AGBNP: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-Resolution Modeling

EMILIO GALLICCHIO, RONALD M. LEVY
Department of Chemistry and Chemical Biology and BIOMAPS Institute of Quantitative Biology, Rutgers University, Piscataway New Jersey 08854

Received 21 August 2003; Accepted 8 October 2003

Abstract: We have developed an implicit solvent effective potential (AGBNP) that is suitable for molecular dynamics simulations and high-resolution modeling. It is based on a novel implementation of the pairwise descreening Generalized Born model for the electrostatic component and a new nonpolar hydration free energy estimator. The nonpolar term consists of an estimator for the solute-solvent van der Waals dispersion energy designed to mimic the continuum solvent solute-solvent van der Waals interaction energy, in addition to a surface area term corresponding to the work of cavity formation. AGBNP makes use of a new parameter-free algorithm to calculate the scaling coefficients used in the pairwise descreening scheme to take into account atomic overlaps. The same algorithm is also used to calculate atomic surface areas. We show that excellent agreement is achieved for the GB self-energies and surface areas in comparison to accurate, but much more expensive, numerical evaluations. The parameter-free approach used in AGBNP and the sensitivity of the AGBNP model with respect to large and small conformational changes makes the model suitable for high-resolution modeling of protein loops and receptor sites as well as high-resolution prediction of the structure and thermodynamics of protein-ligand complexes. We present illustrative results for these kinds of benchmarks. The model is fully analytical with first derivatives and is computationally efficient. It has been incorporated into the IMPACT molecular simulation program.

© 2004 Wiley Periodicals, Inc. J Comput Chem 25: 479–499, 2004

Key words: implicit solvent models; generalized Born model; nonpolar hydration; protein loop modeling; modeling of protein-ligand complexes

Introduction

Hydration phenomena play an important role in virtually every process occurring in aqueous solution. Hydration has a particularly large effect on the thermodynamics of biological processes that involve the breakage or formation of noncovalent bonds. The accurate modeling of hydration thermodynamics is therefore essential for predicting protein structures, ligand binding free energies, and conformational equilibria.[1–5]

Numerous stringent requirements make the development of practically useful solvation free energy models for biological applications very challenging. In order to be applicable to ligand binding affinity prediction, the model should be accurate over a wide range of molecular sizes, from small molecules to large biological macromolecules, and over a wide range of hydrophobic, polar, and ionic functional groups. In order to study protein folding, allosteric reactions, and flexible receptor and ligand docking, the model must be able to describe hydration free energy differences between different molecules as well as different conformations of the same molecule, including large scale protein motions and the motion of only a few atoms. Finally, the model needs to be computationally efficient, and should be expressed in analytical form with analytical gradients for seamless incorporation in a molecular mechanics code to perform conformational sampling and energy optimization calculations. Although models with some of these characteristics exist,[4–12] none of them meets all the above requirements.

Explicit solvent models provide the most detailed and complete description of hydration phenomena.[13] They are, however, computationally demanding not only because of the large number of solvent atoms involved, but also because of the need to average over many solvent configurations to obtain meaningful thermodynamic data. Implicit solvent models[14] offer an attractive alternative

Correspondence to: E. Gallicchio; e-mail: emilio@hpcp.rutgers.edu

Contract/grant sponsor: National Institute of Health; contract/grant number: GM30580.

© 2004 Wiley Periodicals, Inc.

and have been shown to be useful for applications including protein folding[15,16] and protein native fold recognition,[17,18] small molecule hydration free energy prediction,[6,11,19] and binding affinity prediction.[20,21]

In modern implicit solvent models[9] the solvation free energy is typically decomposed into a nonpolar component and an electrostatic component. The nonpolar component corresponds to the free energy of hydration of the uncharged solute, and the electrostatic component, computed using dielectric continuum models, corresponds to the free energy of turning on the solute partial charges.

Dielectric continuum methods account for the electrostatic component by treating the water solvent as a uniform high-dielectric continuum.[22] Methods based on the numerical solution of the Poisson-Boltzmann (PB) equation[23,24] provide a virtually exact representation of the response of the solvent within the dielectric continuum approximation. Their computational complexity is, however, still comparable to explicit solvent models and they are not easily integrated in molecular dynamics simulation programs due to the difficulties associated with calculating the forces associated with the electrostatic polarization energy. Recent advances extending dielectric continuum approaches have focused on the development of Generalized Born (GB) models,[10,25] which have been shown to reproduce with good accuracy PB[12,26,27] and explicit solvent[28] results at a fraction of the computational expense. The development of computationally efficient analytical and differentiable GB methods with gradients based on pairwise descreening schemes[29,30] has made possible the integration of GB models in molecular dynamics packages for biological simulations.[7,31–33]

Despite the fact that nonpolar hydration forces dominate whenever hydrophobic interactions[34] are important, the general availability of accurate models for the nonpolar component of the hydration free energy is lacking. The structure and properties of proteins in water is highly influenced by hydrophobic interactions.[1,2,35,36] Hydrophobic interactions also play a key role in the mechanism of ligand binding to proteins.[8,37–39] Empirical surface area models[40] for the nonpolar component of the solvation free energy are widely used.[6,19,25,41–47] Surface area models are useful as a first approximation; however, deficiencies are observed[43,48,49] that are particularly severe in the context of high resolution modeling and force field transferability.[50]

In this article we develop the Analytical Generalized Born plus Nonpolar (AGBNP) model, an implicit solvent model based on the GB model[25–28,32,51] for the electrostatic component, and on the decomposition of the nonpolar hydration free energy into a cavity component based on the solute surface area and a solute-solvent van der Waals interaction free energy component modeled using an estimator based on the Born radius of each atom.

The electrostatic solvation model (AGB) is a pairwise descreening GB scheme motivated by the model of Hawkins et al.[30] The Born radius of each atom is obtained by summing its pairwise descreening interactions with the other solute atoms. A pairwise descreening interaction is defined as the integral of the $1/r^4$ function centered on one atom over the portion of the van der Waals volume of the other atom that does not intersect the van der Waals volume of the first. Scaling coefficients are assigned to each atom to offset the overcounting of regions of space occupied by more than one atom. The main distinction between AGB and the model of Hawkins et al.[30] is that AGB computes the scaling factors that account for atomic overlaps from the geometry of the molecule rather than introducing them as geometry-independent parameters fit to either experiments or to numerical PB results. This enhances the sensitivity of the AGB model with respect to small conformational changes, making it particularly suitable for high-resolution modeling of protein loops and receptor sites as well as for high-resolution prediction of the structure and thermodynamics of protein-ligand complexes. In AGB the atomic scaling factors are obtained by partitioning the solute volume into volumes assigned exclusively to each atom. The partitioning of the solute volume is performed with an algorithm based on the Poincaré formula, described in the following section, and implemented using a Gaussian overlap approximation to compute the volume of intersection of multiple atomic spheres. The same algorithm is used to compute the solute volume and atomic surface areas.

Previously proposed analytical pairwise descreening GB models[7,31,32,52] require the parameterization of scaling coefficients to treat the effects of atomic overlaps. Parameterized models are unavoidably biased toward their training sets. A parameterization based on native protein structures, for example, reflects the average atomic density of native proteins, which is very high, and yields parameters that are not as transferable to non-native protein structures and therefore are less accurate for protein structure prediction. Pairwise GB models in which the scaling coefficients are independent of molecular geometry[30,31,52] are not optimal for high-resolution modeling of protein-ligand binding, which requires an energy function able to respond to both global conformational changes and smaller atomic rearrangements. The parameter-free approach employed in the AGB model to calculate scaling coefficients is particularly useful when treating unusual functional groups often found when screening large numbers of ligand candidates. The scaling coefficients derive from a training set in which a particular functional group that is not represented may be unsuitable for such a functional group. On the other hand, it is impractical to construct a training set in which all possible functional groups and combinations of functional groups are represented. The parameter-free analytical scheme used by AGB ensures that each atom in any molecule is assigned proper scaling coefficients.

