

Computer-Aided Molecular Design

J. ANDREW McCAMMON

Theoretical chemistry, as implemented on fast computers, is beginning to yield accurate predictions of the thermodynamic and kinetic properties of large molecular assemblies. In addition to providing detailed insights into the origins of molecular activity, theoretical calculations can be used to design new molecules with specific properties. This article describes two types of calculations that show special promise as design tools, the thermodynamic cycle-perturbation method and the Brownian reactive dynamics method. These methods can be applied to calculate equilibrium and rate constants that describe many aspects of molecular recognition, stability, and reactivity.

THE ABILITY TO DESIGN MOLECULES WITH DESIRED CHEMICAL or biochemical activities is fast becoming a reality. This capability partly reflects the introduction of new methods in theoretical chemistry and new computers with the power to apply theory in the analysis of detailed molecular models. These design tools should play a significant role in the search for new drugs, chemical sensors, chromatographic agents, enzymes, pesticides, microelectronic components, and other materials. Given the structure, a desired molecule can be synthesized by traditional methods or by more recent ones such as solid-phase synthesis or genetic manipulation and expression. In addition to providing molecular blueprints, the new theoretical tools provide detailed physical interpretations of molecular activity.

This article describes and illustrates two particular design tools, the thermodynamic cycle-perturbation method and the Brownian reactive dynamics method (1). Both of these methods involve computer simulations of the atomic motion in models of the molecular system of interest.

In its simplest form, the thermodynamic cycle-perturbation method is used to predict how the affinity of one molecule for another (for example, the affinity of a drug for a receptor) depends on the chemical composition of the two molecules. The solvent surroundings are explicitly included in the simulation model. In other applications, the method can be used to estimate the relative free energies of activation (and thus the rate constants) of certain kinds of chemical reactions, the relative folding stabilities of homologous proteins, or other thermodynamic differences.

The Brownian reactive dynamics method is used to predict the rate of initial diffusional encounter between reactant molecules in solution. From a design viewpoint, this rate sets the ultimate limit on the speed of enzymatic and other reactions. If the reactant molecules are such that chemical events develop very rapidly when the reactants come into contact, the net rate of reaction will be equal to the rate of diffusional encounter. Because the rates of such diffusion-controlled reactions can be calculated by Brownian

dynamics methods, one can predict how changes in the reactants will change these rates.

The thermodynamic cycle-perturbation method and the Brownian reactive dynamics method both represent advances in chemical theory. In practice, however, the usefulness of these methods will often depend on the availability of experimental structural data and computational power. For calculations on complicated molecules such as enzymes, one must have available the three-dimensional structure of molecules that are closely related to the design targets. Such calculations also typically require access to supercomputers or minisupercomputers. Fortunately, developments in x-ray crystallography and other structural techniques are providing the needed structural information at an accelerating rate (2), and the power of and access to computers continue to improve.

Thermodynamic Cycle-Perturbation Method

The thermodynamic cycle-perturbation method was developed to answer questions such as the following (3). Given the structure of a ligand-receptor complex, how will the strength of ligand binding change if one group on the ligand is replaced by another? If the complex is that of an enzyme and substrate, how will changes in the substrate affect the rates of the reaction steps that follow binding? If a protein is changed in some manner (such as by site-directed mutation), how do these changes alter the thermodynamic stability of the native conformation of the molecule? And how would changes in an enzyme or other receptor alter ligand-binding properties and subsequent reactivity? Quantitative answers to such questions can be expressed in terms of equilibrium constants or changes in free energy. Thus the affinity of a drug D for a receptor P in the simple association reaction



is given by

$$K = \exp(-\Delta G^0/RT) \quad (2)$$

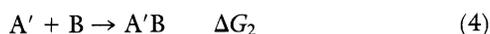
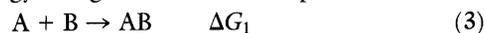
where $K = [DP]/[D][P]$ is the equilibrium constant, ΔG^0 is the change in the standard-state Gibbs free energy, R is the gas constant, and T is the temperature in Kelvin. The Gibbs energy is the appropriate measure of the driving force for reactions at constant temperature and pressure. Helmholtz energy changes ΔA (strictly appropriate for constant temperature and volume) are also commonly used; ΔA is similar to the corresponding ΔG for many reactions in solution. Thermodynamic cycle-perturbation calculations usually involve model systems in which the solutes are effectively at infinite dilution. The free energy differences that are calculated can thus be identified with standard-state differences if the usual biochemical definition of standard states is used (4).

By focusing on differences in homologous processes, the thermodynamic cycle-perturbation method takes advantage of extensive cancellation of large contributions to the changes in thermodynamic functions for a single process. For example, the binding of a drug to a receptor may be accompanied by complicated changes in the conformation, solvation, and other properties of both molecules.

The author is in the Department of Chemistry, University of Houston, Houston, TX 77004.

The changes that accompany the binding of two drugs that differ in a single chemical group may, however, be very similar. The thermodynamic cycle-perturbation method exploits this situation in a natural way by considering the relatively small perturbations of structure and energy associated with replacing one chemical group by another in either the drug-receptor complex (where the complicated changes have already occurred) or in the unbound drug molecule.

