

## RESEARCH ARTICLES

# Discrimination Between Native and Intentionally Misfolded Conformations of Proteins: ES/IS, a New Method for Calculating Conformational Free Energy That Uses Both Dynamics Simulations With an Explicit Solvent and an Implicit Solvent Continuum Model

Yury N. Vorobjev,<sup>1\*</sup> Juan Carlos Almagro,<sup>2</sup> and Jan Hermans<sup>1</sup>

<sup>1</sup>Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina, Chapel Hill, North Carolina

<sup>2</sup>Instituto de Quimica, Universidad Autonoma de Mexico, Mexico D. F., Mexico

**ABSTRACT** A new method for calculating the total conformational free energy of proteins in water solvent is presented. The method consists of a relatively brief simulation by molecular dynamics with *explicit solvent* (ES) molecules to produce a set of microstates of the macroscopic conformation. Conformational energy and entropy are obtained from the simulation, the latter in the quasi-harmonic approximation by analysis of the covariance matrix. The *implicit solvent* (IS) dielectric continuum model is used to calculate the average solvation free energy as the sum of the free energies of creating the solute-size hydrophobic cavity, of the van der Waals solute-solvent interactions, and of the polarization of water solvent by the solute's charges. The reliability of the solvation free energy depends on a number of factors: the details of arrangement of the protein's charges, especially those near the surface; the definition of the molecular surface; and the method chosen for solving the Poisson equation. Molecular dynamics simulation in explicit solvent relaxes the protein's conformation and allows polar surface groups to assume conformations compatible with interaction with solvent, while averaging of internal energy and solvation free energy tend to enhance the precision. Two recently developed methods—SIMS, for calculation of a smooth invariant molecular surface, and FAMBE, for solution of the Poisson equation via a fast adaptive multigrid boundary element—have been employed. The SIMS and FAMBE programs scale linearly with the number of atoms. SIMS is superior to Connolly's MS (molecular surface) program: it is faster, more accurate, and more

stable, and it smooths singularities of the molecular surface. Solvation free energies calculated with these two programs do not depend on molecular position or orientation and are stable along a molecular dynamics trajectory. We have applied this method to calculate the conformational free energy of native and intentionally misfolded globular conformations of proteins (the EMBL set of deliberately misfolded proteins) and have obtained good discrimination in favor of the native conformations in all instances. Proteins 32:399–413, 1998. © 1998 Wiley-Liss, Inc.

**Key words:** protein stability; conformational free energy; structure discrimination; molecular dynamics; molecular surface; continuum solvent model; continuum dielectric model; boundary element method; protein entropy; quasi-harmonic approximation; deliberately misfolded protein structures

## INTRODUCTION

The goal of distinguishing between different compact folded conformations of proteins in terms of a free energy based on a strictly physical potential has

Grant sponsor: National Center for Research Resources of the National Institutes of Health; Grant number: RR08102; Grant sponsor: National Science Foundation; Grant numbers: MCB-9314854 and BIR-9318159.

Yury N. Vorobjev is on leave from the Novosibirsk Institute of Bioorganic Chemistry, Novosibirsk, Russia.

\*Correspondence to: Yury N. Vorobjev, Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina, Chapel Hill, NC 27599–7260.

Received 12 December 1997; Accepted 6 April 1998

been elusive. In fact, the empirical atom–atom potentials commonly used in molecular mechanics models discriminate poorly between correctly folded and misfolded structures.<sup>1–3</sup> Calculations for the EMBL database of deliberately misfolded proteins<sup>4,5</sup> show that heuristic likelihood potentials derived from analysis of known high-resolution protein structures<sup>6</sup> discriminate more effectively. These structure-derived potentials are very helpful for fast structure recognition and for application in the so-called threading problem,<sup>6</sup> but they also have limitations.<sup>7</sup> In particular, the heuristic methods provide only low-resolution structures, which have to be refined by methods based on all-atom descriptions of the energetics.

In this article we report a new method, called ES/IS, to estimate the total free energy of a protein in aqueous solution, which combines molecular dynamics simulation with explicit solvent (ES) with calculation of the solvation free energy for the continuum dielectric, implicit solvent (IS) representation.<sup>8–10</sup> The ES/IS method generates a set of microstates along a 50- to 100-ps molecular dynamics (MD) trajectory and calculates average intraprotein energy and average solvation free energy; the conformational entropy is estimated by quasi-harmonic analysis of the fluctuation covariance matrix. Averaging over microstates enhances the accuracy of the free energy estimation for the macroscopic conformation. We have found the ES/IS method to reliably discriminate the correctly folded protein from deliberately misfolded structures in the EMBL database set.<sup>4</sup> The calculation of the right balance between the intraprotein energy and the free energy of the solvent polarization is found to be the most important element for correct discrimination. Therefore, development of an implicit solvent method that is fast and accurate, and stable to small conformational changes of the solute molecule, is an important element of the method; this in turn depends on the efficiency, accuracy, and stability of methods of calculation of the molecular surface of the solute molecule and solution of the Poisson equation in the dielectric continuum model.

We review two major innovations related to the implicit solvation model: a new method, SIMS (smooth invariant molecular surface), for generating the molecular surface (MS) in terms of small surface elements of known size, and another new method, FAMBE (fast adaptive multigrid boundary element), which uses a boundary element method for calculation of the solvent polarization charge density on the molecular surface and the solvent polarization free energy. These new methods have computational superiority: they are fast, scaling linearly with the number of atoms, and they produce results that are numerically accurate and stable to small changes of the solute molecule's conformation or position. More detailed descriptions of these new methods can be found elsewhere.<sup>11,12</sup> In the next section we formulate ES/IS and review its component terms: the

implicit solvent continuum dielectric model with the new submethods, SIMS and FAMBE, and methods for calculation of hydrophobic solvation free energy and protein conformational entropy.

## THEORY

### Formulation of the ES/IS Method for Calculating Conformational Free Energy

A protein in an aqueous solution undergoes continuous conformational dynamics, and a macroscopic conformation A is an ensemble of microscopic conformations  $\mathbf{x}_{A,i}$ . In the ES/IS method the (conformational) free energy of a protein in solution is calculated as the sum of two parts; one part contains the intramolecular energy and entropy, and the other part contains the free energy of solvation. The free energy  $G_A$ , is presented as

$$G_A = \langle U_m(\mathbf{x}) \rangle_A - TS_{\text{conf},A} + \langle W(\mathbf{x}) \rangle_A \quad (1)$$

where  $\langle \rangle_A$  denotes an average over microconfigurations of the conformation A,  $U_m$  represents the intraprotein conformational energy, and  $S_A$  is the entropy of the conformation A.  $W(\mathbf{x})$  is the solvation free energy, which is computed for each microconformation (for practical reasons, for a fixed-interval subset of microconformations).

In the ES/IS method, a representative set of microscopic configurations  $\mathbf{x}_{A,i}$  of a solute in a solvent is generated by MD simulation with explicit solvent along a relatively short trajectory (50–100 ps). As we shall discuss later,  $W(\mathbf{x})$  is a potential of mean force that can itself be written as a sum of terms for cavity formation, solute–water van der Waals interactions, and electrostatic polarization of solvent by the polar components of the solute. As a result, Eq. 1 becomes

$$G_A = \langle U_{m,\text{sh}} \rangle_A + \langle U_{m,\text{coul}} \rangle_A - TS_{\text{conf},A} + \langle G_{\text{cav}} \rangle_A + \langle G_{s,\text{vdw}} \rangle_A + \langle G_{\text{pol}} \rangle_A \quad (2)$$

where the intramolecular potential energy  $U_m$  has been represented as a sum of short-range (i.e., angle deformation and van der Waals) terms,  $U_{m,\text{sh}}$ , and electrostatic coulombic interactions,  $U_{m,\text{coul}}$ .

Of the six terms in Eq. 1, three, namely,  $U_{m,\text{sh}}$ ,  $U_{m,\text{coul}}$ , and  $G_{s,\text{vdw}}$ , are accumulated as averages during the molecular dynamics simulation; as will be discussed, the free energy,  $G_{s,\text{vdw}}$ , of van der Waals interactions between solute and solvent can be accurately approximated by the potential energy of these interactions:<sup>13</sup>

$$G_{s,\text{vdw}} = U_{s,\text{vdw}} \quad (3)$$

The entropic term,  $TS_{\text{conf},A}$  is estimated in the harmonic approximation from the covariance matrix of the positional fluctuations during the dynamics trajectory. (This term pertains only to the conformational fluctuations of the protein and is not to be

confused with the total entropy, which also contains contributions from solvent polarization and cavity formation.) Finally, the remaining two terms are found with models in which the solvent is treated implicitly, as a continuum. As will be discussed, the free energy for formation of the cavity is well approximated with a term given by the product of the molecular surface and a microscopic surface tension,  $\gamma_v = 69 \text{ cal}/\text{\AA}^2$ ,<sup>14</sup> and the free energy of solvent polarization,  $G_{\text{pol}}$ , is found by modeling the solvent as a continuum dielectric, with Poisson's equation.

### Implicit Continuum Model of Solvent

An accurate calculation of the free energy of fluid systems using computer simulations poses serious difficulties. Unlike a quantity such as energy, which can be expressed as an ensemble average, the free energy is related to the partition function and thus to the volume of accessible phase space.<sup>15–18</sup> Here, the goal is to calculate the free energy of a system of protein molecule plus solvent. As shall be discussed later, it is possible to estimate the free energy of the structurally highly organized protein molecule from the results of a simulation; however, the same is not true of the solvent. Consequently, we adopt an implicit description of solvation and thereby obtain a partition function of the solute in which the interactions with the solvent are represented through a solvation potential (a potential of mean force) that depends explicitly on the solute's coordinates.