The nonpolar hydration free energy estimator proposed in this article is based on the decomposition of the nonpolar free energy into a cavity term, proportional to surface area, and an attractive dispersion energy term, which approximates the continuum solvent solute-solvent van der Waals interaction energy assuming that the solvent density outside the solute is homogeneous.[50] The functional form of the nonpolar hydration free energy proposed here differs in substance from models based on only the solute surface area. This form is motivated by a series of recent studies of small molecules[11,49,53] and macromolecules[50,54] that show that some characteristics of the nonpolar hydration free energy are not captured well by models based only on surface area. As described in the following sections, the nonpolar hydration free energy estimator has been formulated in analytical form. We employ an efficient analytical algorithm to compute the solute surface area based on the same algorithm used to compute the scaling coefficients for the GB hydration free energy calculation. For the van der Waals

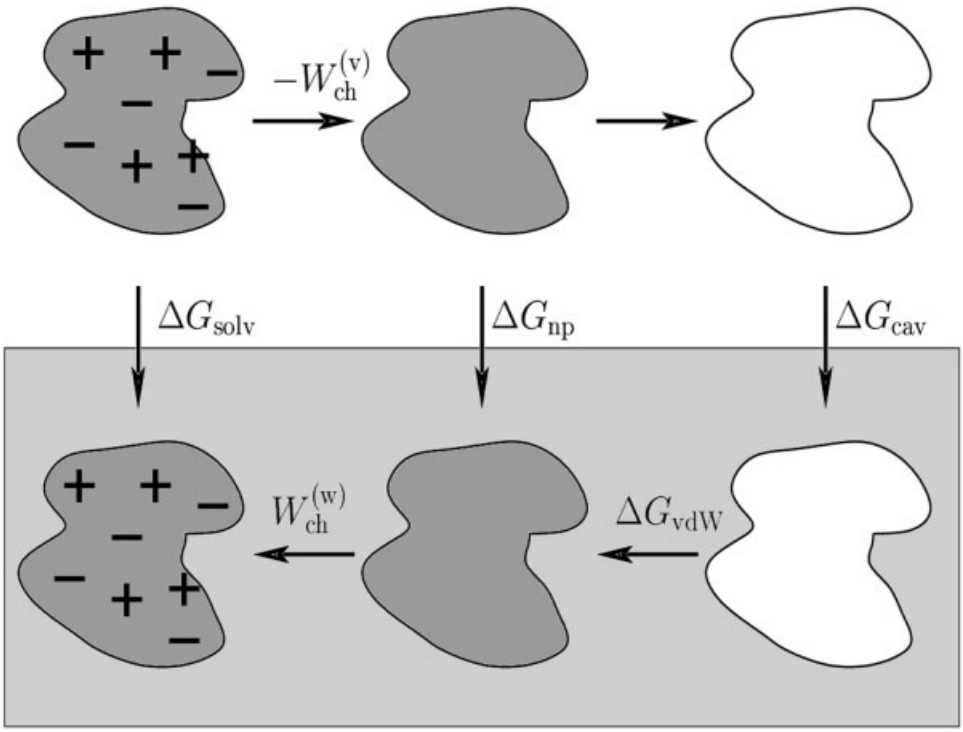


Figure 1. Thermodynamic cycle depicting the decomposition of the solvation free energy into electrostatic and nonpolar components. Starting from the upper left state and moving clockwise, the sequence of steps is: uncharging of the solute in vacuum, removal of the solute-solvent van der Waals intersection in vacuum (there is no free energy change associated with this step because of the lack of solvent molecules), hydration of the solute cavity, establishment of the solute-solvent van der Waals intersections in water, and charging of the solute in water.

hydration free energy term we employ a functional form based on the atomic Born radii.

It has been generally the norm to test the accuracy of GB models against the results of continuum dielectric PB calculations.[7,12,26,27,52] Our approach has been to validate implicit solvent models primarily against experimental data[11,55] and explicit solvent predictions.[28,50,56] The reason behind this choice is our belief that the dielectric continuum assumption for the solvent introduces a more severe approximation than any additional approximation introduced by the GB model. In this article our goal is to design an analytical scheme to generate Born radii and atomic surface areas that respond correctly to conformational changes. We show that the AGBNP analytical estimates of the geometrical quantities used in the AGBNP model (Born radii and atomic surface areas) are indeed in very good agreement with high-accuracy numerical calculations of these same quantities.

To test the stability of native protein structures, we present a series of long molecular dynamics simulations using the AGBNP model. The measured CPU speed of AGBNP is 20 to 26% of the CPU speed in vacuum. To demonstrate the applicability of the AGBNP model to high-resolution structure prediction applications, we also test the ability of the OPLS/AGBNP model in distinguishing the native conformation from a set of high quality decoy sets of protein loops and protein-ligand binding complexes.

Methods

Hydration Free Energy Decomposition

The hydration free energy ΔG_{h} is defined as the free energy change for transferring a molecule from the gas phase to the water solvent phase. The transfer process can be decomposed into a series of steps: first, the atomic partial charges and van der Waals interactions of the solute are removed in vacuum, then the resulting solute cavity is transferred into aqueous solution, and the solute's van der Waals interactions and partial charges are restored.[11,19,25,43,57] This is illustrated in terms of a thermodynamic cycle (see Fig. 1) whereby the solvation free energy can be expressed as

$$\Delta G_{\mathrm{h}} = \Delta G_{\mathrm{elec}} + \Delta G_{\mathrm{np}} = \Delta G_{\mathrm{elec}} + \Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{vdW}} \qquad (1)$$

where ΔG_{elec} is the electrostatic contribution to the solvation free energy, which is the difference between the work, $W_{\mathrm{chg}}^{(\mathrm{w})}$, of charging the solute in solution, and the work, $W_{\mathrm{chg}}^{(\mathrm{v})}$, of charging the solute in vacuum, ΔG_{cav} is the cavity hydration free energy, and ΔG_{vdW} is the free energy for establishing the solute-solvent van der Waals dispersion interactions. Each free energy component is modeled separately by the AGBNP model, as described in detail in the following sections.

Volume and Surface Area Model

The solute volume is described as a set of overlapping spheres of radius R_i centered on the atomic positions $\mathbf{r}_i$. We have implemented a model for calculating the solute volume and solute surface area based on the Poincaré formula for the union of N spheres.[58] According to the Poincaré formula (also known as the inclusion-exclusion principle) the solute volume is given by

$$V = \sum_i V_i - \sum_{i<j} V_{ij} + \sum_{i<j<k} V_{ijk} - \cdots \tag{2}$$

where $V_i = 4\pi R_i^3/3$ is the volume of atom i, V_{ij} is the volume of intersection of atoms i and j (second order intersection), V_{ijk} is the volume of intersection of atoms i, j, and k (third order intersection), and so on. We define the self-volume V_i' of atom i as the volume of atom i, minus one-half of the second order intersection volumes made by atom i and one other atom, plus one-third of the sum of third order intersection volumes made by atom i and two other atoms, and so forth:

$$V_i' = V_i - \frac{1}{2}\sum_j V_{ij} + \frac{1}{3}\sum_{j<k} V_{ijk} + \cdots \tag{3}$$

It can be shown that the summing of the self-volumes of all the atoms yields eq. (2), the total volume of the molecule, thus

$$V = \sum_i V_i' \tag{4}$$

Equation (4) leads to the interpretation of the self-volume of an atom as a measure of the solute volume that belongs exclusively to that atom. Due to the overlaps with other atoms, the self-volume V_i' of an atom is smaller than the actual volume V_i of the atom. The ratio

$$s_i = \frac{V_i'}{V_i} \leq 1 \tag{5}$$

between the self-volume and the volume of the atom measures the fraction of the van der Waals volume of atom i that is considered self-volume; it is used below to evaluate the GB solvation energy of the molecule.

The van der Waals surface area A_i of atom i is given by the derivative of the solute volume with respect to the radius R_i[59]:

$$A_i = \frac{\partial V}{\partial R_i} \tag{6}$$

$$= 4\pi R_i^2 - \sum_j \frac{\partial V_{ij}}{\partial R_i} + \sum_{j<k} \frac{\partial V_{ijk}}{\partial R_i} + \cdots \tag{7}$$

where eq. (7) follows from inserting eq. (2) into eq. (6).

