a specific hydrogen bond) can be used to discriminate good configurations from bad ones. CScore can be used to rank multiple con-

figurations of the same ligand docked with a receptor, or to rank selected configurations of different ligands docked to the same receptor.

These approaches implicitly assume that the observed receptor cavity has some physical stability (i.e., a static view) and is not significantly altered by binding of different ligands. Although there is no guarantee that this is true for any particular case under study, the specificity seen in biological systems argues that a receptor site has some functional significance in imposing its specific steric and electrostatic characteristics in the molecular recognition and selection process. One must always be prepared, however, for binding to sites other than that targeted, and possible exposure of cryptic sites that are not observed in the absence of the ligand (181). The current computational limits in molecular dynamics simulations restrict the chance of uncovering such alternative binding modes in our studies. If we can assume the binding mode of our candidate drug is nearly identical to that of a known compound, however, then we have a legitimate basis for thermodynamic perturbation calculations. Multiple or alternate binding modes remain a major fly in the ointment. Naruto et al. (284) demonstrated a systematic approach to the determination of productive binding modes for mechanism-based inhibitors (Fig. 3.22) that could select starting structures for complexes for molecular dynamics simulations. Combinations of methods, such as Monte Carlo or systematic search, to generate multiple starting configurations for simulations to improve sampling and thermodynamic reliability will increase as adequate computational power to support these hybrid approaches becomes more readily available.

Many technical limitations remain to be overcome before ligand design becomes reliable and routine. Many deficiencies in molecular mechanics previously cited remain that limit reliability. Adequate modeling of electrostatics remains elusive in many experimental systems of interest such as membranes. Newer derivations of force fields, such as MM3 (27, 296 and references therein), CHARMM (297, 298), AMBER/OPLS (157), ECEPP (299), and others (156, 300), are attempting to more accurately represent the experimental data, whereas others include a broader spectrum of chemistry such as metals (29–31,301–305). Combinations of molecular mechanics with quantum chemistry (159, 160, 162, 306) are clearly necessary for problems in which chemical transformations are involved. Rather amazing agreement between calculation and experiment has been reported (165, 307) on the relative stabilities of transition-state structures, although there is some controversy (308) regarding this approach. In any case, this is another area of rapid growth as adequate computational resources become available. Riley et al. (309, 310) found an excellent correlation between the relative stabilities of conformers in manganese complexes of pentaazacrowns and their ability to catalyze the dismutation of superoxide.

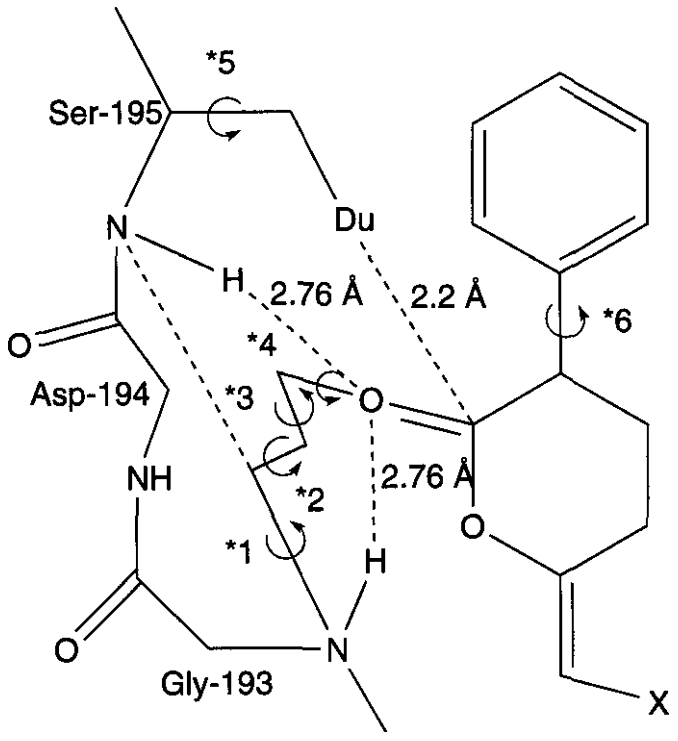


Figure 3.22. Use of systematic search to explore possible binding modes of mechanism-based inhibitors of chymotrypsin (284) by rotation of six bonds (*), which orient carbonyl of substrate relative to hydroxyl (Du) of Ser-195.

3.4 Calculation of Affinity (260)

3.4.1 Components of Binding Affinity (255). The ability to calculate the affinity of prospective ligands based on the known three-dimensional structure of the therapeutic target would allow prioritization of synthetic targets. It would bring quantitation to the qualitative visualization of a potential ligand in the receptor site. Although this problem has been solved in principle, in practice, direct application of molecular mechanics has not yet

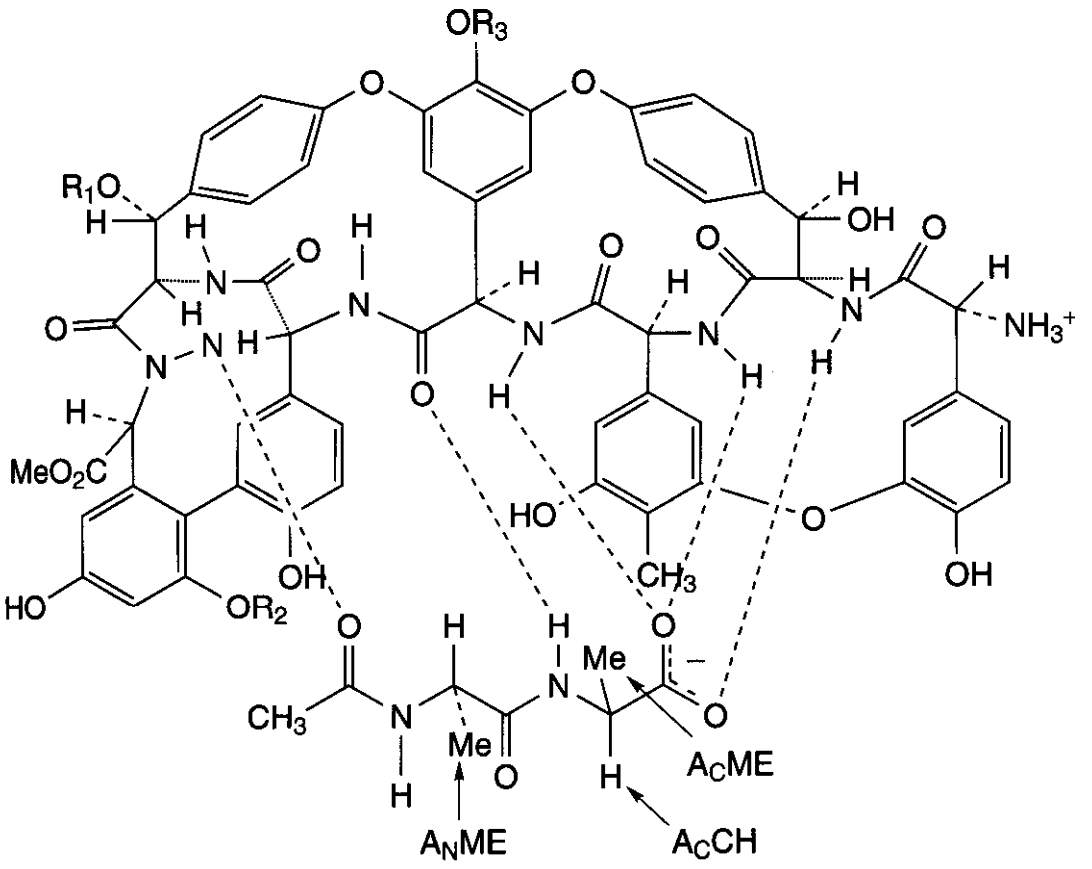


Figure 3.23. Vancomycin-peptide complex used by Williams et al. (311–315) to investigate components of free energy of binding.

proved to be a reliable indicator. The reasons behind this difficulty become more obvious if one dichotomizes the free energy of binding into a logical set of components.

For example, Williams (311–314) used a vancomycin-peptide complex (Fig. 3.23) as an experimental system in which to evaluate the various contributions to binding affinity. A similar analysis for antibody mutants was attempted by Novotny (262).

$$\Delta G = \Delta G_{(\text{trans + rot})} + \Delta G_{\text{rotors}} + \Delta H_{\text{conform}} + \sum \Delta G_i + \Delta G_{\text{vdW}} + \Delta G_{\text{H}}$$

where $\Delta G_{(\text{trans + rot})}$ is the free energy associated with translational and rotational freedom of the ligand. This has an adverse effect on binding of 50–70 kJ/mol (12–17 kcal/mol) at room temperature for ligands of 100–300 Da, assuming complete loss of relative translational and rotational freedom. ΔG_{rotors} is the free energy associated with the number of rotational degrees of freedom frozen. This is 5–6 kJ/mol (1.2–1.6 kcal/mol) per rotatable bond, assuming complete loss of rotational freedom. $\Delta H_{\text{conform}}$ is the strain energy introduced by complex formation (deformation in bond lengths, bond angles, torsional angles, etc. from solution states); Σ AG, is the sum of interaction free energies between polar groups; ΔG_{vdW} is the energy derived from enhanced van der Waals interactions in complex; and ΔG_{H} is the free energy attributed to the hydrophobic effect (0.125 kJ/mol per Å^2 of hydrocarbon surface removed from solvent by complex formation).

Through use of this analysis on the dipeptide-vancomycin system, estimates of the contribution of the hydrogen bonds to binding were made (312) that were considerably higher (−24 kJ/mol, −6 kcal/mol) than those derived experimentally. The most likely source of error is the assumption of complete loss of relative and internal entropy upon binding. In retrospect, Searle and Williams (313) examined the thermodynamics of sublimation of organic compounds without internal rotors, and showed that only 40–70% of theoretical entropy loss occurs on crystallization. This provides an estimate of the entropy loss to be expected on drug-ligand interaction. Applying this correction to the peptide-vancomycin system led (314) to a more conventional view of the hydrogen bond of between −2 and −8 kJ/mol (0.5–2.0 kcal/mol). Because several of the components in the binding energy estimate are directly related to the degree of order of the system (entropy), simulations in solvent may be necessary to quantitate the degree by which the relative motions of the ligand and

protein are quenched and the restriction on rotational degrees of freedom upon complexation. Aqvist (316, 317) developed the linear interaction energy (LIE) method for calculating the ligand-binding free energies from molecular dynamics simulations. Verkhivker et al. (318) developed a hierarchical computational approach to structure and affinity prediction in which dynamics is combined with a simplified, knowledge-based energy function. Despite the focus on short peptides interacting with the SH2 domain with exhaustive calorimetric determination of binding entropy, enthalpy, and heat capacity changes, the overall correlation between computed and experimental binding affinity remained rather modest.

3.4.2 Binding Energetics and Comparisons. Because of the difficulties in calculating binding free energies (see below), attempts to use ΔH as a means of correlation with binding affinities have often appeared in the literature, sometimes meeting with considerable success. These successes, however, are fortuitous and depend on simplifying assumptions as well as the well-known correlation (319) between ΔH and ΔG, which has been suggested as an unusual property of the solvent water. A similar correlation has been observed in nonaqueous systems and relates to higher entropy loss associated with stronger enthalpic interactions (313). It is a common assumption with congeneric series that the desolvation energies and entropic effects will be approximately the same across members of the series. This, often tacit, assumption may hold for most of the series, but complex formation is dependent on the total energetics of the complex, and what may appear a relatively innocuous change in a substituent may trigger a different binding mode in which the ligand has reoriented. This will likely have an impact on desolvation as well as entropic effects, in that the interactions of the majority of the ligand have changed environment.

3.4.3 Atom-Pair Interaction Potentials. Affinities can be calculated based on ligand-receptor atom-pair interaction potentials that are statistical in nature rather than empirical. Muegge and Martin (320) derived these potentials from crystallographic data in the Protein Data Bank, drawing on hundreds or thousands of examples of each interaction type. Grzybowski et al. (321) combined a knowledge-based potential with a Monte Carlo growth algorithm that generated a very potent inhibitor of human carbonic anhydrase (322). The resulting equation for all the atom-pair interactions in a protein-ligand complex can yield free energies directly, given that solvation and entropic terms are treated implicitly.

3.4.4 Simulations and the Thermodynamic Cycle. Given a known structure of a drug-receptor complex with a measured affinity of the ligand, the thermodynamic cycle paradigm allows calculation of the difference in affinity ($\Delta\Delta G$) with a novel ligand. Bash et al. (136) successfully calculated the effect of changing a phosphoramidate group (P-NH) to a phosphate ester (P-O) in transition-state analog inhibitors of thermolysin (Fig. 3.24). The difference in free energy between a benzenesulfonamide and itsp-chloro derivative as an inhibitor of carbonic anhydrase has been calculated (323) as well. This is similar to the original application to enzyme-ligand work on benzamidine inhibitors of trypsin, in which the mutation of a proton to a fluorine was calculated (324). Hansen and Kollman (325) calculated differences in the free energy of binding of an inhibitor of adenosine deaminase as one changes a proton to a hydroxyl group by use of a model of the active site. Other examples (326–328) looked at the difference in binding of two stereoisomers of a transition-state inhibitor of HIV protease (Fig. 3.25) and the affinity of DHFR for methotrexate analogs (329). One obvious conclusion can be drawn: successful applications in the literature deal with relatively minor perturbations to a structure where there is less chance that the binding mode might be altered.

There is at least one example in the literature (330) in which the calculated affinity difference did not agree with the experimental date [binding of an antiviral agent to human rhinovirus HRV-14 and to a mutant virus in which a valine was mutated to a leucine (Fig. 3.26)]. Here a β-branched amino acid (Val) was converted into Leu, which lacks the isopropyl side chain adjacent to the peptide backbone besides the addition of a methyl group.

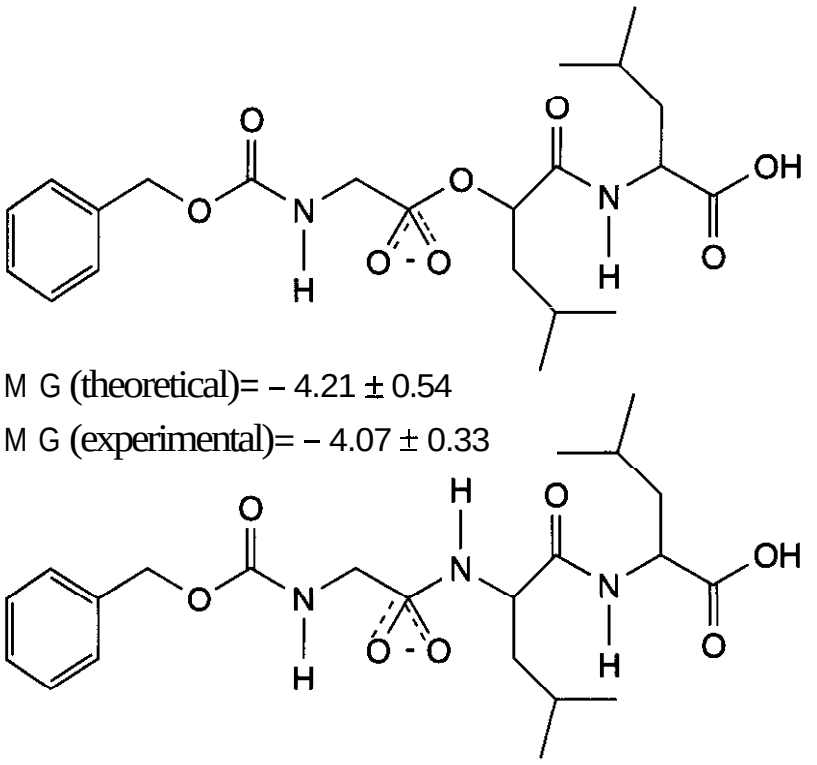


Figure 3.24. Calculated **(136)** difference in affinity ($\Delta\Delta G$) compared with experimental value for two inhibitors of thermolysin.

The differences between calculation and experimental data may be related to rotational isomerism of the side chains that can be explicitly included (331). Despite the successful examples of this approach that appear in the literature, there exists a growing healthy skepticism regarding its general application. In a discussion **(332)** of the application of simulations to prediction of the changes in protein stability attributed to amino acid mutation, problems in adequate sampling, particularly of the unfolded state, as well as difficulties with electrostatics were cited. A review of applications by Kollman **(134)** cites numerous other examples.

3.4.5 **Multiple Binding Modes.** Realistically, congeneric series that can be a useful construct exist only in the mind of the medicinal chemist. The orientation of the drug in the active site depends on a multitude of interactions and a minor perturbation in structure can destabilize the predominant binding mode in favor of another. As examples, detailed

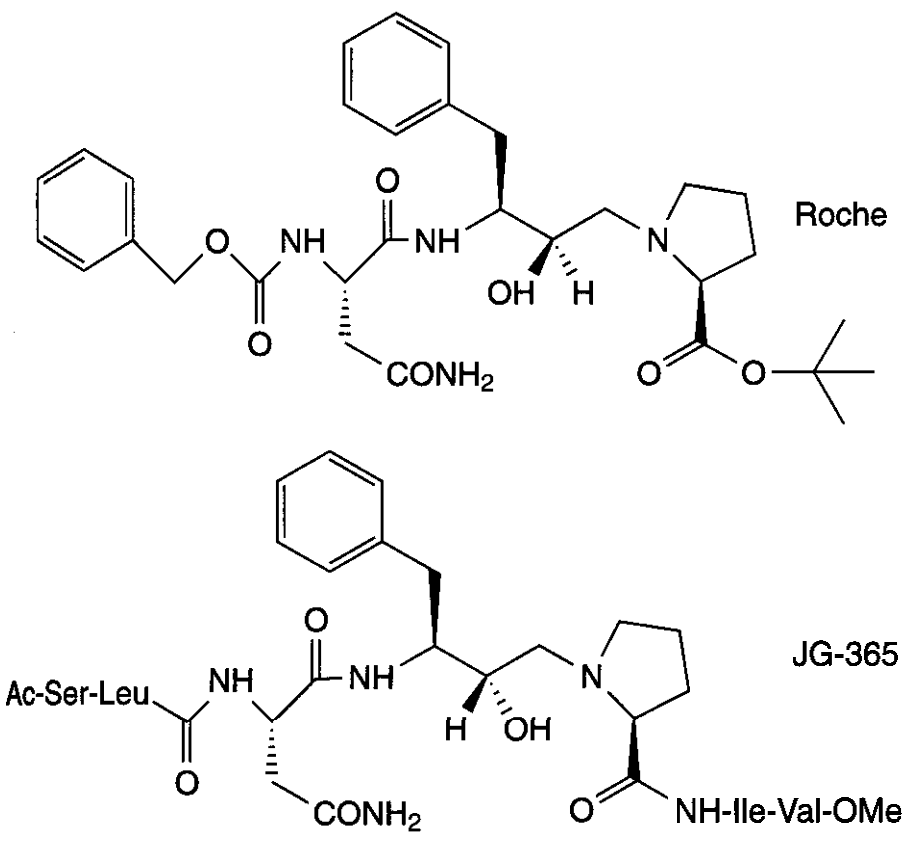


Figure **3.25.** Structures of **JG-365** and Ro **31–8959** in which chirality at crucial transition-state hydroxyl is reversed for optimal binding in the two analogs. An alteration in binding mode was predicted **(333)** to explain this observation that was subsequently confirmed by crystallography.

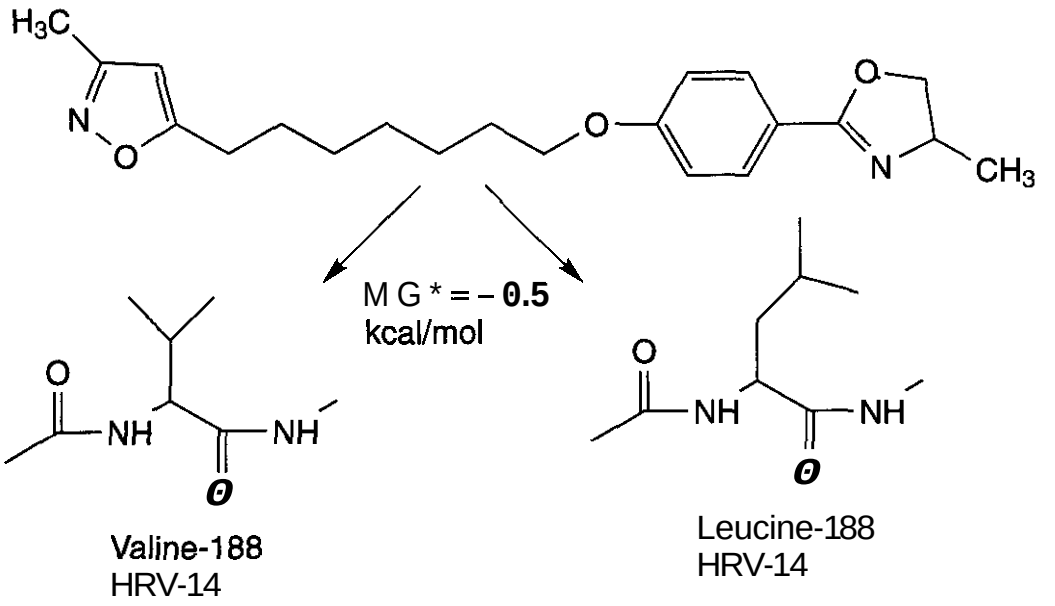


Figure 3.26. Calculated (330) relative affinity of a Sterling-Winthrop antiviral that binds to rhinovirus coat protein (HRV-14) and to the V188L mutant. Biological data indicate that V188L mutation drastically diminishes activity of the antiviral.

analyses of the multiple binding modes shown with thyroxine analogs (334) by transthyretin, a transport protein, and enkephalin analogs (335) by an FAB fragment have been made through crystallography. For this reason, the probability of correct answers with thermodynamic integration studies is directly related to the similarity in structure between the ligand of interest and the reference compound. All three-dimensional methods for predicting affinity require a fundamental assumption about the binding mode (in other words, an orientation rule for aligning compounds in the model). Examination of series of ligands binding to the same site usually includes examples of similar compounds that have different binding modes [e.g., the change in orientation (Fig. 3.25) of the C-terminal portion of the Roche HIV protease inhibitor compared with JG-365] (333). Molecular modeling is currently capable of distinguishing correctly in many cases between alternate binding modes of the same ligand. Many components (desolvation, entropy of binding, etc. of the ligand), which cloud the issue of direct calculation of affinities are constant when comparing binding modes of the same compound and, therefore, do not have to be evaluated. The computational costs of exploring possible binding modes within the active site is nontrivial, however, especially when the protein is capable of reorganizing to expose alternative sites, as was the case for a series of ligands for hemoglobin (181).

In a similar fashion, it is generally assumed from the competitive behavior for binding shown by many agonists and antagonists that they bind at the same site on the receptor (certainly, the simplest hypothesis). Recent studies on G-protein-coupled receptors indicates that agonists and antagonists often have different binding sites, given that mutations in the receptor can affect the binding of one and not the other. An example of such a study on the angiotensin II receptor has been published (336). This story is only beginning to unfold, but appears to be a general phenomenon in G-protein receptors (337, 338). Examples of this phenomenon have been reported with antagonists derived from screening where the structure of antagonist and agonist differ dramatically, but also where the antagonists were obtained by minor structural modification of the natural agonist.

3.5 Protein Structure Prediction

Prediction methods for generating the 3D structure of a protein based on its sequence alone fall into several categories. There are hierarchical methods that predict secondary structures and then attempt to fold those elements together. There are simulation methods that attempt to fold the protein through the use of models of reduced complexity and then refine the prediction by using them to constrain all-atom models. Additionally, there are hybrids of these approaches that rely heavily on heuristics. These methods have been successful in limited cases in the hands of their authors, but have generally been found lacking when tested by others in a more thorough and objective manner. Nevertheless, partial successes indicate that signal has begun to emerge from the smoke and mirrors.

3.5.1 Homology Modeling. Often, the crystal structure of the therapeutic target is not available. but the three-dimensional structure of a homologous protein will have been determined. Depending on the degree of homology between the two proteins, it may be useful to model-build the structure of the unknown protein based on the known structure. Many models (339–341) of the various G-protein-coupled receptors have been built based on homology with bacterial rhodopsin. Models of the three-dimensional structures of human rennin (342) and HIV protease (343,344) were built from crystal structures of homologous aspartyl proteinases as aids to drug design. The known structures of serine proteases have served as templates for models of phospholipase A2 (345) and convertases or subtilases (346). The crystal structure of the MHC class I receptor served to generate a hypothetical model of the foreign antigen-binding site of Class II histocompatibility molecules (347). Models of human cytochrome P450s have been built by homology as well (348).

One of the major difficulties facing construction of such models is the alignment problem that is compounded by multiple insertions and/or deletions. As the number of known homologous sequences increases, the alignment problem is lessened by consensus criteria. Although the interior core of the proteins is often quite similar, significant alterations can occur on surface loops, and much effort has been expended to fold these loops (123, 349). With regard to the utility of such models in drug design, one can expect that they will prove useful conceptually, but that the molecular details required for optimizing specificity, for example, would be deficient. One tries to exploit the often subtle differences that arise from sequence changes, which are reflected in the three-dimensional structure. Models built by homology would be expected to be weakest in those areas in which sequence differences were greatest.

3.5.2 Inverse Folding and Threading (350–353). This is the ultimate in motif recognition. One makes use of the ever-increasing database of known three-dimensional structures to generate a set of 3D folding motifs for proteins. The sequence of an unknown structure is systematically forced to adopt the coordinates of overlapping segments of the 3D motif and its energy evaluated. In essence, the local multibodied interactions induced by the 3D constraints are evaluted with an empirical pseudopotential that has been calibrated on the PDB database (354,355) and that is capable of returning a low energy for native sequences compared with scrambled sequences or protein with other 3D structures. If one cannot discriminate native structures from other folding motifs, then there is little chance that an unknown sequence, which folds in a similar 3D pattern, would be discriminated. The basic assumption is that 3D homology exists between the test sequence and some sequence represented in the motif database. This is not necessarily true, inasmuch as many as 40% of the new structures by crystallography determined have no known 3D homologs. In fact, in an analysis of the genomes of several sequenced microorganisms (356), no more than 12% of the deduced proteins had detectable homology with proteins of known structure. In the CASP competition, however, the most predictive success has been with this approach when a 3D homology existed.

One interesting question that arises is an estimate of the number of protein motifs that exist. One way to approximate this is to as; sume random sampling of protein motif space and then analyze the frequency of new motifs in new crystal structures that leads to a number of approximately 1500 folds (357). Of course, such an estimate is always biased by size of protein, ease of crystallization, abundance, and so forth. Lattice approaches give a maximal estimate of 4000 folds (358). Over 1000 protein structures are known with approximately 120 folds (351).

At a more local level, proteins are generated from a set of architectural building blocks, helices, sheets, turns, and so forth. If one can accurately determine the location of these structural elements within a sequence, then the difficulty of assembly of these components is significantly easier because the degrees of freedom have been drastically reduced. Unfortunately, our ability to accurately determine these elements of secondary structure seems to have peaked at the 75% accuracy level (359, 360).

LINUS. LINUS (Local Independent Nucleating Units of Structure) (361) is an implementation of a hierarchical folding model in which protein sequences are subdivided into overlapping 50-residue fragments to assess the algorithm effectiveness in predicting short- and medium-range interaction as well as to limit computational complexity. The algorithms accumulate favorable structures within a sequence window, and repeat the process as the window is allowed to grow over the sequence. Obviously, this is an embodiment of the principle of hierarchical condensation of local initiation of folding. At the beginning, the segment length is six and the starting conformation set to all extended backbone. Starting at the N-terminus of the segment, three-residue subfragments are perturbed with backbone torsional values from a library to give a trial conformation. If two atoms overlap, the trial conformation is rejected. Otherwise, the energy is evaluated and selection depends on the Metropolis criterion. For each interaction cycle, 6000 iterations of this procedure are performed, 1000 iterations for equilibrium and 5000 samples. Conformations of chain segments that give a high frequency in the sample are frozen and the segment size increased. Backbone atoms and highly simplified side chains are used in the simulations. The simplified energy function has a vdW term, a hydrogen-bonding term, and a backbone torsional term.

Given the arbitrary fragmentation of the protein for computational efficiency, the predicted secondary structures were surprisingly accurate for the five cases examined, with helical and sheet boundaries within two residues of their corresponding native structures. Nevertheless, the rms differences were rather large, from 3 to 9 Å. Certainly, these results are quite encouraging and confirm the ideas from studies on lattices by Dill (362) and others that much of the secondary structure is encoded into local patterns of hydrophobic and polar residues.

GEOCORE (363). Amino acids are represented at the united atom level with explicit polar hydrogens with slightly reduced vdW radii. The approach uses a discrete set of Φ, Ψ values for each residue type: Gly has six, Pro has three, and most others have four or five values. A contact between nonpolar atoms (carbon or sulfur) is worth –0.7 kcal/mol at closest contact and scaled down from there. Buried non-hydrogen-bonding groups get a penalty of 1.5 kcal/mol. Polar conflicts in which two donors or two acceptors are in contact are given a similar penalty. Constraint-based exhaustive search is used (systematic search with limits such that no steric overlap is allowed and that a compact structure is generated), a branch-and-bound method that guarantees that all globally or near-globally optimal conformations will be found, while neglecting less important conformations. The compact structure is guaranteed by a volume constraint about 60% higher than the volume of a native protein of the same size. Side chains are introduced in their most populated rotameric state from the PBD and only changed to an alternate rotamer to avoid a vdW contact. Four proteins were used to test the approach, avian pancreatic polypeptide (1PPT), crambin (1CRN), melittin (2MLT), and apamin (18 residues). Some 190 million conformations were generated for 1PPT, with 8217 having an energy not more than 16 kcal/mol above the optimum found. The conformation with the lowest rms to the native structure was within the 100 lowest energy conformations found, but the true native structure had a lower energy by use of the same energy function than that of any conformer found by 3–10%. This implies that the major problem was conformational sampling, not just an oversimplified potential function.