The implicit solvent model can be consistently derived from a statistical mechanical description of an explicit solvent–solute system. The partition function,  $Z$ , of a solute molecule (coordinates  $\mathbf{x}$ ) in a solvent (coordinates  $\mathbf{y}$ ) can be written as the ratio of the partition functions for solution and pure solvent systems (containing identical numbers of solvent molecules),<sup>19</sup> that is,

$$Z = \frac{\int d\mathbf{x} \int d\mathbf{y} \exp \{-\beta[U_m(\mathbf{x}) + U_{m,s}(\mathbf{x}, \mathbf{y}) + U_{s,s}(\mathbf{y})]\}}{\int d\mathbf{y} \exp \{-\beta U_{s,s}(\mathbf{y})\}}. \quad (4)$$

Here  $U_m(\mathbf{x})$  is the intramolecular potential energy,  $U_{m,s}(\mathbf{x}, \mathbf{y})$  is the potential energy of the solute–solvent interactions, and  $U_{s,s}(\mathbf{y})$  is the potential energy of the solvent–solvent interactions. This can be rewritten as a partition function with solvent-mediated interactions between atoms of the solute molecule:<sup>20,21</sup>

$$Z = \int d\mathbf{x} \exp \{[-\beta[U_m(\mathbf{x}) + W(\mathbf{x})]]\} \quad (5)$$

where

$$\exp[-\beta W(\mathbf{x})] = \frac{\int d\mathbf{y} \exp \{-\beta[U_{m,s}(\mathbf{x}, \mathbf{y}) + U_{s,s}(\mathbf{y})]\}}{\int d\mathbf{y} \exp \{-\beta U_{s,s}(\mathbf{y})\}}. \quad (6)$$

Here  $W(\mathbf{x})$  is the free energy of solvation of the solute molecule with conformation  $\mathbf{x}$ . With this implicit-solvent description, the protein molecule's energy is determined in part by intramolecular forces and in part by an external potential,  $W(\mathbf{x})$ ; this then leads directly to Eq. 1.

The solvation free energy  $W(\mathbf{x})$  can be decomposed into parts<sup>8–10,22</sup> by considering transfer of a solute from gas phase to solvent as a multistep process:

$$W = G_{\text{cav}} + G_{\text{s,vdw}} + G_{\text{pol}} \quad (7)$$

where  $G_{\text{cav}}$  is the free energy of creating the empty solute-sized cavity in the solvent;  $G_{\text{s,vdw}}$  is the free energy of inserting the solute molecule into the solute-sized cavity, that is, turning on the van der Waals interactions between the solute atoms and the solvent molecules; and  $G_{\text{pol}}$  is the free energy of solvent polarization, that is, the excess free energy of charging the solute atoms in the solvent from charges  $q_i$  set to zero to their regular values  $q_i$ , the excess being relative to the same process in vacuo. The terms in Eq. 7 describe elementary physical processes and can be treated individually.

### Free Energy Required to Form the Cavity

In the ES/IS method the free energy of formation of a cavity in solvent of a size necessary to accept the protein molecule is expressed as the product of the surface area and a surface free energy. The rationale for this is briefly examined in what follows.

The free energy  $G_{\text{cav}}$  of cavity formation can be expressed in terms of a process in which the solute is "grown" in the solvent as a consequence of conversion of the interactions between solute and solvent from zero to their regular value, while charges remain zero. Via this process,  $G_{\text{cav}}$  can be estimated by simulations with explicit solvent. However, a calculation of the free energy of forming a cavity the size of a protein by dynamics simulation with thermodynamic integration or perturbation methods has not yet been done; this poses serious computational difficulties because of the slow relaxation of solvent during so-called growth of the solute atoms. The largest peptide for which the solvation energy in water has been calculated from first principles is 14-residue polyalanine.<sup>23</sup> Of course, solute and solute–solvent force fields are not perfectly accurate, and long simulations are needed to attain a high degree of precision. For example, solvation free energies for methane calculated in this manner in TIP4P or SPC water are  $2.5 \pm 0.4$  and  $2.28 \text{ kcal/mol}$ , respectively,<sup>24,25</sup> while the experimental value is  $2.0 \text{ kcal/mol}$ .<sup>26</sup> A simple approach, in which the cavity term is estimated from the cavity surface, appears to give at least as reliable results and may, therefore, be preferable.

It has been found that measured free energies of transfer of hydrocarbons from gas to water vary linearly with the molecular surface of the solvent

molecule, which is also the surface area of the solute cavity,  $A$ .<sup>8,9,21,24,26–30</sup>

$$G_{h,cav} = \gamma_h A + b \quad (8)$$

$\gamma_h$  being a microscopic surface tension for the water–hydrocarbon interface, whose value has been found to be 5 cal/Å<sup>2</sup> in one study<sup>8</sup> but 20 cal/Å<sup>2</sup> in another.<sup>9</sup> The transfer process can be thought of as being performed in two steps, creating the empty cavity and establishing the van der Waals contacts between solute and solvent, which gives a decomposition of the free energy into corresponding terms:

$$G_{h,cav} = G_{cav} + G_{s,vdw} \approx G_{cav} + U_{s,vdw}. \quad (9)$$

This method is more reliable, since the first term on the right side of the equation refers to creation of an empty cavity, and the second term, the solute–solvent van der Waals free energy,  $G_{s,vdw}$ , has been replaced with the average solute–solvent van der Waals energy,  $U_{s,vdw}$ . This is a good approximation, because the free energy of solvent reorganization due to the weak attractive van der Waals interaction with the solute is small.<sup>13</sup> (A perturbation theory of liquid structure provides a theoretical justification of the small dependency of liquid structure during the process of “switching on” the van der Waals forces.<sup>31</sup>) The average solute–solvent van der Waals energy can be calculated easily during a molecular dynamics simulation.

The cavity free energy is then directly proportional to the molecular surface area with constant microscopic surface tension,  $\gamma_v$ , for the water–vacuum interface

$$G_{cav} = \gamma_v A \quad (10)$$

According to Jackson and Sternberg, the best value of the proportionality constant  $\gamma_v$  appears to be a microscopic surface tension of 69 cal/(mol Å<sup>2</sup>).<sup>14</sup> This value can be understood as the macroscopic surface tension of water [ $= 102$  cal/(mol Å<sup>2</sup>)] scaled by a factor equal to the ratio between a planar macroscopic surface area and the larger irregular (molecular) surface area of the molecules exposed on this surface.<sup>14</sup>

Scaled particle theory<sup>14,32–34</sup> provides a theoretical justification for the idea that the free energy,  $G_{cav}$ , is proportional to the surface of the molecular cavity. In simulations with accurate calculation of the long-range electrostatic forces (for the solvent), the free energy of cavity formation in aqueous solution<sup>35</sup> has been found to be proportional to the surface of the cavity, with interfacial surface tension similar to the macroscopic gas–solvent surface tension. [Note, however, that in simulations with spherical cutoff, growth of a large sphere or of an ellipsoid in water produced cavities with different surface-tension coefficients, respectively, 159 and 64 cal/(mol Å<sup>2</sup>).<sup>36,37</sup>]

A question remains about the correct description of the solute molecule’s surface and the definition of the surface area  $A$  to be used in Eq. 10. Obvious choices are the solvent-accessible (SA) surface area (i.e., a surface traced by the center of the solvent-probe sphere when it rolls over the van der Waals spheres of the solute molecule atoms) or the molecular surface of Richards<sup>38</sup> (i.e., the surface that envelops the solvent-excluded volume of the solute molecule). Recent work<sup>14,39,40</sup> has shown that the molecular surface is a physically more appropriate choice and that differences in molecular surface curvature play no role. Calculations of the potential of mean force (PMF) of dimerization of two methane molecules and of the PMF of *n*-butane in water as a function of the C2-C3 dihedral angle reproduce the profiles of the potential of mean force calculated in simulations with explicit solvent, when based on molecular surface area.<sup>14,40,41</sup> On the other hand, similar calculations based on solvent-accessible surface area do not reproduce these profiles.

### Solvent Polarization Free Energy

The calculation of solvent polarization free energy used here is based on a model of two dielectric media separated by a boundary coinciding with the surface of the protein molecule. On the inside, the charge distribution is explicit, and the dielectric constant is low; on the outside, a dielectric continuum with a high dielectric constant represents the aqueous solvent, and the electric polarization of this external continuum in response to the internal charges provides a free energy that can be evaluated numerically, using standard equations of electrostatics. This is now examined in greater detail.

The protein’s charges (charge  $q_i$  is at position  $\mathbf{r}_i$ ) induce in the solvent a charge distribution,  $\rho_{solv}$ . This charge distribution produces an electrostatic potential,  $V_{solv}$

$$\langle V_{solv}(\mathbf{r}_i) \rangle = \int \frac{(\rho_{solv}(\mathbf{r}))}{|\mathbf{r} - \mathbf{r}_i|} d\mathbf{r}. \quad (11)$$

The polarization free energy is found as the work done in a charging process in which the charges of the protein are gradually “turned on” (by changing  $\lambda$  from 0 to 1), as a result of which the solvent electrostatic potential also grows,

$$G_{pol} = \int_0^1 d\lambda \sum_i [q_i \langle V_{solv}(\mathbf{r}_i) \rangle \lambda]. \quad (12)$$

The question is then how to evaluate  $\rho_{solv}$ . In a dynamics simulation with explicit solvent,  $\rho_{solv}$  is identical to the distribution of the (partial) charges of the solvent atoms, and a common approach is to use Eqs. 11 and 12 to compute  $G_{pol}$  with thermodynamic integration or perturbation. However, this approach has not yet been used to calculate the



polarization free energy for a solute molecule the size of a protein.

With a linear response approximation for solvent polarization,<sup>31</sup>  $V_{\text{solv}}$  and  $\rho_{\text{solv}}$  both depend on the extent to which the charges are turned on, that is, both are proportional to  $\lambda$ , which gives

$$G_{\text{pol}} = \frac{1}{2} \sum_i q_i \int \frac{\langle \rho_{\text{solv}}(\mathbf{r}) \rangle}{|\mathbf{r} - \mathbf{r}_i|} d\mathbf{r}. \quad (13)$$

When a linear response is assumed in simulations with explicit solvent, one need do only a simulation for a single protein charge distribution and obtain the corresponding solvent polarization charge density  $\langle \rho_{\text{solv}}(\mathbf{r}) \rangle$ ; there is no need to explicitly perform the charging process of Eq. 12.<sup>42,43</sup> However, calculation of the average solvent potential,  $V_{\text{solv}}$ , in a simulation of lysozyme in explicit solvent showed very slow convergence, so that a simulation time of several nanoseconds would have been required to attain good accuracy.<sup>44</sup>

The validity of the linear response approximation for the solvent reaction potential of an aqueous solvent has been examined by direct simulations of its dependence on  $\lambda$  in molecular dynamics free energy simulations.<sup>45–51</sup> In a majority of simulations of polar and charged molecules,<sup>45,47,48,51</sup> a nearly linear response has been observed for a moderately charged solute, that is, one whose partial atomic charges do not exceed  $\sim 1$  e and whose electrostatic field near the solute surface does not exceed 50 kT/(e Å).<sup>48</sup>

The calculation of the average polarization charge density,  $\langle \rho_{\text{solv}}(\mathbf{r}) \rangle$  can also be done in the framework of macroscopic electrostatics, that is, with an implicit, continuum solvent description. The average electrostatic potential,  $\Phi(\mathbf{r})$  contains contributions from the fixed charges on the protein and the induced charges in the solvent, according to the Poisson equation

$$\nabla^2 \Phi(\mathbf{r}) = -4\pi \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i) - 4\pi \langle \rho_{\text{solv}}(\mathbf{r}) \rangle. \quad (14)$$

With use of standard relations connecting the average polarization charge density  $\langle \rho_{\text{solv}}(\mathbf{r}) \rangle$  with the average polarization, and the polarization with the electric field, one obtains<sup>52</sup>

$$\nabla D(\mathbf{r}) \nabla \Phi(\mathbf{r}) = -4\pi \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i) \quad (15)$$

This last equation contains a position-dependent dielectric constant,  $D(\mathbf{r})$ . If this is known, Eqs. 14 and 15 define the distribution of  $\langle \rho_{\text{solv}}(\mathbf{r}) \rangle$  for a given conformation of the protein, so that  $G_{\text{pol}}$  can be calculated with Eq. 13.