It is impractical to calculate the atomic self-volumes [eq. (3)] by exact evaluation of the spherical intersection volumes $V_{ijkl\cdots}$.[58] The calculation of the atomic self-volumes and surface areas is instead implemented using an algorithm based on the approximate method proposed by Grant and Pickup.[60] According to their method the volume of each atom is described by a Gaussian density function

$$\rho_i(\mathbf{r}) = p \exp[-c_i(\mathbf{r} - \mathbf{r}_i)^2] \tag{8}$$

The overlap volume formed by n spheres is then approximated by the integral of the product of the n corresponding Gaussian functions:

$$V_{12\cdots n} \simeq V_{12\cdots n}^g = \int d^3\mathbf{r}\rho_1(\mathbf{r})\rho_2(\mathbf{r})\cdots\rho_n(\mathbf{r}) \tag{9}$$

which is available in analytic form:

$$V_{12\cdots n}^g = p_{12\cdots n}\exp(-K_{12\cdots n})\left(\frac{\pi}{\Delta_{12\cdots n}}\right)^{3/2} \tag{10}$$

where

$$p_{12\cdots n} = p^n \tag{11}$$

$$K_{12\cdots n} = \frac{1}{\Delta_{12\cdots n}}\sum_{i=1}^{n}\sum_{j=i+1}^{n} c_i c_j r_{ij}^2 \tag{12}$$

and

$$\Delta_{12\cdots n} = \sum_{i=1}^{n} c_i \tag{13}$$

The Gaussian exponent parameter c_i determines the "softness" of the atomic solute density. It is set as

$$c_i = \frac{\kappa}{R_i^2} \tag{14}$$

where κ is a dimensionless parameter that regulates the diffuseness of the Gaussian function. To satisfy the requirement that the integral of $\rho_i(\mathbf{r})$ reproduces the hard-sphere volume $4\pi R_i^3/3$ of the atom, the parameter p is chosen such that

$$p = \frac{4\pi}{3}\left(\frac{\kappa}{\pi}\right)^{3/2} \tag{15}$$

The use of a large κ gives a large p and hence a narrow and tall Gaussian, whereas a small κ gives a short and diffuse Gaussian. In this work we set $\kappa = 2.227$, which results in $p = 2.5$. This value of κ was shown by Grant and Pickup[60] to accurately reproduce the volume and van der Waals surface areas of small molecules and proteins. No further attempt was made in this work to optimize the κ parameter.

GB Model

In the GB model[10] the electrostatic component of the hydration free energy is estimated as

$$\Delta G_{\text{elec}} \simeq \Delta G_{\text{GB}} = u_\varepsilon \sum_{ij} \frac{q_i q_j}{f_{ij}} \tag{16}$$

where

$$u_\varepsilon = -\frac{1}{2}\left(\frac{1}{\varepsilon_{\text{in}}} - \frac{1}{\varepsilon_{\text{w}}}\right) \tag{17}$$

where ε_{in} is the dielectric constant of the interior of the solute, ε_{w} is the dielectric constant of the solvent, q_i and q_j are the charges of atom i and j, and

$$f_{ij} = \sqrt{r_{ij}^2 + B_i B_j \exp(-r_{ij}^2/4B_iB_j)} \tag{18}$$

where B_i and B_j are the Born radii of atoms i and j defined below. The summation in eq. (16) runs for all atom pairs (i, j) including $i = j$. The diagonal $i = j$ terms can be separated from off-diagonal terms $i \neq j$ yielding the equivalent expression

$$\Delta G_{\text{GB}} = u_\varepsilon \sum_i \frac{q_i^2}{B_i} + 2u_\varepsilon \sum_{i<j} \frac{q_i q_j}{f_{ij}} \tag{19}$$

The first term at the right hand side of eq. (19) is the sum of the GB self-energies of the atoms of the molecule, the second term is the sum of the GB pair-energies. The self-energy of atom i, corresponding to the solvation energy of the solute when only the charge of atom i is nonzero, measures the energy of atom i in the reaction field due to the polarization of the solvent induced by the partial charge of atom i in the solute cavity. The self-energy is largest for the atoms that are most exposed to the solvent because they are capable of inducing stronger polarization fields. This effect is captured by the GB model in that atoms exposed to the solvent have smaller Born radii whereas buried atoms tend to have larger Born radii. The pair-energy term corresponds to the dampening of electrostatic interactions in a high dielectric medium due to the screening of the solute charges. The GB equation [eq. (19)] can be shown to be an exact representation of the electrostatic charging free energy of the solute in a continuum dielectric in the two limiting cases of infinite atomic separation and complete atomic overlap.[25]

The solute cavity is described as a set of overlapping spheres of radius R_i centered on the atomic positions $\mathbf{r}_i$. The Born radius of atom i is defined as the radius of the monoatomic solute with partial charge q_i whose continuum dielectric hydration free energy is equal to the self-energy of atom i. The self-energy of atom i is defined as the hydration free energy of the solute when the atomic partial charges of all the solute atoms, except atom i, are set to zero. In the Coulomb field approximation,[31] the expression for the Born radius is given by the integral of $1/r^4$ centered on atom i over the solvent region:

$$\frac{1}{B_i} = \beta_i = \frac{1}{4\pi} \int_{\text{Solvent}} d^3\mathbf{r}\, \frac{1}{(\mathbf{r} - \mathbf{r}_i)^4} \tag{20}$$

The accuracy of the Coulomb field approximation [eq. (20)] has been analyzed using exact analytical models[10,29,31] and accurate numerical PB calculations.[61,62] It has been found to be generally acceptable with the exception of cases with very asymmetric solute geometries, where it tends to overestimate the values of the Born radii. Empirical corrections to Coulomb field approximation have been proposed.[27,62] It has been pointed out that approximations in the integration procedure to obtain the Born radii may actually be of more significance than the Coulomb field approximation itself.[61]

Pairwise Solute Descreening Approximation

By adding and subtracting from eq. (20) the expression for the inverse of the Born radius of a solute composed only of atom i, we obtain a computationally more convenient integral expression for the inverse Born radius[26]:

$$\beta_i = \frac{1}{R_i} - \frac{1}{4\pi} \int_{\Omega_i} d^3\mathbf{r}\, \frac{1}{(\mathbf{r} - \mathbf{r}_i)^4} \tag{21}$$

where Ω_i is the bounded region corresponding to the solute volume excluding the atomic sphere corresponding to atom i. In eq. (21), $1/R_i$ is the inverse Born radius of atom i in the absence of all the other solute atoms. The second term on the right hand side of eq. (21) takes into account the displacement of the solvent dielectric due to the other solute atoms. In the pairwise solute descreening approximation this term is approximated by a pairwise sum[29,30]

$$\beta_i \simeq \beta_i^0 - \frac{1}{4\pi} \sum_{j \neq i} Q_{ji} \tag{22}$$

where $\beta_i^0 = 1/R_i$ and

$$Q_{ji} = \int_{\Omega_{ji}} d^3\mathbf{r}\, \frac{1}{(\mathbf{r} - \mathbf{r}_i)^4} \tag{23}$$

is the integral (available in analytic form, see Appendix B) of $(\mathbf{r} - \mathbf{r}_i)^{-4}$ over the volume, Ω_{ji}, of the atomic sphere corresponding to solute atom j that lies outside atom i (see Fig. 2).

Due to the overcounting of regions that lie inside more than one atomic sphere, the value of the Born radius of atom i would be significantly overestimated if the full descreening integral Q_{ji} is used in eq. (22). To take into account overcounting, scaling coefficients $s_{ji} < 1$ are introduced that reduce the effect of the descreening of atom j on atom i; thereby eq. (22) becomes

$$\beta_i = \beta_i^0 - \frac{1}{4\pi} \sum_{j \neq i} s_{ji} Q_{ji} \tag{24}$$

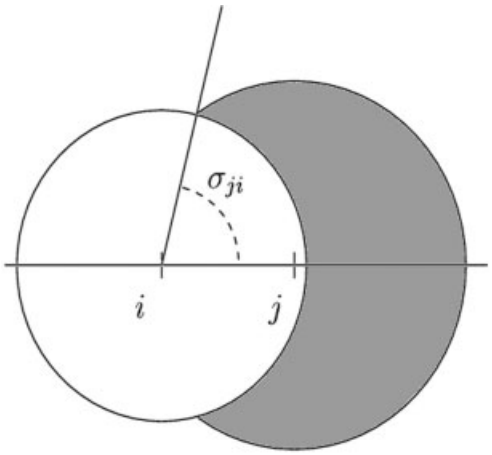


Figure 2. A diagram showing the portion of the volume of atom j (shaded area) over which the function $(\mathbf{r} - \mathbf{r}_i)^{-4}$ is integrated to obtain the contribution [eq. (23)] of atom j to the Born radius of atom i within the pairwise descreening approximation.