Genetic Algorithm. Le Grand and Merz (364) applied the genetic algorithm to a model of proteins using a rotamer library and the AMBER potential function. In a second study, they used a fragment library and a knowledge-based potential function. Sun (365) used a fragment library consisting of di- to pentapeptides and the Sippl potential. He predicted the structures of mellitin, avian pancreatic polypeptide, and apamin (both fragments from apamin and APP were included in the library, so it is not so surprising that the rms agreement for these two was around 1.5 Å). Bowie and Eisenberg (366) used the genetic algorithm with a fragment library of from 9 to 25 residues and their own knowledge-based potential. The fragment most similar to that

of the sequence based on 3D profiles (367) was chosen. They were able to fold 50-residue fragments to within 4.0 Å based on the error in the distance matrix. This avoids the problem of embedding and generating the wrong chirality, which reduces the error estimate.

3.5.3 Contact Matrix. Instead of searching the three-dimensional coordinate space, one can reduce dimensionality by focusing on generating an optimal contact map in 2D (368). The 3D coordinates of a correct contact map can be generated within 1 Å rms for the carbon alphas by distance geometry (369) or other methods (370). By use of the powers of the contact matrix as constraints that limit the contact matrices to compact structures, exploration of various potential interactions between secondary structural elements can be done efficiently. Because of the limited predictive ability of current secondary structure prediction paradigms, a set of plausible inputs to this procedure need to be generated, and the best structures that are derived evaluated further. This may be an efficient low resolution model builder and have some of the computational advantages of the hydrophobic core constraints used by Dill and coworkers. This approach based on geometrical constraints was originally proposed by Kuntz et al. in 1976 (371). The matrices of residue-residue contacts provide, at the very least, a significant partial solution to the prediction of long-range intersegmental contacts through a formalism explicitly describing the structure and some structure-related properties of a protein globule in terms of matrices of residue-residue contacts without explicit knowledge of secondary structure predictions, although they can be a useful source of constraints. In many ways, the success of this approach verifies the conclusions based on lattice models that secondary structures are implicit in the pattern of hydrophobic and hydrophilic residues and the requirements of compactness. The residue-residue contact matrices have some special properties as mathematical objects that can encode the geometrical requirements of compactness; the knowledge of these allows their treatment, starting with the sequence to generate a contact matrix that is consistent with a compact structure. This is done within the framework of a simple and readily formalized geometric model.

The system of intraglobular residue-residue contacts of a protein of N residues may be represented as an N × N matrix of the carbon-alphas, whose elements are ones (contact) or zeros (lack of contact). Any reasonable definition of contact provides ones in the positions (i, i + 1) that correspond to a peptide bond between two adjacent residues in the sequence. The same is true for the residues corresponding to the pair of cysteines forming a disulfide bond (these data may not be available as input and may be used as a test of correct prediction). This set of contacts describes the sequential covalent topology and is a constant part of the contact matrix which does not depend on the spatial structure of the polypeptide chain; however, any additional information on existing intraglobular contacts (e.g., from NMR data or disulfide linkage) can easily be introduced in the constant part A^c of the contact matrix **A:**

$$A^c = \text{const.} \tag{3.1}$$

The number of contacts involving a given residue n_i (the coordination number of the ith residue)

$$n_i = \sum_j a_{ij} \tag{3.2}$$

are assumed to be approximate constants (coordination number) and are determined by a separate algorithm based on residue type and position in the sequence as well as predicted secondary structure.

A very important condition of spatial consistency of any given contact system is defined by the relation

$$b_{ij} = \sum_k a_{ik} a_{kj} \geq c, \quad \text{if } a_{ij} = 1 \tag{3.3}$$

In other words, the squared matrix of **A** should have its elements not less than c at any position where there is a nonzero element in matrix **A.** More generally, there exists a set of specific constraints regulating the relation-

ships of **A** with its powers A^2, A^3, and so forth. These relations are entirely analogous with those known from graph theory for connectivity (adjancency) matrices. The elements of the squared matrix represent the number of paths of length two, the cubed matrix, the number of paths of length three, and so forth. Finally, an obvious property of matrix **A** is its symmetry (for all contact definitions considered so far, if the ith residue is in contact with the jth, the jth residue is in contact with the ith, also).

$$a_{ij} = a_{ji} \tag{3.4}$$

Thus, conditions 3.1–3.4 define the set of matrices **A**, that correspond to spatially consistent, compact structures of protein chains. Besides these general conditions, mainly of geometrical origin, any matrix **A** describing the structure of a real protein molecule should also possess several more specific properties that may be derived from studies of the general properties of protein structures as exemplified in the Brookhaven Protein Databank. The central idea of the approach is to use both the general and specific properties of the contact matrix and its powers for the design of a gain (energy, penalty) function, $\Phi(A)$, so that the task of determining an appropriate intraglobular contact matrix might be formulated as a problem of maximization of $\Phi(A)$,

$$@(A) \rightarrow \max_{A} \tag{3.5}$$

with respect to A under conditions 3.1–3.4. In the simplest and clearest form, $\Phi(A)$ may be expressed in terms of the probabilities of contact between the residues of different types (or groups), q_{ij}. The solution of the problem provides the most probable residue-residue contact matrix **A** in

$$\Phi(A) = \prod_{\text{all contacts}} q_{ij} \rightarrow \max_{A}, \tag{3.6}$$

which is the sense of the maximum likelihood principle. This condition may be rewritten in the form

$$@'(A) = \sum_{\text{all contacts}} \ln q_{ij} = \sum_{i,j} a_{ij} \ln q_{ij} \rightarrow \max_{A} \tag{3.7}$$

It is clear that proper formulation and parameterization of this problem need the analysis of the voluminous experimental data on protein structure to derive the specific properties to be emulated.

This methodology has been used to predict the structure of loops of helical-bundle proteins, given the positions of the connection to the helices (372). Because of the uncertainties in secondary structure predictions that are used as inputs to constrain the search, any single prediction of the method must be viewed with skepticism. Development of scoring functions that discriminate between alternative models at the Ca level of resolution would complement this approach.

Distance Geometry. Aszodi et al. (373–375) explored the use of distance geometry as the metric for comparative modeling of structures. In the CASP2 target set, the methods generated an overall Ca rmsd of 1.85 Å for glutathione transferase based on close homologs with known structure. It had more difficulty with PNS1 and built models based on two different proteins. The correct fold was not obvious based on the CHARMM energy values for the two models.

Neural Networks. PROBE (376) is an integrated suite of neural network modules that predicts folding motif, secondary structure per residue, location of disulfide bonds, and surface accessibility of each residue. No critical assessment of the accuracy of the results from this package was given in the description, but is available for evaluation.

Discrimination Between Folds. Because of the inherent error in potential functions, secondary structure prediction methods, limited sampling, and so forth, one can anticipate that prediction of a variety of alternative structures (perhaps, by several methods) would be more likely to generate a correctly folded structure than any single prediction. The problem then becomes one of discriminating between the correct structure and alterna-

tives that may be very similar in overall quality of fold. Park et al. (377) evaluated the ability of 18 low and medium resolution energy functions to discriminate correct from incorrect folds. Functions that were effective in protein threading were not competitive in discriminating the X-ray structure from ensembles of plausible structures, and vice versa. Obviously, these empirical functions have been derived to optimize their discriminate abilities for a given problem class and the training (selection) sets were different. In other words, the true physics has not been captured by any of the methods. Crippen (378) also raised serious doubts concerning the ability of "empirical" energy functions to identify correctly folded structures based on studies with simple lattice models. Thomas and Dill (379) described an iterative approach ENERGI to generate pairwise residue "energy" scores from the PDB. This is one alternative to the Boltzmann-based pairing frequency analysis used by others (380). The assumption that pairing frequencies are independent is not true based on lattice simulation and, therefore, the underlying assumption of the Boltzmann approach is flawed. The study that used two different sets of proteins to thread was able to classify 88% of 121 proteins having less than 25% homology and no homologs in the training set. The method appears to separate interactive free energies from chain configurational entropies and thus give a more realistic estimate.

4 UNKNOWN RECEPTORS

Until recently, receptors were hypothetical macromolecules whose existence was postulated on the basis of pharmacological experiments. Although recent advances in molecular biology have led to cloning and expression of many of those receptors whose existence was postulated as well as a plethora of subtypes, progress in most cases in defining their three-dimensional structure has yet to provide the medicinal chemist with the necessary atomic detail to design novel compounds. Without detailed information about the three-dimensional nature of the receptor, conventional computationally based approaches, such as molecular dynamics and the Monte Carlo method, are not possible. One can only attempt to deduce an operational model of the receptor that gives a consistent explanation of the known data and, ideally, provides predictive value when considering new compounds for synthesis and biological testing. The utility of such an approach has been demonstrated by Bures et al. (235), who used the pharmacophoric pattern derived for the plant hormone auxin, to find four novel classes of active compounds by searching a corporate three-dimensional database of structures. In many ways, the approach that has evolved is analogous to the American parlor game of 20 questions, in which the medicinal chemist poses the questions in terms of novel three-dimensional chemical structures and attempts to interpret the response of the receptor in a consistent manner. The underlying hypothesis is a structural complementarity between the receptor and compounds that bind. In the same way that the receptor's existence could be deduced based on pharmacological data, some low resolution three-dimensional schematic of the receptor, at least with regard to the active site or binding pocket, can be deduced by analysis of structure-activity data. It is the purpose of this section to summarize the current approaches in use for receptors of unknown three-dimensional structure and evaluate their utility. For purposes of this section, *receptor* is often used in a completely generic sense, including enzymes and DNA, for example, as the macromolecular component (i.e., binding site) of recognition of biologically active small molecules.

4.1 Pharmacophore versus Binding-Site Models

4.1.1 Pharmacophore Models. It is often useful to assume that the receptor site is rigid and that structurally different drugs bind in conformations that present a similar steric and electronic pattern, the pharrnacophore. Most drugs, because of inherent conformational freedom, are capable of presenting a multitude of three-dimensional patterns to a receptor. The pharmacophoric assumption led to a problem statement that logically is composed of two processes. First is the determina-

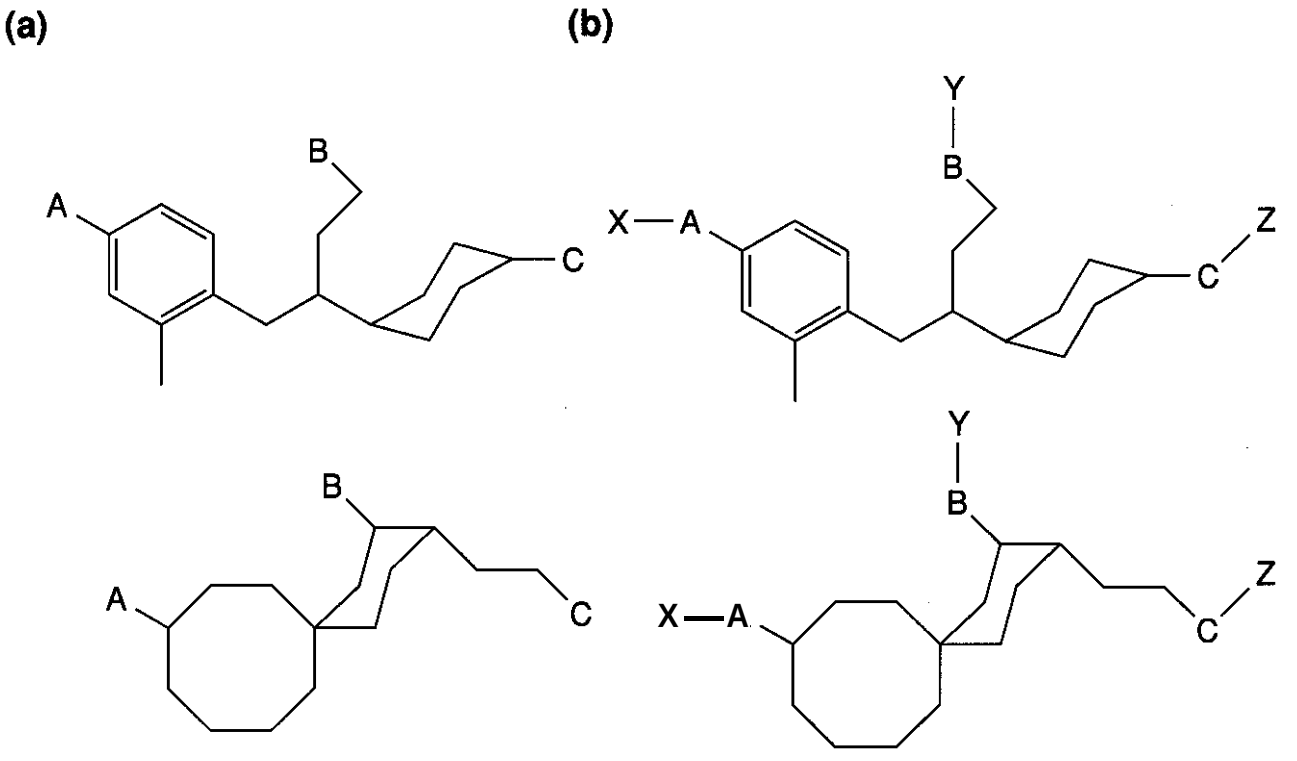


Figure 3.27. (a) Pharmacophore hypothesis with correspondence of functional groups in drugs, **A** = **A',** B = B', C = C'. (b) Binding-site hypothesis by use of drugs with hypothetical binding sites attached (X, Y, and Z overlap).

tion, by chemical modification and biological testing, of the relative importance of different functional groups in the drug to receptor recognition. This can give some indication of the nature of the functional groups in the receptor that are responsible for binding of the set of drugs. Second, a hypothesis is proposed (Fig. 3.27) concerning correspondence, either between functional groups (pharmacophore) in different congeneric series of the drug or between recognition site points postulated to exist within the receptor (binding-site model).

The intellectual framework for use of structure-activity data to extrapolate information regarding the ligand's partner, the receptor, is the concept of the pharmacophore. The pharmacophore, a concept introduced by Ehrlich at the turn of the 20th century, is the critical three-dimensional arrangement of molecular fragments (or distribution of electron density) that is recognized by the receptor and, in the case of agonists, that causes subsequent activation of the receptor upon binding. In other words, some parts of the molecule are essential for interaction, and they must be capable of assuming a particular three-dimensional pattern that is complementary to the receptor to interact favorably. One corollary of the pharmacophoric concept is the ability to replace the chemical scaffold holding the pharmacophoric groups with retention of activity. This is the basis of the current activity (381, 382) in peptidomimetics, in which the amide backbone of peptides has been replaced by sugar rings, steroids (383, 384), benzodiazepines (385), or carbocycles (386, 387) (Fig. 3.28). In the pharmacophoric hypothesis, physical overlap of similar functional groups is assumed; that is, the carboxyl group from compound **A** physically overlaps with the corresponding carboxyl group from compound B and with the bioisosteric tetrazole ring of compound C.

One caveat that must be remembered is the probability of alternate, or multiple, binding modes. The interaction of a ligand with a binding site depends on the free energy of binding, a complex interaction with both entropic and enthalpic components. Simple modifications in structure may favor one of several nearly energetically equivalent modes of interaction with the receptor, and change the correspondence between functional groups that has previously been assumed and supported by experimental data. Changes in binding mode of an antibody FAB fragment to progesterone and its analogs have been shown by crystallography (390,391) of the complexes. For this reason, analysis of agonists as a class is usually preferred, given that the necessity to both

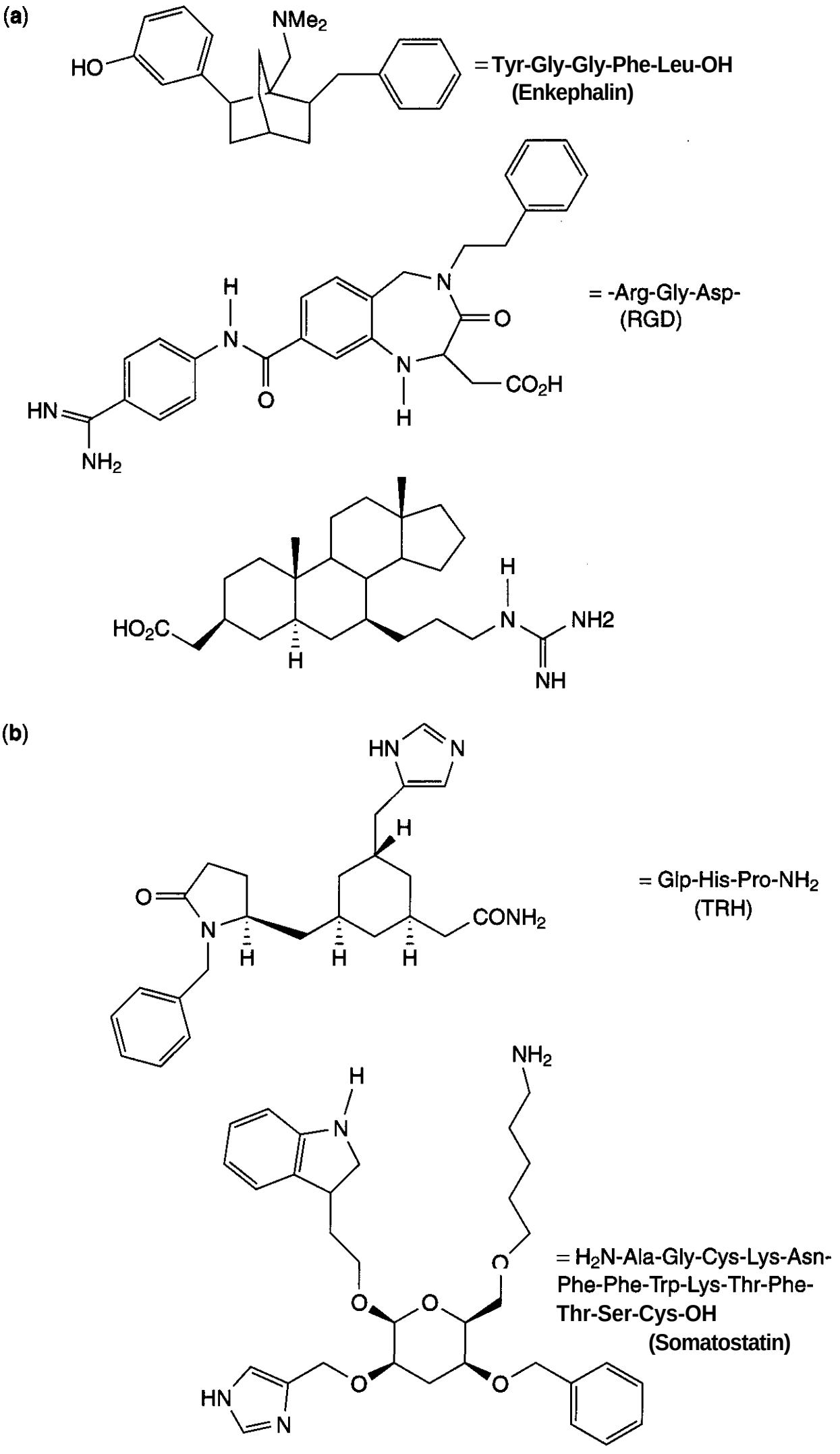


Figure 3.28. Peptidomimetics that have been designed based on iterative introduction of constraints into parent peptide and hypotheses concerning receptor-bound conformation. Enkephalin mimetic (388), RGD platelet GPIIb/IIIa receptor antagonists (384, 385), thyroliberin [TRH (387)], and somatostatin (383,389).For an overview of recent approaches to peptidomimetic design, see the review by Bursavich and Rich (382).

bind and trigger a subsequent transduction event is more restrictive than the simple requirement for binding shared by antagonists (336). Compounds that clearly are inconsistent with models derived from large amounts of structure-activity data may be indicative of such changes in binding mode, and may require a separate structure-activity study to characterize their interaction. Despite its limitations, the pharmacophore approach is often the most appropriate because of lack of detailed information regarding the receptor and can yield useful insights, as seen in the case of clinical success with tyrosine kinase inhibitors (392,393) and other recent examples (394).

4.1.2 Binding-Site Models. One major deficiency in the approach described above is the requirement for overlap of functional groups in accord with the pharmacophoric hypothesis. Although it is true that molecules having functional groups that show three-dimensional correspondence can interact with the same site, it is also true that a particular geometry associated with one site is capable of interacting with equal affinity with a variety of orientations of the same functional groups. One has only to consider the cone of nearly equal energetic arrangements of a hydrogen-bond donor and acceptor to realize the problem. Sufficient examples from crystal structures of drug-enzyme complexes and from theoretical simulation of binding compel the realization that the pharmacophore is a limiting assumption. Clearly, the observed binding mode in a complex represents the optimal position of the ligand in an asymmetric force field created by the receptor that is subject to perturbation from solvation and entropic considerations. Less restrictive is the assumption that the receptor-binding site remains relatively fixed in geometry when binding the series of compounds under study. Experimental support for such a hypothesis can be found in crystal structures of enzyme-inhibitor complexes, where the enzyme presents essentially the same conformation, despite large variations in inhibitor structures; studies of HIV-1 protease complexed with diverse inhibitors support this view (171).

In recent years, therefore, there has been an increasing effort to focus on the groups of the receptor that interact with ligands as being the common features for recognition of a set of analogs. When pharmacophore and binding-site hypotheses are compared, the binding-site model is physicochemically more plausible, in that overlap of functional groups in binding to a receptor is more restrictive than assuming the site remains relatively fixed when binding different ligands. However, the number of degrees of freedom in binding-site hypotheses, represented by the necessary addition of virtual bonds between groups A and X, B and Y, and C and Z in Fig. 3.27, is greater. Additional degrees of freedom complicate subsequent conformational analyses and may preclude any conclusions unless a sufficiently diverse set of compounds is available.

Other approaches to this problem have emphasized comparison of molecular properties rather than atom correspondences. Kato et al. (395) developed a program that allows construction of a receptor cavity around a molecule emphasizing the electrostatic and hydrogen-bonding capabilities. Other molecules can then be fit within the cavity to align them. This is similar in concept to the field-fit techniques available in the CoMFA module of SYBYL, in which the molecular field (electrostatic and steric) surrounding a selected molecule becomes the objective criterion for alignment of subsequent molecules for analysis. An example emphasizing molecular properties in pharmacophoric analysis was given by Moos et al. (396) on inhibitors of cAMP phosphodiesterase II.

4.1.3 Molecular Extensions. If we assume the binding-site points remain fixed and can augment our drug with appropriate molecular extensions that include the binding site (i.e., a hydrogen-bond donor correctly positioned next to an acceptor), we can then examine the set of possible geometrical orientations of site points to see whether one is capable of binding all the ligands. Here, the basic assumption of rigid site points is more reasonable, at least for enzymes that have evolved to catalyze reactions and must, therefore, position critical groups in a specific three-dimensional arrangement to create the correct electronic environment for catalysis. The program checks

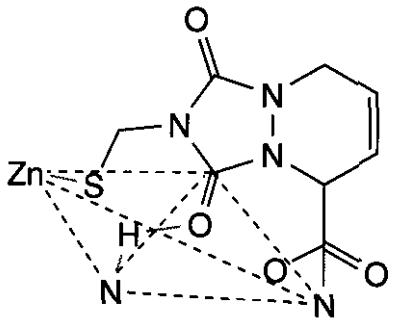


Figure 3.29. The use of active-site models in the Active Analog Approach. The structure shown is one of a series of ACE inhibitors analyzed. The thick gray lines are noncovalent interactions between the inhibitor and active-site points in the enzyme. The dashed lines correspond to the six interatomic distances monitored for each of the inhibitors analyzed.

this hypothesis by determining whether one or more geometrical arrangements of the postulated groups of site points is common to the set of active compounds. Such a geometrical arrangement of receptor groups becomes a candidate binding-site model, which can be evaluated for predictive merit.

In the study of the active site of angiotensin converting enzyme (ACE) by Mayer et al. (397), a binding site model (Fig. 3.29) was used by incorporating the active-site components as parts of each compound undergoing analysis. As an example, the sulfhydryl portion of captopril was extended to include a zinc bound at the experimentally optimal bond length and bond angle for zinc-sulfur complexes (Fig. 3.29). The orientation map (OMAP) (398), which is a multidimensional representation of the interatomic distances between pharmacophoric groups (Fig. 3.30), was based on the distances between binding-site points such as the zinc atom with the introduction of more degrees of torsional freedom to accommodate

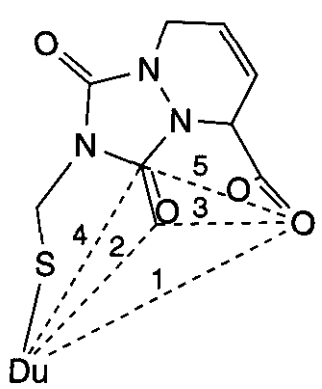


Figure 3.30. Distances used in five-dimensional OMAP used in analysis of ACE inhibitors.

the possible positioning of the zinc relative to ACE inhibitors such as captopril. Analyses of nearly 30 different chemical classes (Fig. 3.31) of ACE inhibitors led to a unique arrangement of the components of the active site postulated to be responsible for binding of the inhibitors. The displacement of the zinc atom in ACE to a location more distant from the carboxyl-binding Arg seen in carboxypeptidase A is compatible with the fact that ACE cleaves dipeptides from the C-terminus of peptides, whereas carboxypeptidase A cleaves single amino acid residues.

Visualization of the OMAP is useful to judge the additional information introduced as each new compound is added (Fig. 3.32). Computationally, it is much more efficient to treat the set of noncongeneric compounds simultaneously (111,399), as we shall see, but reassuring when identical results are obtained if one uses the sequential procedure introducing each molecule in turn, where intermediate results may be visually verified. The use of computer graphics to confirm intermediate processing of data in convenient display modes becomes increasingly more important as the individual computations and numbers of molecules under consideration increase.

4.1.4 Activity versus Affinity. Given a consistent model of either type, a limitation is that one can only ask whether the compound under consideration can present the three-dimensional electronic pattern (pharmacophore) that is the current candidate. In other words, one is limited to predicting the presence or absence of activity, a binary choice. Even the presence of the appropriate pattern is insufficient to ensure biological activity. For example, competition with the receptor for *occupied* space by other parts of the molecule can inhibit binding and preclude activity. We can thus postulate the following conditions for activity:

1. The compound must be metabolically stable and capable of transport to the site for receptor interaction (interpretation of inactive compounds may be flawed by problems with bioavailability).

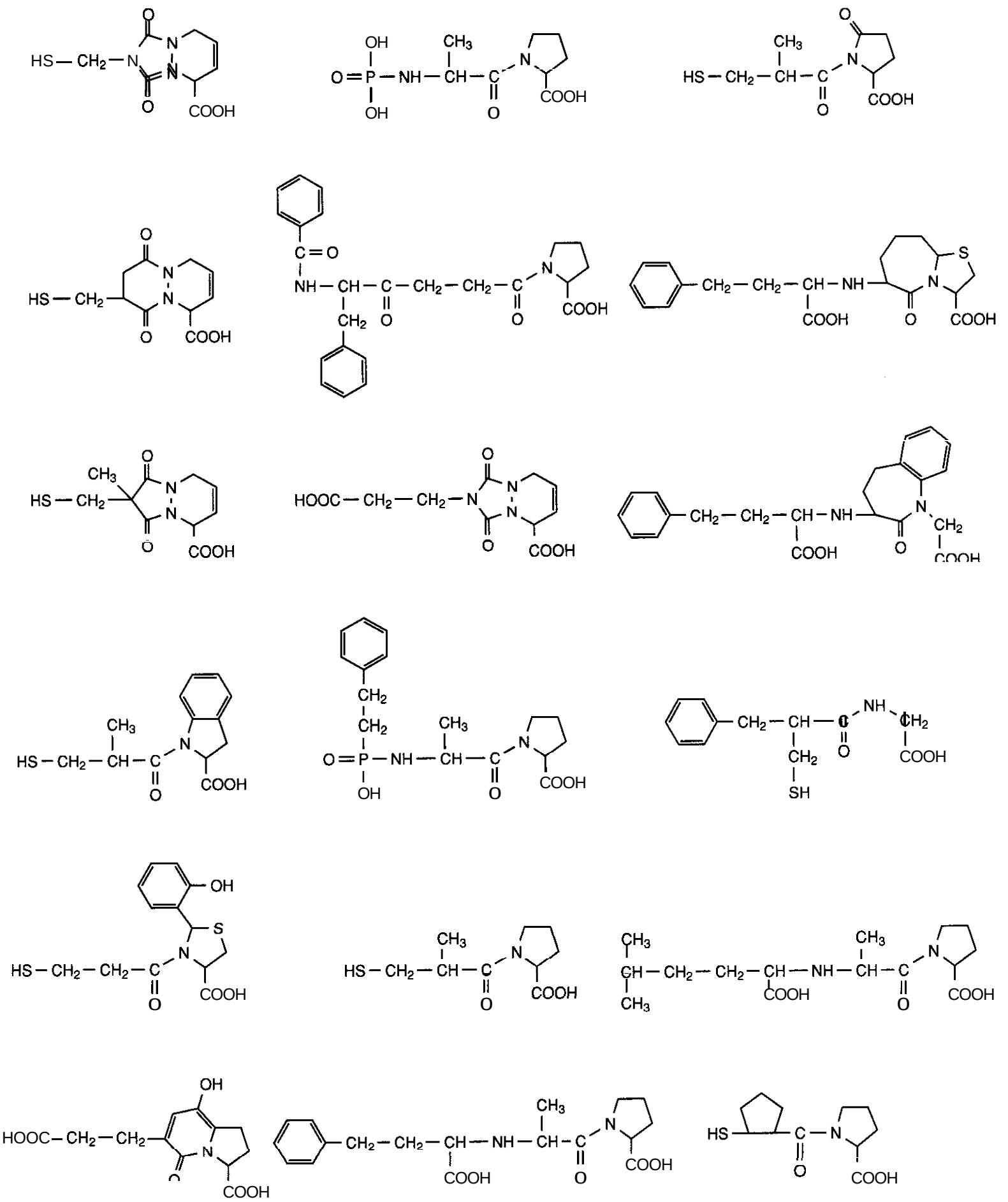


Figure **3.31.** Compounds from different chemical classes of ACE inhibitors used in active-site analysis. Used with permission **(397).**

2. The compound must be capable of assuming a conformation that will present the pharmacophoric or binding-site pattern complementary to that of the receptor.
3. The compound must not compete with the receptor for space while presenting the pharmacophoric or binding-site pattern.