A fundamental question is how to model the distribution of the dielectric constant,  $D(\mathbf{r})$ . Inside the protein molecule's solvent-excluded volume, the dielectric constant  $D_i$  equals 1, because the solvation

free energy  $W(\mathbf{x})$  has to be calculated for a model with fixed and nonpolarizable charge distribution, in a single conformation, according to the current protein force field.<sup>53</sup> In the solvent space it is common practice to use the bulk solvent dielectric constant. Near the water–solute interface, the density of water drops quite abruptly, over a distance of about 0.5 Å, from the bulk density<sup>54</sup> to zero, as has been shown by extensive molecular dynamics simulation of solvent density around proteins.<sup>54</sup> The fluctuation of the microscopic density of water molecules is found not to be large, the maximum peak value near the surface is about 10% higher than bulk water density, and, overall, deviations from bulk water density are small, about 5%. Therefore, a model with a sharp stepwise approximation to the solvent density is reasonable. In any case, the exact choice of where to locate the boundary separating solute and solvent is empirical (see later discussion and Nina and colleagues<sup>45</sup>) and can compensate for deviations of the actual dependence of the dielectric constant from the assumed step function near the protein surface.

The evaluation with Eqs. 14 and 15 of the solvent polarization charge density for a complicated solute charge distribution is done numerically, using finite-element methods. The well-known finite difference method uses elements in three-dimensional space,<sup>8–10,55,56</sup> usually a rectangular three-dimensional box, which includes the solute and a volume of solvent around it; the alternative boundary element (BE) method used here presents a numerical solution of an integral equation over the dielectric boundary, to which the original Poisson equation has been analytically converted,<sup>57</sup> and finds a solution in terms of elements obtained by tessellation of the protein surface;<sup>57–63</sup> however, use of the BE method is restricted to the case (assumed here and in most applications of continuum dielectric models) in which the dielectric constant has one value within the surface and another value outside it and otherwise does not depend on position.

The border between the two dielectric regions remains to be defined. In recent applications of the dielectric continuum model<sup>8–10,45,55–66</sup> both the solvent accessible surface, SA, and the molecular surface, MS, have been used as the dielectric border surface, while the choice of dielectric constant was, at least in some of the studies, not fixed, but instead treated empirically. Obviously, a consistent definition of the dielectric behavior of the continuum and accurate numerical calculation of the dielectric border are important for an accurate and self-consistent application of the dielectric continuum model to problems of macromolecular modeling.

### Calculation of a Smooth Invariant Molecular Surface (SIMS)

Calculation of molecular properties on the MS and integration of a function over the MS requires a numerical representation of the MS as a manifold

$S(s_i, \mathbf{n}_i, \Delta s_i)$ , where  $s_i$ ,  $\mathbf{n}_i$ , and  $\Delta s_i$  are coordinates, normal vector, and area of a small surface element, respectively. The molecular surface is defined as the surface around the protein molecule that excludes any part of a solvent molecule, calculated for a model in which the protein atoms are represented as hard spheres, each with its own radius, and the solvent molecule as a spherical probe having its own radius. Thus defined, the MS contains three types of components, or faces, which are termed "contact," "saddle," and "concave reentrant," according to whether the solvent probe sphere simultaneously touches one, two, or three atoms, respectively.<sup>67,68</sup> The molecular surface of a protein may contain singularities in the direction of the normal vector (i.e., the direction of the normal vector is not continuous in the vicinity of a singular point of the MS). Singularities called cusps and holes appear when the probe can almost, but not quite, pass through a group of two or three atoms of the protein.<sup>69,70</sup> It can be assumed that an MS with smoothed singularities is a better approximation of the dielectric border between high-dielectric solvent and the low-dielectric molecular interior.<sup>70</sup>

An improved numerical representation of surface functions and surface functionals can be achieved if the surface satisfies several conditions: minimum variation of surface element size, adequate smoothing in the vicinity of singularities, stability to variation in the number and size of surface elements, independence to a rotation of the molecule, and stability to change of the molecular conformation.

Several available programs—Connolly's MS program,<sup>71</sup> MSEED,<sup>72</sup> Varshney's program,<sup>73</sup> and Sanner's MSMS<sup>74</sup> program—were not specifically designed for the boundary element application and provide a dot MS of poor quality. The effective MSEED method<sup>72</sup> is not completely general, because it is blind to internal cavities, cannot handle a free saddle surface, and frequently produces results that are unstable to small perturbations of atomic coordinates. Varshney's method<sup>73</sup> lacks a general procedure for handling self-intersecting surfaces. Sanner's MSMS method<sup>74</sup> is primarily designed for surface visualization. It uses the MSEED methodology to define a self-connected network of arcs on the molecular surface and cannot find internal molecular cavities. Connolly's MS program<sup>71</sup> is the most general, but it has an ineffective procedure for removing self-intersecting surface segments, does not provide a sufficiently narrow distribution of surface element sizes, and produces surface area elements that are not invariant to molecular rotation.<sup>75–77</sup>

The GEPOL93 method of Pascual-Ahuir and associates<sup>78</sup> approximates a smooth reentrant face of the MS of the solute molecule by the irregular van der Waals surface of added dummy spheres, which fill the space not accessible to the solvent. The number

and positions of the added dummy spheres are controlled by empirical parameters. For large molecules, like proteins, the GEPOL93 method needs a large number of dummy spheres and requires much time to achieve good accuracy. None of these methods produces a smooth molecular surface. Zauhar suggested a method<sup>70</sup> of solvent probe deformation that completely avoids the occurrence of the self-intersecting singularities in the MS and constructs a smooth molecular surface. A major deficiency of Zauhar's method<sup>70</sup> is that it produces a smooth molecular surface that in terms of topology and atomic areas differs unpredictably from the strict, singular MS.

We have revised Connolly's method of generation of an MS<sup>71</sup> and have developed a new method, SIMS, that (a) produces a near-homogeneous dot distribution, (b) is invariant to molecular rotation and translation, and (c) describes all types of self-intersecting singularities of the MS and contains an exact method to remove the singular regions of the MS to generate a smoothed MS with specified minimal radius of curvature. The SIMS method calculates a dot MS of good numerical quality, which can be used in a variety of implicit continuum models of calculation of solvation effects<sup>14,61,79,80</sup> and for molecular electrostatic calculations with the boundary element method in dielectric continuum models.<sup>12</sup>

A complete description of the SIMS method can be found elsewhere.<sup>11</sup> SIMS revises Connolly's method of calculation of buried saddle faces: the surface area of saddle and reentrant concave faces is calculated analytically, rather than numerically. The method uses rectangle-shaped surface elements that exactly tessellate atomic sphere templates and saddle faces, and the area of each surface element is calculated analytically. The coordinates of the centers of the surface elements (dots) are calculated in a local coordinate system defined by atomic triplets and then are transformed to the global coordinate system. To achieve a high degree of homogeneity of the dot distribution, SIMS uses the method of equal arc length in two orthogonal directions along two lines of main curvature of the surface for all three types of surface faces; this method produces better homogeneity than is obtained by uniform rectangular mapping of saddle faces.<sup>67,70,74</sup> SIMS uses exact conditions to compute coordinates of the surface pieces belonging to self-intersecting parts of the MS and therefore removes these parts with a high degree of accuracy. The SIMS method replaces singular regions by the outward surface of a smoothing sphere that is rolled over the inside of the MS. (A practical choice for the radius of the smoothing sphere is  $\sim 0.4$  Å.)

The area of the molecular surface computed by the SIMS method has been demonstrated to be stable to within 0.1% to a change of density of surface elements in a range from 2 to 25 per Å<sup>2</sup>, whereas Connolly's MS program produces a molecular sur-

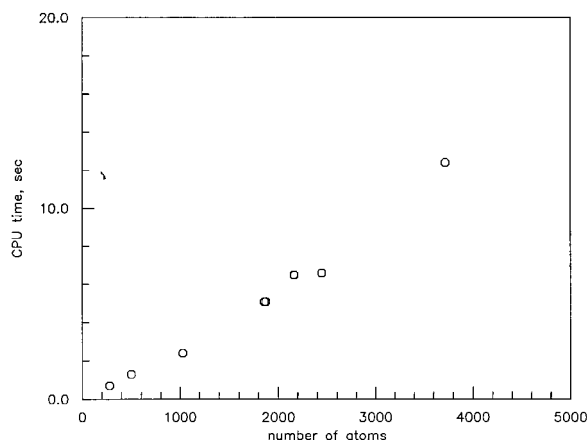


Fig. 1. CPU time of SIMS program as a function of protein size, on SGI Power Onyx (one R10000 processor), for all-atom representation of eight proteins: in order of increasing number of atoms, a 17-residue peptide, 1cbh, eglin, 1b2p, 1p2p, immunoglobulin G1, subtilisin, and bacteriorhodopsin.

face that oscillates erratically by up to 3% as the density is changed. The timing of the SIMS method is somewhat better than the timing of Connolly's method; the CPU time scales as the number of atoms in the molecule (Fig. 1). The SIMS program is available from the authors on request (vorobjev@femto.med.unc.edu) or via the Internet (<http://femto.med.unc.edu/SIMS>).