The scaling coefficients should be chosen such that the pairwise descreening sum in eq. (24) approximates the integral of $(\mathbf{r} - \mathbf{r}_i)^{-4}$ over the solute volume outside atom i [eq. (21)]. The atomic self-volumes [eq. (3)], which measure the amount of effective volume assigned to each atom, are a natural choice in deriving appropriate scaling coefficients. An atom with a small self-volume should descreen other atoms to a lesser extent than an atom with a larger self-volume. The relationship between self-volumes and pairwise descreening scaling coefficients is derived by requiring that the pairwise sum $(1/4\pi) \sum_{j \neq i} s_{ji}\Omega_{ji}$ reproduces the volume of the solute outside atom i. Here Ω_{ji} is defined in analogy with Q_{ji} [eq. (23)] by the integral of the unit function, rather than the function $(\mathbf{r} - \mathbf{r}_i)^{-4}$, over the solute volume outside atom i:

$$\Omega_{ji} = \int_{\Omega_{ji}} d^3\mathbf{r} \tag{25}$$

It can be shown that the identity

$$\Omega_i = \frac{1}{4\pi} \sum_{j \neq i} s_{ji}\Omega_{ji} \tag{26}$$

is obtained when the scaling coefficients s_{ji} are chosen as the self-volume scaling factor for atom j [eq. (5)] when atom i is removed from the solute. With this choice of scaling coefficients, therefore, the best possible uniform coverage of the domain of integration is established.

The geometrical interpretation of this is illustrated in Figure 3. The van der Waals volume of the solute outside atom i is partitioned into volume intersections defined by the boundaries of the spherical atomic volumes. Each intersection belongs to one or more atoms. In eq. (24) the integral of $(\mathbf{r} - \mathbf{r}_i)^{-4}$ over an intersection volume belonging to n atoms is counted n times. The self-volume scaling coefficients in eq. (24) reduce the contribution from each atom to account for the fact that intersections belonging to multiple atoms are integrated over more than once. The self-volume scaling coefficients yield [eq. (26)] the overall volume Ω_i of the domain of integration exactly; therefore, when using the self-volume scaling factors s_{ji}, on average each volume element is weighted equally.

The calculation of the self-volume scaling factors follows the same algorithm used for the calculation of the atomic surface areas. The direct calculation of the self-volume scaling coefficient s_{ji} (the self-volume solute factor of atom j when atom i is removed from the solute) would substantially increase the computationally complexity of the pairwise descreening algorithm. It would entail calculating multiple times the self-volume of atom j, once for each ij pairwise descreening interaction, by removing, from the expression of the self-volume of atom j from eq. (3), all overlap volumes $V_{ijk\cdots}$ of any order that include atoms i and j. To simplify the calculation we adopted the approximate expression

$$s_{ji} \simeq s_j + \frac{1}{2}\frac{V_{ij}}{V_j} \tag{27}$$

based on subtracting from the self-volume of atom j only one-half the V_{ij} direct two-body overlap volume. We found that eq. (27) is a good approximation to s_{ji} and it is exact when intersections of order three or higher are absent.

It is useful to compare the pairwise descreening scheme proposed here with other pairwise descreening methods. To take into account atomic overlaps, Hawkins et al.[30] have introduced a set of

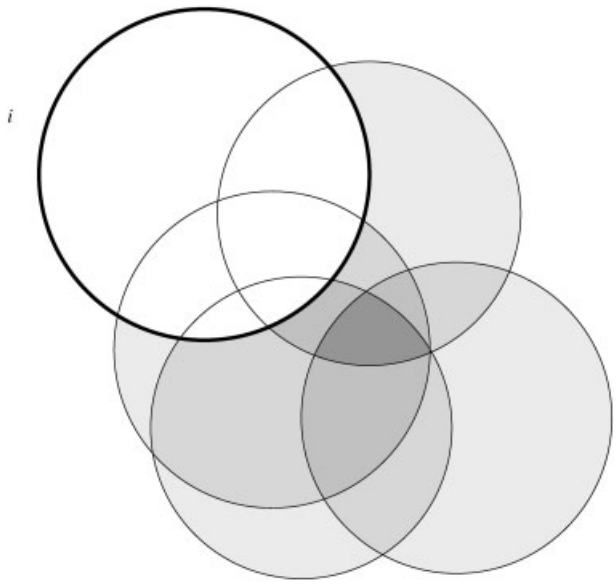


Figure 3. Diagram illustrating the role of volume intersections in the calculation of the Born radius of atom i (upper left, heavy line) using the pairwise descreening approximation and the self-volume rescaling factors [eq. (24)]. Intersections of higher order are represented by darker shades of gray. Volume elements inside intersection regions are overcounted but, when using the self-volume scaling factors s_{ji}, the average weight assigned to each volume element is approximately 1.

constant scaling factors S_j for each atom type (hydrogen, carbon, oxygen, and nitrogen) that reduce the radius of atom j used in the calculation of Q_{ji}. However in the approach of Hawkins et al. the radius scaling parameters S_j do not depend on solute geometry and chemical functionality. According to their model, for example, the radius of two atoms with potentially very different degrees of overlap with other atoms (such as a primary and a tertiary aliphatic carbon) would be reduced by the same amount. Similarly, in the ACE model of Schaefer and Karplus[31] geometry-independent effective integration volumes are assigned to each atom based on the atom type. Values of the effective atomic volumes are available only for functional groups present in proteins. Note, moreover, that in the ACE model the pairwise descreening function is derived with the solute atomic volumes and charge densities described by Gaussian functions. In the AGBNP method proposed here Gaussian atomic volume densities are used solely as a computational device to efficiently compute self-volume scaling factors, whereas the pairwise descreening function Q_{ji} (see Appendix B) is based on using spherical atomic integration volumes. To address atomic overlaps, Qiu et al.[7] have introduced multiplicative parameters P_n to express Q_{ji} that depend on whether atoms i and j are chemically bonded ($n = 2$), or form a bond angle ($n = 3$), or are otherwise nonbonded ($n = 4$). In the model of Qiu et al. the parameters depend only on the direct relationship between atoms j and i and do not take into account the variability of overlapping patterns of atom j with other solute atoms, which themselves may or may not overlap with atom i. Moreover, Qiu et al. make the additional approximation that Q_{ji} can be approximated as V_j/r_{ji}^4. This approximation, based on the assumption that the function $(\mathbf{r} - \mathbf{r}_i)^{-4}$ is constant within the volume of atom j, is expected to be mostly in error for nearest neighbor pairs that contribute the most to the Born radius.

The strategy proposed here for the calculation of the Born radii within the pairwise descreening approximation provides improved accuracy for the Born radii over previously proposed pairwise methods.[7,30,31] In our scheme the pairwise contributions Q_{ij} are calculated exactly taking into account the possible overlap between atoms i and j, and the rescaling parameters s_{ji} we adopt, derived from the self-volume rescaling parameters s_j, are automatically adjusted according to the current solute conformation. This scheme for calculating the Born radii can be applied to molecules with any combination of functional groups without requiring a previous parameterization step. Furthermore, because it is based on parameter-free estimators that describe the actual geometry of the solute rather than in an average way, we believe our model to be particularly well suited for high-resolution modeling.

Nonpolar Model

The nonpolar model adopted in this work differs from most other implicit hydration free energy models in that the nonpolar component ΔG_{np} is subdivided into cavity and solute-solvent van der Waals interaction terms:

$$\Delta G_{\mathrm{np}} = \Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{vdW}} \tag{28}$$

rather than estimated as a whole using a surface area model. This choice is motivated by a recent survey of the solute-water van der Waals interaction energy of proteins and protein-ligand complexes.[50] In the survey it was found that on the fine grained energy scale necessary to predict the high-resolution structure of proteins and protein-ligand complexes, the correlation between ΔG_{vdW} and the solvent accessible surface area of the solute was poor. This decomposition scheme has also been shown to be advantageous in the prediction of experimental hydration free energies of small organic molecules.[11]

The cavity component is described by a surface area model[49,63–65]:

$$\Delta G_{\mathrm{cav}} = \sum_i \gamma_i A_i \tag{29}$$

where the summation runs over solute atoms, A_i is the van der Waals surface area of atom i, and γ_i is the surface tension parameter assigned to atom i.