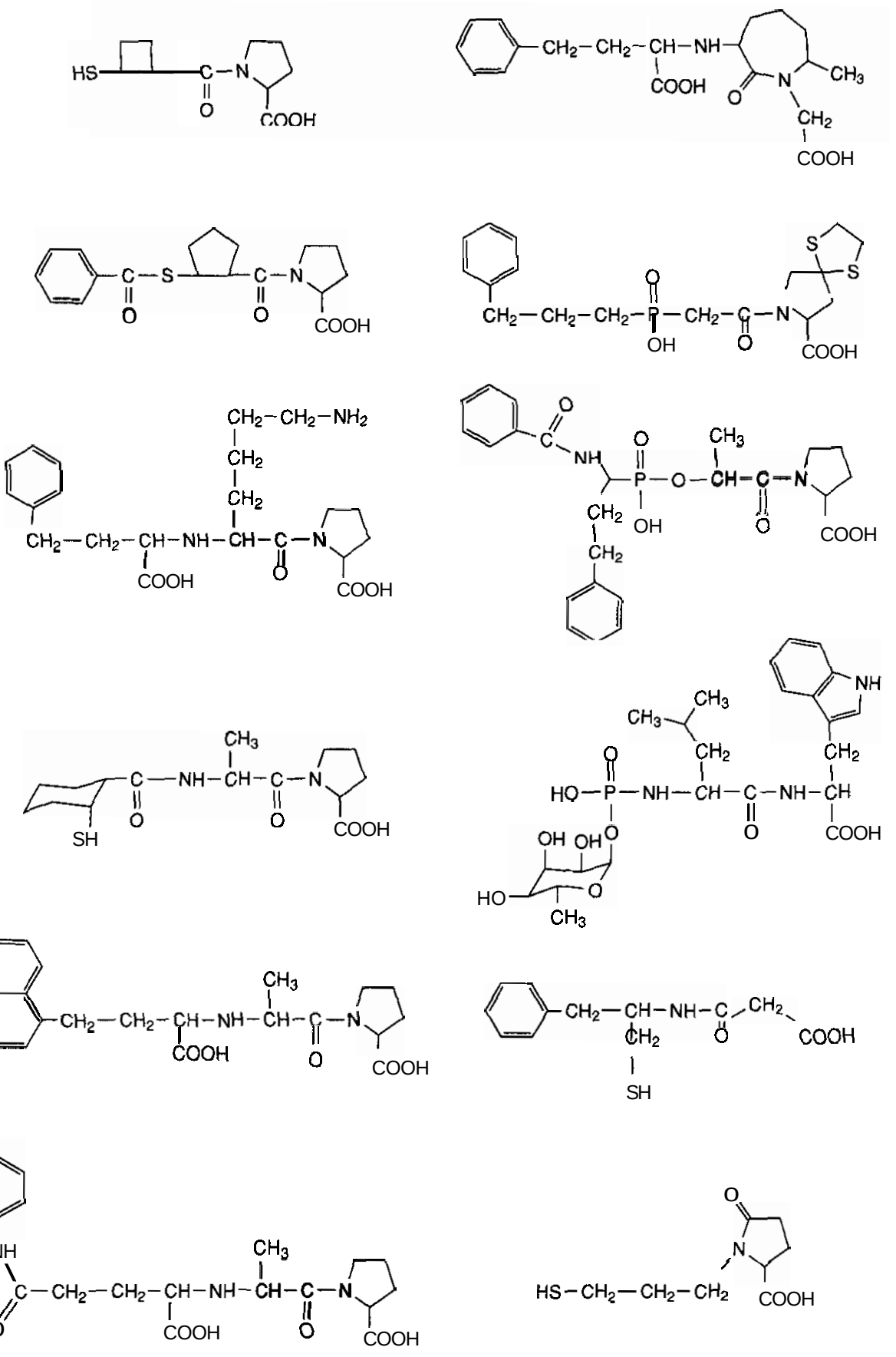


Figure 3.31. (Continued.)

Once these conditions are met, we can attempt to deal with the potency, or binding affinity. This belongs to the domain of **three**-dimensional quantitative structure-activity relationships (**3D-QSARs**) (400) and we **illus**trate the use of a particular variant, **CoMFA** (**187, 401**), on ACE inhibitors at the end of this chapter. Condition 3.3 allows us to utilize compounds capable of presenting the **pharma**cophoric pattern, but incapable of binding, to help **determine** the location of **receptor-occu**pied space in relation to the pharmacophore (receptor-mapping) (402). This allows a crude, low resolution map of the position of the receptor relative to the pharmacophoric elements and indicates in which directions chemical modifications may be productive.

The number and diversity of compounds

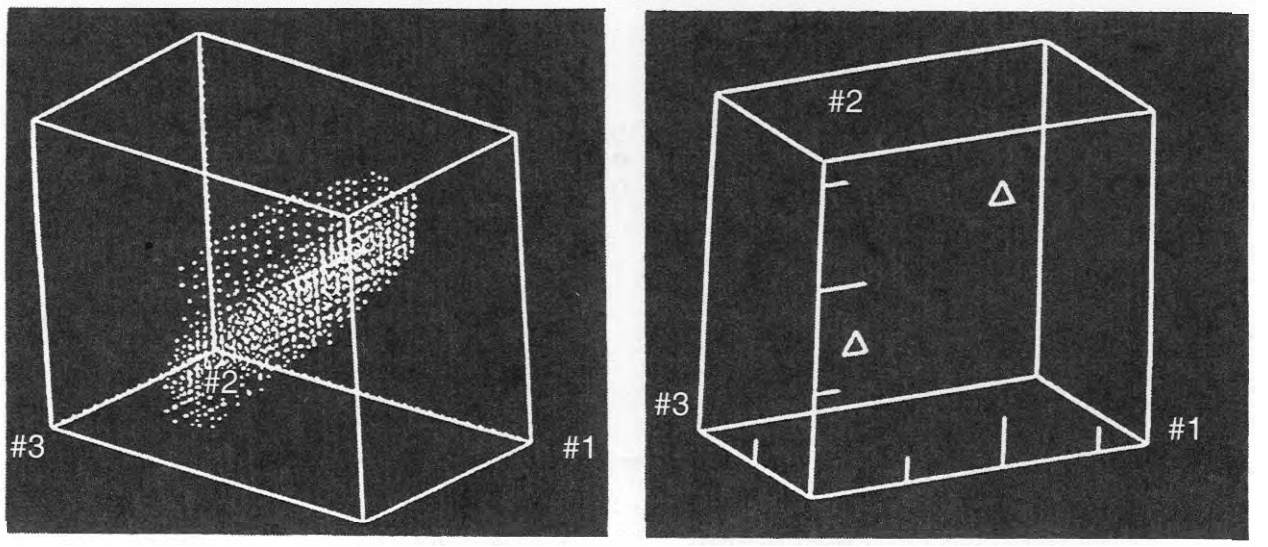


Figure 3.32. Change in ***OMAP*** (projection of three of the five dimensions) as new compounds were introduced to analysis of ACE inhibitors (397). Left is original ***OMAP*** of compound ***1*** (Fig. 3.30). Right is ***OMAP*** after completion of analysis.

available for analysis determine the methodology to be used. If there is a limited data set, then the pharmacophoric approach should be assessed first because of its fewer degrees of freedom. If no pharmacophoric patterns are consistent with the set of analogs, then introduction of logical molecular extensions to enable the active-site approach is warranted. Operationally, one first determines the set of potential pharmacophoric patterns consistent with the set of active analogs [leading to its name of Active Analog Approach (398)]. If there are sufficient data, then a unique pharmacophore, or active-site model, may be identifiable. The basic assumption behind efforts to infer properties of the receptor from a study of structure-activity relations of drugs that bind is the idea of ***complementarity.*** It follows that the stronger the binding affinity, the more likely that the drug fits the receptor cavity and aligns those functional groups that have specific interactions in a way complementary to those of the receptor itself. Certainly, our understanding of intermolecular interactions from studies of known complexes does not dissuade us of this notion, but may make us somewhat skeptical of the naive models that often result from such efforts. Andrews et al. (403) reviewed efforts of this type with regard to CNS drugs.

Clearly, the key to insight relies on chemical modification to determine the relative importance of functional groups for molecular recognition. Often more subtle effects than the simple presence or absence of a group are important and then comparison of molecular properties becomes of interest. A major impediment to analysis is the definition of a common frame of reference by which to align molecules for comparison. This is equivalent to solving the three-dimensional pharmacophoric pattern, and implies that one has distinguished those properties of the molecules under consideration in a manner similar to the receptor. Initial efforts to rationalize structure-activity relationships (SARs) among noncongeneric systems was hampered by an "RMS mentality." That is, a point of view that required atomic centers to align rather than overlap of steric and electronically similar grouping of atoms. An example would be requiring the six atoms of aromatic benzene rings to overlap at each of the six atoms of the ring vertices rather than simple requirements for coincidence and coplanarity that would recognize the torus of electron density that the rings share in common (Fig. 3.33). In congeneric series, the difficulty in assignment of correspondence is less (nonexistent by definition). This allows a variety of approaches, including those based on molecular graph theory (404–407), to detect similarities between molecules that can form the basis of a correlation analysis. Extrapolation outside of the group of congenerically related compounds on which the analysis was based would appear difficult, if not impossible.

Although it is simpler to start an analysis with a congeneric series to identify the recognition elements, diversity in chemical struc-

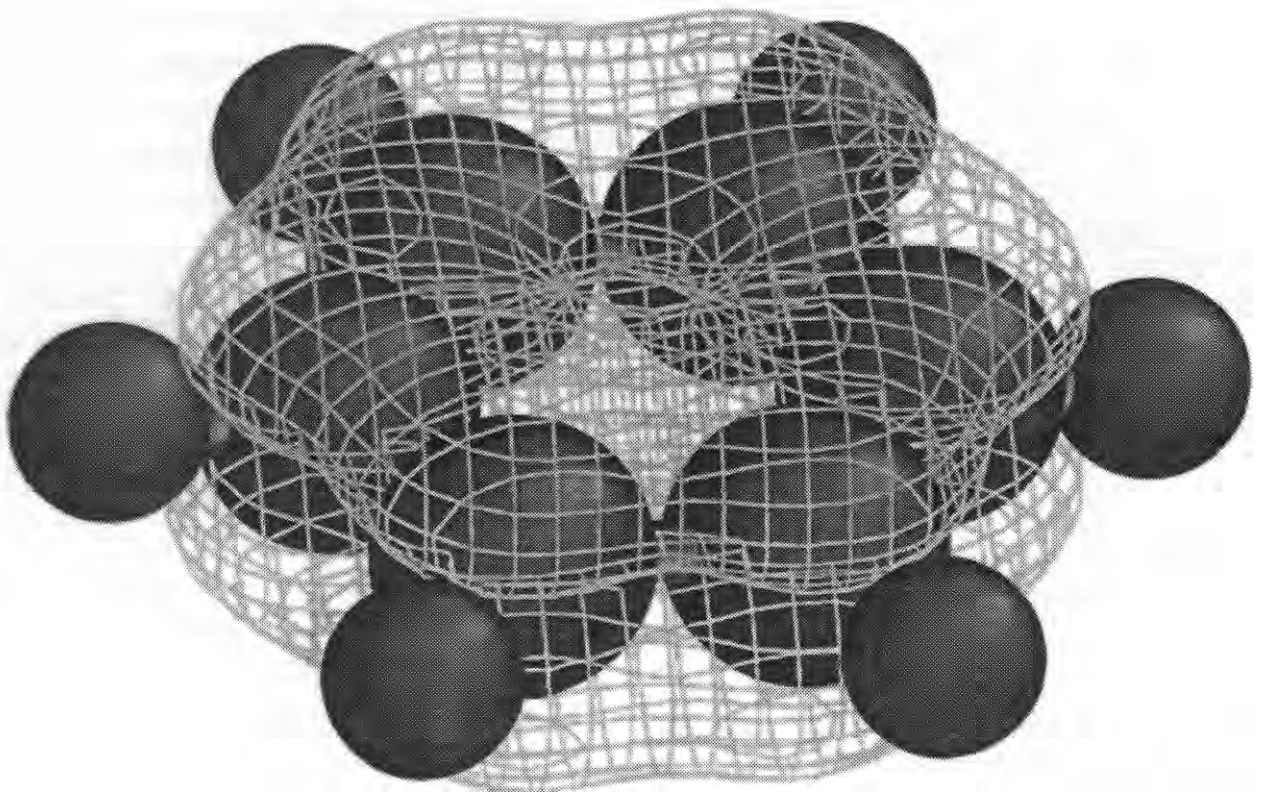


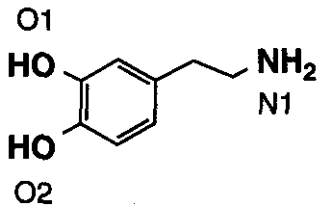
Figure **3.33.** Torus of electron density representing benzene ring. Atom-to-atom correspondences of ring atoms used in normal fitting routines lead to overconstrained fits.

tures implies more information regarding the conformational requirements of the system. **A** congeneric series requires that the basic chemical framework of the molecule remains constant and that groups on the periphery are either modified (e.g., aromatic substitution) or substituted (e.g., tetrazole for carboxyl functional group). Implicit in this concept is the notion that the compounds bind to the receptor in a similar fashion and, therefore, the changes are localized and comparable for each position of modification. Introduction of degrees of freedom in the substituents as well as consideration of differences in properties that are conformationally dependent, such as the electric field, require conformational analysis in an effort to determine the relevant conformation for comparison.

The problem can be divided into two: what are the aspects of the molecules that are in common and that may provide the basis for molecular recognition, and which conformation for each molecule is appropriate to consider. For the first problem, studies on a congeneric series can often yield valuable insight. For determination of the three-dimensional arrangement of the crucial recognition elements, diversity in the chemical scaffolds imloses different constraints on possible three-dimensional patterns and generates an opportunity for determining a unique solution.

4.2 Searching for Similarity

4.2.1 Simple Comparisons. To gain insight into molecular recognition, subtle differences in molecules must be perceived. Comparisons can be divided into two categories: those that are independent of the orientation and position of the molecule and those that depend on a known frame of reference. Simple comparisons deal with properties independent of a reference frame. For example, the magnitude of the dipole moment is frame independent, but the dipole itself is a vectorial quantity dependent on the orientation and conformation of the molecule, Similarly, the bond lengths, valence angles and torsion angles, and interatomic distances are independent of orientation. The distance matrix, composed of the set of interatomic distances (Fig. 3.34), is a convenient representation of molecular structure that is invariant to rotation and translation of the molecule, but which *reflects* changes in internal degrees of freedom. The distance range matrix is an extension (Fig. 3.34) that has two values for each interatomic distance



(a)

O1	2.79	4.74
2.79	O2	6.32
4.74	6.32	N1

(b)

O1	2.79	7.42
2.79	O2	8.05
4.61	5.88	N1

U

L

Figure 3.34. Distance matrix (a) in which unique interatomic distances for a particular conformation of a molecule are stored. Distance range matrix (b) in which ranges of interatomic distances representing conformational flexibilty of molecule are stored. U = upper bound, L = lower bound.

representing the upper and lower limits, or range, allowed for a given interatomic distance arising from the conformational flexibility of the molecule. Crippen (408) developed a procedure that will generate conformations that conform to the constraints represented by such a distance range matrix. This approach is used to generate structures from experimental measurements such as nuclear Overhauser effects in NMR experiments. The use of distance range matrices in the identification of pharmacophoric patterns was initially illustrated by Marshall et al. (398) (Fig. 3.35), and has recently been used by Clark et al. (409) in three-dimensional databases for representing the conformational flexibility of molecules. Pepperrell and Willett (410) examined several techniques for comparing molecules by use of distance matrices. Other descriptors for comparison of pharmacophoric

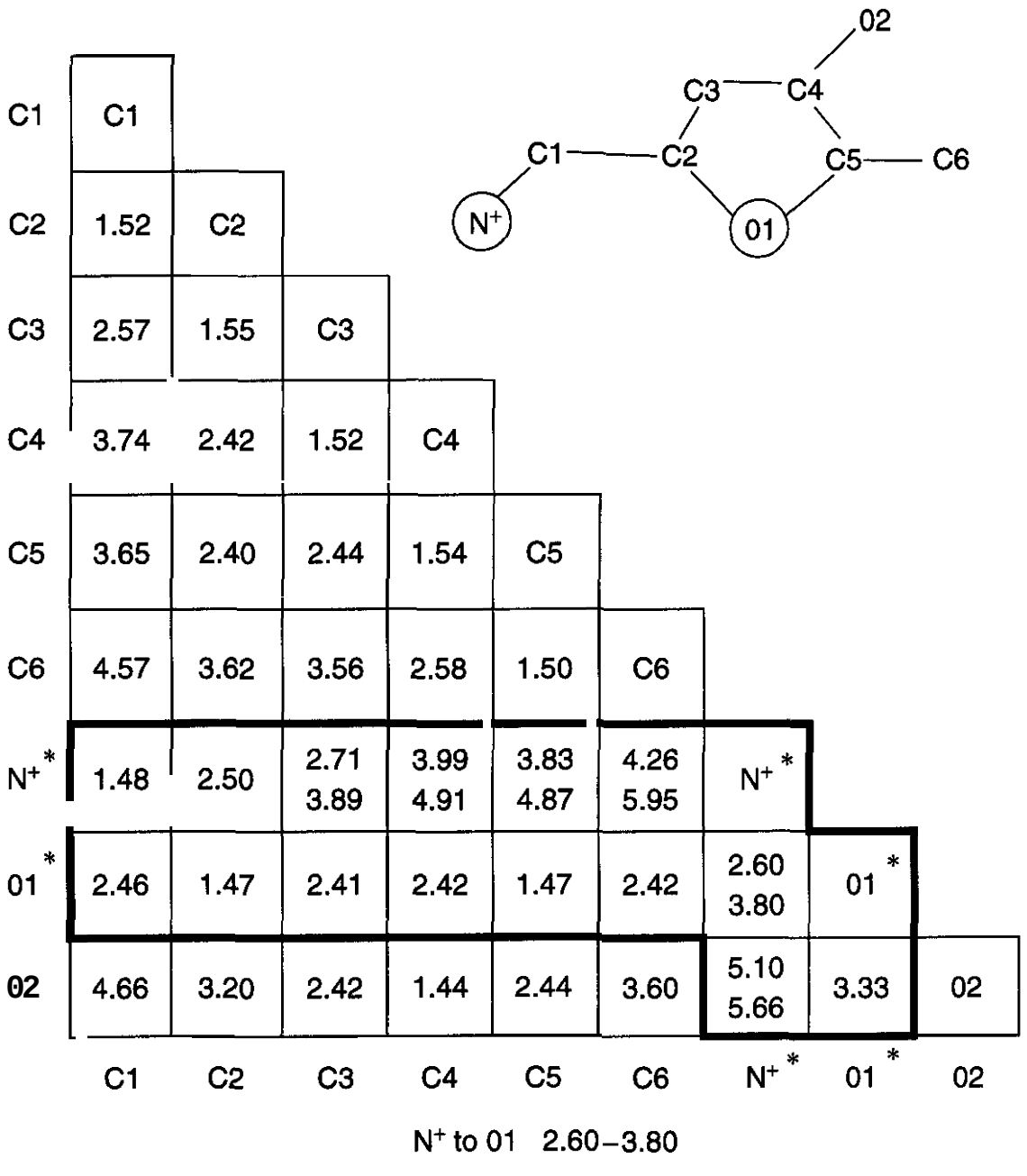


Figure 3.35. Distance range matrices used for illustration of analysis of muscarinic receptors (398). Used with permission.

patterns and retrieval of similar substructures are under active investigation (411).

4.2.2 Visualization of Molecular Properties (412). Although straightforward displays of molecular structure have proved to be extremely useful tools that enable medicinal chemists to visualize molecules and to compare their structural properties in three dimensions, of even greater potential utility is the display of the various chemical and physical properties of molecules in addition to their structures. Such displays allow the comparison not only of molecular shapes and three-dimensional structures, but also of molecular properties such as internal energy, electronic charge distribution, and hydrophobic character. A number of different properties have been displayed (412) in this manner in an effort to gain insight into molecular recognition in a series of compounds.

Among the more useful properties is the electrostatic potential. Any distribution of electrostatic charge, such as the electrons and nulei of a molecule, creates an electrostatic potential in the surrounding space that at any given point represents the potential of the molecule for interacting with an electrostatic charge at that point. This potential is a very useful property for analyzing and predicting molecular reactive behavior. In particular, it has been shown to be an indicator of the sites or regions of a molecule to which an approaching electrophile or nucleophile is initially attracted or from which it is repelled (Fig. 3.36).

The major obstacle to use of electrostatic potentials in the comparison of different molecules has been the sheer volume of information produced. The traditional means of displaying such large amounts of data has been to display the electrostatic potential around a molecule as a two-dimensional contour map. The advent of computer graphics techniques have improved the situation by allowing three-dimensional contour maps to be displayed in color on the graphics screen and manipulated in real time along with a display of the molecule itself. An alternative mode for displaying molecular electrostatic potentials is to employ a dotted surface representation, with the dots taking on an appropriate color according to the electrostatic potential value at the relevant location. Such techniques were initially derived to display empirically determined potentials on the surface of proteins, but have since been used widely to display the electrostatic potentials on sets of small molecules for comparative purposes.

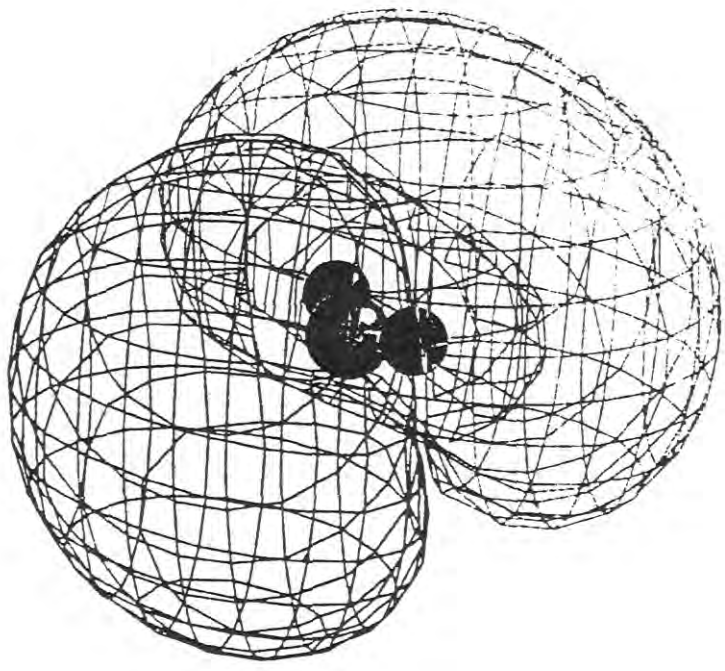


Figure 3.36. Molecular electrostatic potential for water. Positive potential superimposed on right surrounding hydrogens. Negative potential on left surrounding oxygen.

Other graphical uses of the electrostatic potential have been developed by Davis et al. (413), who were able to graphically align cyclic AMP and cyclic GMP, based on the superimposition of their respective electrostatic potential minima, and by Weinstein et al. (414), who oriented 5-hydroxytryptamine and 6-hydroxytryptamine based on the alignment of an electrostatically derived "orientation vector."

In a similar procedure to that described for the display of electrostatic potential, Cohen and colleagues developed a technique whereby the steric field surrounding a molecule can be displayed on a graphics screen as a three-dimensional isopotential contour map (415). The map is generated by calculating the VDW interaction energy between the molecule and a probe atom or molecule placed at varying points around the molecule of interest. This interaction energy is then contoured at specific levels to give the most stable VDW contour lines around the molecule, that is, the contour that represents the most favorable steric position for the probe as it is moved around the target.

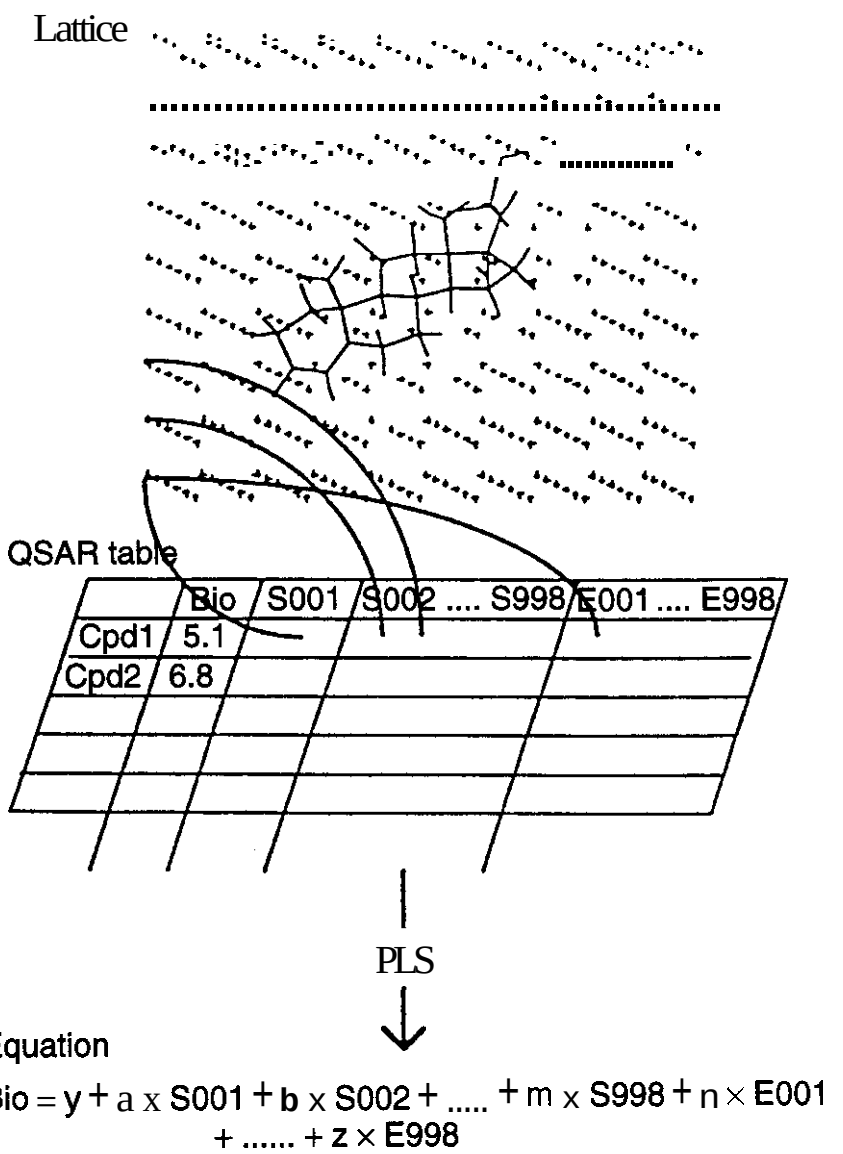


Figure 3.37. Calculation of electrostatic and **VDW** fields surrounding a series of molecules in defined orientations are used as a basis for 3D QSAR correlations in CoMFA (187, 401). Used with permission.

A similar three-dimensional contour representation of a molecule can be obtained for both the electrostatic and steric fields of a molecule within the comparative molecular field analysis (CoMFA) methodology that has been developed by Cramer (187) to investigate 3D-QSARs (400). In this procedure, the molecule is surrounded by a regular lattice of points, at each point of which a van der Waals and an electrostatic interaction energy between the molecule and a probe atom is computed (Fig. 3.37). Isocontours can then be generated around individual molecules, displayed graphically, and they can be statistically compared throughout a series of molecules in an attempt to generate 3D-QSARs and hence to rationalize activity data. This is very similar to the GRID program (186), which uses various probe groups (416) to map potential interactions around a molecule. Inductive logic programming has been combined with CoMFA to develop a new approach (417) to pharmacophore mapping that does not require explicit superimposition of compounds.

In situations where, either from previous QSAR work or from experimental evidence, it is known or suspected that differences in the reactivity of a set of molecules are attributed primarily to their hydrophobic rather than their electrostatic properties; it is probably of more use to compare molecular surfaces that display hydrophobicity or polarity information. Indeed, dotted molecular surfaces color-coded by hydrophobic character have been used very successfully by Hansch and coworkers to rationalize QSARs from several different systems (418,419). This concept has been extended to calculate the hydrophobic field surrounding a molecule by Kellogg and Abraham (420,421) and utilized in CoMFA studies.

4.3 Molecular Comparisons

To compare molecules in a general way, a means of superposition, or correctly orienting the molecules in the same reference frame, must be available. A procedure for positioning an atom in the molecule at the center of the coordinate frame with other atoms positioned

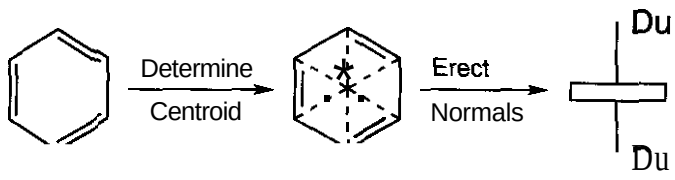


Figure 3.38. Construction of dummy vector perpendicular to plane of aromatic ring at centroid that allows superposition and coincidence of aromatic rings by fitting endpoints (Du) of dummy vector without requiring superposition of ring atoms.

along coordinate axes can be used, or the molecules can be successively fit to one that is used as the standard orientation. Danziger and Dean (422) described an approach that will find geometric similarities in positions of hydrogen-bonded atoms between two molecules. Least-squares-fitting procedures for designated atoms allow selectivity in orienting the molecules with predetermined conformations in the most appropriate manner. Kearsley (423) described an efficient method for fitting a series of molecules when atom-atom associations have been previously defined between members of the series. In some cases, the use of dummy atoms allows geometric superposition of groups such as aromatic rings without requiring superposition of the atoms composing the ring. By defining the centroid of the ring and erecting a normal to the plane of the ring, the dummy atom at the end of the normal and the centroid dummy atom can be used to superimpose the ring on another ring with similar dummy atoms (Fig. 3.38). This method leads to coincidence and coplanarity of the two ring systems without requiring the atoms composing the rings to be coincident. In other words, the rings can be viewed as two toruses of electron density without overemphasizing the positions of the atomic nuclei. In numerous studies [see review by Andrews et al. (403)] of biogenic amine ligands, this method of comparison of the aromatic ring components is essential to allow alignment of the nitrogens.