### Fast Adaptive Multigrid Boundary Element Method (FAMBE)

As discussed, in this article we use the boundary element to find the solution of the Poisson equation. An advantage of the BE method is its invariance to rotation and translation of the solute molecule, and a comparison of multigrid BE and finite difference (FD) methods<sup>57</sup> shows that the BE method exhibits a higher degree of consistency. The main disadvantage of the BE method is its greater complexity.

Recent improved methods of solving the Poisson equation for inhomogeneous dielectric media in problems of physics and engineering have used multigrid and multilevel techniques developed to solve a variety of complex computational problems,<sup>81</sup> and more robust and efficient multigrid FD methods have been developed in consequence.<sup>82–86</sup> Multilevel techniques have been applied also to the iterative BE method,<sup>57,58,62,87</sup> but this technique suffers from slow convergence and is more time consuming than multigrid FD methods.

A new efficient approach to the BE method has been developed by one of us in the fast adaptive multigrid boundary element method.<sup>12</sup> The problem is to solve the following integral equation for the surface polarization charge density  $\sigma(\mathbf{t})$  in terms of the electrostatic fields  $\mathbf{E}_{z_k}$  generated by a set of

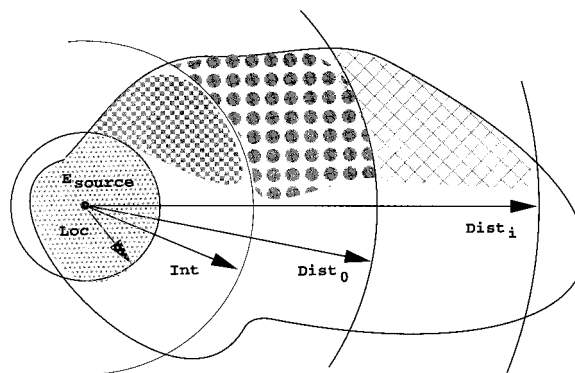


Fig. 2. Tessellation of the molecular surface with multigrid boundary elements. The regions and approximate size of surface elements are as follows: Loc(al) region, within 3 Å of the source, 0.2 Å<sup>2</sup>; Int(ermediate) region, between 3 and 6 Å, 1 Å<sup>2</sup>; Dist(ant)<sub>0</sub> region, between 6 and 12 Å, one surface element per surface atom; Dist(ant)<sub>i</sub> regions, each successive shell twice the thickness of the previous shell, (linear) size of surface elements twice that in the previous shell.

charges,  $z_k$  of the solute:

$$\sigma(\mathbf{t}) = f \int_S \frac{\sigma(\mathbf{s})(\mathbf{t} - \mathbf{s})\mathbf{n}(\mathbf{t}) d\mathbf{s}}{|\mathbf{t} - \mathbf{s}|^3} + \frac{f}{D_i} \sum_k \mathbf{n}_t \mathbf{E}_{z_k}(\mathbf{t}) \quad (16)$$

where  $f = (1/2\pi)(D_i - D_o)/(D_i + D_o)$ ,  $\mathbf{n}(\mathbf{t})$  is the normal vector to the molecular surface at point  $\mathbf{t}$ , and  $D_i$  and  $D_o$  are the dielectric constants inside and outside the surface, respectively.

The FAMBE method splits this into independent minor BE equations, one each for the polarization charge density generated by a single charge (or small group of charges):

$$\sigma_k(\mathbf{t}) = f \int_S \frac{\sigma(\mathbf{s})(\mathbf{t} - \mathbf{s})\mathbf{n}(\mathbf{t}) d\mathbf{s}}{|\mathbf{t} - \mathbf{s}|^3} + \frac{f}{D_i} \mathbf{n}_t \mathbf{E}_{z_k}(\mathbf{t}), \quad k = 1, 2, \dots \quad (17)$$

the surface charge,  $\sigma(\mathbf{t})$ , being the sum of the components  $\sigma_k(\mathbf{t})$ . Each minor BE equation (Eq. 17) is replaced by an approximate, linear equation and is discretized over its own set of *multilevel* BEs:

$$\sigma_k = \mathbf{M}_k \sigma_k + \mathbf{E}_{z_k}. \quad (18)$$

For each charge  $z_k$ , the size of the BEs steadily increases with distance from the source of the molecular electrostatic field (Fig. 2). Hence, for any given charge, the dimensions of the vector of surface charge densities,  $\sigma_k$ , and of the matrix,  $\mathbf{M}_k$ , are much less than the total number of surface elements in the original tessellation. The number of multilevel boundary elements,  $N_{\text{MBE}}$ , that tessellate an MS with area

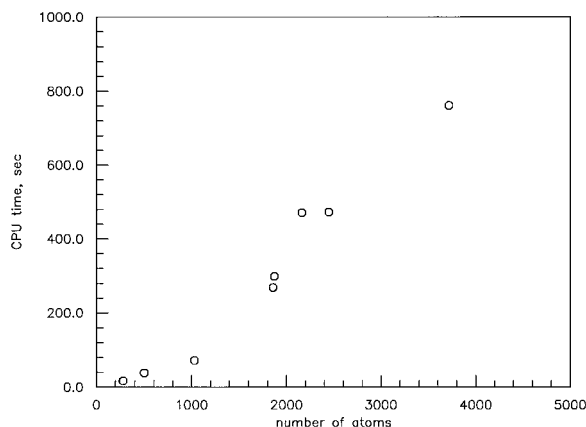


Fig. 3. CPU time of the FAMBE program, as a function of protein size, on SGI Power Onyx (one R10000 processor) for same eight proteins as used in Fig. 1.

$A_S$  scales as

$$N_{\text{MBE}} \approx n_{\text{loc}} \ln (A_S/S_{\text{loc}}) \quad (19)$$

where,  $n_{\text{loc}}$  and  $A_{\text{loc}}$  are an average number of boundary elements and size for the local area with finest tessellation. Each minor matrix equation (Eq. 18) is solved iteratively by the preconditioned bi-conjugate gradient method.<sup>88</sup> Only a few iterations (five or six) are needed to find a solution with a relative accuracy of  $10^{-4}$ . The estimated computational complexity of the FAMBE method scales as

$$\text{complexity} \approx N_z [n_{\text{loc}} \ln (A_S/A_{\text{loc}})]^2 \quad (20)$$

where,  $N_z$  is the number of charges (or charged groups) in the solute molecule. The solvent polarization free energy,  $G_{\text{pol}}$ , is found in analogy with Eq. 13, replacing volume and volume charge density with surface and surface charge density.

Test calculations for several peptides and proteins show that the CPU time of the FAMBE method, with an optimal number of levels of tessellation, scales approximately linearly with the number of atoms of the molecule (Fig. 3; in the limit of very small number of atoms the dependence is quadratic). The free energy calculated with the FAMBE and the SIMS programs varies smoothly with the size of the smallest boundary element; for an element size of 0.5 Å, the free energy is about 3% higher than the size for the extrapolation to zero.<sup>11</sup>

Because of its good numerical qualities, the FAMBE method can be incorporated into molecular dynamics and Monte Carlo simulations, in order to provide energy and forces for an accurate implicit solvent model, with important potential applications, such as simulation of a system undergoing large conformational changes (when relaxation of explicit solvent may become a bottleneck) and simulation of pH-

dependent conformational dynamics (i.e., coupling between ionization equilibria and the protein or protein-drug complex conformation). The FAMBE program is available from the authors on request (vorobjev@femto.med.unc.edu) or via Internet (<http://femto.med.unc.edu/FAMBE>).

### Entropy of a (Stable) Macroscopic Conformation

The conformational entropy of a biomolecule in a stable macroscopic conformation A can be estimated in the quasi-harmonic approximation by analysis of the covariance fluctuation matrix  $\mathbf{C}$ .<sup>89-93</sup>

$$\mathbf{C}_{ij} = \langle (\mathbf{x}_i - \langle \mathbf{x}_i \rangle) (\mathbf{x}_j - \langle \mathbf{x}_j \rangle) \rangle_A \quad (21)$$

where  $\mathbf{x}(t)$  are the atomic coordinates along the MD trajectory. If the conformation is stable, then the molecular motion along the MD trajectory can be approximated as quasi-harmonic to give a statistical description of the real anharmonic intramolecular motions over a large number of microstates of the molecule.<sup>91-93</sup> The eigenvalues  $\lambda_i$  of the (mass-weighted) covariance matrix  $\mathbf{C}$  define frequencies

$$(2\pi\nu_i)^2 = k_B T \lambda_i \quad (22)$$

and positional fluctuations of the normal coordinates. The entropy  $S_A$  of the macrostate A is given by<sup>19</sup>

$$TS_A = \sum_i -k_B T \ln [1 - \exp(-h\nu_i/k_B T)] + h\nu_i/[1 - \exp(-h\nu_i/k_B T)]. \quad (23)$$

Oscillators with low frequency contribute the most to the entropy, while oscillators with large frequency have small positional fluctuations and make insignificant contributions to  $S_A$ . This expression for estimating entropy, which has also been used by others in a similar context,<sup>94,95</sup> is expected to be more robust in practical situations than an expression based on the determinant of the covariance fluctuation matrix.<sup>96,97</sup> (Because the fluctuation matrix has a large number of small eigenvalues, instabilities, due to round-off error, can easily emerge in a calculation of the determinant.)