The general idea of the thermodynamic cycle-perturbation approach is quite simple. Suppose that one is interested in computing the relative free energy change for two different processes



The desired quantity, $\Delta\Delta G = \Delta G_2 - \Delta G_1$, can in principle be obtained from molecular dynamics simulations in which processes 3 and 4 are caused to occur sufficiently slowly in the appropriate solvent surroundings. In molecular dynamics simulations, a computer is used to solve the classical equations of motion for all the atoms in a system for a finite period of time (I). To cause a process such as that in Eq. 3 to occur, an auxiliary force is applied that gradually brings molecules A and B together. Analysis of such simulations with standard methods yields ΔG_1 and ΔG_2 (1, 5). Unfortunately, this direct approach is unworkable except for very simple molecules. The difficulty is that each process must be performed slowly enough so that the system remains in thermodynamic equilibrium. For molecules comprising more than a few atoms, conformation or solvation changes that require more than a few tens of picoseconds can make it difficult to ensure that representative configurations of the system develop during the simulations, which themselves typically cover only 10 to 100 psec. In the thermodynamic cycle-perturbation approach, one considers instead the nonphysical processes



Because processes 3 through 6 form a thermodynamic cycle



the desired relative free energy change $\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$. Analogous relations hold for any other thermodynamic state function, for example, enthalpy or entropy. The quantities ΔG_3 and ΔG_4 can again be calculated by molecular dynamics simulations. Because the changes in the processes represented by Eq. 5 and 6 are typically much smaller and more localized than those in Eqs. 3 and 4, the calculations are greatly simplified.

Free energy changes for nonphysical processes such as Eqs. 5 or 6 can be evaluated with statistical mechanical perturbation theory (1, 5, 6). For example, ΔG_4 can be obtained as

$$\Delta G_4 = -RT \ln \langle \exp(\Delta V/RT) \rangle_{AB} + \Delta G_4^m \quad (8)$$

for small perturbations. In Eq. 8, $\Delta V = V_{AB} - V_{A'B}$, where V_{AB} and $V_{A'B}$ are the potential energies of the systems containing AB and A'B, respectively, and $\langle \rangle_{AB}$ denotes an average over the thermally accessible configurations of the system containing AB. The quantity ΔG_4^m is a contribution due to the difference in mass of AB and A'B (7). It is usually not included in the calculation because it cancels with a similar term in the other half of the thermodynamic cycle; that is, $\Delta G_4^m = \Delta G_3^m$.

The average $\langle \exp(\Delta V/RT) \rangle_{AB}$ can be obtained from molecular dynamics simulations of AB and the surrounding solvent under

isothermal-isobaric conditions in order to obtain ΔG_4 . However, the results usually converge slowly except for very small perturbations. A more general way to obtain a value of ΔG is to express this quantity as

$$\Delta G = \sum_{i=1}^n \Delta G_i \quad (9)$$

where

$$\Delta G_i = -RT \ln \langle \exp[(V_i - V_{i+1})/RT] \rangle_i \quad (10)$$

In the case of ΔG_4 , one has $V_1 = V_{AB}$ and $V_{n+1} = V_{A'B}$; V_i changes gradually (linearly or nonlinearly) from V_{AB} to $V_{A'B}$ when i changes from 1 to $n + 1$. Other methods such as thermodynamic integration can alternatively be used to compute the free energy changes (1, 5).

The example given in Eqs. 3 through 6 involves bimolecular (covalent or noncovalent) association. In principle, the thermodynamic cycle-perturbation approach can also be used to predict reaction rates (by calculating relative free energies of activation), to predict the folding stabilities of globular biopolymers (by calculating relative free energies of unfolding), and in other applications. The results of some early applications are mentioned below. Important theoretical and technical problems must be solved before these methods become fully reliable, however; some of these problems are discussed in the concluding remarks.

Applications of Perturbation Methods

Several studies of molecular recognition and activity based on the thermodynamic cycle-perturbation method have been published during the past two years (8-12). The first application of the method was to predict and rationalize the ion-binding selectivity of an organic host molecule in water (8). Subsequent applications have included studies of the association between enzymes and inhibitors (9-11), of the folding stability of mutant proteins (10), and of the catalytic rates of enzymes (12). Representative examples of some of this work are briefly described here.