4.3.1 **Volume Mapping.** One method of displaying molecular surfaces that retains the ability to transform the display interactively has been developed by Marshall and Barry (424). The procedure involves computing a molecular pseudo-electron density map on a

Atom	Temperature factor	Atomic number
Carbon	60	25
Nitrogen	55	25
Oxygen	50	25
Sulfur	67	35
Phosphorus	70	35
Hydrogen	40	15
Bromine	65	50
Chlorine	60	35
Fluorine	40	40
Iodine	80	53
Sodium	87	85
Potassium	130	85
Calcium	130	85
Lithium	65	50
Aluminum	80	53
Silicon	80	55

Figure 3.39. Set of parameters to generate pseudo-electron density maps of molecules that can be contoured to approximately represent VDW surface (Ho and Marshall, unpublished).

three-dimensional grid that surrounds the molecule whose atoms are replaced by dummy Gaussian atoms. Atom types are characterized by a half-width and an integrated density, chosen so that the Gaussians have a fixed value at a distance equal to the VDW radius (Fig. **3.39).** Such density maps may be contoured in three dimensions to provide a chicken wire-like envelope around the molecule that corresponds to the van der Waals surface.

A concomitant benefit of this technique is that estimates of the molecular surface area and volume are generated as by-products of the contouring routines, whether the surface is being drawn around one or several molecules. Additionally, the generated surfaces and volumes are readily susceptible to logical operations, such as union, intersection, or subtraction, enabling the rapid determination of, for example, union or difference volumes among a series of molecules.

Once one has fixed the molecules in a common frame of reference, then comparison by a variety of techniques becomes feasible. As an example, difference in volume may be important in understanding the lack of seen activity in compounds that appear to possess all the

prerequisites for activity seen in others in the series. In a congeneric series, a significant portion of the molecular structure is common to the molecules under comparison. This common volume that is shared logically should not contribute to differences in activity. By subtraction of the volume shared by two molecules, one obtains a difference map in which the volume occupied by one molecule and not the other remains (398). Correlations between the shared volume and the biological activity of a congeneric series of inhibitors of DHFR have been shown by Hopfinger (425). Simon and his colleagues (426) emphasized the use of both overlapping volume and nonoverlapping volume in QSAR studies in a quantitative methodology, the minimal steric difference, or MTD method. This approach has been enhanced to allow comparison of low energy conformers of each molecule and use of those that are sterically most similar. An application to substrates of acetylcholinesterase illustrates this facility (427).

4.3.2 Field Effects. Once the frame of reference has been established, other properties of molecules, such as the electrostatic field, can be compared as well. Because the electrostatic properties can be sampled on a grid, differences between the values of two molecules can be calculated and a difference map contoured. Such difference maps (428) highlight more clearly the similarities and differences between molecules. Hopfinger (429) integrated the difference between potential fields and showed this parameter to be useful in QSAR studies.

An approach to statistically quantifying the similarity between two molecular electrostatic potential surfaces was developed by Dean and coworkers (430,431) and by Richards and coworkers (215). Here, the previously determined molecular electrostatic potential surfaces are projected outward onto surrounding spheres that provide a common surface of reference, and then statistical analyses are performed over the points on this common surface in an attempt to quantify the similarities or differences between the two molecules under consideration. Burt and Richards (432) introduced flexibility in the comparison of molecules based on their electrostatic potential fields.

4.3.3 Directionality. If one is comparing molecules that share interaction at a common site on a biological macromolecule, it is logical to assume that they may do so by interacting with similar sites in the receptor with optimal interaction shown by molecules with correctly oriented functional groups. If one does not have a three-dimensional model of the receptor from which to deduce potential interactive sites, then one can only attempt to deduce the potential interactive receptor-subsites by examination of the molecules that interact with them. Systematically, one can vary the conformation of a molecule and record the relative orientation of groups postulated, or shown experimentally, to play a dominant role in intermolecular interactions. In this way, one can map out the directionality of interactions of each functional group of the ligand in a common frame of reference. Comparison of these maps can often lead to hypotheses regarding pharmacophoric groups and their correspondence between molecules.

4.3.4 Locus Maps. One can generate a locus plot in coordinate space showing all the potential locations of one group relative to another by fixing one group in a particular orientation as a frame of reference and recording all possible coordinates of the other. An example would be the relative positions of the basic nitrogen to the aromatic ring in compounds such as dopamine interacting with biogenic amine receptors. One must choose the common fragment (in the example, the aromatic ring) of each molecule and its orientation to generate a similar frame of reference, so that the locus of positions of the atom (the basic nitrogen) leads to a meaningful comparison across a series of molecules (Fig. 3.40).

4.3.5 Vector Maps and Conformational Mimicry. Often, one is more interested in accessing the directionality of potential interaction rather than simply looking for overlap of atoms such as the basic nitrogen. In this case, for example, one is interested in determining both the locus of the lone pair of the nitrogen

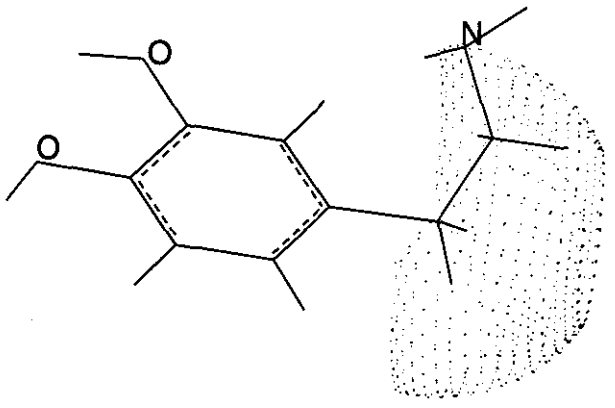


Figure 3.40. Locus of sterically allowed positions of nitrogen atom in dopamine relative to aromatic ring.

and the nitrogen as the ordered pair of coordinates determines a vector in the chosen frame of reference. The resulting plot of the locus of all possible vectors of the nitrogen lone pair constitutes a vector map. The combination of positional information with relative orientation offers considerable insight into potential interactions with a hypothetical receptor. The work of Lloyd and Andrews (233) postulating a common theme in CNS receptors based on an underlying biogenic amine pattern can be rationalized using the vector-map approach.

The use of vector maps is essential to the assessment of conformational mimicry, in that one attempts to determine the statistical probability that the conformation essential for activity will be preserved with a given chemical modification. An example will serve to illustrate this concept and its application. Modification of arnide bonds (introduction of amide isosteres) in peptide drugs to increase metabolic stability may alter the potential accessible conformations. This may preclude the compound containing the isostere from adopting the correct orientation for receptor recognition and activation. In the general case, one has no specific information regarding which particular conformation is biologically relevant and can only assess whether the chemical modification mimics the amide bond in its conformational effects. This can be quantitatively assessed by the comparison of the percentage of vectors of the vector map of the parent amide bond that can be found in a comparable vector map of the analog.

Work by Zabrocki et al. (433) on the use of 1,5-disubstituted tetrazole rings as surrogates for the cis-amide bond illustrates this application. The linear dipeptide, acetyl-Ala-Ala-methylamide, with the amide bond between the two alanine residues in the cis-conformation, and the tetrazole analog, acetyl-AlaΨ[CN_4]Ala-methylamide, were modeled using the coordinates derived from diketopiperazines for the cis-amide bond or from the crystal structure of the cyclic tetrazole dipeptide. A systematic, or grid, search, which determines the sterically allowed conformations by systematically varying the torsional degrees of freedom, was used to generate a Ramachandran plot for each of the pairs of backbone torsional angles (Φ, Ψ) associated with each amino acid residue. The rigid geometry approximation was used with the set of scaled VDW radii, shown by Iijima et al. (109) to reproduce the experimental crystal data for proteins and peptides. When the cis-amide dipeptide model was calculated, the orientations of the C^{α}-C^{β} bond of Ala-1 with the methylamide fixed as a frame of reference were recorded for each sterically allowed conformation (Fig. 3.41). Use of the same orientation of the methylamide in the tetrazole allowed the

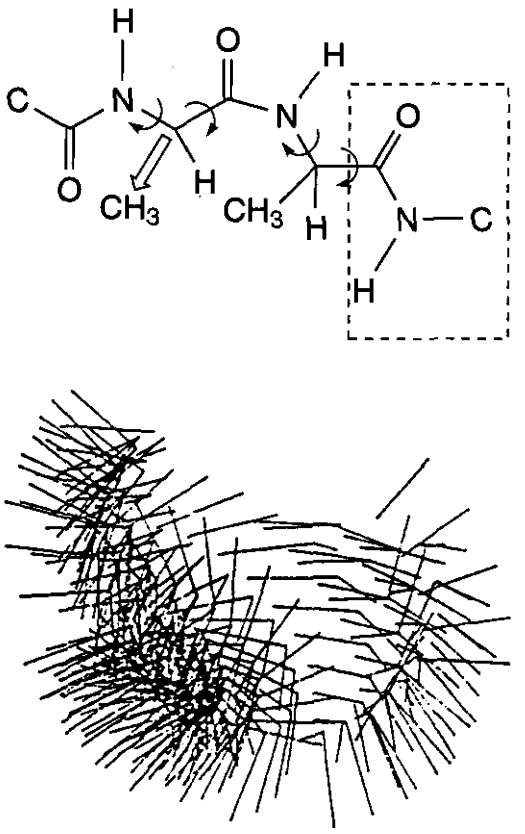


Figure 3.41. Vector map of the orientations of the C^{α}-C^{β} bond of Ala[1], with the methylamide fixed as a frame of reference of the dipeptide Ac-Ala-Ala-NH-CH, in which the central amide bond was cis (433). Used with permission.

program to determine which vectors, or orientations of the Ala-1 side chain relative to the methylamide, were common to both dipeptides. Alternatively, the acetyl group was used as the fixed frame of reference and the side-chain orientation of Ala-2 was used to monitor conformational mimicry. Because the quantitative results were essentially the same, the measurement of mimicry was shown to be independent of the chosen frame of reference. A torsional increment of 10 degrees was used, and a side-chain vector was assumed to correspond if both the carbon-a and carbon-β were within 0.2 Å of the coordinates of another vector. The percentage of orientations available to the analog that are available to the parent is referred to as the conformational mimicry index. For the tetrazole surrogate of the *cis*-amide bond, the conformational mimicry index is 88% [the number of vectors (747) common to both the tetrazole and cis-amide divided by the total number of vectors (849) allowed for the cis-amide]. The tetrazole analog has more conformational freedom than the cis-amide model with 33,359 conformers allowed compared to 14,912 allowed for the cis-amide of the 36^4 (or 1,679,616) possible conformations. This difference was easily visualized in plots of the vector maps for the two dipeptides.

A more recent example of the use of vector maps to evaluate conformational similarity is an application to β-turn mimetics by Ball et al. (434,435). This led to a recognition that many of the various turn types described in peptides based on their backbone dihedral angles lead to quite similar topographical arrangements of the side chains. A new parameter, β [the dihedral angle formed by the backbone atoms $C_{(1)}$-$\alpha C_{(2)}$-$\alpha C_{(3)}$-$N_{(4)}$], was described (Fig. 3.42) that more readily facilitated comparison of the topography of the system.

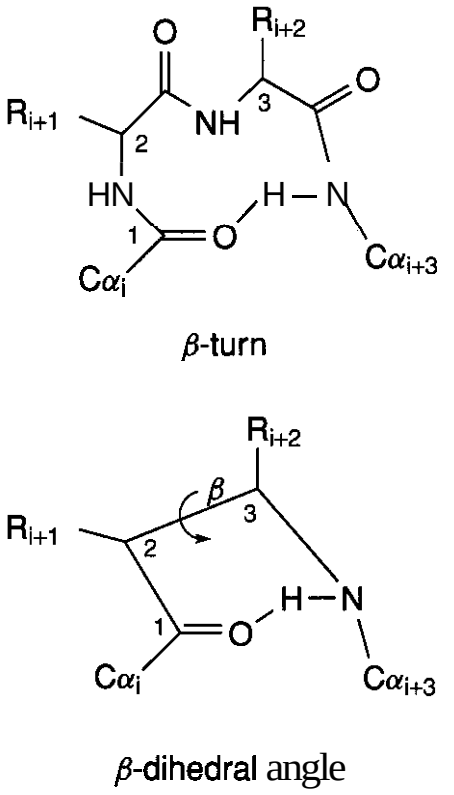


Figure 3.42. Definition of new parameter β, the dihedral angle between the backbone atoms (&I-d&,-&(„-N(„ of peptides, used to describe the topography of reverse turns **(434,435).**

4.4 Finding the Common Pattern

If one assumes that a common binding mode exists for two or more compounds, then one can use the computer to verify the geometric feasibility of the assumption. One needs to determine whether it is possible for the two molecules to present a common geometric arrangement of the designated "important" functional groups for recognition. There are two distinct approaches to this problem. The first that is associated with minimization methodology focuses on the existence issue. Is there a conformation that is energetically accessible to each of the molecules under consideration that will place the designated functional groups in a similar orientation? The second approach attempts to systematically enumerate all possible conformations and thereby derive all possible orientations or patterns to determine the set of patterns shared by the compounds under study. The latter approach, when it can be applied, can directly address the question of uniqueness of the common pattern.

The search for the global minimum, or complete set of low energy minima, on a potential surface is a common problem in science and engineering that does not have a general solution. Numerous approaches in chemistry have been used: most commonly stochastic methods such as distance geometry (408), molecular dynamics, and Monte Carlo sampling. Although distance geometry and molecular dynamics are widely used in the elucidation of solution conformations from NMR data, they have problems in conformational sampling and homogeneous treatment of data from

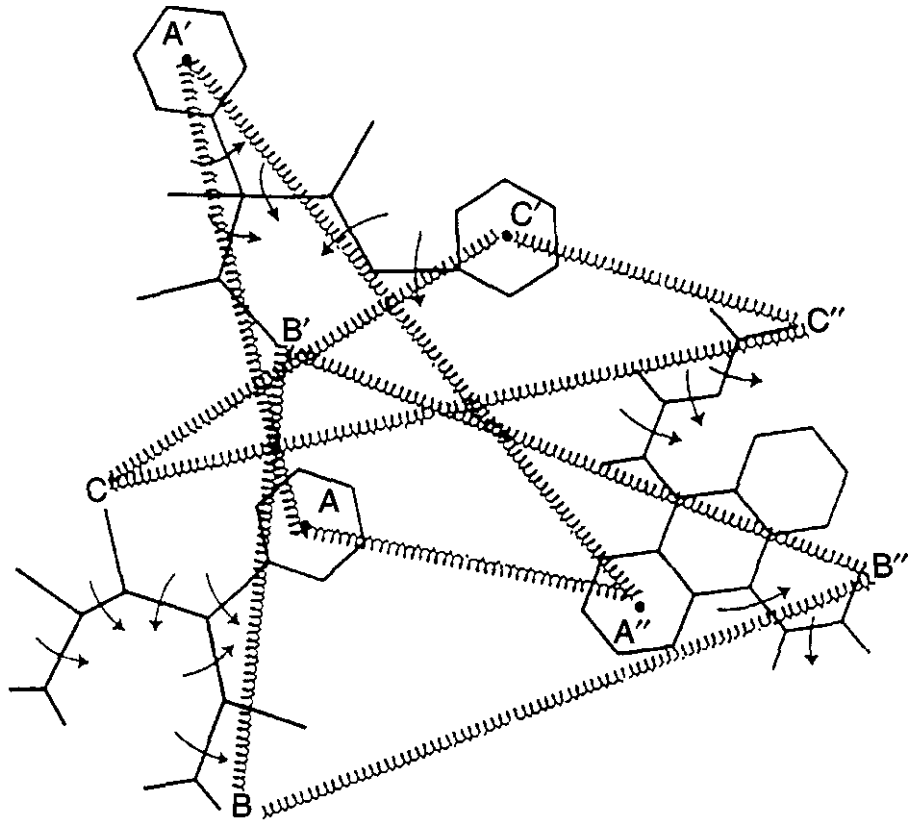


Figure 3.43. Simultaneous minimization of molecules to force overlap of pharmacophoric groups **A,** B, and **C.** Springs represent constraints between groups and only interatomic forces evaluated.

rigid and mobile domains. In general, the difficulties with most methods are similar to those seen with minimization procedures. If one is in the area of the global minimum, then one is likely to converge to that solution. Otherwise, one will be trapped in some local minimum. In contrast, systematic search methods are algorithmic, so that all sterically allowed conformations are generated at the selected torsional grid parameters. Systematic search methods, therefore, do not have problems in sampling and are path independent, but are combinatorial in complexity, which may limit the fineness of the sample grid and thus compromise the results. Only in small systems such as cycloalkane rings (121) and small peptides (90, 436) have the potential energy hypersurfaces been mapped.

4.4.1 Constrained Minimization. In cases where one has internal degrees of freedom, besides the six associated with position and orientation. the use of constrained minimization procedures becomes a useful technique. Often the standard molecule for comparison has a fixed conformation and the molecule to be fitted has internal degrees of freedom. Several groups have published methods for dealing with this problem. In case one has simultaneous degrees of freedom in both the molecule to be fitted and the target, a different approach with simultaneous minimization of all variables is recommended (Fig. 3.43).

The combination of molecular mechanics with flexible minimization routines allows penalty functions to be assigned to force geometrical correspondence of groups, whereas individual molecules have their internal energy evaluated, but are invisible to the other molecules under consideration. A program has been described (437) with this capability and its use illustrated on histamine antagonists by Naruto et al. (438). Template forcing allows one molecule to be set up as a template and another molecule to be constrained to overlap in a specified manner. The strain energy involved in forcing correspondence gives an upper-bound estimate of the distortion energy required, given that the results depend on the initial-problem definition.

An alternative approach uses the distance geometry paradigm, in which all the constraints are combined to form the distance matrix from which energetically feasible conformations of the set of molecules are sought mathematically. Sheridan et al. (439) demonstrated this approach on acetylcholine analogs that are muscarinic agonists. Both of these approaches ask the same question and suffer from the same limitations, and differ only in computational technique. Each suffers from the local minima problem, in that each uses a

minimization technique, and the results will be dependent on the starting geometries of the initial set of molecules. Both have the advantage that the unique constraints imposed by particular molecules enter consideration at an early stage and minimize comparison of conformations.

Another variant recently reported by Hodgkin et al. (440) uses a Monte Carlo search procedure to generate candidate pharmacophoric patterns. A reduced force-field parameter set is used initially to lower energy barriers between conformations to ensure greater configurational sampling. Candidate pharmacophores are then refined to produce low energy conformations of molecules overlaid in a common binding mode. Application to antagonists of the human platelet-activating factor led to a consistent binding model for a set of five diverse structures when active-site hydrogen-bonding groups were postulated. Barakat and Dean (441, 442) utilized simulated annealing to optimize structure matching by minimizing the difference matrix between the two molecules. A somewhat similar approach is that of Perkins and Dean (443), who used simulated annealing to search conformational space followed by cluster analysis for each molecule, with subsequent comparison of a small number of diverse conformers between different molecules.

4.4.2 Systematic Search and the Active Analog Approach. Once the existence of a common pattern has been determined, then the issue of uniqueness needs to be addressed. The Active Analog Approach (398) uses a systematic search to generate the set of sterically allowed conformations based on a grid search of the torsional variables at a given angular increment. For each sterically allowed conformation, a set of distances between the postulated pharmacophoric groups are measured. The set of distances, each of which represents a unique pharmacophoric pattern, constitutes an OMAP. Each point of the OMAP is simply a submatrix of the distance matrix and, as such, is invariant to global translation and rotation of the molecule. If the initial assumption is valid, that the same binding mode of interaction, or pharmacophoric pattern, is common to the set of molecules under consideration, then the OMAP for each active molecule must contain the pattern encrypted in the set of distances. By logically intersecting the set of OMAPs, one can determine which patterns are common to all molecules (444). In other words, all potential pharmacophoric patterns consistent with the activity of the set of molecules can be found by this simple manipulation of OMAPs, and the question of uniqueness addressed directly (Fig. 3.44).

A good example is the work of Nelson et al. (445) on the receptor-bound conformation of morphiceptin. Based on structure-activity data, the tyramine portion and phenyl ring of residue three of morphiceptin, Tyr-Pro-Phe-Pro-NH,, were postulated to be the pharmacophoric groups responsible for recognition and activation of the opioid preceptor. It was assumed further that the aromatic rings bound to the receptor in the different analogs were coincident and coplanar. A series of active analogs with a variety of conformationally constrained amino acid analogs in positions two and three were analyzed. A unique conformation was found for the two most constrained analogs that allowed overlap of the Phe and Tyr portions of the molecules (Fig. 3.45). In this case, a five-dimensional orientation map with distances between the nitrogen and normals to the two aromatic rings was used in the analysis.

The Active Analog Approach (Fig. 3.46) is appropriate for the unknown receptor problem, given that no objective criteria function, such as potential energy, can be used *a priori* in the absence of information regarding the receptor. Adequate sampling of the potential surface to ensure that the complete set of local minima is found is still problematic because of the phenomenon known as "grid tyranny." This relates to the fact that the combinatorial explosion that results by decreasing the increment of the torsion angles scanned limits one to a finite increment for a given problem, say, 10° for a seven-rotatable bond problem. Because the energetics of the system is very sensitive to interatomic distances, a conformation generated at the 10'' increment may be sterically disallowed, but very close to a minimum. Relaxation of the structure might find the relevant conformation, for example, by allowing a torsional angle to vary by 1°. Im-

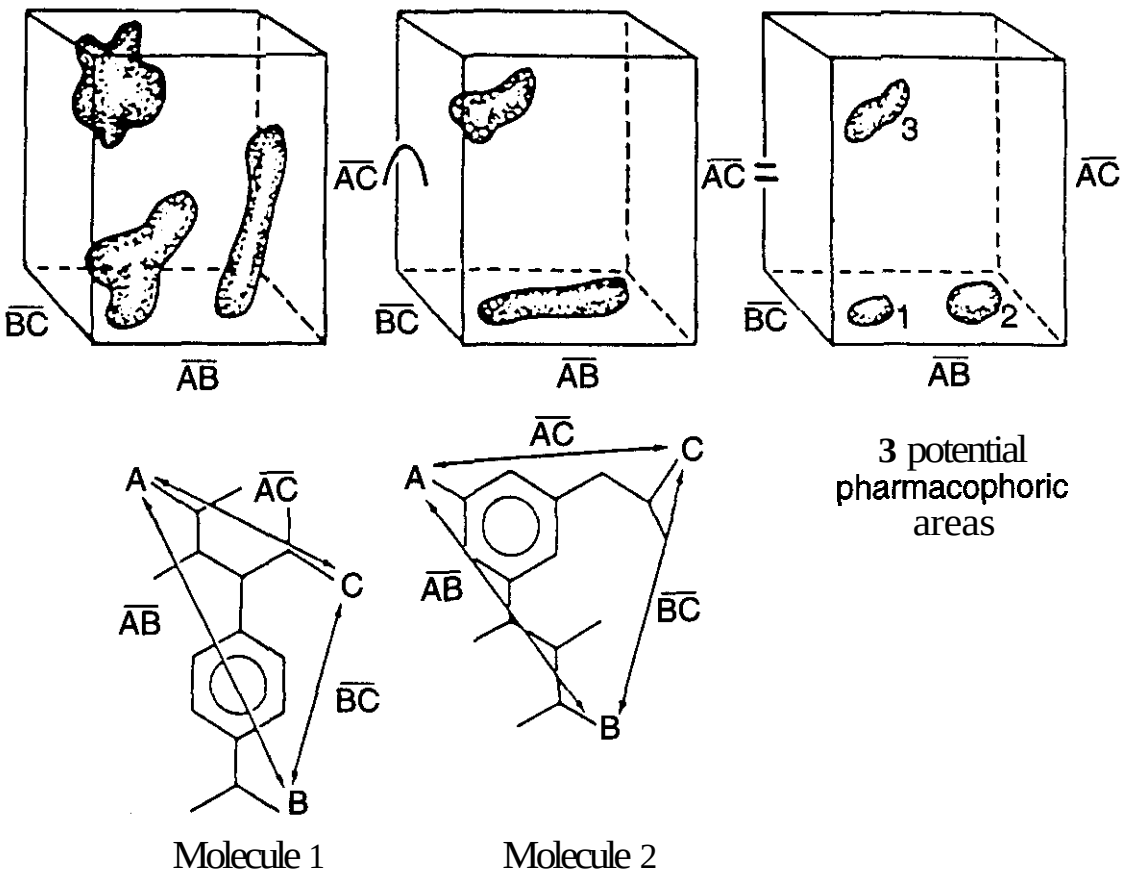


Figure 3.44. OMAPs generated for two molecules can be logically intersected to determine which three-dimensional patterns are common.

provements in algorithms described in the following section have helped to overcome this problem.

4.4.3 Strategic Reductions of Computational Complexity. Logically, the Active Analog Approach can be conceived as sequentially determining all the sterically allowed conformations for each molecule under consideration, generation of an OMAP from those conformations, and logical intersection of the OMAPs to determine the common pharmacophoric patterns. A simple analysis will easily convince one that this is not feasible because of the computational complexity of the problem. For example, the set of 28 ACE inhibitors' (Fig. 3.31), analyzed by Mayer et al. (397), have a total of 163 torsional degrees of freedom that have to be explored to find a common pattern, as seen in Table 3.1. If we were to determine all possible conformations for each molecule at 10° torsional scan, the scan parameter (s) = 10° and the number of torsional increments r = 360°/s, or 36. For each molecule, there are r^n possibilities to be examined. For the set of molecules there are $(6 \times 36^3) + (7 \times 36^5) + (3 \times 36^6) + (5 \times 36^7) + (6 \times 36^8) + (1 \times 36^9)$ possible conformations to be generated and examined. If one compares each conformation of each molecule with all the conformations of the other molecules to find possible correspondences, the combinatorials of the problem explode and one reaches the same level of complexity as a complete conformational search of a peptide of 30 residues at a 10° scan (not currently feasible).

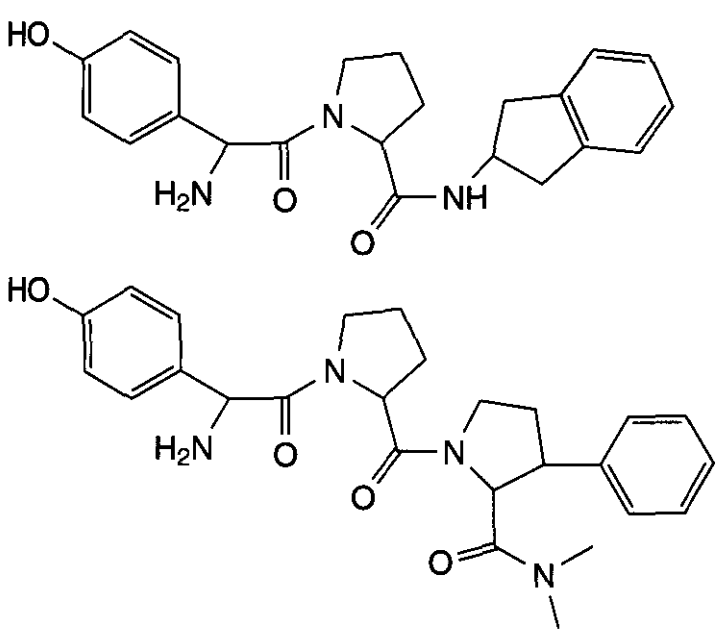


Figure 3.45. Conformations of two constrained analogs of morphiceptin in which aromatic rings of Tyr^1 and Phe^3 are overlapped (445).