### APPLICATION TO NATIVE AND MISFOLDED PROTEINS

#### Simulation Protocol

To calculate the free energy, we used the following protocol with the program SigmaX for each native and misfolded protein structure: (a) optimize the structure by one hundred cycles of conjugate gradient minimization of the energy in vacuum; (b) place



**TABLE I. Averages Along Short MD Trajectories for Each of the Native and Misfolded Structures (energies in kcal/mol)\***

Native misfolded	$U_{\text{geom}}$	$U_{\text{m,vdw}}$	$U_{\text{m,coul}}$	$U_{\text{s,vdw}}$	$G_{\text{cav}}$	$G_{\text{pol}}$	$-TS$	$G_{\text{ES/IS}}$	MS (Å <sup>2</sup> )
1bp2	1,173	-665	-2,542	-170	447	-1,594	-19	-3,370	6,481
1bp2-on-2paz	1,195	-669	-2,593	-170	412	-1,409	-18	-3,255	5,974
2paz	1,357	-677	-2,218	-80	395	-1,755	-18	-2,998	5,730
2paz-on-1bp2	1,247	-559	-2,485	-168	464	-1,455	-20	-2,976	6,729
1cbh	307	-183	-375	-85	132	-282	-8	-495	1,913
1cbh-on-1ppt	308	-161	-341	-98	151	-335	-8	-484	2,185
1ppt	383	-176	-425	-82	175	-854	-9	-988	2,535
1ppt-on-1cbh	393	-158	-611	-91	181	-621	-10	-917	2,617
1p2p	1,208	-752	-2,634	-167	398	-1,515	-16	-3,479	5,763
1p2p-on-1rn3	1,207	-671	-2,281	-176	448	-1,798	-18	-3,289	6,487
1sn3	603	-343	-1,259	-71	242	-1,148	-13	-1,990	3,502
1sn3-on-2ci2	651	-304	-1,623	-112	255	-803	-13	-1,949	3,701
1sn3-on-2cro	623	-308	-1,571	-122	251	-812	-14	-1,953	3,638
1rn3	1,155	-727	-1,914	-122	409	-1,952	-19	-3,171	5,930
1rn3-on-1p2p	1,191	-631	-2,191	-185	441	-1,666	-20	-3,061	6,387
2ci2	677	-315	-1,564	-82	252	-1,087	-12	-2,131	3,649
2ci2-on-2cro	721	-274	-1,393	-103	287	-1,212	-15	-1,990	4,156
2ci2-on-1sn3	722	-277	-1,528	-82	281	-1,094	-15	-1,993	4,073
2cro	668	-350	-926	-84	231	-1,342	-13	-1,816	3,349
2cro-on-1sn3	692	-317	-682	-108	260	-1,493	-15	-1,663	3,774
2cro-on-2ci2	672	-286	-937	-146	284	-1,298	-15	-1,727	4,111

\*  $U_{\text{geom}}$ , energy of deformation of bond angles and dihedral angles;  $U_{\text{m,vdw}}$ , intramolecular van der Waals energy (Lennard-Jones potential);  $U_{\text{m,coul}}$ , intramolecular coulomb energy;  $U_{\text{s,vdw}}$ , protein-solvent van der Waals energy;  $G_{\text{cav}}$ , free energy of creating the hydrophobic cavity;  $G_{\text{pol}}$ , free energy of solvent polarization;  $S$ , conformational entropy estimated from the fluctuation matrix for the C $^{\alpha}$  atoms, Eq. 23.  $G_{\text{ES/IS}}$ , total free energy computed via the ES/IS method, Eq. 2; MS, molecular surface in Å<sup>2</sup>.

the protein in the center of a water box with the size extending a minimum of 6 Å from the protein surface; (c) optimize the water structure by one hundred cycles of conjugate gradient minimization of the energy; (d) equilibrate the system by 3- to 5-ps dynamics simulations successively at each of the temperatures 50, 100, 150, 200, 250, and 300 K; (e) perform a final 5- to 10-ps equilibration in the N,P,T ensemble; and (f) perform a 50-ps equilibrium MD simulation in the N,P,T ensemble (P = 1 bar, T = 300 K). The calculations use the Shake method to constrain bond lengths, an 8-Å cutoff distance on non-bonded interactions, and a “Berendsen” thermostat.<sup>98</sup>

Snapshots at 1-ps intervals were collected and served as a representative set of microstates of the protein. The solvation free energy of each was calculated with the continuum solvent model, with use of the PARSE set of atomic radii.<sup>8</sup> The covariance matrix was calculated for all C $^{\alpha}$  atoms from this same set of protein microstates, after each had been superimposed on the first microstate to eliminate rigid-body translation and rotation of the protein.<sup>91–93</sup> The use of only C $^{\alpha}$  atoms for the calculation of the covariance matrix is a further approximation made to limit the size of the matrix and make the calculation of eigenvalues more manageable; we plan in future work to investigate in particular the consequences of neglecting side chain motions.

### Free Energies of the EMBL Set of Misfolded Protein Structures

We have applied this new method of free energy calculation to the so-called EMBL set of deliberately misfolded protein structures.<sup>4</sup> These misfolded protein models were generated by selecting pairs of proteins with known structures with an equal number of residues. Using the backbone conformation of the X-ray crystal structures, the sequences were swapped and the side chain structure optimized using a fast Monte Carlo algorithm with a simple energy function. The misfolded models were then subjected to five hundred steps of steepest descent energy minimization. The models were named AAA-on-BBB, where AAA is the protein data bank (PDB) identifier for the sequence and BBB is the PDB identifier of the backbone coordinates.<sup>4</sup>

The terms of the total free energy (*cf.* Eq. 2) have been calculated for families of the native and misfolded structures and are shown in Table I. In all cases the calculated conformational free energy,  $G_{\text{ES/IS}}$ , is lowest for the native fold, that is, with this formulation of the free energy, each native structure is correctly found to be more stable than misfolded structures of the same sequence. Figure 4 shows how in one case the total conformational free energy (all terms in Eq. (2), except the entropic term) varies along the equilibrium trajectories for native and two

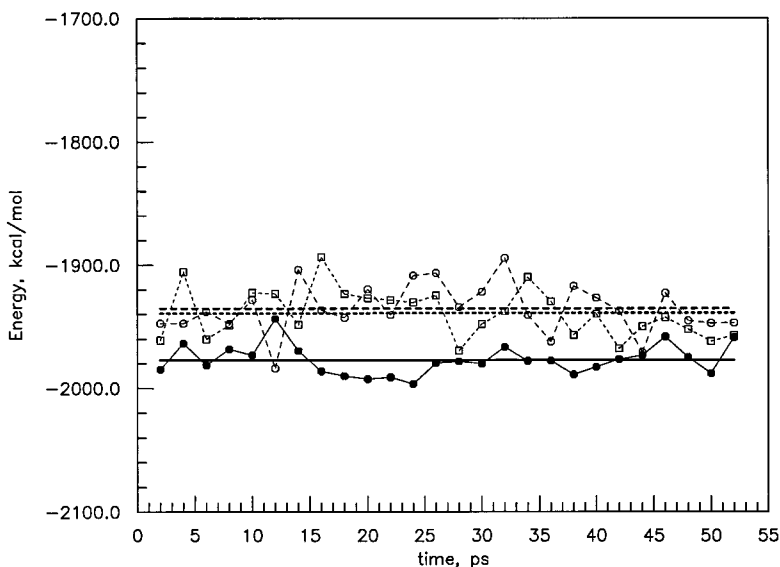


Fig. 4. Total free energy along molecular dynamics trajectories for the protein 1sn3 (= native conformation, filled circles and solid line) and for misfolded conformations 1sn3-on-2cro (open circles and dashed line) and 1sn3-on-2ci2 (open squares and dotted line).

misfolded structures; it is obvious that the differences in the mean values of the free energy, which for this case are 42 and 37 kcal/mol, are highly significant.

### Component Terms

Table II shows for each native–misfolded pair the difference relative to the native fold of several energy terms: the free energy,  $G_{\text{ES/IS}}$ ; the average total energy of protein plus solvent in molecular dynamics simulations,  $U_{\text{MD}}$ ; the intramolecular energy of the protein,  $U_{\text{m}}$ ; and the intramolecular energy less the intramolecular coulomb energy (i.e., the sum of the energy for geometry deformation and the van der Waals energy),  $U_{\text{m,sh}}$ . An interesting result is that the sum of the terms for deformation of the local geometry (bond angles and dihedral angles) and the internal van der Waals energy of the misfolded proteins, by being consistently higher, also discriminates in favor of the native fold.

Figure 5 shows the total MD potential energy of the systems, that is, the sum of intra- and intermolecular energies for protein and explicit solvent in the MD simulations.

The mean van der Waals solute–solvent energy,  $U_{\text{s,vdw}}$ , calculated over the MD trajectory of each of the native and misfolded structures is shown in Figure 6 as a function of the mean molecular surface area. An approximately linear correlation is found, that is,

$$U_{\text{s,vdw}} = \gamma_{\text{vdw}} A \quad (24)$$

where  $A$  is the molecular surface area. The value of the coefficient  $\gamma_{\text{vdw}}$ , 0.0276 kcal/(mol Å<sup>2</sup>). However, the root mean square deviation is quite large, ~17 kcal/mol, and this supports our use of Eqs. 9 and 10,

**TABLE II. Average (Free) Energy Differences Along Short MD Trajectories for Each of the Native and Misfolded Structures (kcal/mol)<sup>†</sup>**

Native misfolded	Number of residues	$\Delta G_{\text{ES/IS}}$	$\Delta U_{\text{MD}}$	$\Delta U_{\text{m}}$	$\Delta U_{\text{m,sh}}$
1bp2	123				
1bp2-on-2paz		115	-123	-34	17
2paz	123				
2paz-on-1bp2		21	-25	-259	8
1cbh	36				
1cbh-on-1ppt		10	-3	58	23
1ppt	36				
1ppt-on-1cbh		71	-83	-157	28
1p2p	124				
1p2p-on-1rn3		190	172	434	80
1sn3	65				
1sn3-on-2ci2		42	18	-257	87
1sn3-on-2cro		37	103	-276	55
1rn3	124				
1rn3-on-1p2p		110	127	-144	132
2ci2	65				
2ci2-on-2cro		141	48	256	85
2ci2-on-1sn3		138	153	120	84
2cro	64				
2cro-on-1sn3		152	146	301	57
2cro-on-2ci2		88	187	57	68
discrimination (%)		100	66	50	100

<sup>†</sup> $G_{\text{ES/IS}}$ , total conformational free energy;  $U_{\text{MD}}$ , total energy of protein and solvent in MD simulation;  $U_{\text{m}}$ , intramolecular energy of the protein in the same MD simulation;  $U_{\text{m,sh}}$ , ditto, less the coulomb energy.

which explicitly contain  $U_{\text{s,vdw}}$ , for the cavity free energy, rather than Eq. 8, which does not.