In the ion-binding study, the relative free energy of binding Cl^- and Br^- to the macrotricyclic molecule SC24 (Fig. 1) (13) in water was calculated (8). The processes are described by Eqs. 3 through 6 with $A = \text{Cl}^-$, $A' = \text{Br}^-$, and $B = \text{SC24}$. The perturbation technique was used to compute the Helmholtz free energy changes ΔA_3 and ΔA_4 . All simulations were performed with the canonical ensemble (that is, constant temperature, volume, and number of particles) at 300 K. For the computation of ΔA_3 , the system

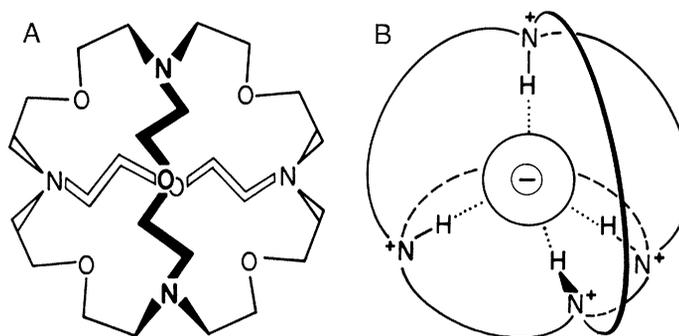
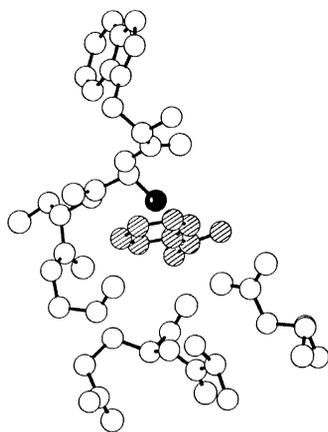


Fig. 1. (A) Structure of the synthetic receptor molecule SC24 (13). (B) Schematic diagram of the complex formed between SC24 and a chloride or bromide ion (the negatively charged sphere in the center of the complex). The ability of the receptor to recognize or selectively bind one ion versus another in solution is quantitatively defined by the ratio of the equilibrium constants for binding the two ions. This ratio has been calculated by the thermodynamic cycle-perturbation method (8).

Fig. 2. The inhibitor benzamidine (hatched spheres) bound in the specificity pocket of the enzyme trypsin. The methyl group (solid sphere) added in the mutation of Gly²¹⁶ to Ala²¹⁶ projects toward the amidinium group. The resulting steric conflict causes the binding affinity to be reduced from that of the wild-type enzyme. The ratio of equilibrium constants for inhibitor binding has been calculated by the thermodynamic cycle-perturbation method (1, 9, 10).



consisted of Cl⁻ and 214 water molecules in a cubic box of length 18.62 Å with periodic boundary conditions. The system for calculation of ΔA_4 comprised SC24, Cl⁻, and 191 water molecules in a cubic box as above. The initial structure for the SC24 complex with Cl⁻ was taken from an x-ray crystallographic study (14). The potential energy function for the system was that from the GROMOS molecular modeling program, supplemented as described elsewhere (9, 15). In replacing Cl⁻ by Br⁻, the essential change is an increase of the ionic radius by about 0.15 Å. The molecular dynamics simulations were performed with the program AMBER (16). All hydrogen masses were increased to 10 amu to allow a long dynamics time step, $\Delta t = 4$ femtoseconds, by slowing the librational motions of groups containing hydrogen. This mass adjustment has no effect on equilibrium properties of a classical system but does result in a more efficient sampling of configurations.

After extensive equilibration of each system, simulations were continued for 30 psec. Configurations were saved every 0.1 psec, and were used to calculate ΔA_3 and ΔA_4 as outlined above. The perturbations are small enough that statistically reliable values ($\Delta A_3 = 3.35 \pm 0.15$ kcal/mol, $\Delta A_4 = 7.50 \pm 0.20$ kcal/mol) could be obtained with Eq. 8 rather than the more general Eqs. 9 and 10. The predicted relative free energy of binding ($\Delta\Delta A = 4.15$ kcal/mol) is in good agreement with the experimental result that was subsequently made available (4.3 kcal/mol) (17). Examination of ΔA_3 and ΔA_4 suggests that selective binding of Cl⁻ to SC24 is due to the highly favorable interaction of Cl⁻ with the receptor, which more than compensates for the unfavorable free energy of desolvation of Cl⁻ versus Br⁻. The more favorable interaction of Cl⁻ with the receptor relative to Br⁻ arises because the Br⁻ anion is slightly too large to be comfortably accommodated in the relatively rigid SC24 molecule.

The error estimates given above for ΔA_3 and ΔA_4 were obtained by comparing the results from several 10-psec simulations. Because each simulation is of finite length and thus provides a sample of finite size, slight differences are found in the results of different simulations. These error estimates do not include certain systematic errors that may occur, as described in the concluding remarks.

The benzamidine-trypsin system was chosen for initial studies of biomolecular recognition with the thermodynamic cycle-perturbation method (9, 10). This choice was motivated by several factors. Benzamidine is one of the simplest inhibitors of trypsin. An x-ray structure of the enzyme-inhibitor complex is available (18), in which it is apparent that the benzene moiety of the inhibitor contacts several hydrophobic groups in the walls of the specificity pocket of the enzyme, and that the positively charged amidinium moiety forms a salt bridge with the carboxyl group of an aspartic acid residue at the base of this pocket. Thermodynamic data are available for the binding of benzamidine with various substituent groups to

trypsin (19). Also, a variety of modified trypsins have been produced by genetic manipulations; these and other modified forms of the enzyme are obvious candidates for model studies of ligand binding (20).