One is not interested in the conformational

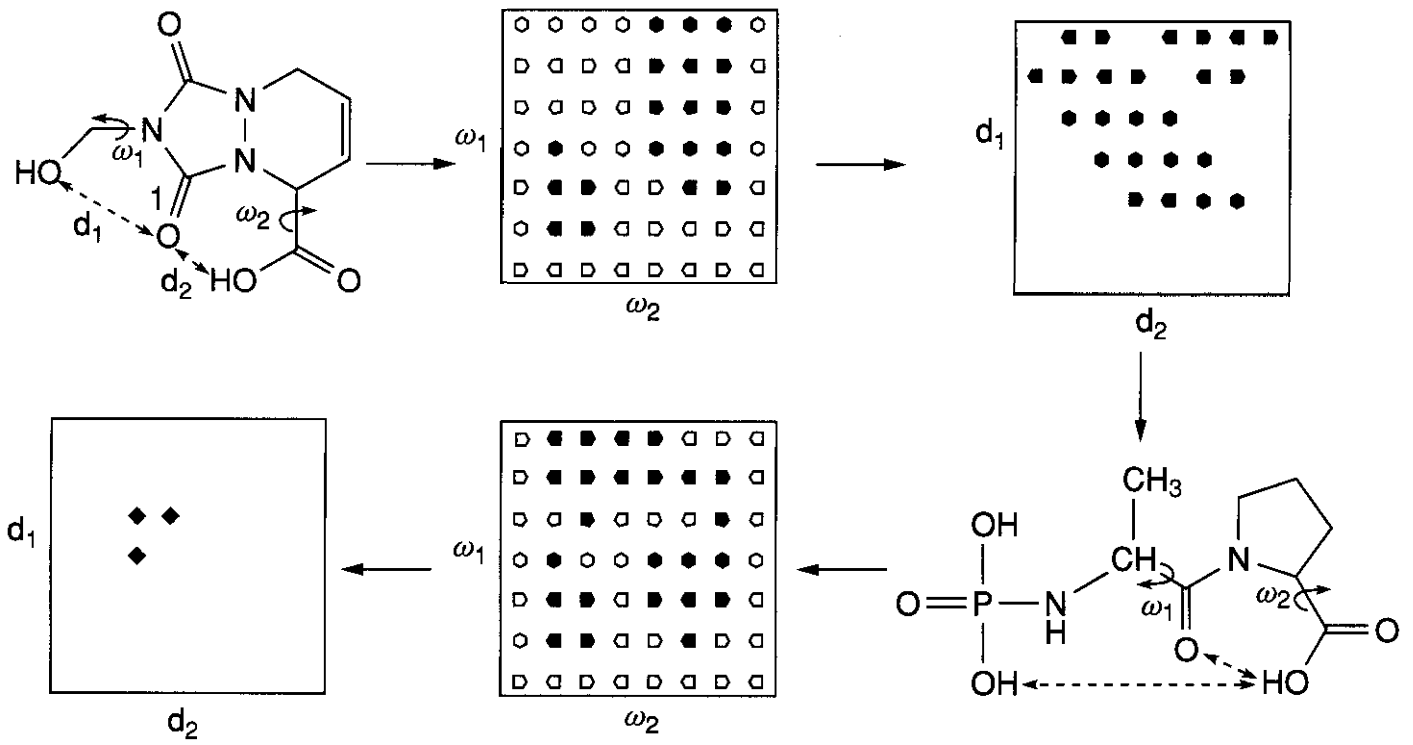


Figure 3.46. The flow of information in the Active Analog Approach (111,399).Sterically allowed conformations (represented by filled circles on the ω_1, ω_2 torsional grid) of a molecule are determined and the distances (d_1, d_2, etc.) between pharmacophore elements are recorded for each. The resulting OMAP is used to constrain the next molecule in the series. Ideally, once all of the molecules have been evaluated, only a single point or cluster of points remains in the OMAP.

hyperspace of the set of the inhibitors, but rather the three-dimensional patterns common to the total set of inhibitors. Many conformations of a molecule often map into one three-dimensional pattern. Transformation of the multidimensional conformational hyperspace in a smaller-dimensioned OMAP space reduces the number of objects for comparison. If one starts with the most constrained inhibitor (fewest torsional degrees of freedom) and determined an OMAP for it, then one can use the upper and lower distance bounds as constraints for searches for the next molecule. In other words, one looks only where there are possible solutions to the problem. A more advanced approach simply examines each candidate solution from the initial OMAP to see whether all the other molecules are capable of presenting the same pattern. By changing the focus to the hypothesis of a common three-dimensional pattern, a more efficient approach has been devised (Fig. 3.46) (399).Clearly, the algorithms that one chooses to do the problem are important.

Table 3.1 Degrees of Torsional Freedom to Specify ACE Active Site Geometry

Degrees of Freedom (n)	Number of Molecules	Total
3	6	18
5	7	35
6	3	18
7	5	35
8	6	48
9	1	9
Totals	28	163

4.4.4 **Alternative Approaches.** A conceptually similar approach to receptor mapping has been taken by Ghose and Crippen (446–449), who used the distance geometry method to analyze site points and drug interactions. A site model was postulated with some initial estimates of force constants between the appropriate portion of the ligand and the site point. The binding energy for a particular binding mode can be calculated:

$$E_{\text{calcd}} = cE_c + Sx_{ti,tm}$$

where E_c is the conformational energy, c is a coefficient to be fit, x is the interaction of a site point i with the bound ligand point m, which depends on their types. The novel aspect of

this approach was the use of distance geometry to generate a variety of conformers binding within the postulated site and then finding a set of force constants between the postulated site points and ligand points that will predict the affinities of the compounds in the data set when bound in their optimal manner. With a site model of 11 attractive site points and 5 repulsive ones for DHFR, Ghose and Crippen (447) were able to derive force constants that fit 62 molecules, with an $R^2 = 0.90$, and predict the activity of 33 molecules, with an $R^2 = 0.71$. The compounds, however, are essentially an extended congeneric series because the core recognition portion of the inhibitor, the pyrimidine ring, is common to all the compounds.

Linschoten et al. (450) extended Crippen's method by use of lipophilicity to describe the binding of parts of the ligand to lipophilic areas of the receptor. Through the use of only a nine-point model of the turkey erythrocyte β-receptor and six energy parameters, they successfully modeled 58 compounds. Distance geometry approaches to receptor-site modeling have been reviewed (449,451).

Simon and his coworkers have developed (426) a quantitative 3D-QSAR approach, the minimal steric (topologic) difference (MTD) approach. Oprea et al. (452) compared MTD and CoMFA on affinity of steroids for their binding proteins and found similar results. Snyder and colleagues (453) developed an automated method for pharmacophore extraction that can provide a clear-cut distinction between agonist and antagonist pharmacophores. Klopman (404, 454) developed a procedure for the automatic detection of common molecular structural features present in a training set of compounds. This has been used to produce candidate pharmacophores for a set of antiulcer compounds (404). Extensions (454) of this approach allow differentiation between substructures responsible for activity and those that modulate the activity.

Bersuker and Dimoglo (455) described a matrix-based approach that combines geometric and electronic features of a molecule, the electron-topological approach. For each molecule, an electron-topological matrix of congruity (ETMC) is constructed based on a conformer selected by conformational analysis. The ETMC is essentially an interatomic distance matrix (Fig. 3.47), with the diagonal elements containing an electronic structural parameter (atomic charge, polarizability, HOMO energy, etc.). Off-diagonal elements for two atoms that are chemically bonded are used to store information regarding the bond (bond order, polarizability, etc.). Matrices for active compounds in a series are then searched for common features that are not shared by inactive compounds. The successful examples cited are predominately for small, relatively rigid structures where the conformational parameter does not confuse the analysis.

Martin et al. (456) developed a strategy for determining both the bioactive conformation and a superposition rule for each active molecule in a data set. In DISCO, a set of low energy conformers for each molecule is processed to locate atoms within the molecule and extensions for binding-site points for superposition. A clique-finding algorithm then finds superpositions containing at least one conformation of each molecule and a user-specified minimum number of site points.

Unlike methods that are limited to a precomputed set of rigid conformers, GASP (Genetic Algorithm Similarity Program) (457) allows full conformational flexibility of ligands. GASP employs a genetic algorithm for determining the correspondence between functional groups in different molecules and the alignment of these groups in a common geometry for receptor binding. For a set of ligands, GASP automatically identifies rotatable bonds and pharmacophore elements such as rings and potential hydrogen-bonding sites. A population of chromosomes is randomly constructed, where each chromosome represents a possible alignment of all the molecules. Chromosomes encode the torsion settings for rotatable bonds as well as the intermolecular mapping of elements. The fitness score of a particular alignment is the weighted sum of three terms: the number and similarity of overlaid elements, the common volume of all the molecules, and the internal van der Waals energy of each molecule. Using a mutation or crossover operator, child chromosomes are produced. Those with improved fitness scores replace the least-fit members of the existing population. The calculation terminates when

O_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	H_{17}
–0.23	5.09	5.22	5.69	6.59	4.22	2.82	2.30	1.75	3.54	4.74	4.96	6.11	6.98	7.09	6.84	1.89
	–0.01	1.00	2.52	2.48	2.54	2.97	4.33	5.14	5.15	4.96	3.79	6.29	6.45	6.79	7.50	6.23
		0.05	0.99	0.99	0.97	2.56	3.86	4.95	4.37	3.89	2.57	5.11	5.38	5.41	6.41	5.95
			–0.01	2.52	2.48	3.24	4.48	5.46	5.04	4.62	3.40	5.79	6.31	5.72	5.09	6.41
				–0.01	2.56	3.85	4.99	6.20	5.17	4.33	2.94	5.21	5.09	5.40	6.68	7.44
					0.05	1.42	2.46	3.69	2.84	2.49	1.42	3.87	4.37	4.38	5.06	4.61
						–0.03	1.40	2.39	2.43	2.83	2.40	4.36	5.07	5.05	5.20	3.40
							0.01	1.09	1.40	2.45	2.78	3.87	4.79	4.75	4.33	2.17
								0.18	2.40	3.69	4.13	4.98	5.94	5.89	5.12	0.93
									–0.03	1.41	2.40	2.58	3.64	3.60	2.92	2.70
										0.05	1.42	0.97	2.51	2.50	2.57	4.10
											–0.04	2.55	3.05	3.07	3.88	4.83
												0.05	0.99	1.00	0.99	5.16
													–0.01	2.53	2.49	6.22
														–0.01	2.49	6.07
															–0.01	4.97
																0.03

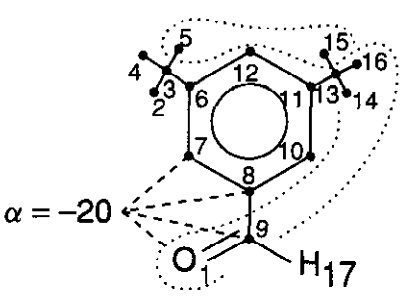


Figure 3.47. The electron-topological matrix of congruity (ETMC) for a 17-atom fragment proposed by Bersuker and Dimoglo (455) to encode geometrical and electronic features of molecules.

the fitness of the population fails to improve by a specified amount, or when the preset number of genetic operations is completed. GASP produces several sets of alignments and their associated pharmacophore elements.

4.4.5 Receptor Mapping. One can attempt to decipher physical properties of the receptor by use of data from both active and inactive analogs. Interpretation of results requires some understanding of the interactions between ligand and receptor that underlie molecular recognition. Oprea and Kurunczi (458) reviewed these interactions in the context of receptor mapping. A basic assumption is that a compound that contains the correct pharmacophoric elements and has the capability of positioning them correctly should be active. Compounds with these attributes that are inactive must be incapable of binding to the receptor in the correct orientation; that is, steric overlap with the receptor must occur. By calculating the combined volume of the active analogs superimposed in the correct orientation, one has mapped space that cannot be occupied by the receptor and that must be available for binding. Inactive compounds mentioned above should possess novel volume requirements, some portion of which is likely to overlap with that occupied by the receptor. As an example of receptor mapping, Sufrin et al. (402) showed with amino acid analogs of methionine, which inhibited the enzyme, methionine:adenosyl transferase, that the data for a set of rigid amino acid inhibitors required the postulation of competition between the inactive analogs and the enzyme for a particular volume of space (Fig. 3.48). Summation of the volume requirements for the set of compounds, when oriented on the amino acid framework, yielded a minimum space from which the receptor could be excluded. Each amino acid had the necessary binding elements, but several were inactive. Each of the inactive analogs required extra volume not required by the active analogs and shared a small common unique volume whose occupancy by the enzyme would be sufficient to rationalize their inactivity.

Klunk et al. (459) used separate receptor

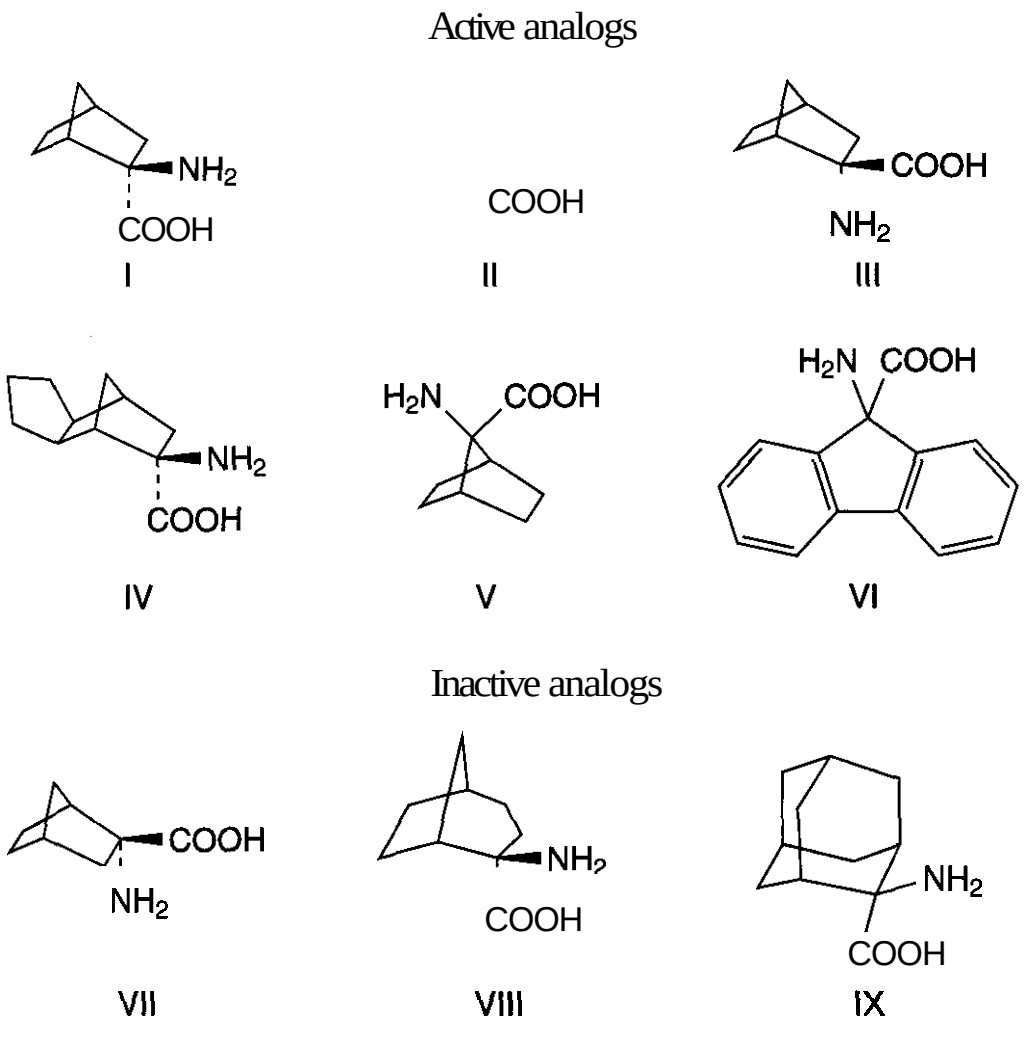


Figure 3.48. Example of receptor mapping of set of enzyme inhibitors that **can** be aligned on common amino acid framework. Set of inactive compounds all require common novel volume when compared with active compounds (402). Used with permission.

mapping of two different chemical classes of ligands to support the hypothesis that they bound to the same site. Calder et al. (460) argued that a successful correlative CoMFA model for 36 compounds of six chemical classes of GABA inhibitors indicated that the alignments used were significant. In some cases, comparison of volume maps for two receptors have allowed optimization of activity at one receptor with respect to the other. The work of Hibert et al. (461, 462), through the use of receptor mapping to increase the selectivity of a lead compound for the 5-HT_{1A} receptor over the a,-adrenoreceptor, has resulted in clinical trials for a novel chemical class. This steric-mapping approach has become relatively popular, and numerous examples appear in current journals (463) on a regular basis.

Although there are several feasible algorithms to deal with unions of molecular volumes, the use of pseudoelectron density functions calibrated to reproduce VDW radii (424) with three-dimensional contouring to represent the surface has allowed mathematical manipulation of the density associated with each lattice point to allow for union, intersection, and subtraction of volumes. Analytical representation of molecular volumes by Connolly (464, 465) and solvent-accessible surfaces by Kundrot et al. (466) may be an alternative that would allow optimization of volume overlap, for example, by minimizing the difference in volume between two structures. The solvent-accessible surface area can be used to approximate the free energy of hydration and a rapid, numerical procedure for its calculation has been reported (467).

4.4.6 Model Receptor Sites. One of the first visualizations of a receptor model is that of Beckett and Casey (468) for the opiate receptor published in 1954. Because morphine and many other compounds active at this receptor are essentially rigid, the model did not have to address the interaction of myriad numbers of flexible, naturally occurring opioid ligands, such as endorphins and enkephalin, which were only subsequently discovered. The model receptor had an anionic site to bind the charged nitrogen, a hydrophobic flat surface with a cleft to bind the phenyl ring, and a hydrophobic hydrocarbon bridge seen in morphine. Kier (469) published a number of pa-

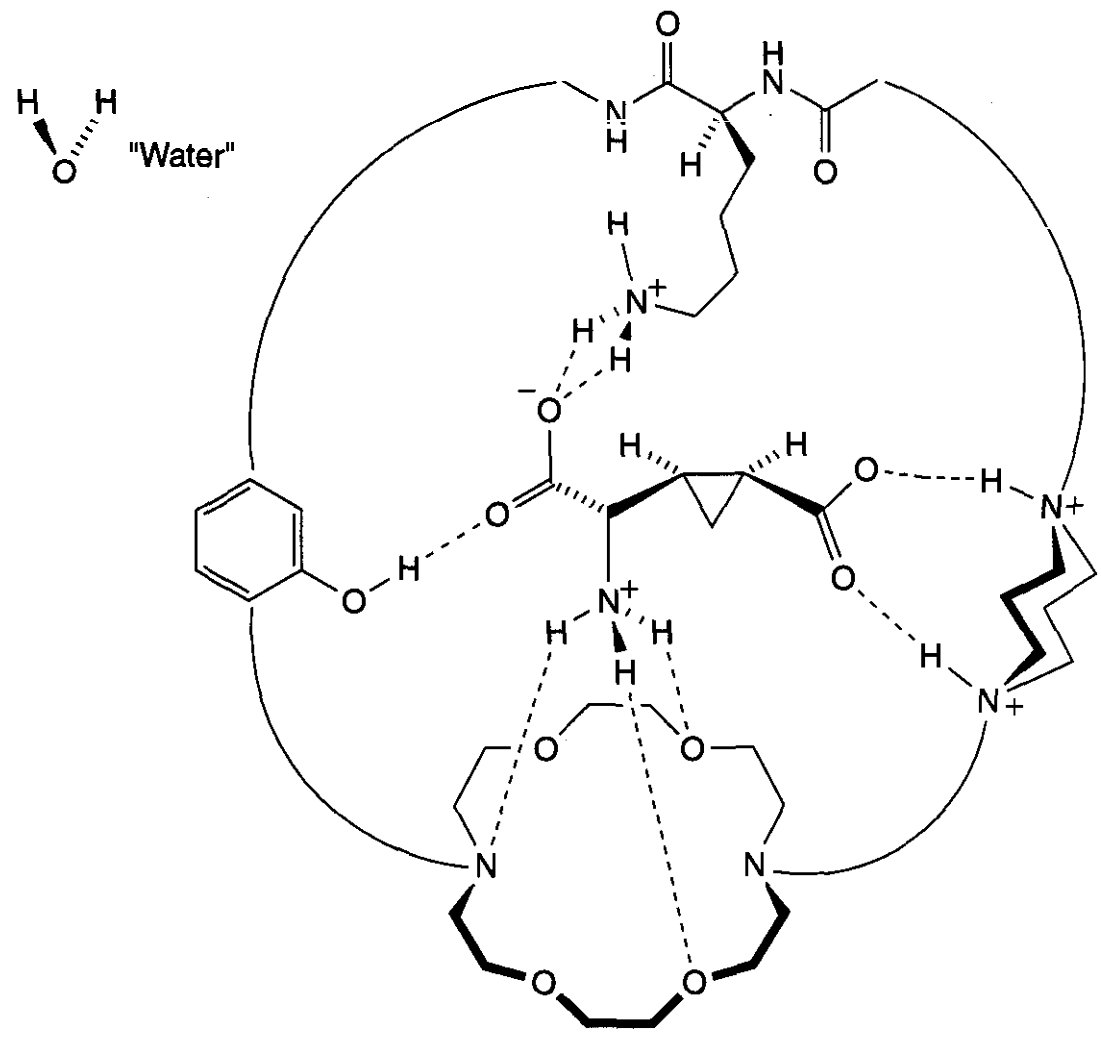


Figure 3.49. Peptidic pseudoreceptor used to calculate affinity of NMDA agonists and antagonists (453). Used with permission.

pers attempting to define the pharmacophore based on semiempirical molecular orbital calculations of in *vacuo* minimum-energy conformations. Although his basic concepts were valid, his emphasis on the global minima in *vacuo* limited his scope of applicability.

Humber et al. (470) used semirigid antipsychotic drugs, the so-called neuroleptics, which antagonize CNS dopamine transmission and displace dopamine from its receptor, to formulate a geometrical arrangement of receptor groups to rationalize their activity. Olson et al. (471) used this model to design a novel stereospecific dopamine antagonist and successfully predicted its stereochemistry.

Because we are reasonably convinced the receptor is a protein, construction of hypothetical sites from amino acid fragments and calculation of affinity for these sites should correlate with observed affinity, assuming that the type of interactions and their geometry is represented by the site in some reasonable manner. An individual fragment such as an indole ring from tryptophan does a good job of simulating a flat hydrophobic surface. Holtje and Tintelnot (472) constructed a site for chloramphenicol from arginine and histidine by varying the distances of the amino acid from its postulated binding position and finding the optimal distance for correlation with observed affinity for the ribosome. Peptidic pseudoreceptors have been constructed (453) that correctly rank-order glutamate NMDA agonists and antagonists (Fig. 3.49).

An intermediate between unknown receptors and ones where the three-dimensional structure is known are models based on homology. For the medicinal chemist, the G-protein receptors have been of intense interest and numerous models (339,340,461,473) of the various receptor types have been developed based on their presumed three-dimensional homology with bacteriorhodopsin (474). Mechanisms of signal transduction (475) and differences between agonists and antagonists (476) have even been rationalized based on such models. Nordvall and Hacksell (341) recently combined the construction of such a model for the muscarinic m1 receptor with constraints derived from steric mapping of muscarinic agonists. By adding the experimental constraints from ligand binding, a qualitative model was derived that was able to reproduce experimentally derived stereoselectivities.

4.4.7 Assessment of Model Predictability. Because it is unlikely that there will be sufficient structure-activity data to uniquely define a model at atomic resolution in competition with crystallography, justification for model building must come from its potential predictive power and possible insight into the receptor-drug interaction before detailed three-dimensional information from either crystal structure or NMR studies. Certainly, the questions regarding the ability of a proposed drug to bind to the active site without steric conflict with the receptor can be addressed by the methods outlined above in a qualitative manner. The resolution of our receptor models is too crude, however, to subject them to molecular mechanics estimates of affinities. There are alternative paradigms, however, based on pattern recognition techniques in which a set of analogs and their activities are used, along with their physicochemical parameters, to generate a mathematical model that relates the values of the physicochemical parameters for a given analog with its activity. One such paradigm is comparative molecular field analysis (CoMFA), which combines the three-dimensional electrostatic and steric fields surrounding the analogs with powerful statistical techniques, partial least squares (PLS) (477) and cross-validation, to generate predictive models if a set of orientation rules are available for aligning the molecules for comparison and prediction. Alternative methods for assessing similarity and their use in QSAR schemes have been compared (215) with CoMFA. Another approach is the use of neural nets that learn to "see" patterns in much the same way as our own nervous system processes information. Examples of the use of this pattern-recognition approach include classification of mechanism of action for cancer chemotherapy (478) and QSAR studies of DHFR inhibitors (479, 480) and carboquinones (481). Machine learning has also been applied (482) to the QSAR problem. Trimethoprim analogs were successfully analyzed for their inhibition of DHFR and similar results to the original Hansch results were obtained. It is not clear that this paradigm could be applied to noncongeneric series, at least as outlined.

What appears crucial to such studies is the choice of training set, which encompasses as much of parameter space as one is likely to use in the predictive mode as well as tests of the predictive ability of resulting models. Given that one is dealing with a situation in which the number of variables is larger (often several times) than the number of observations, linear regression models are not applicable because chance correlations are highly probable. The use of cross-validation allows selection of correlations that are predictive in a self-consistent manner within the training set. This does not mean to imply that such internally self-consistent models have predictive power outside of the training set, or extremely close congeners.

DePriest et al. (483, 484) applied the CoMFA methodology to a series of 68 ACE (angiotensin-converting enzyme) inhibitors representing 28 different chemical classes. Through use of the binding-site geometry determined by Mayer et al. (397), a CoMFA model with a statistically significant cross-validated R^2 and considerable predictive ability for inhibitors outside of the training set was derived. Because the geometry of the ACE inhibitors was determined computationally by an active-site analysis rather than experimentally, a comparison of the results of the ACE series against thermolysin inhibitors, for which there were crystallographic data to explicitly define the binding-site geometry and the resulting alignment rules, was made, given that thermolysin is also a zinc-containing metallopeptidase with numerous similarities between ACE and thermolysin. Their results give strong support to both the Active Analog Approach (398) used to define the alignment rule for the ACE series and the CoMFA methodology itself. In the absence of an experimentally known active-site geometry, correlations were derived that explain as much as 84% of the variance in activities among a set of 68 diverse ACE inhibitors by use of CoMFA steric and electrostatic potentials plus a zinc indicator variable (Fig. 3.50). If the set of 68 ACE inhibitors was divided into three classes and correlations are derived for each class, CoMFA parameters alone explain 79–99% of the variance in activities. It was notable that statistically significant correla-

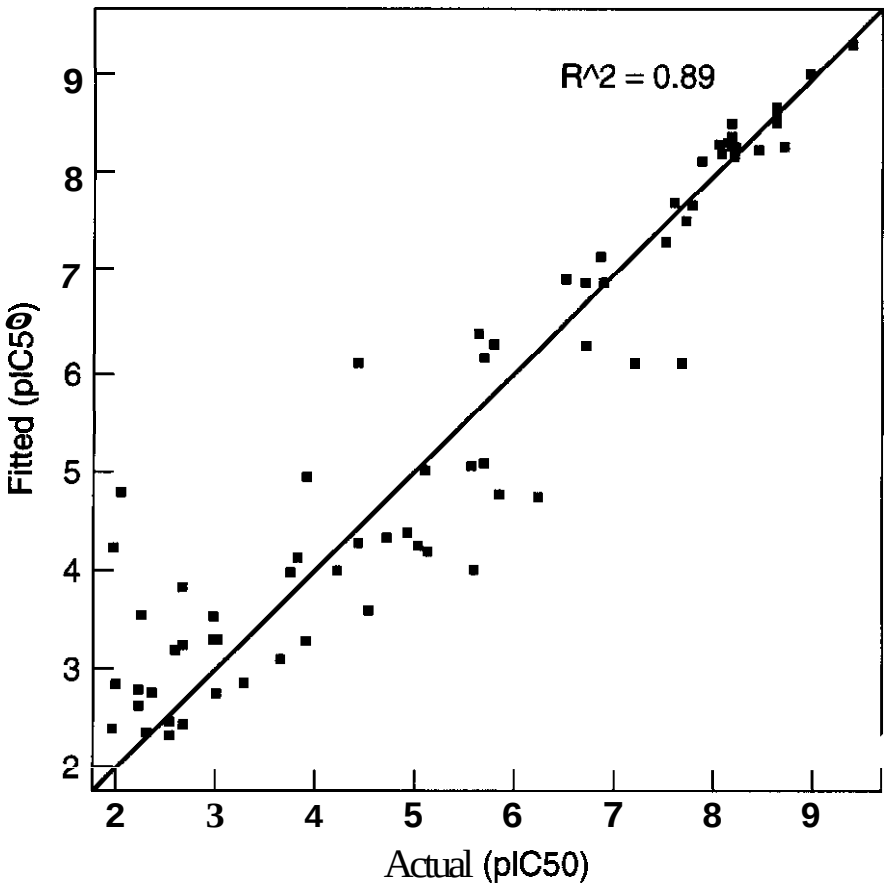


Figure 3.50. Plot of experimental versus predicted inhibition constants for 68 ACE inhibitors used in derivation of CoMFA model for the ACE active site (484). This plot shows the self-consistency of the model. Used with permission.

tions were found, in spite of the fact that CoMFA does not explicitly consider hydrophobicity or solvation. In further support of the active-site paradigm, the cross-validated results of the ACE series were equivalent to those of the thermolysin series (cross-validated $R^2 = 0.65$ to 0.70), for which the alignment rule was defined by crystallographic data.

The predictions for molecules outside the training sets are a valid test of the predictive ability of the model, rather than just a confirmation of self-consistency of the derived model. In other words, statistical analysis alone does not answer the question of a chance correlation (485) for the training set. One must investigate lateral correlations such as predictability. The predictive correlations presented by DePriest et al. (483;484) represent a total of 66 diverse inhibitors that were not chosen as analogs of compounds present in the training set, but by selecting published papers on three different chemical classes and testing all compounds in the papers [predictive $r^2 = 0.46$ for the set of 66 compounds predicted, which had not been included in the training set for the ACE model with a zinc indicator of 10 (Fig. 3.51)]. The "predictive" r^2 was based only on molecules not included in the training set and was defined as

$$\text{predictive } r^2 = (\text{SD} - \text{“press”})/\text{SD}$$

where SD is the sum of the squared deviations between the affinities of molecules in the test set and the mean affinity of the training set molecules, and "press" is the sum of the squared deviations between predicted and actual affinity values for every molecule in the test set. It should be obvious from the equation that prediction of the mean value of the training set for each member of the test set would yield a predictive $r^2 = 0.35$ out of the 66 predicted molecules had residuals less than one log value with a predictive r^2 value for the collective set of these 35 test molecules of 0.90. Of the 31 inhibitors with residuals greater than 1.0, 8 were carboxylates, 12 were phosphates, and 11 were thiols. Clearly, no single class of inhibitors dominated the distribution of residuals. Considering both the composition and the method of selection of the test data sets (range of activities over 7 log units), the fact that more than 50% of the molecules were predicted with correlations greater than $r^2 = 0.90$ lends strong support to the use of CoMFA as a tool for QSAR development.