The free energy term related to the protein conformational mobility,  $-TS_{\text{conf}}$ , is seen to make a very small contribution to the free energy difference between native and misfolded structures. Figure 7

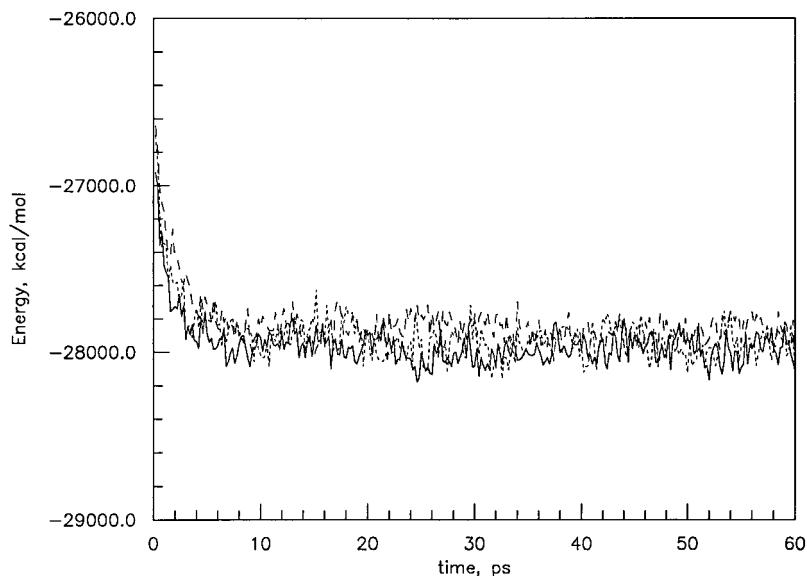


Fig. 5. Total potential energy along molecular dynamics trajectories for the protein 1sn3 (= native conformation, solid line) and for misfolded conformations 1sn3-on-2cro (dashed line) and 1sn3-on-2ci2 (dotted line).

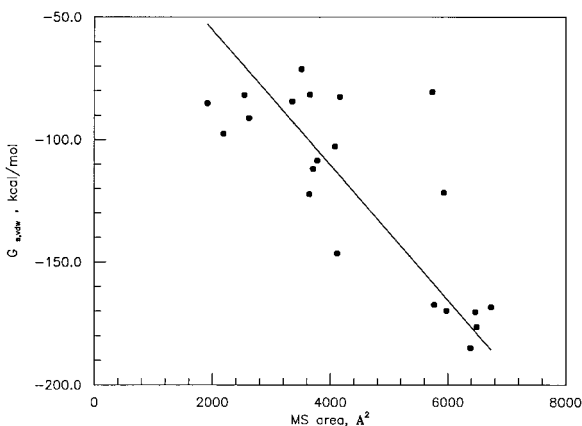


Fig. 6. Average protein-solvent van der Waals energy,  $U_{s,vdw}$  of the 21 structures versus average molecular surface.

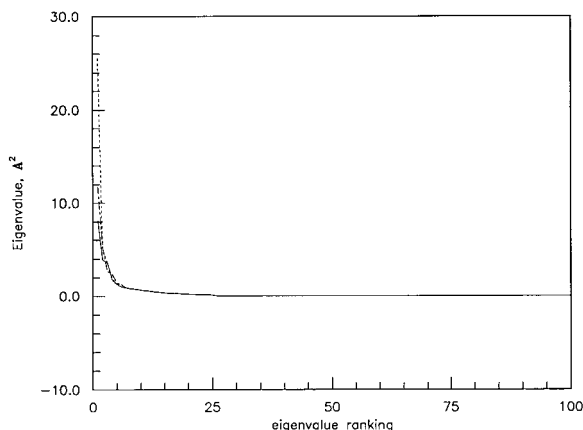


Fig. 7. Eigenvalues of the covariance fluctuation matrix for 1sn3 (= native structure, solid line) and for the misfolded conformations 1sn3-on-2cro (dashed line) and 1sn3-on-2ci2 (dotted line).

shows a plot of the one hundred largest eigenvalues for the covariance matrices of 1sn3 in its native and in two misfolded structures in 50-ps simulations. (These are results for the  $C^\alpha$  atoms only.) It can be seen that only a small fraction of the modes have appreciable amplitude, as has been noted by others.<sup>89,99,100</sup> Significant conformation dependence shows up only in the modes with high amplitudes, which typically correspond to global motions.

We have found a large anticorrelation and compensatory effect for the polarization free energy,  $G_{pol}$ , and the intraprotein electrostatic energy,  $U_{m,coul}$ , for all studied proteins and conformations along their MD trajectories. A typical example of a strong anticorrelation is shown in Figure 8 for a 200-ps simulation of eglin and 2,076 water molecules with periodic boundary conditions. One sees that the sum  $U_{m,coul} + G_{pol}$  fluctuates much less than the individual components. The correlation coefficient between  $U_{m,coul}$  and

$G_{pol}$  is  $-0.86$  for eglin, and for the EMBL set of proteins it is  $-0.8 \pm 0.1$ .

## DISCUSSION

The new ES/IS method has a 100% success rate in predicting the stability of a native protein structure over a misfolded one. Predictions based on the total energy in MD simulation with explicit water,  $U_{MD}$ , as shown in Table II, succeed only part of the time. The point is that the free energy, but not the energy, should be able to discriminate between different conformations of the same protein. Also, the intramolecular energy,  $U_m$  in Table II, is not a reliable predictor of stability, as was noted in the original studies of intentionally misfolded proteins.<sup>1,2,4</sup>

As can be seen from Table I, the free energy of solvent polarization,  $G_{pol}$ , contributes hundreds of kilocalories per mole to the difference in free energy

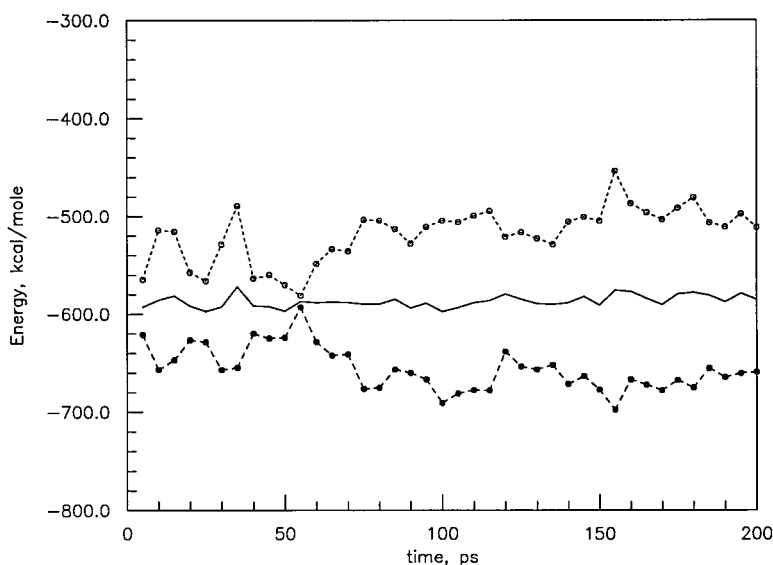


Fig. 8. Internal electrostatic energy  $U_{m,coul}$  (open circles) and free energy of polarization,  $G_{pol}$  (filled circles) and their sum (solid line) along a molecular dynamics trajectory of eglin in water solvent calculated in a box of 2,076 water molecules with periodic boundary conditions.

between native and misfolded conformations, most often, but not always, in favor of the native fold. As mentioned, changes of the intraprotein electrostatic energy,  $U_{m,coul}$ , and the solvent polarization free energy,  $G_{pol}$ , tend to compensate each other. A strong compensation between intramolecular gas-face potential energy and solvation energy has also been observed in a 10-ns MD simulation of the peptide AYPYD in explicit water solvent.<sup>101</sup> A proper balance between these energies is needed to reliably predict the greater stability of the native structures.

The free energy of creation of the hydrophobic cavity,  $G_{cav}$ , and van der Waals solute-solvent interaction,  $U_{s,vdw}$ , contribute differences of only tens of kilocalories per mole, that is, one order magnitude less than the electrostatic interactions. As noted, we observed that the sum of geometric and van der Waals energies, that is, what one might call the internal "packing" energy of the protein,  $U_{m,sh}$  in Table II, correlates correctly with whether the conformation is native or misfolded. A reasonable hypothesis is that this is a natural consequence of the artificial method chosen to construct the misfolded conformations, which results in a mismatched set of amino acid side chains in the interior of the molecule having to be reconfigured into an optimal packing arrangement. The resulting quality of local packing, which has been pointed out to be a major determinant of a protein's unique structure,<sup>102,103</sup> cannot be expected to be as good as that in the native protein, where the side chains have been naturally selected to fit the requirements of packing.

Correct discrimination between native and misfolded protein structures achieved by our new ES/IS method requires use of a free energy function and evaluation of intra- and intermolecular interactions with an accurate physics-based model. Early at-

tempts to discriminate native and misfolded protein structures using the energy calculated with similar atom-atom potentials showed poor discrimination.<sup>1,2</sup> Previous calculations for the EMBL set performed with physics-based potentials, including empirical solvation free energy (proportional to exposed surface area), also failed to discriminate correctly;<sup>5</sup> apparently the potentials used in these calculations were incomplete. The early poor performance of physics-based potentials triggered the development and application of low-resolution potentials of mean force based on statistical analysis of packing of protein residues in known protein structures.<sup>6</sup> Heuristic pairwise likelihood potentials<sup>104-108</sup> provide good discrimination<sup>107</sup> (94% and 100% success rates) for the EMBL misfolded set,<sup>4</sup> and at the second Asilomar meeting on the Critical Assessment of Protein Structure Prediction (CASP-2),<sup>109</sup> it was generally recognized that methods directly based on exploiting known protein structures performed better than any of the *ab initio* methods. However, the resolution of the likelihood potentials tends to be limited. For example, residue-based likelihood potentials do not explicitly take into account such details as the backbone-backbone hydrogen bond network and will fail to provide a penalty for disruption of the backbone-backbone hydrogen bonds without formation of new hydrogen-bonded connections.<sup>110</sup> Hence, it is important to have available an alternative method that is solidly based in physics and formulates the stability problem in terms of a conformational free energy function computed for a high-resolution potential.