The first step of the thermodynamic calculation was to perform a molecular dynamics simulation of benzamidine-inhibited trypsin in water. The system is a very large one, comprising 4785 water molecules plus the inhibited enzyme in a box of dimensions 49.15 by 54.43 by 64.28 Å. Hydrogen atoms that were capable of participating in hydrogen bonds were included explicitly. The total number of atoms in the system was 16,384. Periodic boundary conditions were used; the box dimensions were chosen to be large enough so that all solute atoms were separated from the closest atoms of image solutes by at least four layers of solvent. The dynamics calculations were performed on a supercomputer with a vectorized version of the GROMOS program (15). After starting with the x-ray structure for the complex and a bulk solvent configuration, the system was relaxed and equilibrated at 300 K during a period of 16.3 psec. A subsequent 28.8-psec simulation was performed with the system coupled to a constant temperature bath at 300 K.

To compare the binding of differently substituted benzamidines to trypsin or of benzamidine to differently substituted enzymes, the simulation of the complex needed to be supplemented with simulations of the separated inhibitor or enzyme, respectively, in water. Such simulations have been performed (9, 10). Significant results of the analysis of these simulations include the following. The para-fluoro analog of benzamidine binds somewhat less strongly to trypsin than does benzamidine itself ($\Delta\Delta A \approx 0.9$ kcal/mol). Unlike the ion-binding system described above, selectivity in the present case primarily reflects solvation effects, as it is less difficult to desolvate benzamidine than its parafluoro derivative ($\Delta A_3 \approx -0.8$ kcal/mol). The effects upon benzamidine binding of the mutation of Gly²¹⁶ to Ala²¹⁶ in trypsin have also been examined (Fig. 2). The mutant enzyme was predicted to have the lower affinity for benzamidine ($\Delta\Delta A \approx 1.3$ kcal/mol), primarily as a result of steric crowding in the binding site due to the methyl group added in the mutation. All of the net changes in free energies of binding are in accord with experimental data (9).

Exciting applications of the thermodynamic cycle-perturbation method to other enzymes have also been reported recently. These include successful predictions of the effects of deleting a hydrogen-bonding group in an inhibitor of the enzyme thermolysin (11) and, in work by Rao *et al.* (12), of the change in the free energy of activation (or rate constant) for cleavage of a tripeptide substrate due to a mutation in the enzyme subtilisin. The method has also been used to predict the relative free energies of hydration of a variety of molecules (7, 21).

Brownian Reactive Dynamics Method

The Brownian reactive dynamics method was developed to calculate the rate at which reactant molecules diffusing in solution would collide with the appropriate orientations for reaction (22–24). If chemical events within the encounter complex unfold so quickly that the overall rate of reaction is just equal to this rate of encounter, the reaction is said to be diffusion controlled (24, 25). A wide variety of reactions exhibit diffusion control, including many reactions between ions or free radicals, and certain biological redox and enzyme-catalyzed reactions. The rates of diffusional encounter between real molecules are influenced by a number of complicated factors (1). Typically, only part of the molecular surface is active. Electrostatic, hydrodynamic, and other interactions between the

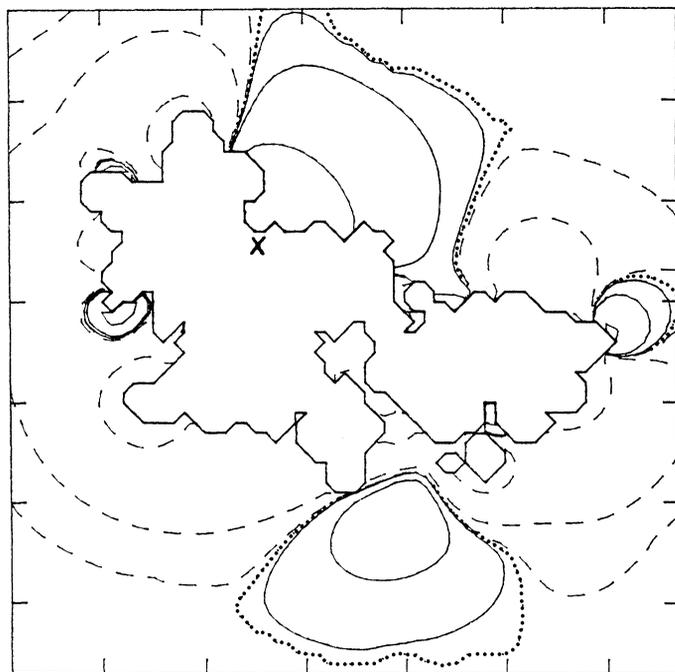


Fig. 3. The electrostatic potential energy field surrounding the enzyme superoxide dismutase in an electrolyte solution (0.1M ionic strength); the distance between tic marks on the axes is 12.5 Å. Brownian reactive dynamics calculations show how the catalytic rate of the enzyme is increased by the effect of this field in steering the diffusion of substrate (O_2^-) molecules toward the active sites, one of which is indicated by X (1, 24, 30–32). The contours correspond to surfaces at which the effective potential energy of O_2^- is ± 0.006 , ± 0.06 or ± 0.6 kcal/mol; dashed, dotted, and solid lines indicate positive, zero, and negative energies, respectively. Contours with the largest magnitudes are closest to the protein, whose surface is represented by the heavy solid line.

reaction partners will generally produce attractive or repulsive forces that can increase or decrease the rate of diffusional encounter. Structural fluctuations in the reactant molecules may result in a time-dependent reactivity upon contact. Effects such as these can influence reactions in a variety of interesting ways. For example, reactants that have complementary distributions of electric charge may be “steered” into productive orientations during diffusional encounters, which can result in increased reaction rates.