Use of CoMFA as a predictive tool for receptors of known three-dimensional structure has also been explored. Klebe and Abraham

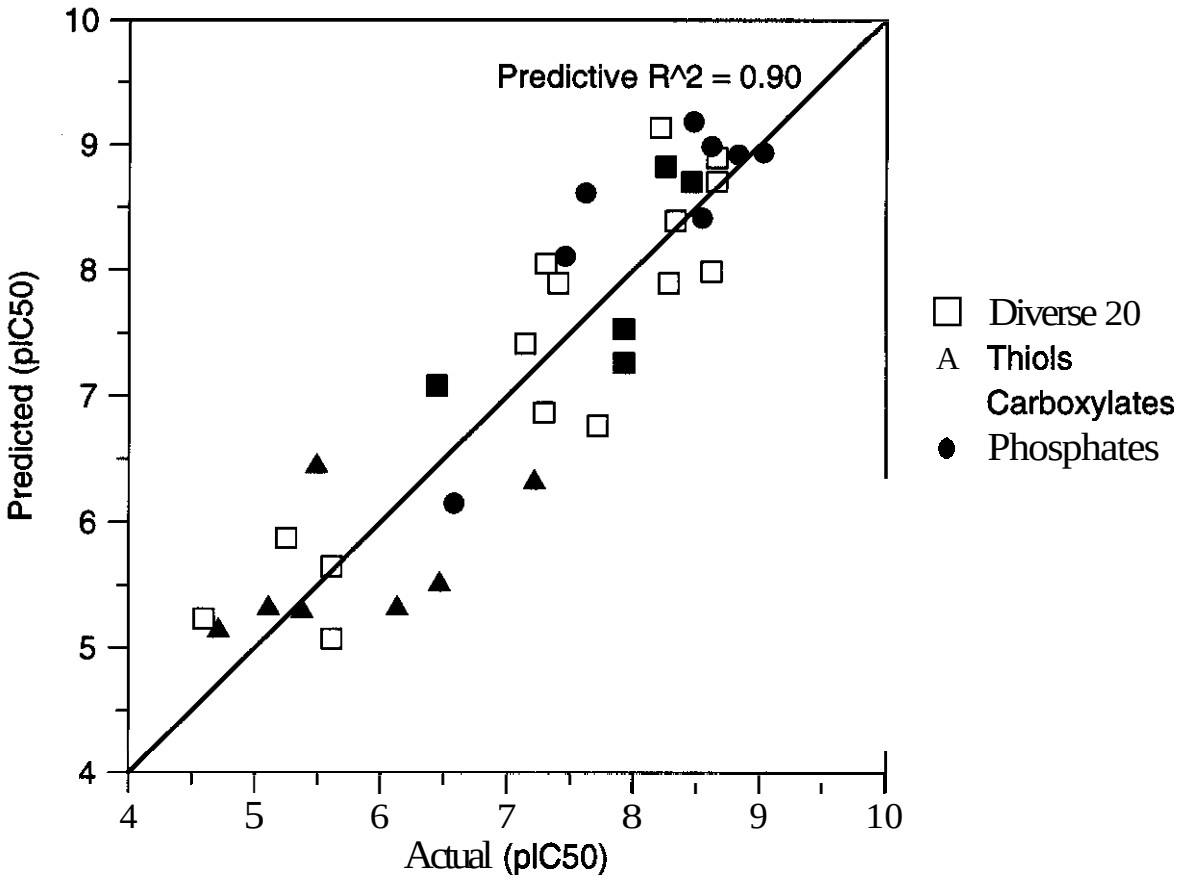


Figure 3.51. Plot of experimental versus predicted inhibition constants for 35 ACE inhibitors not used in derivation of CoMFA model (484). This plot indicates the predictability of the model. Used with permission.

(486) used two enzymes (thermolysin and renin) as well as antiviral activity against human rhinovirus, where the coat-protein receptor is known, to calibrate CoMFA methodology. They concluded that only enthalpies of binding and not binding affinities were predicted by CoMFA. Waller et al. (264) developed a predictive CoMFA model for the binding affinities of HIV-protease inhibitors based on crystal structures of complexes. Initial analysis of the 59 molecules in the training set representing five structurally diverse classes (hydroxyethylamine, statine, norstatine, ketoamide, and dihydroxyethylene) of transition-state protease inhibitors yielded a correlation with a cross-validated r^2 value of 0.786. To evaluate the predictive ability of this model, a test set of 18 additional inhibitors (487) was used that represented another class of transition-state isostere, hydroxyethylurea. The model expressed good predictive ability for the test set of hydroxyethylurea compounds ($r^2_{pred} = 0.624$) with all compounds predicted within 1.06 log unit (1.4 kcal/mol in binding affinity) of their actual activities, with an average absolute error of 0.58 log units (0.8 kcal/mol) across a range of 3.03 log units (Fig. 3.52). Predictions from this CoMFA model of HIV protease are being used to prioritize synthesis of *de* novo-designed HIV-protease inhibitors not included in development of the model.

Crippen developed a method (488) to objectively model the binding of small ligands to receptors, given the experimentally determined affinities of a set of ligands. The procedure, Vorom, used Voronoi polyhedra to generate the simplest geometrical model of the binding site. In a recent application to DHFR inhibitors (489), only eight analogs were used in the training set to derive the model and the affinities of 23/39 of the test set molecules were correctly predicted, with an average relative error of 0.83 kcal/mol for the remaining compounds.

5 CONCLUSIONS

Rapid advances in molecular and structural biology have provided ample therapeutic targets characterized in three dimensions. Tools to exploit this information are being rapidly developed and several strategies for *de* novo design of ligands, given an active site, are un-

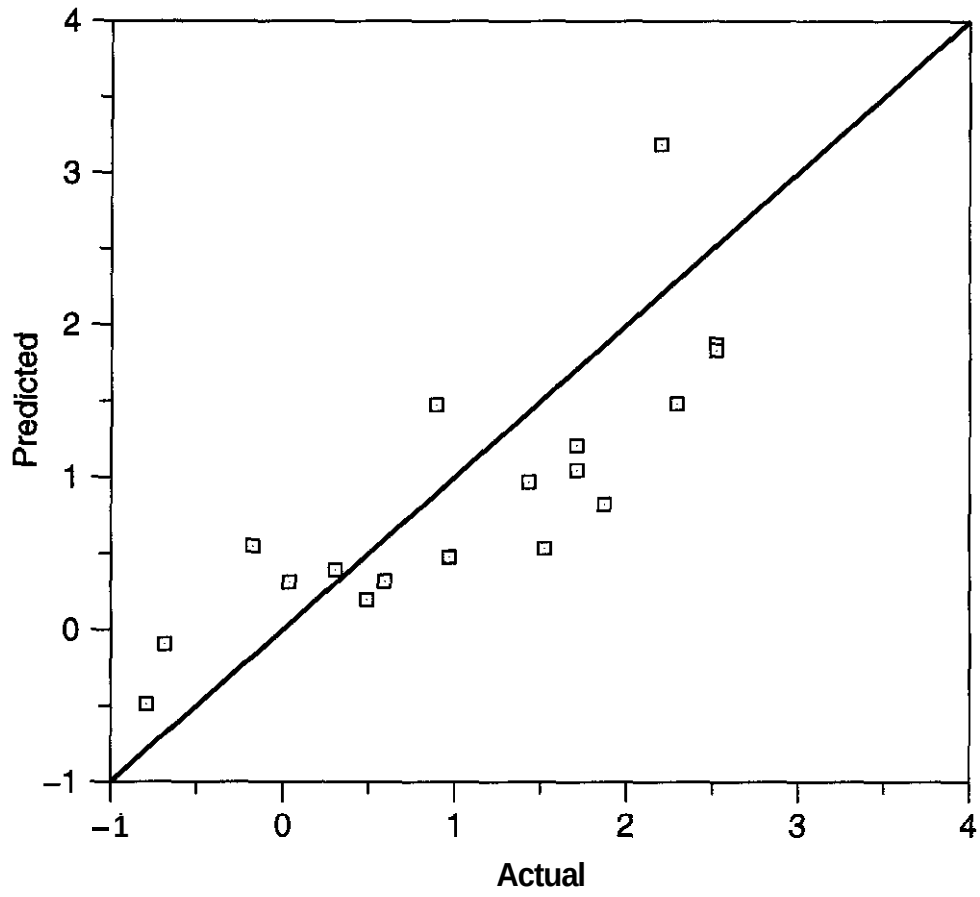


Figure 3.52. Plot of experimental versus predicted inhibition constants for 18 HIV-1 protease inhibitors not used in derivation of CoMFA model (264). This plot indicates the predictability of the model.

der investigation. It is already clear, however, that iterative approaches are necessary because of the lack of precision in predicting affinities for bound ligands. Molecular mechanics and computer graphics are essential components for design of novel ligands, and rapid progress in evolving a useful set of tools is apparent.

The ultimate goal in comparison of molecules with respect to their biological activity is insight into the receptor and its requirements for recognition and activation. Conjecture regarding the receptor is often a necessary part of rationalizing a set of structure-activity data. Although the problem of characterizing the active site of an unknown macromolecule indirectly is certainly challenging, the analysis of structure-activity data of a set of ligands, especially if their structural variety is wide, allows useful models of active sites to be developed. There are numerous caveats that must be acknowledged, however, such as flexibility of the receptor, multiple binding modes for ligands, and lack of uniqueness of most models because of limited experimental observations. Success in using these methods would appear to be increasing. This reflects both technological advances as well as insight into the problem and algorithmic improvements in our analytical approaches.

The game of 20 questions with receptors has progressed with experience. Ambiguity in interpretation of results and multiple models clearly reflect the uncertainties inherent in this indirect approach. Nevertheless, the absence of direct experimental data in many biological systems of intense therapeutic interest make this the only game available for many. It is hoped that the next decade will see further progress in our ability to extract three-dimensional information from structure-activity studies on unknown receptors.

This perspective has examined the approaches to molecular modeling and drug design and emphasized their limitations. The reader should be aware. however. that these tools are daily used on many problems of therapeutic interest with increasing success. This is clearly witnessed by publications of such studies in almost every issue of current major journals. For specific application areas, such as RNA (490, 491), DNA (492–496), membrane (497–507), or peptidomimetic modeling (382, 508–513), the reader is referred to the literature. The prediction of molecular properties, such as log P and correlation between substructures and metabolism. has led to a dramatic increase in efforts to correlate adsorption, distribution (514), metabolism (515–517), and elimination (ADME) with chemical

structure (518–522). In addition, the advent of combinatorial chemistry has focused modeling efforts on prioritizing compounds (523–528) for high throughput screening based on chemical diversity (529–531), druglike properties (532, 533), predicted oral bioavailability (534, 535), and so forth.

6 ACKNOWLEDGMENTS

The work and influence of many talented collaborators as well as the National Institute of Health for grant support are gratefully acknowledged. Although my former colleagues' names are prominent in the references cited, their contributions are numerous and individual citations are avoided because of probable omissions. The author apologizes to many contributors to the field whose efforts have not been adequately recognized in this overview, the result of a somewhat arbitrary citation of references. Space and time limitations preclude a more thorough discussion of many important aspects.

REFERENCES

1. M. Salzmann, K. Pervushin, G. Wider, H. Senn, and K. Wuthrich, *J. Biomol. NMR*, **14**, 85-88 (1999).
2. P. J. Hajduk, R. P. Meadows, and S. W. Fesik, *Q. Rev. Biophys.*, **32**, 211-240 (1999).
3. P. J. Hajduk, G. Sheppard, D. G. Nettesheim, E. T. Olejniczak, S. B. Shuker, R. P. Meadows, D. H. Steinman, G. M. Carrera, P. A. Marcotte, J. Severin, K. Walter, H. Smith, E. Gubbins, R. Simmer, T. F. Holzman, D. W. Morgan, S. K. Davidsen, J. B. Summers, and S. W. Fesik, *J. Am. Chem. Soc.*, **119**, 5818-5827 (1997).
4. L. M. McDowell and J. Schaefer, *Curr. Opin. Struct. Biol.*, **6**, 624-629 (1996).
5. R. Ishima and D. A. Torchia, *Nat. Struct. Biol.*, **7**, 740-743 (2000).
6. L. M. McDowell, M. A. McCarrick, D. R. Studelska, W. J. Guilford, D. Arnaiz, J. L. Dallas, D. R. Light, M. Whitlow, and J. Schaefer, *J. Med. Chem.*, **42**, 39104918 (1999).
7. J. M. Moore, *Biopolymers*, **51**, 221-243 (1999).
8. J. Fejzo, C. A. Lepre, J. W. Peng, G. W. Bemis, Ajay, M. A. Murcko, and J. M. Moore, *Chem. Biol.*, **6**, 755-769 (1999).
9. C. J. Dinsmore, M. J. Bogusky, J. C. Culberson, J. M. Bergman, C. F. Homnick, C. B. Zartman, S. D. Mosser, M. D. Schaber, R. G. Robinson, K. S. Koblan, H. E. Huber, S. L. Graham, G. D. Hartman, J. R. Huff, and T. M. Williams, *J. Am. Chem. Soc.*, **123**, 2107-2108 (2001).
10. J. Hajdu, R. Neutze, T. Sjogren, K. Edman, A. Szoke, R. Wilmouth, and C. M. Wilmot, *Nat. Struct. Biol.*, **7**, 1006-1012 (2000).
11. A. Perrakis, R. Morris, and V. S. Lamzin, *Nat. Struct. Biol.*, **6**, 458-463 (1999).
12. C. R. Beddell, Ed., *The Design of Drugs to Macromolecular Targets*, John Wiley & Sons, New York, 1992.
13. I. D. Kuntz, *Science*, **257**, 1078-1082 (1992).
14. P. W. Finn and L. E. Kavraki, *Algorithmica*, **25**, 347-371 (1999).
15. D. Joseph-McCarthy, *Pharmacol. Ther.*, **84**, 179-191 (1999).
16. P. G. Mezey, *J. Mol. Model.*, **6**, 150-157 (2000).
17. E. F. Meyer, S. M. Swanson, and J. A. Williams, *Pharmacol. Ther.*, **85**, 113-121 (2000).
18. V. Schnecke and L. A. Kuhn, *Perspect. Drug Discov. Des.*, **20**, 171-190 (2000).
19. U. Burkert and N. L. Allinger in M. C. Caserio, Ed., *Molecular Mechanics, ACS Monograph 339*, Vol. 177, American Chemical Society, Washington, DC, 1982.
20. J. P. Bowen and N. L. Allinger in K. B. Lipkowitz and D. B. Boyd, Eds., *Revisions in Computational Chemistry*, VCH, New York, 1991, pp. 81-98.
21. A. R. Leach, *Molecular Modelling: Principles and Applications*, 2nd ed., Prentice Hall, New York, 2001, 744 pp.
22. M. Clark, R. Cramer, and N. Van Opdenbosch, *J. Comput. Chem.*, **10**, 982-1012 (1989).
23. J. G. Vinter, A. Davis, and M. R. Saunders, *J. Cornput.-Aided. Mol. Des.*, **1**, 31–51 (1987).
24. D. N. J. White and M. J. Bovill, *J. Chem. Soc. Perkin Trans. 2*, **12**, 1610–1623 (1977).
25. K. Gundertofte, J. Palm, I. Pettersson, and A. Stamvik, *J. Comput. Chem.*, **12**, 200-208 (1991).
26. W. L. Jorgensen and J. Gao, *J. Am. Chem. Soc.*, **110**, 4212-4216 (1988).
27. J.-H. Lii and N. L. Allinger, *J. Comput. Chem.*, **12**, 186-199 (1991).
28. J. Aqvist and A. Warshel, *J. Am. Chem. Soc.*, **112**, 2860 (1990).
29. R. D. Hancock, *Acc. Chem. Res.*, **23**, 253-257 (1990).

30. A. Vedani and D.W. Huhta, J. Am. Chem. Soc., 112,4759–4767 (1990).
31. V. S. Allured, C. M. Kelly, and C. R. Landis, J. Am Chem. Soc., 113, 1–12 (1991).
32. A. E. Carlsson, Phys. Rev. Lett., 81, 477–480 (1998).
33. A. E. Carlsson and S. Zapata, Biophys. J., 81, 1–10 (2001).
34. A. T. Hagler, E. Hugler, and S. Lifson, J. Am. Chem. Soc., 96,5319 (1974).
35. S. C. Harvey, Proteins, 5, 78–92 (1989).
36. M. E. Davis and J. A. McCammon, Chem. Rev., 90,509–521 (1990).
37. W. F. van Gunsteren and H. J. C. Berendsen, Angew. Chem. Znt. Ed. Engl., 29, 992–1023 (1990).
38. C. E. Dykstra, Chem. Rev., 93, 2339–2353 (1993).
39. A. J. Stone and M. Alderton, Mol. Phys., 56, 1047–1064 (1985).
40. M. J. Dudek and J. W. Ponder, J. Comput. Chem., 16,791–816 (1995).
41. C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, J. Phys. Chem., 97, 10269–10280 (1993).
42. D. E. Williams in K. B. Lipkowitz and D. B. Boyd, Eds., Revisions in Computational Chemistry, VCH, New York, 1991, pp. 219–271.
43. R. J. Loncharich and B. R. Brooks, Proteins, 6, 32–45 (1989).
44. J. Guenot and P. A. Kollman, J. Comput. Chem., 14,295–311 (1993).
45. K. Tasaki, S. McDonald, and J. W. Brady, J. Comput. Chem., 14,278–284 (1993).
46. J. Shimada, H. Kaneko, and T. Takada, J. Comput. Chem., 14,867–878 (1993).
47. H. Schreiber and O. Steinhauser, Chem. Phys., 168, 75–89 (1992).
48. H. Schreiber and O. Steinhauser, Biochemistry, 31,5856–5860 (1992).
49. P. E. Smith and B. M. Pettit, J. Chem. Phys., 95,8430–8441 (1991).
50. G. E. Marlow, J. S. Perkyns, and B. M. Pettit, Chem. Rev., 93,2503–2521 (1993).
51. M. Whitlow and M. M. Teeter, J. Am. Chem. Soc., 108, 7163–7172 (1986).
52. M. K. Gilson, K. A. Sharp, and B. H. Honig, J. Comput. Chem., 9, 327–335 (1987).
53. A. Nicholls and B. Honig, J. Comput. Chem., 12,435–445 (1991).
54. M. Schaefer and C. Froemmel, J. Mol. Biol., 216, 1045–1066 (1990).
55. W. G. Richards, P. M. King, and C. A. Reynolds, Protein Eng., 2, 319–327 (1987).
56. R. A. Pierotti, Chem. Rev., 76,717–726 (1976).
57. G. L. Pollack, Science, 251, 1323–1330 (1991).
58. R. J. Zauhar and R. S. Morgan, J. Comput. Chem., 9,171–187 (1988).
59. J. Tomasi, R. Bonaccorsi, R. Cammi, et al., Theochem. J. Mol. Struct., 80,401–424 (1991).
60. D. A. Liotard, G. D. Hawkins, G. C. Lynch, C. J. Cramer, and D. G. Truhlar, J. Comput. Chem., 16,422–440 (1995).
61. K. Sharp, J. Comput. Chem., 12, 454–468 (1991).
62. W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, Chem. Soc., 112, 6127–6129 (1990).
63. C. A. Schiffer, J. W. Caldwell, P. A. Kollman, and R. M. Stroud, Mol. Simul., 10, 121–149 (1993).
64. P. F. W. Stouten, C. Frommel, H. Nakamura, and C. Sander, Mol. Simul., 10, 97–120 (1993).
65. R. J. Zauhar, J. Comput. Chem., 12, 575–583 (1991).
66. A. J. Stone, Mol. Phys., 56, 1065–1082 (1985).
67. S. Kuwajima and A. Warshel, J. Phys. Chem., 94,460–466 (1990).
68. R. A. Sorensen, W. B. Liau, L. Kesner, and R. H. Boyd, Macromolecules, 21, 200–208 (1988).
69. J. Caldwell, L. X. Dang, and P. A. Kollman, J. Am. Chem. Soc., 112,9144–9147 (1990):
70. C. J. Cramer and D. G. Truhlar, J. Am. Chem. Soc., 113,8305–8311 (1991).
71. C. J. Cramer, J. Am. Chem. Soc., 113, 8552–8554 (1991).
72. C. J. Cramer and D. G. Truhlar, Science, 256, 213–217 (1992).
73. C. J. Cramer and D. G. Truhlar, J. Comput. Chem., 13,1089–1097 (1992).
74. G. Rauhut, T. Clark, and T. Steinke, J. Am. Chem. Soc., 115,9174–9181 (1993).
75. F. Franks in F. Franks, Ed., Water, A Comprehensive Treatise, Vol. 1, Plenum Press, New York, 1975.
76. F. H. Stillinger, Science, 209,451–457 (1980).
77. L. R. Pratt, Ann. Rev. Phys. Chem., 36, 433–449 (1985).
78. J. P. M. Postma, H. J. C. Berendsen, and J. R. Haak, Faraday Symp. Chem. Soc., 17, 55–67 (1982).
79. B. G. Rao and U. C. Singh, J. Am. Chem. Soc., 111,31253133 (1989).

80. I. Ohmine and H. Tanaka, *Chem. Rev.*, 93, 2545–2566 (1993).
81. W. L. Jorgensen, J. Gao, and C. Ravimohan, *J. Phys. Chem.*, 89,34703473 (1985).
82. N. Muller, *Trends Biochem. Sci.*, 17,459–463 (1992).
83. L. X. Dang, J. E. Rice, J. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, 113, 2481–2486 (1991).
84. A. K. Rappe and W. A. Goddard III, *J. Phys. Chem.*, 95,3358–3363 (1991).
85. R. Czerminski and R. Elber, *Int. J. Quantum Chem. Quantum Chem. Symp.*, 24, 167–186 (1990).
86. C. Choi and R. Elber, *J. Chem. Phys.*, 94, 751–760 (1991).
87. S. E. Huston and G. R. Marshall, *Biopolymers*, 34, 74–90 (1994).
88. R. V. Pappu, R. K. Hart, and J. W. Ponder, *J. Phys. Chem. B*, 102,9725–9742 (1998).
89. L. Piela, *Collect. Czech. Chem. Commun.*, 63, 1368–1380 (1998).
90. R. K. Hart, R. V. Pappu, and J. W. Ponder, *J. Comput. Chem.*, 21,531–552 (2000).
91. M. Saunders, K. N. Houk, Y.-D. Wu, W. C. Still, M. Lipton, G. Chang, and W. C. Guida, *J. Am. Chem. Soc.*, 112,1419–1427 (1990).
92. M. Saunders, *J. Am. Chem. Soc.*, 109, 3150–3152 (1987).
93. J. T. Ngo and M. Karplus, *J. Am. Chem. Soc.*, 119,56575667 (1997).
94. R. V. Pappu, G. R. Marshall, and J. W. Ponder, *Nat. Struct. Biol.*, 6, 50–55 (1999).
95. K. R. Mackenzie, J. H. Prestegard, and D. M. Engelman, *Science*, 276, 131–133 (1997).
96. J. H. Holland, *Sci. Am.*, July, 66–72 (1992).
97. S. Forrest, *Science*, 261, 872–878 (1993).
98. P. Willett, *Trends Biotechnol.*, 13, 516–521 (1995).
99. J. E. Devillers, *Genetic Algorithms in Molecular Modeling*, Academic Press, New York, 1996.
100. D. B. McGarrah and R. S. Judson, *J. Comput. Chem.*, 14,1385–1395 (1993).
101. R. S. Judson, Y. T. Tan, E. Mori, C. Melius, E. P. Jaeger, A. M. Treasurywala, and A. Mathiowetz, *J. Comput. Chem.*, 16, 1405–1419 (1995).
102. R. P. Meadows and P. J. Hajduk, *J. Biomol. NMR*, 5, 41–47 (1995).
103. B. Waszkowycz, D. E. Clark, D. Frenkel, J. Li, C. W. Murray, B. Robson, and D. R. Westhead, *J. Med. Chem.*, 37,3994–4002 (1994).
104. C. M. W. Ho and G. R. Marshall, *J. Comput.-Aided Mol. Des.*, 7,623–647 (1993).
105. A. W. R. Payne and R. C. Glen, *J. Mol. Graphics*, 11, 74–91 (1993).
106. H. A. Scheraga in K. B. Lipkowitz and D. B. Boyd, Eds., *Revisions in Computational Chemistry*, VCH, New York, 1992, pp. 73–142.
107. T. Schlick in K. B. Lipkowitz and D. B. Boyd, Eds., *Revisions in Computational Chemistry*, VCH, New York, 1992, pp. 1–71.
108. D. D. Beusen, E. F. B. Shands, S. F. Karasek, G. R. Marshall, and R. A. Dammkoehler, *THEOCHEM*, 370, 157–171 (1996).
109. H. Iijima, J. B. Dunbar, Jr., and G. R. Marshall, *Proteins*, 2,330339 (1987).
110. I. Motoc, R. A. Dammkoehler, and G. R. Marshall in N. Trinajstic, Ed., *Mathematic and Computational Concepts in Chemistry*, Ellis Horwood, Chichester, UK, 1986, pp. 222–251.
111. R. A. Dammkoehler, S. F. Karasek, E. F. B. Shands, and G. R. Marshall, *J. Cornput.-Aided Mol. Des.*, 3, 3–21 (1989).
112. N. *Go* and H. A. Scheraga, *Macromolecules*, 3, 178–187 (1970).
113. A. R. Leach in K. B. Lipkowitz and D. B. Boyd, Eds., *Revisions in Computational Chemistry*, VCH, New York, 1991, pp. 1–55.
114. S. K. Burt and J. Greer, *Ann. Rep. Med. Chem.*, 23,285–294 (1988).
115. D. M. Ferguson and D. J. Raber, *J. Am. Chem. Soc.*, 111,4371–4378 (1989).
116. M. Saunders, *J. Comput. Chem.*, 10, 203–208 (1989).
117. M. Saunders, *J. Comput. Chem.*, 12, 645–663 (1991).
118. M. Saunders and H. A. Jimenez-Vazquez, *J. Comput. Chem.*, 14,330–348 (1993).
119. M. Saunders and N. Krause, *J. Am. Chem. Soc.*, 112,1791–1795 (1990).
120. A. V. Shah and D. P. Dolata, *J. Cornput.-Aided Mol. Des.*, 7, 103–124 (1993).
121. I. Kolossvary and W. C. Guida, *J. Am. Chem. Soc.*, 115,2107–2119 (1993).
122. H.-J. Boehm, G. Klebe, T. Lorenz, T. Mietzner, and L. Siggel, *J. Comput. Chem.*, 11, 1021–1028 (1990).
123. A. E. Howard and P. A. Kollman, *J. Med. Chem.*, 31,1669–1675 (1988).
124. M. Lipton and W. C. Still, *J. Comput. Chem.*, 9, 343–355 (1988).
125. D. D. Beusen, R. D. Head, J. D. Clark, W. C. Hutton, U. Slomczynska, J. Zabrocki, M. T. Leplawy, and G. R. Marshall in C. H. Schnei-

der and A. N. Eberle, Eds., ***The Solution NMR Structures of Emerimicins III and IV Determined Using the New Program, MACROSEARCH,*** ESCOM Scientific, Leiden, Netherlands, **1993,** pp. **79–80.**

126. M. P. Allen and D. J. Tildesley, ***Computer Simulation of Liquids,*** Oxford Science Publications, Oxford, UK, **1989,** p. **385.**

127. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, ***J. Chem. Phys.,*** **21, 1087 (1953).**

128. J. A. McCammon and S. C. Harvey, ***Dynamics of Protein and Nucleic Acids,*** Cambridge University Press, Cambridge, UK, **1987,** p. **234.**

129. G. Zhang and T. Schlick, ***J. Comput. Chem.,*** **14,1212–1233 (1993).**

130. T. Schlick and W. K. Olson, ***Science,*** **257, 1110–1115 (1992).**

131. D. S. Goodsell and A. J. Olson, ***Proteins,*** **8,** 195–**202 (1990).**

132. W. L. Jorgensen, ***Acc. Chem. Res.,*** **22,184–189 (1989).**

133. D. L. Beveridge and F. M. DiCapua in W. van Gunsteren and P. K. Weiner, Eds., ***Computer Simulation of Biomolecular Systems,*** ESCOM Science, Leiden, Netherlands, **1989,** pp. **1–26.**

134. P. Kollman, ***Chem. Rev.,*** **93,2395–2417 (1993).**

135. W. L. Jorgensen, ***J. Phys. Chem.,*** **87,** 5304–**5314 (1983).**

136. P. A. Bash, U. C. Singh, F. K. Brown, R. Langridge, and P. A. Kollman, ***Science,*** **235,** 574–**576 (1987).**

137. P. A. Kollman and K. M. Merz, ***Acc. Chem. Res.,*** **23,** 246–252 **(1990).**

138. T. P. Lybrand, J. A. McCammon, and G. Wipff, ***Proc. Natl. Acad. Sci. USA,*** **83,** 833–835 **(1986).**

139. J. Hermans, R. H. Yun, and A. G. Anderson, ***J. Comput. Chem.,*** **13,429–442 (1992).**

140. J. Hermans, ***Curr. Opin. Struct. Biol.,*** **3,** 270–**276 (1993).**

141. R. Elber and M. Karplus, ***J. Am. Chem. Soc.,*** **112,9161–9175 (1990).**

142. D. J. Tobias, J. E. Mertz, and C. L. Brooks III, ***Biochemistry,*** **30,6054–6058 (1991).**

143. D. J. Tobias and C. L. Brooks III, ***Biochemistry,*** **30,6059–6070 (1991).**

144. D. J. Tobias, S. F. Sneddon, and C. L. Brooks III, ***J. Mol. Biol.,*** **216, 783–796 (1990).**

145. S. F. Sneddon, D. J. Tobias, and C. L. Brooks III, ***J. Mol. Biol.,*** **209, 817–820 (1989).**

146. D. J. Tobias, S. F. Sneddon, and C. L. Brooks III in R. Lavery, J.-L. Rivail, and J. Smith, Eds., Advances in Biomolecular Simulations, American Institute of Physics Conference Proceedings No. **239,** Obernai, France, **1991,** pp. **174–199.**

147. M. L. Smythe, S. E. Huston, and G. R. Marshall, ***J. Am. Chem. Soc.,*** **115, 11594–11595 (1993).**

148. M. L. Smythe, S. E. Huston, and G. R. Marshall, ***J. Am. Chem. Soc.,*** **117,** 5445–5452 **(1995).**

149. G. H. Loew and S. K. Burt in C. A. Ramsden, Ed., ***Quantitative Drug Design,*** Pergamon Press, Oxford, UK, **1990,** pp. **105–123.**

150. S. L. Price and N. G. J. Richards, ***J. Comput.-Aided Drug Des.,*** **5,** 41–54 **(1991).**

151. U. C. Singh and P. A. Kollman, ***J. Comput. Chem.,*** **5, 129 (1984).**

152. B. H. Besler, K. M. Merz, Jr., and P. A. Kollman, ***J. Comput. Chem.,*** **11,431–439 (1990).**

153. G. Rauhut and T. Clark, ***J. Comput. Chem.,*** **14,** 503–509 **(1993).**

154. J. G. Vinter and M. R. Saunders in D. J. Chadwick and K. Widdows, Eds., ***Host-Guest Molecular Interactions: From Chemistry to Biology,*** John Wiley & Sons, Chichester, UK, **1991,** pp. **249–265.**

155. C. A. Hunter and J. K. M. Sanders, ***J. Am. Chem. Soc.,*** **112,5525–5534 (1990).**

156. U. Dinur and A. T. Hagler in K. B. Lipkowitz and D. B. Boyd, Eds., ***Revisions in Computational Chemistry,*** VCH, New York, **1991,** pp. **99–164.**

157. J. Pranata, S. G. Wierschke, and W. I. Jorgensen, ***J. Am. Chem. Soc.,*** **113, 2810–2819** (1991).

158. J. Tirado-Rives and W. L. Jorgensen, ***J. Am. Chem. Soc.,*** **112,2773–2781 (1990).**

159. A. Alex and T. Clark, ***J. Comput. Chem.,*** **13, 704–717 (1992).**

160. J. Aqvist and A. Warshel, ***Chem. Rev.,*** **93, 2523–2544 (1993).**

161. M. J. Field, P. A. Bash, and M. Karplus, ***J. Comput. Chem.,*** **11,700–783 (1990).**

162. A. Warshel, ***Computer Modeling of Chemical Reactions in Enzymes and Solutions,*** John Wiley & Sons, New York, **1991,** p. **236.**

163. V. Daggett, S. Schroder, and P. Kollman, ***J. Am. Chem. Soc.,*** **113,** 8926–8935 **(1991).**

164. P. R. Andrews and D. A. Winkler in G. Jolles and K. R. H. Wooldridge, Eds., ***Drug Design: Fact or Fantasy?,*** Academic Press, New York, **1984,** pp. **145–174.**

165. J. E. Eksterowicz and K. N. Houk, Chem. Rev., 93, 2439-2461 (1993).

166. K. Appelt, R. J. Bacquet, C. A. Bartlett, C. L. J. Booth, S. T. Freer, M. A. M. Fuhry, M. R. Gehring, S. H. Herrmann, E. F. Howland, C. A. Janson, T. R. Jones, C.-C. Kan, V. Kathardekar, K. K. Lewis, G. P. Marzoni, D. A. Mathews, C. Mohr, E. W. Moomaw, C. A. Morse, S. J. Oatley, R. C. Ogden, M. R. Reddy, S. H. Reich, W. S. Schoettin, W. W. Smith, M. D. Varney, J. E. Villafranca, R. W. Ward, S. Webber, S. E. Webber, K. M. Welsh, and J. White, J. Med. Chem., 34, 1925–1934 (1991).