The solute-solvent van der Waals free energy term is modeled by the expression

$$\Delta G_{\mathrm{vdW}} = \sum_i \alpha_i \frac{a_i}{(B_i + R_w)^3} \tag{30}$$

where α_i is an adjustable dimensionless parameter on the order of 1 and

$$a_i = -\frac{16}{3}\pi\rho_w\varepsilon_{iw}\sigma_{iw}^6 \tag{31}$$

where $\rho_w = 0.33428\ \text{Å}^{-3}$ is the number density of water at standard conditions, and σ_{iw} and ε_{iw} are the OPLS force field[66] Lennard-Jones interaction parameters for the interaction of solute atom i with the oxygen atom of the TIP4P water model.[67] If σ_i and ε_i are the OPLS Lennard-Jones parameters for atom i

$$\sigma_{iw} = \sqrt{\sigma_i\sigma_w} \tag{32}$$

$$\varepsilon_{iw} = \sqrt{\varepsilon_i\varepsilon_w} \tag{33}$$

where $\sigma_w = 3.15365$ Å, and $\varepsilon_w = 0.155$ kcal/mol are the Lennard-Jones parameters of the TIP4P water oxygen. In eq. (30) B_i is the Born radius of atom i and R_w is a parameter corresponding to the radius of a water molecule.

Equation (30) is motivated by the following arguments. The free energy associated with turning on the solute-solvent van der Waals interactions is well approximated by the average solute-solvent van der Waals interaction energy U_{vdW}[49,68,69]:

$$\Delta G_{\mathrm{vdW}} \simeq U_{\mathrm{vdW}} = \sum_i U_{\mathrm{vdW}}(i) \tag{34}$$

where $U_{\mathrm{vdW}}(i)$ is the van der Waals interaction energy of atom i with the water solvent. We have shown that $U_{\mathrm{vdW}}(i)$ can be accurately estimated by integrating over the solvent region, the attractive part of the solute-solvent Lennard-Jones interaction potential of atom i assuming homogeneous solvent density ρ_w outside the solute[50]:

$$U_{\mathrm{vdW}}(i) \simeq -\rho_w \int_{\mathrm{Solvent}} d^3\mathbf{r}\, \frac{4\varepsilon_{iw}\sigma_{iw}^6}{(\mathbf{r}-\mathbf{r}_i)^6} \tag{35}$$

By setting in analogy with eq. (20)

$$\frac{1}{C_i^3} = \frac{3}{4\pi} \int_{\mathrm{Solvent}} d^3\mathbf{r}\, \frac{1}{(\mathbf{r}-\mathbf{r}_i)^6} \tag{36}$$

where C_i is an effective Born radius evaluated by means of the function $1/r^6$ rather than the standard $1/r^4$ function, we obtain

$$U_{\mathrm{vdW}}(i) \simeq -\frac{16\pi\rho_w\varepsilon_{iw}\sigma_{iw}^6}{3C_i^3} \tag{37}$$

For a spherical geometry the $1/r^6$ effective Born radius C_i is equal to the Born radius B_i. Equation (30) is then obtained by approximating C_i in eq. (37) by the Born radius. The Born radius is then further augmented by the radius of a water molecule, R_w, to take into account the fact that, for the purpose of evaluating the van der Waals solute-solvent energy, the distance of closest approach between the atomic centers of atom i and of a water molecule is $R_i + R_w$, where R_i is the van der Waals radius of atom i. R_w here is set to 1.4 Å.

The nonpolar model proposed here has evolved from the nonpolar model proposed by Gallicchio et al.[11] that was shown, together with the SGB polar model,[27] to predict with high accuracy the experimental hydration free energies of a set of organic molecules. The present model differs from the model of Gallicchio et al.[11] with respect to the definition of surface area and with respect to the dependence of $U_{\mathrm{vdW}}(i)$ on the Born radius. In the nonpolar model presented here and in ref. 11 the functional form of $U_{\mathrm{vdW}}(i)$ reflects the fact that the solute-solvent van der Waals interaction energy of atom i decreases with the degree of burial of the atom as measured by the Born radius. However, in the model of Gallicchio et al.[11] the dependence of $U_{\mathrm{vdW}}(i)$ on the Born radius was chosen on an ad hoc basis, whereas the functional form used here is motivated by the physical arguments presented.

Numerical Implementation

This section describes the numerical implementation of the algorithms described in the previous section to compute surface areas and Born radii, which are the core ingredients of the AGBNP model. The AGBNP model has been implemented within the IMPACT[70,71] molecular simulation program.

Volume and Surface Area

The self-volumes of hydrogen atoms are ignored. For the calculation of surface areas the radii of heavy atoms are set as described in the section Atomic Radii for the Surface Area Calculation.

Atomic surface areas are computed using the Gaussian surface model of Grant and Pickup[60] described in the section Volume and Surface Area Model. The derivatives with respect to atomic radii, R_i, in eq. (7) are obtained in analytic form from eqs. (10–13)

$$\frac{\partial V_{12\cdots n}^g}{\partial R_i} = \frac{2\kappa}{R_i^3}\left(\frac{3}{2\Delta_{12\cdots n}} + \left|\mathbf{r}_i - \mathbf{r}_{12\cdots n}^c\right|^2\right) V_{12\cdots n}^g \tag{38}$$

where

$$\mathbf{r}_{12\cdots n}^c = \frac{1}{\Delta_{12\cdots n}} \sum_{j=1}^{n} c_i \mathbf{r}_i \tag{39}$$

is the coalescence center of the n Gaussians.

The search and calculation of intersection volumes is implemented using the "depth-first" algorithm.[72] First a neighbor list for each atom i is constructed containing the atom indexes $j > i$ that satisfy the distance criterion $r_{ij} < R_i + R_j + \delta R$, where δR is a distance offset chosen so that all of the intersections of non-negligible volume in which atom i is participating are calculated. We found that a value of $\delta R = 0.5$ Å gives converged surface areas and stable derivatives of the cavity free energy ΔG_{cav}. For each atom i the search starts with the two-body intersection volume, V_{ij_1}, between atom i and its first neighbor. The contribution of V_{ij_1} to the self-volumes [eq. (3)] and surface areas [eq. (7)] of atoms i and j_1 is calculated. If V_{ij_1} is smaller than a certain small threshold value $V_{\min}$ (here set to 0.01 Å^3), the two-body intersection volume V_{ij_2} between atom i and its second neighbor is examined (the search ends if j_1 is the only neighbor of atom i). If, instead, V_{ij_1} is larger than the threshold value the three-body intersection volume $V_{ij_1j_2}$ between atom i and its first two neighbors is examined. The process described for V_{ij_1} is then repeated for $V_{ij_1j_2}$, leading to either the intersection volume of lower order V_{ij_2} (if j_2 is the last neighbor), the intersection of same order $V_{ij_1j_3}$ (if $V_{ij_1j_2}$ is smaller than $V_{\min}$), or the intersection of higher order $V_{ij_1j_2j_3}$ (if $V_{ij_1j_2}$ is larger than $V_{\min}$). The search then continues in this fashion until all intersection volumes between atom i and its neighbors are examined. The process is then started for the next atom until all of the atoms have been processed. During this process the contributions of each intersection volume $V_{i_1\cdots i_n}$ to the self-volumes and surface areas of atoms $i_1 \cdots i_n$ are collected.

For completely or nearly completely buried atoms the Gaussian overlap approximation sometimes leads to small negative surface areas. The occurrence of negative surface areas is prevented by multiplying the raw surface areas from eq. (7) by the switching function

$$f_a(A_i) = \begin{cases} \dfrac{A_i^2}{a^2 + A_i^2} & A_i > 0 \\ 0 & A_i \le 0 \end{cases} \tag{40}$$

The function $f_a(A_i)$ goes smoothly from 1 for large surface areas to zero for small or negative surface areas. The switch occurs around $A_i = a$. The switching function has negligible effect for exposed atoms with surface areas much larger than a. We set $a = 5$ Å^2.

Appendix A describes the implementation of the calculation of the derivatives of the cavity free energy $\Delta G_{\mathrm{cav}} = \Sigma_i\, \gamma_i f_a(A_i) A_i$

with respect to atomic positions, a necessary step to obtain the forces due to the cavity free energy.

Born Radii

The inverse Born radii β_i are obtained from eqs. (24) and (27). First the self-volumes V'_i are calculated as described in the section Volume and Surface Area Model. The self-volume scaling actors s_i are then obtained by dividing the self-volumes by the atomic volumes V_i [eq. (5)]. The pairwise descreening sum [the sum at the right hand side of eq. (24)] runs over heavy atoms only, that is, hydrogen atoms do not participate in descreening other atoms. The Born radii of hydrogen atoms are obtained from eq. (24) in which s_{ji} is replaced by s_j (the two-body overlap between a heavy atom j and a hydrogen atom i is set to zero).