In the Brownian reactive dynamics method, simulations of the diffusional motion of reactant molecules are performed and the trajectories are then analyzed to determine the mechanism and rate of reactive encounter. The model systems used in the simulations can be constructed to include virtually any feature of real molecules. Also, any appropriate algorithm can be used to simulate the motion of the reactants. Most studies to date have made use of the Ermak-McCammon algorithm, in which successive positions of a set of particles are sampled from probability distributions that represent short-time solutions to the diffusion equation for the particles (26).

The calculation of a rate constant for reactive encounter is in principle straightforward for a dilute solution, where it is sufficient to consider the dynamics of one reactant particle relative to a fixed reaction partner. The rate constant can then be written as

$$k = k_D(b)\beta_\infty \quad (11)$$

Here, $k_D(b)$ is the steady-state rate at which mobile particles first strike a spherical surface of radius b around a target particle, and β_∞ is the probability that a reactant starting at $r = b$ will react with the target rather than escape. If b is chosen large enough that the reactant pair interactions are approximately centrosymmetric for $r > b$, then $k_D(b)$ can be determined analytically. For example, $k_D(b)$

is given by the familiar Smoluchowski result

$$k_D(b) = 4\pi bD \quad (12)$$

where D is the relative diffusion constant for the reactant pair if the interaction force between the reactants vanishes for $r > b$.

The quantity β_∞ reflects the complicated interactions that occur at short range and is calculated from the simulated trajectories. In principle, one could perform a number of trajectory calculations, each with a mobile particle starting at a randomly chosen point on the $r = b$ surface, and obtain β_∞ as that fraction of the trajectories that lead to reaction rather than escape. This straightforward approach has to be modified, however, because of the difficulty of determining the ultimate fate of particles that have not reacted at the end of a simulation of finite length. Given the nature of Brownian motion, even a mobile particle that has diffused a large distance away from the target could ultimately return toward the target and react. In practice, therefore, trajectories are truncated at a radius $q > b$ and the increase in the rate constant due to trajectories that would have returned and reacted is determined as an analytic correction. The formula that is used in the simplest calculations is then (22)

$$k = k_D(b)\beta[1 - (1 - \beta)\Omega]^{-1} \quad (13)$$

where β is the uncorrected probability of reaction in the presence of the truncation surface and $\Omega = k_D(b)/k_D(q)$.

Thus, to calculate the rate constant for a bimolecular diffusion-controlled reaction in dilute solution, one need only compute a number of trajectories of one reactant diffusing in the vicinity of the other, fixed reactant. Trajectories are initiated at $r = b$ and terminated upon reaction or upon reaching $r = q$. The fraction of trajectories that react is β , and this quantity yields the rate constant through Eq. 13. Analysis of the trajectories also provides information on the mechanistic details of the reaction, for example, whether reactants tend to be steered into productive collision geometries during the diffusional encounter.

A variety of extensions of these basic ideas have also been developed. These extensions require somewhat more elaborate calculations but can lead to substantial increases in efficiency in the rate constant calculations for certain systems (23, 24).

Applications of Brownian Methods

The Brownian reactive dynamics method has been used successfully to calculate reaction rates in two rather different systems. One system involves the diffusion-controlled reaction of the small substrate molecule superoxide (O_2^-) catalyzed by the enzyme superoxide dismutase (SOD), and the other involves the diffusion-controlled association of two proteins (cytochrome c and cytochrome c peroxidase) to form an electron transfer complex. In both systems, substantial rate enhancements result from electrostatic interactions that steer the reactants toward optimal reaction geometries during diffusional encounter.