167. P. J. Goodford, J. Med. Chem., 27, 557-564 (1984).

168. C. R. Beddell, Chem. Soc. Rev., 13, 279-319 (1984).

169. R. Wootton in C. R. Beddell, Ed., The Design of Drugs to Macromolecular Targets, John Wiley & Sons, New York, 1992, pp. 49-83.

170. L. F. Kuyper, B. Roth, D. P. Baccanari, R. Ferone, C. R. Beddell, J. N. Champness, D. K. Stammers, J. G. Dann, F. E. Norrington, D. J. Baker, and P. J. Goodford, J. Med. Chem., 28, 303-311 (1985).

171. K. Appelt, J. Cornput.-Aided Mol. Des., 1, 23-48 (1993).

172. M. von Itzstein, W.-Y. Wu, G. B. Kok, M. S. Pegg, J. C. Dyason, B. Jin, T. V. Phan, M. L. Smythe, H. E. White, S. W. Oliver, P. M. Colman, J. N. Varghese, D. M. Ryan, J. M. Woods, R. C. Bethell, V. J. Hotham, J. M. Cameron, and C. R. Penn, Nature, 363, 418-423 (1993).

173. J. W. Liebeschuetz, S. D. Jones, P. J. Morgan, C. W. Murray, A. D. Rimmer, J. M. Roscoe, B. Waszkowycz, P. M. Welsh, W. A. Wylie, S. C. Young, H. Martin, J. Mahler, L. Brady, and K. Wilkinson, J. Med. Chem., 45, 1221-1232 (2002).

174. K. E. Lind, Z. Du, K. Fujinaga, B. M. Peterlin, and T. L. James, Chem. Biol., 9, 185-193 (2002).

175. M. Miller, M. Jaskolski, J. K. M. Rao, J. Leis, and A. Wlodawer, Nature, 337, 576-579 (1989).

176. M. Miller, B. K. Sathyanarayana, A. Wlodawer, M. V. Toth, G. R. Marshall, L. Clawson, L. Selk, J. Schneider, and S. B. H. Kent, Science, 246, 1149-1152 (1989).

177. R. L. Stanfield, M. Takimoto-Kamimura, J. M. Rini, A. T. Profy, and I. A. Wilson, Structure, 1, 83-93 (1993).

178. S. W. Fesik, J. Med. Chem., 34, 2938-2945 (1991).

179. G. Otting, *Curr.* Opin. Struct. Biol., 3, 760–768 (1993).

180. S. O. Smith, *Curr.* Opin. Struct. Biol., 3, 755–759 (1993).

181. M. F. Perutz, G. Fermi, D. J. Abraham, C. Poyart, and E. Bursaux, J. Am. Chem. *Soc.*, 108, 1064-1078 (1986).

182. A. S. Mehanna and D. J. Abraham, Biochemistry, 29, 3944-3954 (1990).

183. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin, J. Mol. Biol., 161, 269 (1982).

184. R. Voorintholt, M. T. Kosters, G. Vegter, G. Vriend, and W. G. J. Hol, J. Mol. Graphics, 7, 243-245 (1989).

185. C. M. W. Ho and G. R. Marshall, J. *Comput.*-Aided *Mol.* Des., 4, 337–354 (1990).

186. P. J. Goodford, J. *Am.* Chem. *Soc.*, 28, 849–856 (1985).

187. R. D. Cramer III, D. E. Patterson, and J. D. Bunce, J. Am. Chem. *Soc.*, 110, 5959-5967 (1988).

188. R. D. Cramer III and M. Milne, The Lattice Model: A General Paradigm for Shape-Related Structure/Activity Correlation, in Proceedings of the 19th National Meeting of the American Chemical Society, American Chemical Society, Washington, DC, 1979.

189. A. Miranker and M. Karplus, Proteins, 11, 29-34 (1991).

190. A. Caflisch, A. Miranker, and M. Karplus, J. Med. Chem., 36, 2142-2167 (1993).

191. P. K. Weiner, C. Landridge, J. M. Blaney, R. Schaefer, and P. A. Kollman, Proc. Natl. Acad. Sci. USA, 79, 3754–3758 (1982).

192. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, J. S. Profeta, and P. Weiner, J. *Am.* Chem. *Soc.*, 106, 765-784 (1984).

193. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, J. Comput. Chem., 7, 230-252 (1986).

194. F. H. Allen, J. E. Davies, J. J. Galloy, O. Johnson, O. Kennard, C. F. Macrea, E. M. Mitchell, G. F. Mitchell, J. M. Smith, and D. G. Watson, J. Chem. *Inf.* Comput. Sci., 31, 187-204 (1991).

195. E. E. Abola, F. C. Bernstein, and T. F. Koetzle in P. S. Glaeser, Ed., The Role of Data in Scientific Progress, Elsevier, New York, 1985.

196. P. R. Andrews, E. J. Lloyd, J. L. Martin, and S. L. A. Munro, J. Mol. Graphics, 4, 41-45 (1986).

197. R. S. Pearlman, Chem. Des. Auto. News, 2, 1 (1987).

198. R. S. Pearlman, *CONCORD User's Manual,* Tripos Associates, St. Louis, MO, **1992.**

199. R. S. Pearlman, ***Chem. Des. Auto. News,* 8, 3–15 (1993).**

200. R. **P.** Sheridan, A. Rusinko **III,** R. Nilakantan, and R. Venkataraghavan, ***Proc. Natl. Acad. Sci. USA,* 86,8165–8169 (1989).**

201. **P.** Gund, W. T. Wipke, and R. Langridge, *Comput. Chem. Res. Ed. Technol.,* **3,** 5–21 (1974).

202. P. Gund, *Prog. Mol. Subcell. Biol.,* **11,117–143 (1977).**

203. A. **M.** Lesk, ***Commun. ACM,* 22, 221–224 (1979).**

204. S. E. Jakes and P. Willett, **J.** ***Mol. Graphics, 4,* 12–20 (1986).**

205. S. **E.** Jakes, N. Watts, P. Willett, D. Bawden, and J. D. Fisher, **J.** ***Mol. Graphics,* 5, 41–48** (1997).

206. P. **A.** Bartlett, G. T. Shea, S. **J.** Telfer, and S. Waterman in S. M. Roberts, Ed., ***Molecular Recognition: Chemical and Biological*** Prob***lems,*** Royal Society of Chemistry, London, **1989,** pp. **182–196.**

207. J. H. Van Drie, D. Weininger, and Y. C. Martin, **J.** ***Cornput.-Aided Mol. Des.,* 3,225–251 (1989).**

208. R. P. Sheridan, R. Nilakantan, **A.** I. Rusinko, N. Bauman, K. S. Haraki, and R. Venkataraghavan, **J.** ***Chem. Inf. Comput. Sci.,* 29, 255–260 (1989).**

209. Molecular Design, ***MACCS-3D,*** Molecular Design Ltd., San Leandro, CA, **1993.**

210. Chemical Design, ***CHEM-X,*** Chemical Design Ltd., Oxford OX2 OJB, UK, **1993.**

211. ***UNITY User's Manual,*** Tripos, Inc., St. Louis, MO, **2002.**

212. **Y.** C. Martin, M. G. Bures, and P. Willett in K. Lipkowitz and D. Boyd, Eds., ***Revisions in Computational Chemistry,*** VCH, New York, **1990,** pp. **213–263.**

213. **Y. C.** Martin, **J.** ***Med. Chem.,* 35, 2145–2154 (1992).**

214. A. I. Rusinko, R. P. Sheridan, R. Nilakantan, K. S. Haraki, N. Bauman, and R. Venkataraghavan, **J.** ***Chem. Inf. Comput. Sci.,* 29, 251–255 (1989).**

215. A. C. Good, S. J. Peterson, and W. G. Richards, ***J. Med. Chem.,* 36,2929–2937 (1993).**

216. R. **S.** Pearlman in H. Kubinyi, Ed., ***3D QSAR in Drug Design: Theory, Methods and Applications,*** ESCOM Scientific, Leiden, Netherlands, **1993,** pp. **41–79.**

217. J. Sadowski and J. Gasteiger, ***Chem. Rev.,* 93, 2567–2581 (1993).**

218. R. L. DesJarlais, R. P. Sheridan, G. L. Seibel, J. S. Dixon, I. D. Kuntz, and R. Venkataraghavan, **J.** ***Med. Chem.,* 31,722–729 (1988).**

219. R. L. DesJarlais, G. L. Seibel, I. **D.** Kuntz, P. S. F'urth, J. C. Alvarez, P. R. Ortiz de Montellano, D. L. Decamp, L. M. Babe, and C. S. Craik, ***Proc. Natl. Acad. Sci. USA,* 87, 6644–6648 (1990).**

220. R. L. DesJarlais, R. P. Sheridan, J. S. Dixon, I. D. Kuntz, and R. Venkataraghavan, ***J. Med. Chem.,* 29,2149–2153 (1986).**

221. H. J. Bohm, ***J. Cornput.-Aided Mol. Des.,* 6, 61–78 (1992).**

222. H. J. Bohm, ***J. Cornput.-Aided Mol. Des.,* 6, 593–606 (1992).**

223. P. L. Chau and P. M. Dean, **J.** ***Comput-Aided. Mol. Des.,* 6, 385–396 (1992).**

224. P. L. Chau and P. M. Dean, ***J. Comput-Aided. Mol. Des.,* 6,397–406** (1992).

225. **P. L.** Chau and P. M. Dean, ***J. Comput-Aided. Mol. Des.,* 6,407–426 (1992).**

226. C. M. W. Ho and G. R. Marshall, **J.** ***Comput.-Aided Mol. Des.,* 7, 3–22 (1993).**

227. G. Klebe, **J.** ***Mol. Med.,* 78,269–281** (2000).

228. H. **J.** Bohm and M. Stahl, ***Med. Chem. Res.,* 9, 445–462 (1999).**

229. H. **J.** Bohm and M. Stahl, ***Curr. Opin. Chem. Biol., 4,* 283–286 (2000).**

230. **R.** A. Lewis and P. M. Dean, ***Proc. R. Soc. Lond. B,* 236,141–162 (1989).**

231. R. A. Lewis and P. M. Dean, ***Proc. R. Soc. Lond.' B,* 236,125–140 (1989).**

232. Y. Nishibata and *A.* Itai, ***Tetrahedron,* 47, 8985–8990** (1991).

233. **Y.** Nishibata and A. Itai, ***J. Med. Chem.,* 36, 2921–2928 (1993).**

234. D. A. Pearlman and M. A. Murko, ***J. Comput. Chem.,* 14,1184–1193 (1993).**

235. M. G. Bures, C. Black-Schaefer, and G. Gardner, **J.** ***Cornput.-Aided Mol. Des.,* 5, 323–334 (1991).**

236. **J. M.** Blaney and J. S. Dixon, *Perspect. Drug. Discov. Des.,* **1, 301419 (1993).**

237. D. L. Bodian, R. B. Yamasaki, R. L. Buswell, J. F. Stearns, J. M. White, and I. D. Kuntz, ***Biochemistry,* 32,** 2967–2978 **(1993).**

238. C. S. Ring, E. Sun, J. H. McKerrow, G. K. Lee, P. J. Rosenthal, I. D. Kuntz, and F. E. Cohen, ***Proc. Natl. Acad. Sci. USA,* 90,** 3583–3587 **(1993).**

239. B. K. Shoichet, R. M. Stroud, D. V. Santi, I. D. Kuntz, and K. M. Perry, ***Science,* 259,** 1445–**1450 (1993).**

240. I. Halperin, B. Ma, H. Wolfson, and R. Nussinov, Proteins, 47, 409-443(2002).

241. E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Afialo, and I. A. Vakser, Proc. Natl. Acad. Sci. USA, 89, 2195-2199(1992).

242. I. A. Vakser and C. Aflalo, Proteins, 20, 320–329 (1994).

243. I. A. Vakser, Protein Eng., 9, 37-41 (1996).

244. I. A. Vakser, Biopolymers, 39, 455-464(1996).

245. H. A. Gabb, R. M. Jackson, and M. J. E. Sternberg, J. Mol. Biol., 272, 106-120 (1997).

246. M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, J. Mol. Biol., 261, 470-489 (1996).

247. S. R. Krystek, Jr., R. E. Bruccoleri, and J. Novotny, Znt. J. Pept. Protein Res., 38, 229-236 (1991).

248. J. Aqvist, J. Comput. Chem., 17, 1587-1597 (1996).

249. A. N. Jain, J. Cornput.-Aided Mol. Des., 10, 427-440 (1996).

250. A. Alex and P. Finn, THEOCHEM, 398, 551–554 (1997).

251. R. C. Wade, A. R. Ortiz, and F. Gago, Perspect. Drug Discov. Des., 9-11, 19-34 (1998).

252. R. M. A. Knegtel and P. D. J. Grootenhuis, Perspect. Drug Discov. Des., 9-11, 99-114 (1998).

253. M. D. Eldridge, C. W. Murray, T. R. Auton, G.V. Paolini, and R. P. Mee, J. Cornput.-Aided Mol. Des., 11, 425-445 (1997).

254. M. K. Gilson, J. A. Given, and M. S. Head, Chem. Biol., 4, 87–92 (1997).

255. M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, Biophys. J., 72, 1047-1069 (1997).

256. T. I. Oprea and G. R. Marshall, Perspect. Drug Discov. Des., 9-11, 35-61 (1998).

257. M. K. Holloway, Perspect. Drug Discov. Des., 9-11, 63–84 (1998).

258. H. J. Bohm, J. Cornput.-Aided Mol. Des., 12, 309–323 (1998).

259. T. Hansson, J. Marelius, and J. Aqvist, J. Comput.-Aided Mol. Des., 12, 27–35 (1998).

260. G. R. Marshall, R. H. Head, and R. Ragno in E. Di Cera, Ed., Thermodynamics in Biology, Oxford University Press, Oxford, UK, 2000, pp. 87-111.

261. S. Krystek, T. Stouch, and J. Novotny, J. Mol. Biol., 234, 661-679 (1993).

262. J. Novotny, R. E. Bruccoleri, and F. A. Saul, Biochemistry, 28, 4735-4749(1989).

263. R. D. Head, M. L. Smythe, T. I. Oprea, C. L. Waller, S. M. Green, and G. R. Marshall, J. Am. Chem. Soc., 118, 39594969(1996).

264. C. L. Waller, T. I. Oprea, A. Giolitti, and G. R. Marshall, J. Med. Chem., 36, 4152–4160 (1993).

265. N. Prattibiraman, M. Levitt, T. E. Ferrin, and R. Langridge, J. Comput. Chem., 6, 432(1985).

266. E. C. Meng, B. K. Shoichet, and I. D. Kuntz, J. Comput. Chem., 13, 505-524(1992).

267. S. Makino and I. D. Kuntz, J. Comput. Chem., 18, 1812-1825(1997).

268. T. J. A. Ewing and I. D. Kuntz, J. Comput. Chem., 18, 1175-1189(1997).

269. B. Sandak, R. Nussinov, and H. J. Wolfson, J. Comp. Biol., 5, 631-654(1998).

270. C. A. Baxter, C. W. Murray, D. E. Clark, D. R. Westhead, and M. D. Eldridge, Proteins, 33, 367-382 (1998).

271. D. M. Lorber and B. K. Shoichet, Protein Sci., 7, 938-950(1998).

272. Y. Sun, T. J. A. Ewing, A. G. Skillman, and I. D. Kuntz, J. Cornput.-Aided Mol. Des., 12, 597–604 (1998).

273. R. Mangoni, D. Roccatano, and A. Di Nola, Proteins, 35, 153-162 (1999).

274. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, J. Comput. Chem., 19, 1639-1662(1998).

275. M. Liu and S. M. Wang, J. Cornput.-Aided Mol. Des., 13, 435-451(1999).

276. S. Makino, T. J. A. Ewing, and I. D. Kuntz, J. Cornput.-Aided Mol. Des., 13, 513-532 (1999).

277. M. Rarey, B. Kramer, and T. Lengauer, Bioinformatics, 15, 243-250(1999).

278. R. M. A. Knegtel and M. Wagener, Proteins, 37, 334-345 (1999).

279. M. L. Lamb, K. W. Burdick, S. Toba, M. M. Young, K. G. Skillman, X. Q. Zou, J. R. Arnold, and I. D. Kuntz, Proteins, 42, 296-318(2001).

280. R. Abagyan and M. Totrov, Curr. Opin. Chem. Biol., 5, 375–382 (2001).

281. H. Claussen, C. Buning, M. Rarey, and T. Lengauer, J. Mol. Biol., 308, 377-395(2001).

282. C. A. Baxter, C. W. Murray, B. Waszkowycz, J. Li, R. A. Sykes, R. G. A. Bone, T. D. J. Perkins, and W. Wylie, J. Chem. Inf. Comput. Sci., 40, 254-262 (2000).

283. N. Ota and D. A. Agard, J. Mol. Biol., 314, 607-617 (2001).

284. S. Naruto, I. Motoc, G. R. Marshall, S. B. Daniels, M. J. Sofia, and J. A. Katzenellenbogen, J. Am. Chem. Soc., 107, 5262-5270 (1985).

285. K. P. Clark and Ajay, **J.** Comput. Chem., **16, 1210–1226 (1995).**

286. G. M. Verkhivker, P. A. Rejto, D. K. **Gehlhaar,** and S. T. Freer, Proteins, **25,342353 (1996).**

287. D. Q. McDonald and W. C. Still, **J. *Am.*** Chem. *Soc.*, **116,11550–11553 (1994).**

288. F. Guarnieri and W. C. Still, **J.** Comput. Chem., **15,1302–1310 (1994).**

289. D. Q. McDonald and W. C. Still, **J. *Am.*** Chem. *Soc.*, **118,2073–2077 (1996).**

290. Z. **R.** Wasserman and C. N. **Hodge,** Proteins, **24**, 227–237 **(1996).**

291. **J.** Desmet, I. A. Wilson, M. Joniau, M. De**maeyer,** and I. Lasters, ***FASEB*** **J., 11,** 164–172 (1997).

292. B. L. King, S. Vajda, and C. Delisi, FEBS Lett., **384**, 87–91 (1996).

293. **D.** S. Goodsell, H. Lauble, C. D. Stout, and A. **J.** Olson, Proteins, **17, 1–10 (1993).**

294. **R. X.** Wang and S. M. Wang, **J.** Chem. *Inf. Com*put. Sci., **41,1422–1426 (2001).**

295. P. S. Charifson, **J. J.** Corkery, M. A. Murcko, and W. P. Walters, **J.** Med. Chem., **42, 5100–5109 (1999).**

296. N. L. **Allinger,** Z.-q. S. Zhu, and K. Chen, **J. *Am.*** Chem. *Soc.*, **114,6120–6133 (1992).**

297. B. **R.** Brooks, R. E. **Bruccoleri,** B. D. Olafson, D. **J.** States, S. Swaminathan, and M. **Karplus, J.** Comput. Chem., **4, 187–217 (1983).**

298. F. A. **Momany** and R. **Rone, J.** Comput. Chem., **13,888–900 (1992).**

299. G. Nemethy, M. S. Pottle, and H. A. Scheraga, **J.** Phys. Chem., **87,1883–1887 (1983).**

300. T. A. Halgren, *J.* ***Am.*** Chem. *Soc.*, **114,** 7827–**7843 (1992).**

301. P. S. Charifson, R. G. **Hiskey,** L. G. Pedersen, and L. F. Kuyper, **J.** Comput. Chem., **12**, 899–**908 (1991).**

302. S. C. Hoops, K. W. Anderson, and K. M. Merz, Jr., **J. *Am.*** Chem. *Soc.*, **113,8262–8270 (1991).**

303. C. **J. Casewit,** K. S. Colwell, and A. K. Rappe, *J.* Am. Chem. *Soc.*, **114**, 10035–10046 (1992).

304. C. **J. Casewit,** K. S. Colwell, and A. K. Rappe, **J.** Am. Chem. *Soc.*, **114**, 10046–10053 (1992).

305. **A. K.** Rappe, C. **J. Casewit,** K. S. Colwell, W. A. Goddard III, and W. M. **Skiff, J. *Am.*** Chem. *Soc.*, **114**, 10024–10035 **(1992).**

306. Y.-D. **Wu** and K. N. Houk, **J. *Am.*** Chem. *Soc.*, **114,1656–1661 (1992).**

307. K. Houk, **J.** A. Tucker, and A. Dorigo, *Acc.* Chem. Res., **23,107–113 (1990).**

308. **F.** M. Menger and M. **J. Sherrod, J. *Am.*** Chem. *Soc.*, **112,8071–8075 (1990).**

309. D. P. Riley, P. **J.** Lennon, W. L. Neumann, and R. H. Weiss, **J. *Am.*** Chem. *Soc.*, **119,** 6522–**6528 (1997).**

310. K. **Aston,** N. Rath, A. **Naik,** U. Slomczynska, **O. F. Schall,** and D. P. Riley, Inorg. Chem., **40, 1779–1789 (2001).**

311. D. H. Williams, Aldrichimica *Acta*, **24, 71–80 (1991).**

312. A. **J.** Doig and D. H. Williams, **J. *Am.*** Chem. *Soc.*, **114,338–343 (1992).**

313. M. S. Searle and D. H. Williams, **J.** Am. Chem. *Soc.*, **114,10690–10697 (1992).**

314. M. S. Searle, D. H. Williams, and U. Gerhard, **J. *Am.*** Chem. *Soc.*, **114,10697–10704 (1992).**

315. **D.** H. Williams and B. **Bardsley,** *Perspect.* Drug Discov. Des., **17, 43–59 (1999).**

316. M. Graffner-Nordberg, K. Kolmodin, **J.** Aqvist, S. F. Queener, and A. Hallberg, *J.* Med. Chem., **44,2391–2402 (2001).**

317. **J.** Aqvist, V. B. Luzhkov, and B. O. Brandsdal, *Acc.* Chem. Rev., **35**, 358–365 **(2002).**

318. G. M. Verkhivker, D. Bouzida, D. K. Gehlhaar, P. A. Rejto, L. Schaffer, S. **Arthurs,** A. B. **Col**son, S. T. Freer, V. Larson, B. A. **Luty,** T. **Mar***rone,* and P. W. Rose, **J.** Med. Chem., **45**, 72–89 **(2002).**

319. R. Lumry and S. Rajender, Biopolymers, **9, 1125–1227 (1970).**

320. **I.** Muegge and Y. C. Martin, **J.** Med. *Chem.*, **42**, **791–804 (1999).**

321. B. A. **Grzybowski,** A. V. Ishcheno, **J.** Shimada, and E. I. Shakhnovich, *Acc.* Chem. Res., **35,** 261–269 **(2002).**

322. B. A. **Grzybowski,** A. V. Ishcheno, C.-Y. Kim, G. Topolov, R. Chapman, D. W. Christianson, G. M. Whitesides, and E. I. Shakhnovich, Proc. *Natl. Acad.* Sci. USA, **99,1270–1273 (2002).**

323. **K.** M. Merz, Jr., M. A. Murcko, and P. A. Kollman, **J. *Am.*** Chem. *Soc.*, **113, 4484–4490 (1991).**

324. C. F. Wong and J. A. **McCammon, J.** Am. Chem. *Soc.*, **108,3830–3832 (1986).**

325. L. M. Hansen and P. A. Kollman, **J.** Comput. Chem., **11,994–1002 (1990).**

326. B. G. Rao, R. F. **Tilton,** and U. C. Singh, **J. *Am.*** Chem. *Soc.*, **114,4447–4452** (1992).

327. W. **E.** Harte, **Jr.** and D. L. Beveridge, **J. *Am.*** Chem. *Soc.*, **115**, 3883–3886 (1993).

328. **D.** M. **Ferguson,** R. **J. Radmer,** and P. A. Kollman, **J.** Med. Chem., **34,2654–2659** (1991).

329. J. J. McDonald and C. L. Brooks III, J. Am. Chem. Soc., 114,2062–2072 (1992).

330. T. P. Lybrand and J. A. McCammon, J. Comput.-Aided Mol. Des., 2,259–266 (1988).

331. W. R. Cannon, J. D. Madura, R. P. Thummel, and J. A. McCammon, J. Am. Chem. Soc., 115, 879–884 (1993).

332. S. Yun-yu, A. E. Mark, W. Cun-Xin, H. Fuhua, J. C. Berendsen, and W. F. van Gunsteren, Protein Eng., 6, 289–295 (1993).

333. D. H. Rich, C.-Q. Sun, J. V. N. Vara Prasad, M. V. Toth, G. R. Marshall, P. Ahammadunny, M. D. Clare, R. D. Mueller, and K. Houseman, J. Med. Chem., 34,1222–1225 (1991).

334. P. De La Paz, J. M. Burridge, S. J. Oatley, and C. C. F. Blake in C. R. Beddell, Ed., The Design of Drugs to Macromolecular Targets, John Wiley & Sons, New York, 1992, pp. 119–172.

335. A. B. Edmundson, J. N. Herron, K. R. Ely, X.-M. He, D. L. Harris, and E. W. Voss, Jr., Philos. Trans. R. Soc. Lond. Biol., 323, 495–509 (1989).

336. C. Bihoreau, C. Monnot, E. Davies, B. Teutsch, K. E. Bernstein, P. Corvol, and E. Clauser, Proc. Natl. Acad. Sci. USA, 90, 5133–5137 (1993).

337. T. M. Fong, R. R. C. Huang, and C. D. Strader, J. Biol. Chem., 267,25664–25667 (1992).

338. U. Gether, T. E. Johansen, R. M. Snider, I. Lowe, J. A., S. Nakanishi, and T. W. Schwartz, Nature, 362,345–348 (1993).

339. M. F. Hibert, S. Trumpp-Kallmeyer, A. Bruinvels, and J. Hoflack, Mol. Pharmacol., 40, 8–15 (1991).

340. M. F. Hibert, S. Trumpp-Kallmeyer, J. Hoflack, and A. Bruinvels, Trends Pharmacol. Sci., 14, 7–12 (1993).

341. G. Nordvall and U. Hacksell, J. Med. Chem., 36, 967–976 (1993).

342. T. L. Blundell, B. L. Sibanda, M. J. E. Sternberg, and J. M. Thornton, Nature, 326, 347–352 (1987).

343. L. H. Pearl and W. R. Taylor, Nature, 329, 351–354 (1987).

344. I. T. Weber, Proteins, 7, 172–184 (1990).

345. L. M. Balbes and F. I. Carroll, Med. Chem. Res., 1, 283–288 (1991).

346. R. J. Siezen, W. M. de Vos, J. A. M. Leunissen, and B. W. Dijkstra, Protein Eng., 4, 719–737 (1991).

347. J. H. Brown, T. Jardetzky, M. A. Saper, B. Samraoui, P. J. Bjorkman, and D. C. Wiley, Nature, 332,845–850 (1988).

348. L. M. H. Koymans, N. P. E. Vermeulen, A. Baarslag, and G. M. Donne-op den Kelder, J. Cornput.-Aided Mol. Des., 7,281–289 (1993).