For each heavy atom pair ji the descreening function Q_{ji} is evaluated (see Appendix B), and s_{ji} is then computed from eq. (27) by adding $V^g_{ij}/2V_j$ to s_j. The quantity $s_{ji}Q_{ji}/4\pi$ is then subtracted from the current value of β_i starting from the isolated initial value $1/R_i$. When using nonbonded distance cutoffs, interacting atom pairs are stored in a neighbor list that lists atom pairs ji with $j > i$. In this case for each heavy atom pair ji both $s_{ji}Q_{ji}/4\pi$ and $s_{ij}Q_{ij}/4\pi$ are evaluated and subtracted from β_i and β_j, respectively. For pairs in which j is a heavy atom and i a hydrogen atom only the quantity $s_jQ_{ji}/4\pi$ is evaluated and subtracted from the inverse Born radius, β_i, of the hydrogen atom. For maximum efficiency, the two-body overlap volumes V^g_{ij}, used to calculate s_{ji} for each ji interacting atom pair are retrieved from a memory cache filled during the computation of the self-volumes, rather than recalculated. The amount of memory required for storing the two-body volumes cache scales as $N_H \times N_b$ where N_H is the number of heavy atoms and N_b the average number of overlapping neighbors (as defined in the last section) for each atom. Because N_b is a small number (10–20 typically) memory consumption for the two-body volume cache is normally not problematic.

In some instances, such as when the input conformation is from an unrefined X-ray structure with unphysical atomic overlaps, the pairwise descreening scheme produces unreasonably large or, worse, negative Born radii. The occurrence of unreasonably large or negative Born radii is prevented by filtering the inverse Born radii β_i, given by the pairwise descreening formula [eq. (24)], by the function

$$B_i^{-1} = f_b(\beta_i) = \begin{cases} \sqrt{b^2 + \beta_i^2} & \beta_i > 0 \\ b & \beta_i \le 0 \end{cases} \quad (41)$$

where $b^{-1} = 50$ Å. The filter function eq. (41) is designed to prevent the occurrence of negative Born radii or Born radii larger than 50 Å. The goal of the filter function is simply to increase the robustness of the algorithm in limiting cases. The filter function has negligible effect for the most commonly observed Born radii smaller than 20 Å.

Appendixes C and D describe the calculation of the derivatives with respect to atomic positions of the generalized Born polar hydration free energy term ΔG_{GB} [eqs. (18) and (19)] and of the van der Waals component ΔG_{vdW} [eq. (30)] of the nonpolar hydration free energy. The derivatives of both free energy terms are complicated by the dependence of the Born radii on the solute conformation.

Selection of Parameters

The parameters of the AGBNP model are the atomic radii used in the surface area calculation, the atomic radii that define the solute volume for the purpose of calculating Born radii, and the surface tension and van der Waals parameters γ_i and α_i in eqs. (29) and (30). The atomic radii for the calculation of the Born radii are generally set from the corresponding OPLS Lennard-Jones σ parameters adjusted in order to better reproduce the Free Energy Perturbation (FEP) explicit solvent electrostatic charging free energies for a set of small molecules.[28] The atomic radii used in the surface area calculation are obtained by increasing by 0.5 Å the atomic radii for the Born radii calculation (see the section Atomic Radii for the Surface Area). The surface tension parameters are set to 117 cal/mol/Å^2, obtained from fitting eq. (29) to the hydration free energies of alkane cavities[49] using the present definition of the solute surface area. This value is intermediate between the values of the surface tension of alkane cavities measured using the molecular surface area (γ = 139 cal/mol/Å^2)[73] or the accessible surface area (γ = 72 cal/mol/Å^2).[49] The values of the α parameters have been set so as to reproduce as best as possible, using eq. (30), the solute-solvent van der Waals energies of individual atoms of a large set of protein conformations, peptides, and small molecules obtained from a continuum solvent model[50] designed to reproduce the results of explicit solvent simulations. Table 1 lists the atomic radii and nonpolar parameters of the model.

Atomic Radii for the Born Radii Calculation

The atomic radii used in the calculation of the Born radii are generally set from the Lennard-Jones OPLS σ parameters:

$$R_i = \frac{\sigma_i}{2} \quad (42)$$

The atomic radii of some atomic types were adjusted to better reproduce the charging free energies of a set of small molecules calculated in explicit solvent using the FEP method.[28] Table 1 lists the OPLS and atomic radii for the atom types found in proteins and the small molecule database. When using van der Waals radii to describe the solute volume, small crevices between the atomic spheres not occupied by water molecules are effectively considered as high-dielectric solvent regions.[12,26] This leads to underestimation of the Born radii, particularly for buried atoms. We are currently exploring methods to overcome this limitation.

Figure 4 illustrates the accuracy of the inverse Born radii, $1/B_i$, obtained by the AGB pairwise descreening method described in the section GB Model [eq. (24)] by comparing them to accurate estimates obtained by evaluating the integral in eq. (21) numerically on a grid (see Appendix E for details of the grid integration method). The inverse Born radii are proportional to the self-energies [see eq. (19)]. Figure 4a compares the analytical and numerical inverse Born radii of a set of small molecules and a diverse set of conformations of peptides and proteins. Figure 4b compares the differences, $\Delta B_i^{-1} = B_i^{-1}$ (Unbound) $- B_i^{-1}$ (Bound), of the inverse Born radii between the unbound and bound forms of a set of ligands. The agreement between the analytical

Table 1. List of the van der Waals Radii, R, Used in the Computation of the Born Radii, and the Nonpolar Parameters for a Series of Atom Types.

Atom type	OPLS $\sigma/2$[a]	R[a]	γ[b]	α
sp2 Carbon (carbonyl)	1.875	1.875	117	0.80
sp2 Carbon (alkene)	1.775	1.775	117	0.80
sp2 Carbon (aromatic)	1.775	1.650	117	0.80
sp3 Carbon	1.750	1.650	117	0.70
sp Carbon (alkyne)	1.750	1.750	117	0.75
sp Carbon (nitrile)	1.650	1.750	117	1.15
Aliphatic hydrogen	1.250	1.250	0[c]	0.80
Aromatic hydrogen	1.210	1.250	0[c]	0.80
sp2 Nitrogen	1.625	1.525	117	0.75
sp3 Nitrogen	1.650	1.650	117	0.75
Nitrogen (nitro group)	1.625	1.625	117	0.75
Oxygen (nitro group)	1.480	1.480	117	0.75
sp Nitrogen	1.600	1.700	117	0.75
sp3 Oxygen (alcohol)	1.560	1.480	117	0.70
sp3 Oxygen (ether)	1.500	1.500	117	0.80
sp2 Oxygen (carbonyl)	1.480	1.480	117	0.80
Phosphorus (phosphate)	1.870	1.870	117	0.85
Sulfur	1.800	1.800	117	0.75
sp3 Nitrogen (ammonium)	1.625	1.625	117	0.75
sp2 Oxygen (carboxyl)	1.480	1.480	117	0.80
Hydrogen on heteroatoms	1.210	1.210	0[c]	0[d]

The nonpolar parameters for atom types not listed are set as $\gamma = 117$ cal/mol/Å^2 and $\alpha = 1$. The van der Waals radii of the atom types not listed are set to $\sigma/2$ where σ is the OPLS Lennard-Jones σ parameter. The values of the atomic radii used for the surface area calculation are those listed plus 0.5 Å, except for the radii of hydrogen atoms that are set to zero.

[a]In Å.

[b]In cal/mol/Å^2.

[c]Hydrogen atoms do not contribute to the cavity hydration free energy.

[d]Hydrogen atoms bonded to heteroatoms do not contribute to the van der Waals hydration free energy because their OPLS Lennard-Jones ε parameter is zero.

and numerical inverse Born radii is excellent. The comparison of the inverse Born radii differences shown in Figure 4b is a *very stringent* test of the method; it shows that the analytical Born radii respond correctly to conformational changes such as those occurring in binding/unbinding transitions. It should be stressed that the agreement between the numerical and analytical Born radii shown in Figure 4 has been obtained without resorting to parameter fitting. No adjustable parameters are used in the analytical formulation used to calculate the Born radii. All the parameters in the pairwise descreening function [eq. (24)] are calculated directly from the solute geometry rather than by fitting parameters to reflect the average geometric properties of particular functional types. It is therefore expected that the same level of accuracy as that shown in Figure 4 can be achieved for molecules with any combination of functional groups.