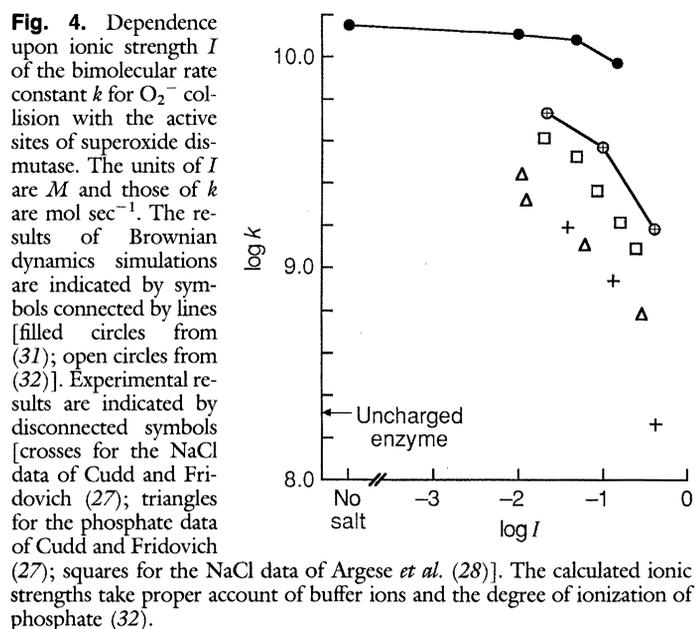
The reaction rate of the SOD- O_2^- system is quite high and decreases with increasing ionic strength of the solvent (27, 28). Because the active sites represent a small fraction (~0.1%) of the surface area of the dimeric enzyme, and the electrostatic charges of the enzyme (-4) and substrate (-1) are of the same sign, it has been recognized that the approach of substrate to the active sites is likely facilitated by the anisotropic electrostatic potential field around the enzyme (27–29). These results were qualitatively confirmed in the first Brownian dynamics simulations of simple models of the SOD- O_2^- system (30). Current studies of SOD reactivity involve fairly realistic models of the molecular shapes and interactions (31, 32). In the calculations by Allison and co-workers (24, 32), the topography

of the enzyme surface, including the important channels leading to the catalytic centers, was modeled by construction of a 1 Å resolution exclusion grid based on the x-ray structure of the enzyme (33). Also, the electrostatic potential field is evaluated on a 1.25 Å resolution grid with finite difference methods to solve the linearized Poisson-Boltzmann equation (Fig. 3) (34). This electrostatic model reflects the detailed topography of the enzyme, the difference in dielectric constants of the enzyme and solvent, and the concentration of salt dissolved in the solvent. The calculated rate of O_2^- reaction is in reasonable agreement with experiment, as is the variation of the rate with salt concentration (Fig. 4). The rate constants obtained by Sharp *et al.* (31) are about twice those of Allison *et al.* (32), but they also display the correct variation with salt concentration at high ionic strength. Because the methods used by the two groups are similar, the differences in results probably reflect differences in the details of the models used (for example, in the effective size of the superoxide ion). More importantly, the level of agreement among these independent theoretical and experimental results indicates that Brownian dynamics is becoming a reliable approach for predicting the rates of diffusion-controlled reactions of complicated molecules.

The recent study by Northrup *et al.* (35) of cytochrome c association with cytochrome c peroxidase makes use of topographic and electrostatic models similar to those described above for the larger of the two proteins (peroxidase), which is treated as a stationary target. Cytochrome c is represented as a rigid array of 34 charges corresponding to the ionized residues. Each charge interacts with the peroxidase as an independent small test charge (that is, the interior of cytochrome c is approximated as an extension of the solution to simplify the calculation of electrostatic forces and torques). The steric interactions with peroxidase are determined by the positions of a set of surface atoms, and the array translates and rotates independently with appropriate diffusion constants. Stringent conditions for productive association were defined in terms of the relative distance and orientation of the heme groups in the two proteins. Two reactive regions at the peroxidase surface were found to satisfy these conditions, a primary region near Asp³⁴ and a secondary region near Asp¹⁴⁸. Good agreement with experimental data for the association rate and its dependence on salt concentration was obtained for this model. The electrostatic forces and torques were essential to account for the rate. At a univalent salt concentration of 0.1M, and with a requirement of heme plane alignment to within 60° for reaction, the calculated rate constants in the presence and absence of the electrostatic interactions are $5.4 \times 10^8 M^{-1} \text{sec}^{-1}$ and $0.41 \times 10^8 M^{-1} \text{sec}^{-1}$, respectively; the corresponding experimental rate constant is $4.75 \times 10^8 M^{-1} \text{sec}^{-1}$ (36).

Concluding Remarks

The attentive reader will have noticed that no molecules were actually designed in the work described here. The design potential of the thermodynamic cycle-perturbation and Brownian reactive dynamics methods should be clear, however. These new methods can, in favorable cases, provide accurate predictions concerning the thermodynamic and kinetic effects of changes in molecular composition. The reliability of the predictions can be expected to depend on the system considered and the extent of the compositional changes. Consider, for example, a ligand and receptor that can bind in different orientations with only a small difference in free energy. Replacement of one functional group on the ligand by another could lead to a reversal of the favored orientation, yet this reversal may not be seen in the simulation if there is a sizable energy barrier to motion from the first orientation to the second. Additional



theoretical work is needed to handle problems like this one, in which the system to be compared differ by an extensive structural rearrangement.