The ES/IS method owes its success to two additional features, development of a stable "physically consistent" continuum dielectric method based on the SIMS and FAMBE techniques for calculating the



MS surface and the solvation free energy. Stability and consistency, in turn, allow statistical enhancement of the accuracy of the calculation by averaging over a set of microstates of the stable conformation. (Stability and consistency required resolution of problems with the method used to solve the Poisson equation, which have not yet been entirely overcome for the finite-difference method.<sup>64,65,111</sup>)

A number of possible improvements of the ES/IS method have suggested themselves. The method's accuracy depends on the form and parameters of the molecular dynamics force field and of the continuum dielectric model. The values of parameters of the molecular dynamics force field are adjusted to obtain the best possible agreement between model and experiment for a number of carefully selected measured molecular properties.<sup>112</sup> Parameters required for the continuum dielectric model are partial atomic charges and atomic radii. The charges are the same as used in the molecular mechanics force field; sets of atomic radii have been derived by direct comparison of model and experimental hydration free energies.<sup>8,45</sup>

The results obtained with the continuum dielectric model have been found to be moderately sensitive to the values of the radii determining the placement of the dielectric interface.<sup>45</sup> As an alternative to a direct empirical fit of these "Born" radii, these radii may be defined by fitting the solvent polarization free energy of the dielectric continuum model to the "exact" solvent polarization free energies, which can be obtained via thermodynamic integration.<sup>17,45</sup> Molecules for which a transfer free energy cannot be measured experimentally can be included in such a fit between two theoretical methods. Another source of inconsistency with the atomic force field is the free energy of cavity creation,  $G_{\text{cav}}$ . The validity of the empirical Eq. 10 can also be checked for molecules of any size and shape, and the best parameter,  $\gamma_{\text{cav}}$  can be found for the particular force field. By this approach, the implicit continuum solvent model can be rendered consistent with the solvent-solute force field of the explicit solvent model. Such a refinement of the ES/IS model is in progress for the CEDAR force field<sup>53</sup> that has been used in the present study.

Another shortcoming of the ES/IS model is that coupling between the protein conformation and the ionization state of ionizable amino acid residues is ignored. Consideration of this coupling was found to be important in Monte Carlo modeling of the folding of a short 17-residue peptide in aqueous solvent.<sup>113</sup> The calculation of the free energy of ionization equilibrium for a given protein structure, though complicated and time consuming, will ultimately be an inevitable part of the process of estimating the stability of protein conformations or protein-drug complexes from first principles.

## REFERENCES

- Novotny, J., Brucoleri, R.E.E., Karplus, M. An analysis of incorrectly folded protein models: Implications for structure prediction. *J. Mol. Biol.* 177:788-818, 1984.
- Novotny, J., Rashin, A.A., Brucoleri, R.E. Criteria that discriminate between native proteins and incorrectly folded models. *Proteins* 4:19-30, 1988.
- Wang, Y., Zhang, H., Li, W., Scott, R.A. Discriminating compact nonnative structures from the native structure of globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* 92:709-713, 1995.
- Moult, J.A. Pro-Star: The protein potential site. Evaluation of potentials. (<http://prostar.Carb.nist.Gov/cgi-bin/PEvITbl1.Cgi>) 1997.
- Aybelj, F., Moult, J. Determination of the conformation of folding initiation sites in proteins by computer simulation. *Proteins* 23:129-141, 1995.
- Jernigan, R.L., Bahar, I. Structure derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* 6:195-209, 1996.
- Thomas, P.D., Dill, K. Statistical potentials extracted from protein structures: What is wrong with them? *J. Mol. Biol.* 257:457-469, 1996.
- Sitkoff, D., Sharp, K.A., Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978-1988, 1994.
- Simonson, T., Brünger, A. Solvation free energies estimated from macroscopic continuum theory: An accuracy assessment. *J. Phys. Chem.* 98:4683-4694, 1994.
- Honig, B., Sharp, K., Yang, A.S. Macroscopic models of aqueous solutions: Biological and chemical applications. *J. Phys. Chem.* 97:1101-1109, 1993.
- Vorobjev, Y., Hermans, J. SIMS: Computation of a smooth invariant molecular surface. *Biophys. J.* 73:722-732, 1997.
- Vorobjev, Y.N., Scheraga, H.A. A fast adaptive multigrid boundary element method for macromolecular electrostatics in a solvent. *J. Comp. Chem.* 18:569-583, 1997.
- Tomasi, J., Persico, M. Molecular interactions in solution: An overview of methods based on continuum distribution of the solvent. *Chem. Rev.* 94:2027-2094, 1994.
- Jackson, R.M., Sternberg, J.E. Application of scaled particle theory to model the hydrophobic effect: Implications for molecular association and protein stability. *Protein Eng.* 7:371-383, 1994.
- Frenkel, D. Molecular dynamic simulations of statistical mechanical systems. In: "Proceedings of the Enrico Fermi Summer School, Varenna, 1985." Ciccotti, G., Hoover, W.G. (eds.). Amsterdam: North-Holland, 1986:151-164.
- Helms, V., Wade, R.C. Free energies of hydration from thermodynamic integration: Comparison of molecular mechanics force fields and evaluation of calculation accuracy. *J. Comp. Chem.* 18:449-462, 1997.
- Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* 93:2395-2417, 1993.
- Radmer, R.J., Kollman, P.A. Free energy calculation methods: A theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes. *J. Comp. Chem.* 18:902-919, 1997.
- Hill, T.L. "An Introduction to Statistical Thermodynamics." New York: Dover, 1986.
- Chandler, D., Pratt, L.R. Statistical mechanics of chemical equilibria and intramolecular structures of non-rigid molecules in condensed phases. *J. Chem. Phys.* 65:2925-2940, 1976.
- Ben-Naim, A. Solvent effects on protein association and protein folding. *Biopolymers* 29:567-596, 1990.
- Langlet, J., Claverie, P., Caillet, J., Pullman, A. Improvements of the continuum model. I. Application to the calculation of the vaporization thermodynamic quantities of non-associated liquids. *J. Phys. Chem.* 92:1617-1631, 1988.
- Wang, L., O'Connell, T., Tropsha, A., Hermans, J. Energetic decomposition of the  $\alpha$ -helix-coil equilibrium of a dynamic model system. *Biopolymers* 39:479-489, 1996.

24. Jorgensen, W.L., Blake, J.F., Buckner, J.K. Free energy of TIP4P water and the free energy energies of hydration of CH<sub>4</sub> and Cl<sup>-</sup> from statistical perturbation theory. *Chem. Phys.* 129:193–200, 1989.
25. Matubayasi, N., Reed, L.H., Levy, R.M. Thermodynamics of the hydration shell. I. Excess energy of a hydrophobic solute. *J. Phys. Chem.* 98:10640–10649, 1994.
26. Ben-Naim, A., Marcus, Y. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* 81:2016–2027, 1984.
27. Guillot, B., Guissani, Y., Bratos, S.J. A computer simulation study of hydrophobic hydration of rare gases and of methane. I. Thermodynamic and structural properties. *J. Chem. Phys.* 95:3643–3648, 1991.
28. Hermann, R.B. Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem.* 76:2754–2759, 1972.
29. Chothia, C.H. Hydrophobic bonding and accessible area in proteins. *Nature* 248:338–339, 1974.
30. Reynolds, J.A., Gilbert, D.B., Tanford, C. Empirical correlation between hydrophobic free energy and cavity surface area. *Proc. Natl. Acad. Sci. U.S.A.* 71:2925–2927, 1974.
31. Hansen, J.P., McDonald, I.R. "Theory of Simple Liquids." New York: Academic Press, 1986.
32. Linford, R.G., Powell, R.J., Hildebrand, J.H. A comparison of the Gibbs energy and entropy of interfaces water-*n*-hexane and water-perfluorotributylamine. *J. Phys. Chem.* 74:3024–3025, 1970.
33. Reiss, H., Frisch, H.L., Helfand, E., Lebovitz, J.L. Aspects of the statistical thermodynamic of real fluids. *J. Chem. Phys.* 32:119–124, 1960.
34. Pierotti, R.A. A scaled particle theory of aqueous and non-aqueous solutions. *Chem. Rev.* 76:717–726, 1976.
35. Hummer, G., Pratt, L.R., Garcia, A.E. Free energy of ionic hydration. *J. Phys. Chem.* 100:1206–1215, 1996.
36. Wallqvist, W., Berne, B.J. Molecular dynamics study of the dependence of water solvation free energy on solute curvature and surface area. *J. Phys. Chem.* 99:2885–2892, 1995.
37. Wallqvist, W., Berne, B.J. Computer simulation of hydrophobic hydration forces on stacked plates at short range. *J. Phys. Chem.* 99:2893–2899, 1995.
38. Richards, F.M. Areas, volumes, packing, and protein structures. *Ann. Rev. Biophys. Bioeng.* 6:151–176, 1977.
39. Tunon, I., Silla, E., Pascual-Ahuir, J.L. Molecular surface area and hydrophobic effect. *Protein Eng.* 5:715–716, 1992.
40. Hermann, R.B. Modeling hydrophobic solvation of non-spherical systems: Comparison of use of molecular surface area with accessible surface area. *J. Comp. Chem.* 18:115–125, 1997.
41. Pangali, C., Rao, M., Berne, B.J. A Monte Carlo simulation of the hydrophobic interaction. *J. Chem. Phys.* 71:2975–2981, 1979.
42. Aqvist, J., Medina, C., Samuelsson, J.E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* 7:385–391, 1994.
43. Carlson, H.A., Jorgensen, W.L. An extended linear response method for determining free energies of hydration. *J. Phys. Chem.* 99:10667–10673, 1995.
44. Figueirido, F., Del Buono, G.S., Levy, R.M. Dielectric behavior in the interior of proteins: Result from a large-scale molecular dynamic simulation of solvated lysozyme. *Biophys. J.* 72:A216–A216, 1997.
45. Nina, M., Beglov, D., Roux, B. Atomic radii for continuum electrostatic calculations based on molecular dynamics free energy simulations. *J. Phys. Chem.* 101:5239–5248, 1997.
46. Roux, B., Yu, H.A., Karplus, M. Molecular basis for the Born model of ion solvation. *J. Phys. Chem.* 94:4683–4688, 1990.
47. Levy, R.M., Belhadj, M., Kitchen, D.B. Gaussian fluctuation formula for electrostatic free energy changes in solution. *J. Chem. Phys.* 95:3627–3633, 1991.
48. Jayaram, B., Fine, R., Sharp, K., Honig, B. Free energy calculations of ion hydration: An analysis of the Born model in terms of microscopic simulations. *J. Phys. Chem.* 93:4320–4327, 1989.
49. Rick, S.W., Berne, B.J. The aqueous solvation of water: A comparison of continuum methods with molecular dynamics. *J. Am. Chem. Soc.* 116:3949–3954, 1994.
50. Hummer, G., Pratt, L.R., Garcia, A.E. Hydration free energy of water. *J. Phys. Chem.* 99:14188–14194, 1995.
51. Aqvist, J., Hansson, T. On the validity of electrostatic linear response in polar solvent. *J. Phys. Chem.* 100:9512–9521, 1996.
52. Landau, L.D., Lifshitz, E.M. "Course of Theoretical Physics." Vol. 8: "Electrodynamics of Continuous Media." Translated from the Russian. Oxford: Pergamon Press, 1988.
53. Hermans, J. Sigma documentation. (URL: <http://femto.med.unc.edu/SIGMA>) University of North Carolina: 1995.
54. Lounnas, V., Pettitt, B.M., Phillips, G.N. A global model of the protein-solvent interface. *Biophys. J.* 66:601–614, 1994.
55. Sharp, K.A., Honig, B. Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Chem.* 19:301–332, 1990.
56. Madura, J.D., Davis, M.E., Gilson, M.K., Wade, R.C., Luty, B.A., McCammon, J.A. Biological application of electrostatic calculations and Brownian dynamics simulations. *Rev. Comp. Chem.* 5:229–267, 1994.
57. Bharadwaj, R., Windemuth, A., Sridharan, S., Honig, B., Nicholls, A. The fast multipole boundary element method for molecular electrostatics: An optimal approach for large systems. *J. Comp. Chem.* 16:898–913, 1995.
58. Rashin, A.A. Hydration phenomena, classical electrostatics, and the boundary element method. *J. Phys. Chem.* 94:1725–1733, 1990.
59. Rashin, A.A., Young, L., Topol, I.A. Quantitative evaluation of hydration thermodynamics with continuum model. *Biophys. Chem.* 51:359–374, 1994.
60. Yoon, B.J., Lenhoff, A.M. A boundary element method for molecular electrostatics with electrolyte effects. *J. Comp. Chem.* 11:1080–1086, 1990.
61. Juffer, A.H., Botta, E.F.F., Bert, A.M., van Keulen, B.A.M., van der Ploeg, A., Berendsen, H.J.C. The electric potential of a macromolecule in a solvent: A fundamental approach. *J. Comp. Phys.* 97:144–171, 1991.
62. Vorobjev, Y.N., Grant, J.A., Scheraga, H.A. A combined iterative and boundary element approach for solution of the nonlinear Poisson-Boltzmann equation. *J. Am. Chem. Soc.* 114:3189–3196, 1992.
63. Zhou, H.X. Macromolecular electrostatic energy within the nonlinear Poisson-Boltzmann equation. *J. Chem. Phys.* 100:3152–3162, 1994.
64. Brucoleri, R.E., Novotny, J., Davis, M.E. Finite-difference Poisson-Boltzmann calculations: Increased accuracy achieved by harmonic dielectric smoothing and charge anti-aliasing. *J. Comp. Chem.* 18:268–276, 1997.
65. Froloff, N., Windemuth, A., Honig, B. On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide interactions. *Protein Sci.* 6:1293–1301, 1997.
66. Zauhar, R.J., Morgan, R.S. The rigorous computation of the molecular electric potential. *J. Comp. Chem.* 9:171–187, 1988.
67. Connolly, M.L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* 16:548–558, 1983.
68. Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221:709–713, 1983.
69. Connolly, M.L. Computation of molecular volume. *J. Am. Chem. Soc.* 107:1118–1124, 1985.
70. Zauhar, R.J. SMATR: A solvent-accessible triangulated surface generator for molecular graphics and boundary element applications. *J. Comput. Aided Mol. Des.* 9:149–159, 1995.
71. Connolly, M.L. QCPE Program No. 429, 1983.
72. Perrot, G.B., Cheng, B., Gibson, K.D., et al. MSEED: A program for rapid analytical determination of accessible surface areas and their derivatives. *J. Comp. Chem.* 13:1–11, 1992.