Atomic Radii for the Surface Area Calculation

The Gaussian-based model of Grant and Pickup[60] is designed for van der Waals surface areas. van der Waals surface areas, however, are very insensitive to molecular geometry and consequently are not useful for modeling the nonpolar solvation free energy. We have observed for example only minor changes of the van der Waals surface areas for folding proteins and for binding ligands to protein receptors. This is due to the fact that the distance between two adjacent nonbonded atoms tends to be near the minimum of the Lennard-Jones interaction potential. At that distance the van der Waals surfaces of the two atoms are not touching.

Although solvent accessible surface areas (SASAs), obtained by setting the atomic radii to the van der Waals radii augmented by the radius of a water molecule (typically 1.4 Å), are optimally suited for nonpolar free energy estimators, their calculation is problematic in the context of the Gaussian model of Grant and Pickup. This is due to the fact that larger atomic radii produce many more intersections and cause large atomic overlaps that are beyond the range of applicability of the Gaussian overlap approximation to the volume of spherical intersections. Weiser et al.[72] have reported that a parameterized model based on the method of Grant and Pickup can be used to calculate accessible surface areas, albeit at a computational cost about 100 times larger than calculating the van der Waals surface areas. We found, however, that the surface area obtained by augmenting the van der Waals radii by only 0.5 Å, a fraction of the radius of a water molecule, offers a good compromise between computational complexity and sensitivity to molecular conformation. We are currently exploring the possibility of extending the method of Grant and Pickup[60] for the calculation of molecular surface areas.

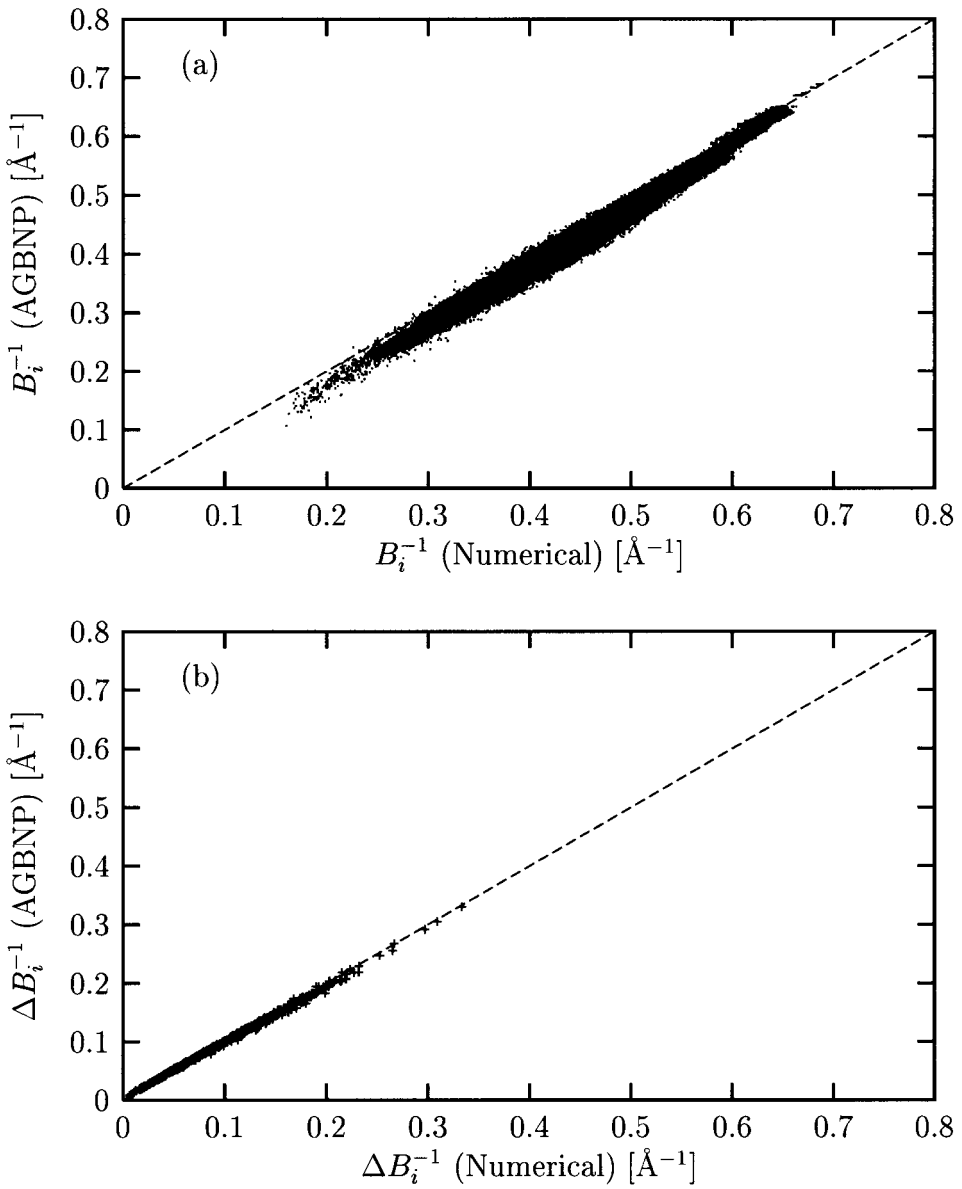


Figure 4. (a) Comparison between the numerical and AGBNP estimates of the inverse Born radii for the atoms of a set of 200 small molecules,[11] seven low-energy conformations of the Ace-GEWTYDDAT-KTFTVTE-Nme octadecapeptide, and a set of 30 native and misfolded protein structures from the 1lz1 (116 residues), 1ctf (68 residues), and 2cro (65 residues) decoy sets[78,79] (43,135 data points). (b) Comparison between the numerical and AGBNP estimates of the inverse Born radius changes upon unbinding of the atoms of the ligands from the following ligand-protein complexes (PDB id's): 1bkm, 2clr, 1dwc, 1aq7, a HEPT analog (H01) complexed with HIV reverse transcriptase (1rt1) from ref. 80, 2bpx, 1hpv, 1htg, and 1hvj.

In our implementation of the Gaussian solute volume and atomic surface area model,[60] the atomic radii are those listed in Table 1 augmented by 0.5 Å; hydrogen atoms are assigned zero radii. Figure 5a illustrates the accuracy of the surface areas obtained using the Gaussian model for a set of organic molecules[11] and proteins by comparing them to exact numerical surface areas.[74] The agreement between analytical and numerical surface areas is very good. As shown in Figure 5b, the agreement between the analytical and numerical estimates of the changes upon binding of the atomic surface areas of several ligands is also very good. This shows that the analytical surface area model is able to reproduce equally well absolute atomic surface areas and variations of surface areas due to conformational changes.