Other limitations of the methods described here should also be mentioned. These include the implicit assumption in studies of mutant proteins that the mutations will not lead to substantial unfolding and inactivation of the protein. In principle, the thermodynamic cycle-perturbation method can be used to predict shifts in the folding-unfolding equilibrium, but practical calculations will require a better understanding of the structures of unfolded proteins than is currently available (10). All of the types of calculations discussed here are also limited by the accuracy of the underlying potential energy functions. For example, the types of functions that are commonly used at present do not explicitly allow for fluctuations in the polarization of electronic distributions within atoms and molecules, and may therefore not be quantitatively reliable for the study of certain very strong electrostatic interactions.

Viewed in a positive light, limitations such as those described above represent deeply interesting research problems (1). Moreover, the methods that are presently available are sufficient to begin design work on a wide variety of interesting problems. For example, the availability of x-ray structural data for drug-virus complexes (37) has stimulated thermodynamic cycle-perturbation studies of these systems (38). Also, Brownian dynamics calculations of the effects of mutations in SOD upon its diffusion-controlled rate constant are in progress with a view toward enhancing the activity of this protein in medical and industrial applications.

REFERENCES AND NOTES

1. J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, Cambridge, 1987).
2. W. G. J. Holt, *Angew. Chem. Int. Ed. Engl.* **25**, 767 (1986); H. Fruhbeis, R. Klein, H. Wallmeier, *ibid.* **26**, 403 (1987).
3. B. L. Tembe and J. A. McCammon, *Comput. Chem.* **8**, 281 (1984).
4. K. E. Van Holde, *Physical Biochemistry* (Prentice-Hall, Englewood Cliffs, NJ, 1971).
5. M. Mezei and D. L. Beveridge, *Ann. N.Y. Acad. Sci.* **482**, 1 (1986).
6. J. P. M. Postma, H. J. C. Berendsen, J. R. Haak, *Faraday Symp. Chem. Soc.* **17**, 55 (1982).
7. T. P. Lybrand, I. Ghosh, J. A. McCammon, *J. Am. Chem. Soc.* **107**, 7793 (1985).
8. T. P. Lybrand, J. A. McCammon, G. Wipff, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 833 (1986).
9. C. F. Wong and J. A. McCammon, *J. Am. Chem. Soc.* **108**, 3830 (1986).
10. _____, in *Structure, Dynamics and Function of Biomolecules*, A. Ehrenberg, R. Rigler, A. Graslund, L. Nilsson, Eds. (Springer-Verlag, Berlin, 1987), p. 51.