349. R. E. Bruccoleri and M. Karplus, Biopolymers, 26,137–168 (1987).

350. D. Jones and J. Thornton, J. Cornput.-Aided Mol. Des., 7,439–456 (1993).

351. J. U. Bowie and D. Eisenberg, Curr. Opin. Struct. Biol., 3,437–444 (1993).

352. S. J. Wodak and M. J. Rooman, Curr. Opin. Struct. Biol., 3, 247–259 (1993).

353. M. J. Sippl, J. Cornput.-Aided Mol. Des., 7, 473–501 (1993).

354. S. Miyazawa and R. C. Jernigan, Macromolecules, 18, 534–552 (1985).

355. S. H. Bryant and C. E. Lawrence, Proteins, 16, 92–112 (1993).

356. D. Frishman and H. W. Mewes, Nat. Struct. Biol., 4,626–628 (1997).

357. C. Chothia, Nature, 357,543–544 (1992).

358. P.-A. Lindgard and H. Bohr in H. Bohr and S. Bunak, Eds., Protein Folds, CRC Press, Boca Raton, FL, 1996, pp. 98–102.

359. P. E. Boscott, G. J. Barton, and W. G. Richards, Protein Eng., 6, 261–266 (1993).

360. D. Frishman and P. Argos, Proteins, 27, 329–335 (1997).

361. R. Srinivasan and G. D. Rose, Proteins, 22, 81–99 (1995).

362. K. A. Dill, H. S. Chan, and K. Yue, Macromol. Symp., 98,615–617 (1995).

363. K. Yue and K. A. Dill, Protein Sci., 5,254–261 (1996).

364. S. M. Le Grand and K. L. Merz, Jr. in S. M. Le Grand and K. L. Merz, Jr., Eds., The Protein Folding Problem and Tertiary Structure Prediction, Birkhauser, Boston, 1994, pp. 109–124.

365. S. Sun, Protein Sci., 2, 762–785 (1993).

366. J. U. Bowie and D. Eisenberg, Proc. Natl. Acad. Sci. USA, 91,4436–4440 (1994).

367. J. U. Bowie, K. Zhang, M. Wilmanns, and D. Eisenberg, Methods Enzymol., 266, 598–616 (1996).

368. S. G. Galaktionov and G. Marshall, Molecular Graphics and Drug Design: 27th Hawaii International Conference on System Sciences, IEEE Computer Society Press, Washington, DC, 1994.

369. S. Saitoh, T. Nakai, and K. Nishikawa, Proteins, 15, 191–204 (1993).

370. M. Vendruscolo, E. Kussell, and E. Domany, Fold. Des., **2,295–306 (1997).**

371. I. D. Kuntz, G. M. Crippen, P. A. Kolman, and D. Kimelman, J. Mol. Biol., **106, 983–994** (1976).

372. S. Galaktionov, G. V. Nikiforovich, and G. R. Marshall, Biopolymers, **60,** 153–168 **(2001).**

373. A. Aszodi and W. R. Taylor, Fold. Des., **1,** 325–334 (1996).

374. A. Aszodi, R. E. J. Munro, and W. R. Taylor, Fold. Des., **2, S3–S6 (1997).**

375. A. Aszodi and W. R. Taylor, Comput. Chem., **21,** 13–23 (1997).

376. S. R. Holbrook, I. Dubchak, and S.-H. Kim, Biotechniques, **14, 984–989 (1993).**

377. B. H. Park, E. S. Huang, and M. Levitt, J. Mol. Biol., **266, 831–846 (1997).**

378. G. M. Crippen and V. N. Maiorov in H. Bohr and S. Bunak, Eds., Protein Folds, CRC Press, Boca Raton, FL, **1996,** pp. **189–201.**

379. P. D. Thomas and K. A. Dill, J. Mol. Biol., **257, 457–469 (1996).**

380. S. Miyazawa and R. L. Jernigan, *J. Mol.* Biol., **256,623–644 (1996).**

381. V. J. Hruby, W. Qui, T. Okayama, and V. A. Soloshonok, Methods *Enzymol.*, **343, 91–123** (2002).

382. M. *G.* Bursavich and D. H. Rich, J. Med. Chem., **45,541–558 (2002).**

383. R. Hirschmann, K. C. Niwlaou, S. Pietranico, J. Salvino, E. M. Leahy, P. A. Sprengeler, G. Furst, and A. B. Smith III, J. Am. Chem. *Soc.*, **114,9217–9218 (1992).**

384. R. Hirschmann, P. A. Sprengeler, T. Kawasaki, J. W. Leahy, W. C. Shakespeare, and A. B. Smith III, *J. Am.* Chem. *Soc.*, **114,** 9699–**9701 (1992).**

385. T. W. Ku, F. E. Ali, L. S. Barton, J. W. Bean, W. E. Bondinell, J. L. Burgess, J. F. Callahan, R. R. Calvo, L. Chen, D. S. Eggelston, J. S. Gleason, W. F. Huffman, S. M. Hwang, D. R. Jakas, C. B. Karash, R. M. Keenan, K. D. Kopple, W. H. Miller, K. A. Newlander, A. Nichols, M. F. Parker, C. E. Peishoff, J. M. Samanen, I. Uzinskas, and J. W. Venslavsky, *J. Am.* Chem. *Soc.*, **115,** 8861–8862 **(1993).**

386. G. L. Olson, H.-C. Cheung, M. E. Voss, D. E. Hill, M. Kahn, V. S. Madison, C. M. Cook, J. Sepinwall, and G. Vincent, Concepts and Progress in the Design of *Peptide* Mimetics: Beta Turns and Thyrotropin Releasing Hormone (Biotechnology USA *1989)*, Conference Management Corporation, Norwald, CT, **1989,** pp. **348–360.**

387. G. L. Olson, D. R. Bolin, M. P. Bonner, M. Bos, C. M. Cook, D. C. Fry, B. J. Graves, M. Hatada, D. E. Hill, M. Kahn, V. S. Madison, V. K. Rusiecki, R. Sarabu, J. Sepinwall, G. P. Vincent, and M. E. Voss, J. Med. Chem., **36,** 3039–**3049 (1993).**

388. P. C. Belanger and C. Dufresne, Can. J. Chem., **64,1514–1520 (1986).**

389. R. Hirschmann, K. C. Nicolaou, S. Pietranico, E. M. Leahy, J. Salvino, B. Arison, M. A. Cichy, P. G. Spoors, W. C. Shakespeare, P. A. Sprengeler, P. Hamley, A. B. Smith III, T. Reisine, K. Raynor, L. Maechler, C. Donaldson, W. Vale, R. M. Friedinger, M. R. Cascieri, and C. D. Strader, **J. Am.** Chem. *Soc.*, **115,12550–12568 (1993).**

390. J. H. Arevalo, E. A. Stura, M. J. Taussig, and I. A. Wilson, J. Mol. Biol., **231,103–118(1993).**

391. J. H. Arevalo, M. J. Taussig, and I. A. Wilson, Nature, **365,859–863 (1993).**

392. P. Traxler, J. Green, H. Mett, U. Sequin, and P. Furet, J. Med. Chem., **42, 1018–1026 (1999).**

393. P. Traxler, G. Bold, E. Buchdunger, G. Caravatti, P. Furet, P. Manley, T. O'Reilly, J. Wood, and J. Zimmermann, Med. Res. Rev., **21,** 499–512 (2001).

394. R. Bureau, C. Daveu, J. C. Lancelot, and S. Rault, J. Chem. *Inf.* Comput. Sci., **42,429–436** (2002).

395. Y. Kato, A. Itai, and Y. Iitaka, Tetrahedron Lett., **43,5229–5236 (1987).**

396. W. H. Moos, C. C. Humblet, I. Sircar, C. Rithner, R. E. Weishaar, J. A. Bristol, and A. T. McPhail, J. Med. Chem., **30, 1963–1972** (1987).

397. D. Mayer, C. B. Naylor, I. Motoc, and G. R. Marshall, J. Cornput.-Aided Mol. Des., **1,** 3–16 (1987).

398. G. R. Marshall, C. D. Barry, H. E. Bosshard, R. A. Dammkoehler, and D. A. Dunn in E. C. Olsen and R. E. Christoffersen, Eds., Computer-Assisted Drug Design, American Chemical Society, Washington, DC, **1979,** pp. 205–**226.**

399. R. A. Dammkoehler, S. F. Karasek, E. F. B. Shands, and G. R. Marshall, Constrained Search of Conformational Hyperspace: Segmentation and Parallelism, Abstr. 204th ACS National Meeting, American Chemical Society, Washington, DC, **1992.**

400. G. R. Marshall and R. D. Cramer III, Trends Pharmacol. Sci., **9,** 285–289 (1988).

401. R. D. Cramer III and S. B. Wold, Comp. Mol. Field Anal. *(CoMFA)*, **5,388** (editorial) **(1991).**

402. *J. R. Sufrin, D. A. Dunn, and G. R. Marshall, Mol. Pharmacol.,* **19, 307313 (1981).**

403. *P. R. Andrews, E. J. Lloyd, J. L. Martin, S. L. Munro, M. Sadek, and M. G. Wong in A. S. V.* Burgen, *G. C. K. Roberts,* and M. *S.* Tute, *Eds., Molecular Graphics and Drug Design,* Elsevier *Science, Amsterdam,* **1986,** *pp.* 216–255.

404. *G.* Klopman *and S. Srivastava, Mol. Pharmacol.,* **37,** 958–965 (1989).

405. *G. Klopman and M. L. Dimayuga, J. Comput.-Aided Mol. Des.,* **4, 117–130** (1990).

406. *G. Rum and W. C. Herndon, J. Am. Chem. Soc.,* **113,** 9055–9060 **(1991).**

407. *C. Silipo and A. Vittoria in C. A.* Ramsden, *Ed., Quantitative Drug Design, Pergamon Press, Oxford, UK,* **1990,** *pp.* 153–204.

408. *G. M. Crippen in D.* Bawden, *Ed., Distance Geometry and Conformational Calculations (Chemometrics Research Studies), Vol.* **1,** *John* Wiley & *Sons, Chichester, UK,* **1981.**

409. *D. E. Clark, P. Willett, and P. W. Kenny, J. Mol. Graphics,* **10, 194–204 (1992).**

410. *C. A. Pepperrell and P. Willett, J. Comput.-Aided Mol. Des.,* **5,455–474 (1991).**

411. *A. R. Poirette, P. Willett, and F. H. Allen, J. Mol. Graphics,* **11,** 2–14 **(1993).**

412. *G. R. Marshall and C. B.* Naylor *in C. A.* Ramsden, *Ed., Quantitative Drug Design, Pergamon Press, Oxford, UK,* **1990,** *pp.* **431–458.**

413. A. *Davis, B. H.* Warrington, *and J. G. Vinter, J. Cornput.-Aided Mol. Des.,* **1,** 97–120 (1987).

414. *H. Weinstein, R. Osman, S. Topiol, and J. P. Green, Ann. N. Y. Acad. Sci.,* **367, 434–448 (1981).**

415. N. *C. Cohen in B. Testa, Ed., Advances in Drug Research, Academic Press, New York,* **1985,** *pp.* **40–144.**

416. R. *C. Wade, K. J. Clark, and P. J. Goodford, J. Med. Chem.,* **36,140–147 (1993).**

417. *N. Marchand-Geneste, K. A. Watson, B. K.* Alsberg, *and R. D. King, J. Med. Chem.,* **45,** 399–409 **(2002).**

418. *C. Hansch, J. Mcclarin, T. Klein, and R.* Langridge, *Mol. Pharmacol.,* **27, 493–498 (1995).**

419. *C. Hansch,* T. *Klein, J.* McClarin, *R.* Langridge, *and N. W.* Cornell, *J. Med. Chem.,* **29, 615–620 (1986).**

420. *G. E. Kellogg, S. F. Semus, and D. J. Abraham, J. Comput.-Aided Mol. Des.,* **5,545–552 (1991).**

421. *G. E. Kellogg and D. J. Abraham, J. Mol. Graphics,* **10,212–217 (1992).**

422. D. *J. Danziger and P. M. Dean, J. Theor. Biol.,* **116,215–224 (1985).**

423. *S. K. Kearsley, J. Comput. Chem.,* **11,** 1187–1192 **(1990).**

424. G. *R. Marshall and C. D. Barry, Functional Representation of Molecular Volume for Computer-Aided Drug Design, Abstr. Amer.* Cryst. *Assoc., Honolulu, HI,* **1979.**

425. *A. J. Hopfinger, J. Med. Chem., 2,* **7196–7206 (1980).**

426. *Z. Simon, A. Chiriac, S. Holban, D. Ciubotariu, and G. I. Mihalas, Minimum Steric Difference, Research Studies Press, Letchworth, UK,* **1984.**

427. *D. Ciubotariu,* E. *Deretey, T. I. Oprea, T. I. Sulea, Z. Simon, L.* Kurunczi, *and A. Chiriac, Quant. Struct.-Act. Relat.,* **12,367–372 (1993).**

428. *H.-D. Holtje and S.* Marrer, *J. Cornput.-Aided Mol. Des.,* **1,** 23–30 **(1987).**

429. *A. J. Hopfinger, J. Med. Chem.,* **26, 990–996 (1983).**

430. *S. Namasivayam and P. M. Dean, J. Mol. Graphics,* **4,** 46 (1986).

431. *P. L. Chau and P. M. Dean, J. Mol. Graphics,* **5, 97 (1987).**

432. *C. Burt and W. G. Richards, J. Cornput.-Aided Mol. Des.,* **4,231–238** (1990).

433. *J.* Zabrocki, G. *D. Smith, J. B.* Dunbar, *Jr., H. Iijima, and G. R. Marshall, J. Am. Chem. Soc.,* **110,5875–5880** (1988).

434. J. *B. Ball, R. A. Hughes, P. F.* Alewood, *and P. R. Andrews, Tetrahedron,* **49, 34673478 (1993).**

435. J. *B. Ball and P. F.* Alewood, *J. Mol. Recognit.,'* **3,** 55–64 **(1990).**

436. G. *V.* Nikiforovich, *K. E. Kover, W. J. Zhang, and G. R. Marshall, J. Am. Chem. Soc.,* **122, 32623273 (2000).**

437. J. Labanowski, *I. Motoc, C. B.* Naylor, *D. Mayer, and R. A. Dammkoehler, Quant. Struct.-Act. Relat.,* **5, 138–152 (1986).**

438. *S.* Naruto, *I. Motoc, and G. R. Marshall, Eur. J. Med. Chem.,* **20,529–532** (1985).

439. *R.* P. Sheridan *and R. Venkataraghavan, J. Cornput.-Aided Mol. Des.,* **1,243–256**(1987).

440. *E. E.* Hodgkin, *A. Miller, and M. Whittaker, J. Comput.-Aided Mol. Des.,* **7,515–534**(1993).

441. *M. T. Barakat and P. M. Dean, J. Comput.-Aided Mol. Des.,* **4,295–316 (1990).**

442. *M. T. Barakat and P. M. Dean, J. Comput.-Aided Mol. Des.,* **4,** 317–330 **(1990).**

443. *T.* D. J. Perkins *and P. M. Dean, J. Comput.-Aided Mol. Des.,* **7,173–182 (1993).**

444. *I. Motoc, J.* Labanowski, *C. B.* Naylor, *D. Mayer, and R. A. Dammkoehler, Quant. Struct.-Act. Relat.,* **5, 99–105 (1986).**

445. R. D. Nelson, D. I. Gottlieb, T. M. Balasubramanian, and G. R. Marshall in R. S. Rapaka, G. Barnett, and R. L. Hawks, Eds., ***Opioid Peptides: Medicinal Chemistry,*** NIDA Office of Science, Rockville, MD, 1986, pp. 204-230.

446. A. K. Ghose and G. M. Crippen, ***J. Med. Chem.,*** 27, 901-914 (1984).

447. A. K. Ghose and G. M. Crippen, ***J. Med. Chem.,*** 28, 333-346 (1985).

448. A. K. Ghose and G. M. Crippen in C. A. Ramsden, Ed., ***Quantitative Drug Design,*** Pergamon Press, Oxford, UK, 1990, pp. 716-733.

449. A. K. Ghose and G. M. Chippen, ***Mol. Pharmacol.,*** 37, 725-734 (1990).

450. M. R. Linschoten, T. Bultsma, A. P. IJzerman, and H. Timmerman, ***J. Med. Chem.,*** 29, 278–286 (1986).

451. G. M. Donne-op den Kelder, ***J. Cornput.-Aided Mol. Des.,*** 1, 257-264 (1987).

452. T. I. Oprea, D. Ciubotariu, T. I. Sulea, and Z. Simon, ***Quant. Struct.-Act. Relat.,*** 12, 21-26 (1993).

453. J. P. Snyder, S. N. Rao, K. F. Koehler, A. Vedani, and R. Pellicciari in C. G. Wermuth, Ed., ***Trends in QSAR and Molecular Modelling 92,*** ESCOM Scientific, Leiden, Netherlands, 1993, pp. 44-51.

454. G. Klopman, ***Quant. Struct.-Act. Relat.,*** 11, 176-185 (1992).

455. I. B. Bersuker and A. S. Dimogo in K. B. Lipkowitz and D. B. Boyd, Eds., ***Revisions in Computational Chemistry,*** VCH, New York, 1991, pp. 423-460.

456. Y. C. Martin, M. G. Bures, E. A. Danaher, J. DeLazzer, I. Lico, and P. Pavlik, A., ***J. Comput.-Aided Mol. Des.,*** 7, 83–102 (1993).

457. G. Jones, P. Willett, and R. C. Glen, ***J. Mol. Biol.,*** 245, 4343 (1995).

458. T. I. Oprea and L. Kurunczi in N. Voiculetz, I. Motoc, and Z. Simon, Eds., ***Specific Interactions and Biological Recognition Processes,*** CRC Press, Boca Raton, FL, 1993, pp. 295-326.

459. W. E. Klunk, B. L. Kalman, J. A. Ferrendelli, and D. F. Covey, ***Mol. Pharmacol.,*** 23, 511-518 (1982).

460. J. A. Calder, J. A. Wyatt, D. A. Frenkel, and J. E. Casida, ***J. Cornput.-Aided Mol. Des.,*** 7, 45–60 (1993).

461. M. F. Hibert, R. Hoffmann, R. C. Miller, and A. A. Carr, ***J. Med. Chem.,*** 33, 1594-1600 (1990).

462. M. F. Hibert, M. W. Gittos, D. N. Middlemiss, A. K. Mir, and J. R. Fozard, ***J. Med. Chem.,*** 31, 1087-1093 (1988).

463. A. W. Schmidt and S. J. Peroutka, ***Mol. Pharmacol.,*** 36, 505–511 (1989).

464. M. L. Connolly, ***Science,*** 221, 709-713 (1983).

465. M. L. Connolly, ***J. Appl. Crystallogr.,*** 16, 548–558 (1983).

466. C. E. Kundrot, J. W. Ponder, and F. M. Richards, ***J. Comput. Chem.,*** 12, 402-409 (1991).

467. S. M. Le Grand and K. M. Merz, Jr., ***J. Comput. Chem.,*** 14, 349-352 (1993).

468. A. H. Beckett and A. F. Casey, ***J. Pharm. Pharmacol.,*** 6, 986-999 (1954).

469. L. B. Kier and H. S. Aldrich, ***J. Theor. Biol.,*** 46, 529-541 (1974).

470. L. G. Humber, F. T. Bruderlin, A. H. Philipp, M. Gotz, and K. Voith, ***J. Med. Chem.,*** 22, 761–767 (1979).

471. G. L. Olson, H. C. Cheung, K. D. Morgan, J. F. Blount, L. Todaro, L. Berger, A. B. Davidson, and E. Boff, ***J. Med. Chem.,*** 24, 1026-1034 (1981).

472. H.-D. Holtje and M. Tintelnot, ***Quant. Struct.-Act. Relat.,*** 3, 6–9 (1984).

473. W. C. Probst, L. A. Snyder, D. J. Schuster, J. Brosius, and S. C. Sealfon, ***DNA Cell Biol.,*** 11, 1-20 (1992).

474. S. Trumpp-Kallmeyer, J. Hoflack, A. Bruinvels, and M. Hibert, ***J. Med. Chem.,*** 35, 3448–3462 (1992).

475. D. Timms, A. J. Wilkinson, D. R. Kelly, K. J. Broadley, and R. H. Davies, ***Znt. J. Quantum Chem. Quantum Biol. Symp.,*** 19, 197–215 (1992).

476. D. Zhang and H. Weinstein, ***J. Med. Chem.,*** 36, 934-938 (1993).

477. B. L. Bush and R. B. Nachbar, Jr., ***J. Comput.-Aided Mol. Des.,*** 7, 587-619 (1993).

478. J. N. Weinstein, K. W. Kohn, M. R. Grever, V. N. Viswanadhan, L. V. Rubeinstein, A. P. Monks, D. A. Scudiero, L. Welch, A. D. Koutsoukos, A. J. Chiausa, and K. D. Paull, ***Science,*** 258, 447-451 (1992).

479. T. A. Andrea and H. Kalayeh, ***J. Med. Chem.,*** 34, 2824-2836 (1991).

480. S.-S. So and W. G. Richards, ***J. Med. Chem.,*** 35, 32014207 (1992).

481. I. V. Tetko, A. I. Luik, and G. I. Poda, ***J. Med. Chem.,*** 36, 811-814 (1993).

482. R. D. King, S. Muggleton, R. A. Lewis, and M. J. E. Sternberg, ***Proc. Natl. Acad. Sci. USA,*** 89, 11322-11326 (1992).

483. S. A. DePriest, E. F. B. Shands, R. A. Dammkoehler, and G. R. Marshall in C. Silipo and A. Vittoria, Eds., ***QSAR: Rational Approaches to***

the Design of Bioactive Compounds, Elsevier *Science, Amsterdam,* **1991,** *pp.* **405–414.**

484. S. *A.* DePriest, *D. Mayer, C. B.* Naylor, *and G. R. Marshall, J. Am. Chem. Soc.,* **115,** 5372–5384 **(1993).**

485. C. *Hansch, Acc. Chem. Res.,* **26, 147–153** (1993).

486. *G.* Klebe *and U. Abraham, J. Med. Chem.,* **36, 70–80 (1993).**

487. *D. P.* Getman, *G. A.* DeCrescenzo, *R. M. Heintz, K. L. Reed, J. J. Talley, M. L. Bryant, M. Clare, K. A. Houseman, J. J. Marr, R.* A. *Mueller, M. L. Vazquez, H.-S. Shieh, W. C.* Stallings, *and R. A. Stegeman, J. Med. Chem.,* **36,288–291 (1993).**

488. G. *M. Crippen, J. Comput. Chem.,* **8,943–955** (1987).

489. *M. P. Bradley and G. M. Crippen, J. Med. Chem.,* **36,** 3171–3177 **(1993).**

490. *F. Major, M. Turcotte, D. Gautheret, G.* Lapalme, *E.* Fillion, *and R. Cedergren, Science,* **253,** 1255–1260 **(1991).**

491. *D. Gautheret and R. Cedergren, FASEB J.,* **7, 97–105 (1993).**

492. *P. A. Greenidge, T. C. Jenkins, and S. Neidle, Mol. Pharmacol.,* **43,** 982–988 **(1993).**

493. *M.* G. *Cardozo and A. J. Hopfinger, Mol. Pharmacol.,* **40, 1023–1028 (1991).**

494. *M.* J. *J. Blommers, C. B.* Lucasius, *G.* Kateman, *and R. Kaptein, Biopolymers,* **22, 45–52 (1992).**

495. A. *G. Palmer* III *and D.* A. *Case, J. Am. Chem. Soc.,* **114,9059–9067 (1992).**

496. *K. Boehncke, M. Nonella, K. Schulten, and A. H. J. Wang, Biochemistry,* **30, 5465–5475 (1991).**

497. J. *Xingand H. L. Scott, Biochem. Biophys. Res. Commun.,* **165,** 1–6 **(1989).**

498. *T. R.* Stouch, *K. B. Ward,* A. Altieri, *and A. T. Hagler, J. Comput. Chem.,* **12, 1033–1046 (1991).**

499. *H.* L. *Scott and S.* Kalaskar, *Biochemistry, 28,* **3687–3691 (1989).**

500. *P. S.* O'Shea *and R.* Matela, *Biochem. Soc. Trans.,* **14,1119–1120 (1986).**

501. *D. M.* Kroll *and G. Gompper, Science,* **255, 968–971 (1992).**

502. *L. I.* Krishtakik, *V. V. Topolev, and Y. I.* Kharkats, *Biophysics,* **36,257–262** (1991).

503. *E. Egberts and H. J. C. Berendsen, J. Chem. Phys.,* **89,3718–3732 (1988).**

504. *R. P. Mason, D. G. Rhodes, and L. G. Herbette, J. Med. Chem.,* **34,869–877 (1991).**

505. *L. G. Herbette in C. G. Wermuth, Ed., Trends in QSAR and Molecular Modelling 92,* ESCOM *Scientific,* Leiden, *Netherlands,* **1993,** *pp.* **76–85.**

506. *H. Heller, M. Schaeffer, and K. Schulten, J. Phys. Chem.,* **97,8343–8360 (1993).**

507. *W. Im and B.* Roux, *J. Mol. Biol.,* **319,** 1177–**1197 (2002).**

508. *T. Kataoka,* D. *D.* Beusen, *J. D. Clark, M.* Yodo, *and G. R. Marshall, Biopolymers,* **32,** 1519–**1533 (1992).**

509. *G.* R. *Marshall, Tetrahedron,* **49,** 3547–3558 (1993).

510. *G. V.* Nikiforovich *and G. R. Marshall, Biochem. Biophys. Res. Commun.,* **195, 222–228 (1993).**

511. *G.* V. Nikiforovich *and V. J.* Hruby, *Biochem. Biophys. Res. Commun.,* **194,** 9–16 **(1993).**

512. *G. Nikiforovich and G. R. Marshall, Int. J. Pept. Protein Res.,* **42, 171–180 (1993).**

513. *G.* V. Nikiforovich *and G. R. Marshall, Int. J. Pept. Protein Res.,* **42, 181–193 (1993).**

514. *P.* Poulin *and F. P. Theil, J. Pharm. Sci.,* **91, 1358–1370 (2002).**

515. *G. M.* Keseruu *and L. Molnar, J. Chem. Inf. Comput. Sci.,* **42,437–444** (2002).

516. *H. van de Waterbeemd, Curr. Opin. Drug Discov. Dev.,* **5, 33–43** (2002).

517. J. Langowski *and A. Long, Adv. Drug Deliv. Rev.,* **54,** 407–415 **(2002).**

518. *S.* Ekins *and J. Rose, J. Mol. Graph. Model.,* **20,305309 (2002).**

519. *T.* I. *Oprea,* I. *Zamora, and A. L. Ungell, J. Comb. Chem.,* **4,258–266 (2002).**

520. *H. E. Selick,* A. *P. Beresford,* and M. *H.* Tarbit, *Drug Discov. Today,* **7, 109–116 (2002).**

521. A. P. *Li and M.* Segall, *Drug Discov. Today,* **7, 25–27 (2002).**

522. *A. Kulkarni, Y. Han, and A. J. Hopfinger, J. Chem. Inf. Comput. Sci.,* **42,331342** (2002).

523. *R. D. Brown, M. Hassan, and M.* Waldman, *J. Mol. Graph. Model.,* **18,** 427–437,537 (2000).

524. O. Roche, *P. Schneider, J. Zuegge, W. Guba, M.* Kansy, *A. Alanine, K. Bleicher, F. Danel, E. M. Gutknecht, M. Rogers-Evans, W.* Neidhart, *H.* Stalder, *M.* Dillon, *E.* Sjogren, *N. Fotouhi, P.* Gillespie, *R.* Goodnow, *W. Harris, P. Jones, M.* Taniguchi, *S. Tsujii,* W. *von der* Saal, *G.* Zim-

mermann, and G. Schneider, J. Med. Chem., **45,137–142 (2002).**

525. *O. Llorens, J. J. Perez, and H. O. Villar, J. Med. Chem.,* **44,2793–2804** (2001).

526. **A.** *Cheng,* **D.** *J. Diller, S. L. Dixon, W. J. Egan, G. Lauri, and K. M. Merz, Jr., J. Comput. Chem.,* **23,172–183 (2002).**

527. *T.* **I.** *Oprea, J. Cornput.-Aided Mol. Des.,* **14, 251–264 (2000).**

528. *T. Olsson and T. I. Oprea, Curr. Opin. Drug Discov. Dev.,* **4, 308–313 (2001).**

529. **D.** *Gorse and R. Lahana, Curr. Opin. Chem. Biol.,* **4,287–294 (2000).**

530. *M.* **J. Valler** *and D. Green, Drug Discov. Today,* **5,286–293 (2000).**

531. **Y.** *C. Martin,* ***Farmaco*****, 56, 137–139** (2001).

532. *J.* **Xu** *and J. Stevenson, J. Chem. Inf. Comput. Sci.,* **40,** 1177–1187 **(2000).**

533. *J.* **S.** *Mason and B. R.* **Beno,** *J. Mol. Graph. Model.,* **18,438–451,538 (2000).**

534. *T.* **I.** *Oprea and J. Gottfries, J. Mol. Graph. Model.,* **17,261–274,329 (1999).**

535. **A. K.** *Mandagere, T. N. Thompson, and K. K. Hwang, J. Med. Chem.,* **45, 304–311 (2002).**