73. Varshney, A., Brooks, F.P., Wright, W.V. Computing smooth molecular surface. *IEEE Comput. Graphics Appl.* 14:19–25, 1994.
74. Sanner, M.F., Olson, A.J., Spehner, J.C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* 38:305–320, 1996.
75. Besler, B.H., Merz, K.M., Kollman, P.A. Atomic charges derived from semi-empirical methods. *J. Comp. Chem.* 11:431–439, 1990.
76. Merz, K.M. Analysis of a large data base of electrostatic potential derived atomic charges. *J. Comp. Chem.* 13:749–767, 1992.
77. Vorobjev, Y.N. Performance of the Connolly MS, MSEED and Varshney methods with the boundary element method. Unpublished results, 1995.
78. Pascual-Ahuir, J.L., Silla, E., Tunon, I. GEPOL: An improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. *J. Comp. Chem.* 15:1127–1138, 1994.
79. Juffer, A.H.F., Eisenhaber, Hubbard, S.J., Walter, D., Argos, P. Comparison of atomic solvation parametric sets: Applicability and limitations in protein folding and binding. *Protein Sci.* 4:2499–2509, 1995.
80. Meyer, A.Y. The size of molecules. *Chem. Soc. Rev.* 15:449–474, 1986.
81. Douglas, C.C. Multigrid methods in science and engineering. *Comput. Sci. Eng.* 3:55–68, 1996.
82. McKenney, A., Greengard, L. A fast Poisson solver for complex geometries. *J. Comp. Phys.* 118:348–355, 1995.
83. Holst, M., Kozack, R.E., Saied, F., Subramaniam, S. Treatment of electrostatic effects in proteins: Multigrid-based Newton iterative method for solution of the full nonlinear Poisson-Boltzmann equation. *Proteins* 18:231–245, 1994.
84. Holst, M., Saied, F. Numerical solution of the nonlinear Poisson-Boltzmann equation: Developing more robust and efficient methods. *J. Comp. Chem.* 16:337–364, 1995.
85. Zhou, Z., Payne, P., Vasquez, M., Kuhn, N., Levitt, M. Finite-difference solution of the Poisson-Boltzmann equation: Complete elimination of self-energy. *J. Comp. Chem.* 17:1344–1351, 1996.
86. Goel, N.S., Gang, F., Ko, Z. Electrostatic field in inhomogeneous dielectric media: Indirect boundary element method. *J. Comp. Phys.* 118:172–179, 1995.
87. Zauhar, R.J., Varnek, A. A fast and space-efficient boundary element method for computing electrostatics and hydration effects in large molecules. *J. Comp. Chem.* 17:864–877, 1996.
88. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. "Numerical recipes in C." Cambridge: Cambridge University Press, 1988.
89. Amadei, A., Linsen, A.B.M., Berendsen, H.J.C. Essential dynamics of proteins. *Proteins* 17:412–425, 1993.
90. Garcia, A.E. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68:2696–2699, 1992.
91. Brooks, B.R., Janezic, D., Karplus, M. Harmonic analysis of large systems. I. Methodology. *J. Comp. Chem.* 16:1522–1542, 1995.
92. Janezic, D., Brooks, B.R. Harmonic analysis of large systems. II. Comparison of different protein models. *J. Comp. Chem.* 16:1543–1553, 1995.
93. Janezic, D., Venable, R.M., Brooks, B.R. Harmonic analysis of large systems. III. Comparison with molecular dynamics. *J. Comp. Chem.* 16:1554–1566, 1995.
94. Steinberg, I.Z., Scheraga, H.A. Entropy changes accompanying association reactions of proteins. *J. Biol. Chem.* 238:172–181, 1963.
95. Tidor, B., Karplus, M. The contribution of vibrational entropy to molecular association. *J. Mol. Biol.* 238:405–414, 1994.
96. Karplus, M., Kushick, J.N. Method for estimating the configurational entropy of macromolecules. *Macromolecules* 14:325–332, 1981.
97. Brady, G.P., Sharp, K.A. Entropy in protein folding and in protein-protein interactions. *Curr. Opin. Struct. Biol.* 7:215–221, 1997.
98. Allen, M.P., Tildesley, D.J. "Computer simulation of liquids." New York: Oxford Press, 1994.
99. Garcia-Moreno, B.E., Dwyer, J.J., Gittis, A.G., Lattman, E.E., Spencer, D.S., Stites, W.E. Experimental measurement of the effective dielectric in the hydrophobic core of protein. *Biophys. Chem.* 64:211–224, 1997.
100. van Aalten, D.M.F., de Groot, B.L., Findlay, J.B.C., Berendsen, H.J.C., Amadei, A. A comparison of techniques for calculating protein essential dynamics. *J. Comp. Chem.* 18:169–181, 1997.
101. Demchuk, E., Bashford, D., Gippert, G.P. Thermodynamics of reverse turn motif: Solvent effects and side-chain packing. *J. Mol. Biol.* 270:305–317, 1997.
102. Singh, J., Thornton, J.M. SIRIUS: An automated method of analysis of preferred packing arrangements between protein groups. *J. Mol. Biol.* 211:595–615, 1990.
103. Pattabiraman, N., Ward, K.B., Fleming, P.J. Occluded molecular surface: Analysis of protein packing. *J. Mol. Recognit.* 8:334–344, 1995.
104. Sippl, M.J. Calculation of conformational ensembles from potential of mean force: An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* 213:859–883, 1990.
105. Casari, G., Sippl, M.J. Structure-derived hydrophobic potential: Hydrophobic potential derived from X-ray structures of globular proteins is able to identify native folds. *J. Mol. Biol.* 224:725–732, 1992.
106. Kocher, J.P., Rooman, M.J., Wodak, S.J. Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. *J. Mol. Biol.* 235:1598–1613, 1994.
107. Brauer, A., Beyer, A. An improved pair potential to recognize native protein folds. *Proteins* 18:254–261, 1994.
108. Miyazawa, S., Jernigan, R.L. Residue-residue potentials with favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* 256:623–644, 1996.
109. Moulton, J., Hubbard, T., Bryant, S.H., Fidelis, K., Pedersen, J.T. Critical assessment of methods of protein structure prediction (CASP) Round II. *Proteins Suppl.* 1:2–6, 1997.
110. Honig, B., Cohen, F.E. Adding backbone to protein folding: Why proteins are polypeptides. *Fold. Des.* 1:R17–R20, 1996.
111. Vajda, S., Sippl, M., Novotny, J. Empirical potentials and functions for protein folding and binding. *Curr. Opin. Struct. Biol.* 7:222–228, 1997.
112. Moulton, J. Comparison of database potentials and molecular mechanics force fields. *Curr. Opin. Struct. Biol.* 7:194–199, 1997.
113. Ripoll, D.R., Vorobjev, Y.N., Liwo, A., Vila, J.A., Scheraga, H.A. Coupling between folding and ionization equilibria: Effects of pH on the conformational preferences of polypeptides. *J. Mol. Biol.* 264:770–783, 1996.