11. P. A. Bash *et al.*, *Science* **235**, 574 (1987).
12. S. N. Rao, U. C. Singh, P. A. Bash, P. A. Kollman, unpublished results. An interesting study of catalytic rates in wild-type and mutant trypsins has also been reported by A. Warshel and F. Sussman, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3806 (1986); there, perturbation methods were used to calculate free energy changes for physical processes, rather than the nonphysical processes considered in thermodynamic cycle-perturbation calculations.
13. J. M. Lehn, *Science* **227**, 849 (1985).
14. B. Metz, J. M. Rosalky, R. Weiss, *Chem. Commun.* **14**, 533 (1976).
15. J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, J. P. M. Postma, *Biopolymers* **23**, 1513 (1984).
16. P. K. Weiner and P. A. Kollman, *J. Comput. Chem.* **2**, 287 (1981).
17. E. Kaufmann, thesis, Université Louis Pasteur, Strasbourg, France (1979).
18. W. Bode and P. Schwager, *J. Mol. Biol.* **98**, 693 (1975).
19. M. Mares-Guia, D. L. Nelson, E. Rogana, *J. Am. Chem. Soc.* **99**, 2331 (1977).
20. C. S. Craik *et al.*, *Science* **228**, 291 (1985).
21. W. L. Jorgensen and C. Ravimohan, *J. Chem. Phys.* **83**, 3050, (1985); C. L. Brooks, *J. Phys. Chem.* **90**, 6680 (1986); T. P. Straatsma, H. J. C. Berendsen, J. P. M. Postma, *J. Chem. Phys.* **85**, 6720 (1986); U. C. Singh, F. K. Brown, P. A. Bash, P. A. Kollman, *J. Am. Chem. Soc.* **109**, 1607 (1987); P. A. Bash, U. C. Singh, R. Langridge, P. A. Kollman, *Science* **236**, 564 (1987); C. L. Brooks, *J. Chem. Phys.* **86**, 5156 (1987).
22. S. H. Northrup, S. A. Allison, J. A. McCammon, *J. Chem. Phys.* **80**, 1517 (1984).
23. S. A. Allison, N. Srinivasan, J. A. McCammon, S. H. Northrup, *J. Phys. Chem.* **88**, 6152 (1984); S. A. Allison, S. H. Northrup, J. A. McCammon, *J. Chem. Phys.* **83**, 2894 (1985); S. H. Northrup, M. S. Curvin, S. A. Allison, J. A. McCammon, *ibid.* **84**, 2196 (1986).
24. J. A. McCammon, R. J. Bacquet, S. A. Allison, S. H. Northrup, *Faraday Discuss. Chem. Soc.* **83**, 000 (1987).
25. D. F. Calef and J. M. Deutch, *Annu. Rev. Phys. Chem.* **34**, 493 (1983); O. G. Berg and P. H. von Hippel, *Annu. Rev. Biophys. Chem.* **14**, 131 (1985); J. A. McCammon, S. H. Northrup, S. A. Allison, *J. Phys. Chem.* **90**, 3901 (1986); J. Keizer, *Chem. Rev.* **87**, 167 (1987).
26. D. L. Ermak and J. A. McCammon, *J. Chem. Phys.* **69**, 1352 (1978).
27. A. Cudd and I. Fridovich, *J. Biol. Chem.* **257**, 11443 (1982).
28. E. Argese, P. Viglino, G. Rotilio, M. Scarpa, A. Rigo, *Biochemistry* **26**, 3224 (1987).
29. E. D. Getzoff, J. A. Tainer, P. K. Weiner, P. A. Kollman, J. S. Richardson, D. C. Richardson, *Nature (London)* **306**, 287 (1983).
30. S. A. Allison and J. A. McCammon, *J. Phys. Chem.* **89**, 1072 (1985); T. Head-Gordon and C. L. Brooks, *ibid.* **91**, 3349 (1987).
31. K. Sharp, R. Fine, B. Honig, *Science* **236**, 1460 (1987).
32. S. A. Allison, R. J. Bacquet, J. A. McCammon, *Biopolymers*, in press.
33. J. A. Tainer, E. D. Getzoff, K. M. Beem, J. S. Richardson, D. C. Richardson, *J. Mol. Biol.* **160**, 181 (1982).
34. J. Warwicker and H. C. Watson, *ibid.* **157**, 671 (1982); I. Klapper, R. Hagstrom, R. Fine, K. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **1**, 47 (1986).
35. S. H. Northrup, J. O. Boles, J. C. L. Reynolds, *J. Phys. Chem.*, in press.
36. C. H. Kang, D. L. Brautigan, N. Osheroff, E. Margoliash, *J. Biol. Chem.* **253**, 6502 (1978).
37. T. J. Smith *et al.*, *Science* **233**, 1286 (1986).
38. T. P. Lybrand, W. F. Lau, J. A. McCammon, B. M. Pettitt, in *Protein Structure and Design*, vol. 69 (new series) of *UCLA Symposia on Molecular and Cellular Biology*, D. Oxender, Ed. (Liss, New York, in press).
39. Many of the ideas and results presented here were developed in collaboration with S. A. Allison, R. J. Bacquet, W. F. Lau, T. P. Lybrand, M. H. Mazar, S. H. Northrup, B. M. Pettitt, and C. F. Wong. S. A. Allison, B. Honig, P. A. Kollman, and S. H. Northrup kindly provided preprints of unpublished work. P. Kollman, S. Northrup, and A. Treasurywala made helpful suggestions after reviewing an early draft of this article. The AMBER and GROMOS programs were generously provided by P. A. Kollman and W. F. van Gunsteren, respectively. This work was supported in part by the National Science Foundation, the National Institutes of Health, the Robert A. Welch Foundation, the Texas Advanced Technology Research Program, and the National Center for Supercomputing Applications. The author is the recipient of the 1987 George H. Hitchings Award from the Burroughs Wellcome Fund.

New Perspectives in Cell Adhesion: RGD and Integrins

ERKKI RUOSLAHTI AND MICHAEL D. PIERSCHBACHER

Rapid progress has been made in the understanding of the molecular interactions that result in cell adhesion. Many adhesive proteins present in extracellular matrices and in the blood contain the tripeptide arginine-glycine-aspartic acid (RGD) as their cell recognition site. These proteins include fibronectin, vitronectin, osteopontin, collagens, thrombospondin, fibrinogen, and von Willebrand factor. The RGD sequences of each of the adhesive proteins are recognized by at least one member of a family of structurally related receptors, integrins, which are heterodimeric proteins with two membrane-spanning subunits. Some of these receptors bind to the RGD sequence of a single adhesion protein only, whereas

others recognize groups of them. The conformation of the RGD sequence in the individual proteins may be critical to this recognition specificity. On the cytoplasmic side of the plasma membrane, the receptors connect the extracellular matrix to the cytoskeleton. More than ten proved or suspected RGD-containing adhesion-promoting proteins have already been identified, and the integrin family includes at least as many receptors recognizing these proteins. Together, the adhesion proteins and their receptors constitute a versatile recognition system providing cells with anchorage, traction for migration, and signals for polarity, position, differentiation, and possibly growth.

THE ATTACHMENT OF CELLS TO THEIR SURROUNDINGS IS important in determining cell shape and in maintaining proper cell function and tissue integrity. Such binding helps anchor cells and provides positional signals that direct cellular traffic and differentiation. Most cells possess multiple mechanisms for binding to the structures that surround them. For example, they can bind to extracellular matrices (1) or to other cells (2).

Extracellular matrices are made up of an insoluble meshwork of protein and carbohydrate that is laid down by cells and that fills most of the intercellular spaces. Matrices in different locations in the body consist of different combinations of collagens, proteoglycans,

Cancer Research Center, La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037.