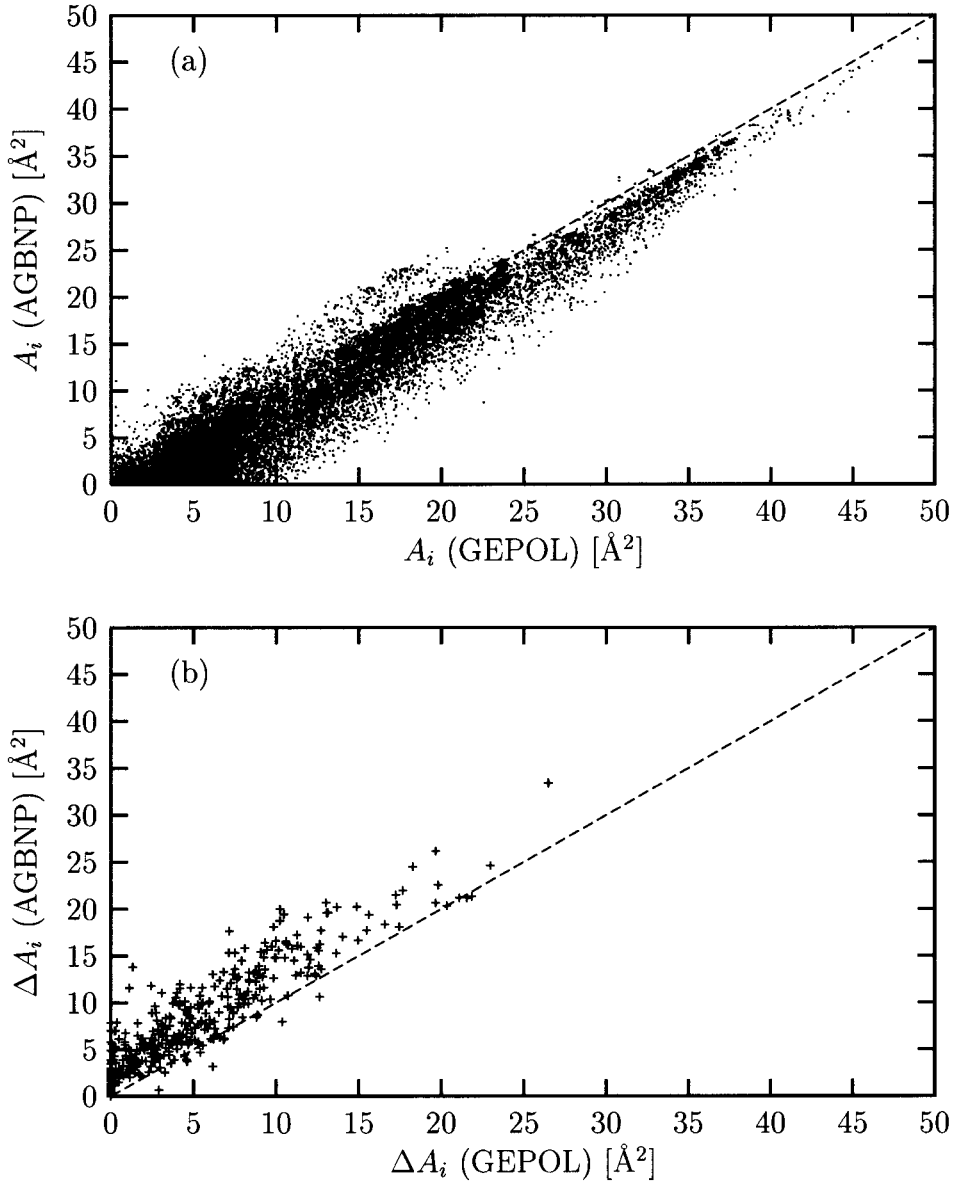


Figure 5. (a) Comparison between the numerical (GEPOL) and AGBNP estimates of the accessible surface areas (R_w = 0.5 Å) for the heavy atoms of a set of 200 small molecules,[11] seven low-energy conformations of the Ace-GEWTYDDATKTFTVTE-Nme octadecapeptide, and a set of 30 native and misfolded protein structures from the 1lz1 (116 residues), 1ctf (68 residues), and 2cro (65 residues) decoy sets[78,79] (20,115 data points). (b) Comparison between the numerical (GEPOL) and AGBNP estimates of the accessible surface area changes (R_w = 0.5 Å) of unbinding of the heavy atoms of the ligands from the following ligand-protein complexes (PDB id's): 1bkm, 2clr, 1dwc, 1aq7, a HEPT analog (H01) complexed with HIV reverse transcriptase (1rt1) from ref. 80, 2bpx, 1hpv, 1htg, and 1hvj.

Performance

Table 2 shows the results of a performance test on representative systems using the AGBNP method described in the previous sections. The test consists of 50 steps of steepest descent energy minimization, followed by 250 steps of constant temperature molecular dynamics, and 250 steps of constant energy molecular dynamics with a 1 fs time step. The same test was performed in vacuum and in implicit solvent with AGBNP. AGBNP is found to be from 3.84 to 5 times slower than the corresponding calculation in vacuum. Most of the extra computation time is due to the calculation of Born radii, pair GB interaction energies, and, to a lesser extent, atomic surface areas. The speed of AGBNP relative to vacuum improves as the system size increases due to the fact

Table 2. Performance Test Results Using the AGBNP Model.

System	*N*	Cutoff[b]	Time[c]	Cache size[d]	$\log(\overline{\Delta E})$[e]	*R*[e]
1gb1(41–56)[a]	256		5.00 × Vac.	0.3 Mb	−2.50	0.046
1gpt	711		4.88 × Vac.	2.3 Mb	−2.98	0.006
1lz1	1803		4.73 × Vac.	13.2 Mb	−2.25	0.043
1lz1	1803	13.0	4.66 × Vac.	6.5 Mb	−3.01(−1.83)	0.074(1.080)
2clr	5983	13.0	3.84 × Vac.	26.9 Mb	−3.51(−2.20)	0.029(0.402)

The test consists of 50 steps of steepest descent energy minimization, followed by 250 steps of constant temperature molecular dynamics, and 250 steps of constant energy molecular dynamics with a 1 fs time step. *N* is the number of atoms and "Cutoff" is the residue-based nonbonded cutoff radius (when not indicated all nonbonded interactions are calculated). "Cache size" is the size of the memory cache used to store intermediate values for the calculation of the gradients (see Appendix C). The parameters $\log(\overline{\Delta E})$ and *R* measure energy conservation[75] during the constant energy section of the trajectory. They are defined as $\log(\overline{\Delta E}) = (1/M) \sum_k |(E_k - E_0)/E_0|$, where *M* is the number of MD steps and E_k is the total energy at step *k*, and $R^2 = \langle \delta^2 E \rangle / \langle \delta^2 KE \rangle$ is the ratio of the variance of the total energy and the variance of the kinetic energy. Energy is considered as being conserved to an acceptable degree when $\log(\overline{\Delta E})$ is around −2.5 or less and *R* is around 0.05 or less.

[a]In parentheses is the chain segment included in the calculation.

[b]Nonbonded cutoff distance in Å if applicable.

[c]CPU time as a multiple of the CPU time for the same system in vacuum.

[d]Mb = 10^6 bytes.

[e]The corresponding value in vacuum is indicated in parentheses.

that the cost of calculating surface areas decreases relative to the cost of calculating nonbonded interaction energies.

When using a nonbonded distance cutoff the memory requirements remain modest even for the largest system tested (26.9 Mb for 2clr, 5983 atoms). The quality of energy conservation [as measured by the $\log(\overline{\Delta E})$ and *R* indicators[75]] is comparable to vacuum except when a nonbonded distance cutoff is used, in which case energy conservation with AGBNP is found to be consistently superior compared to vacuum. This is due to effective charge-charge interactions at the cutoff distance, which are still significant in vacuum but become negligibly small in implicit solvent due to dielectric screening. This suggests that the application of a nonbonded distance cutoff, although questionable in vacuum, is appropriate in combination with GB.[32] The combined results for speed and accuracy, as reported in Table 2 and Figure 4, of AGBNP are among the best of the GB implicit solvent models to date.[7,12,27,31,52]

Molecular Dynamics of Proteins

To test the stability of native protein conformations with AGBNP, we have conducted 2 ns constant temperature (T = 300 K) molecular dynamics simulations on three small proteins: the *C*-terminal domain of the L7/L12 50 S ribosomal protein (1ctf), γ-1-H thionin (1gpt), and 434 Cro protein (2cro). Starting from their experimental conformation, we have monitored (see Figure 6) the root mean square deviation (RMSD) from the experimental structure as a function of time. In all these cases the native structure remains stable; the measured RMSD deviations are comparable to the experimental resolution. Visual inspection of the final structure reveals that all native secondary and tertiary structures are preserved after 2 ns of molecular dynamics. Molecular dynamics trajectories obtained in vacuum show instead rapid divergence from the native protein conformations.

Protein Loops and Protein-Ligand Complex Decoys

Protein decoys[5] have been extensively used to test scoring functions aimed at protein folding and protein homology modeling. To demonstrate the applicability of the AGBNP model to high-resolution structure prediction applications, we tested the OPLS/AGBNP model on high quality decoy sets of protein loops and protein-ligand binding complexes. Specifically, we have tested the ability of the OPLS/AGBNP model to recognize the native conformation among a set of non-native conformations (decoys). Decoy conformations are scored based on their OPLS/AGBNP energy, and the quality of the energy model is tested by verifying that the energy of the native conformation is one of the lowest in the decoy set and that the decoy of lowest energy is structurally similar to the native.

The protein loop decoy sets we have examined have been selected from a large database of loop decoy sets generated by Jacobson et al.[76] Jacobson et al. have generated the loop decoy database using the PLOP program; PLOP generates loop conformations by exhaustive grid enumeration of the loop backbone torsional angles. Loop conformations with steric clashes are rejected. Loop sidechain conformational predictions have also been performed using the PLOP program.[76] We have extracted from the database of Jacobson et al. the decoy sets corresponding to loops between 7 and 12 residues in length in single-chain monomeric proteins of 150 residues or less, free of large prosthetic groups. This selection resulted in the 60 loop decoy sets listed in Table 3, for a total of approximately 43,000 loop conformations. These loop

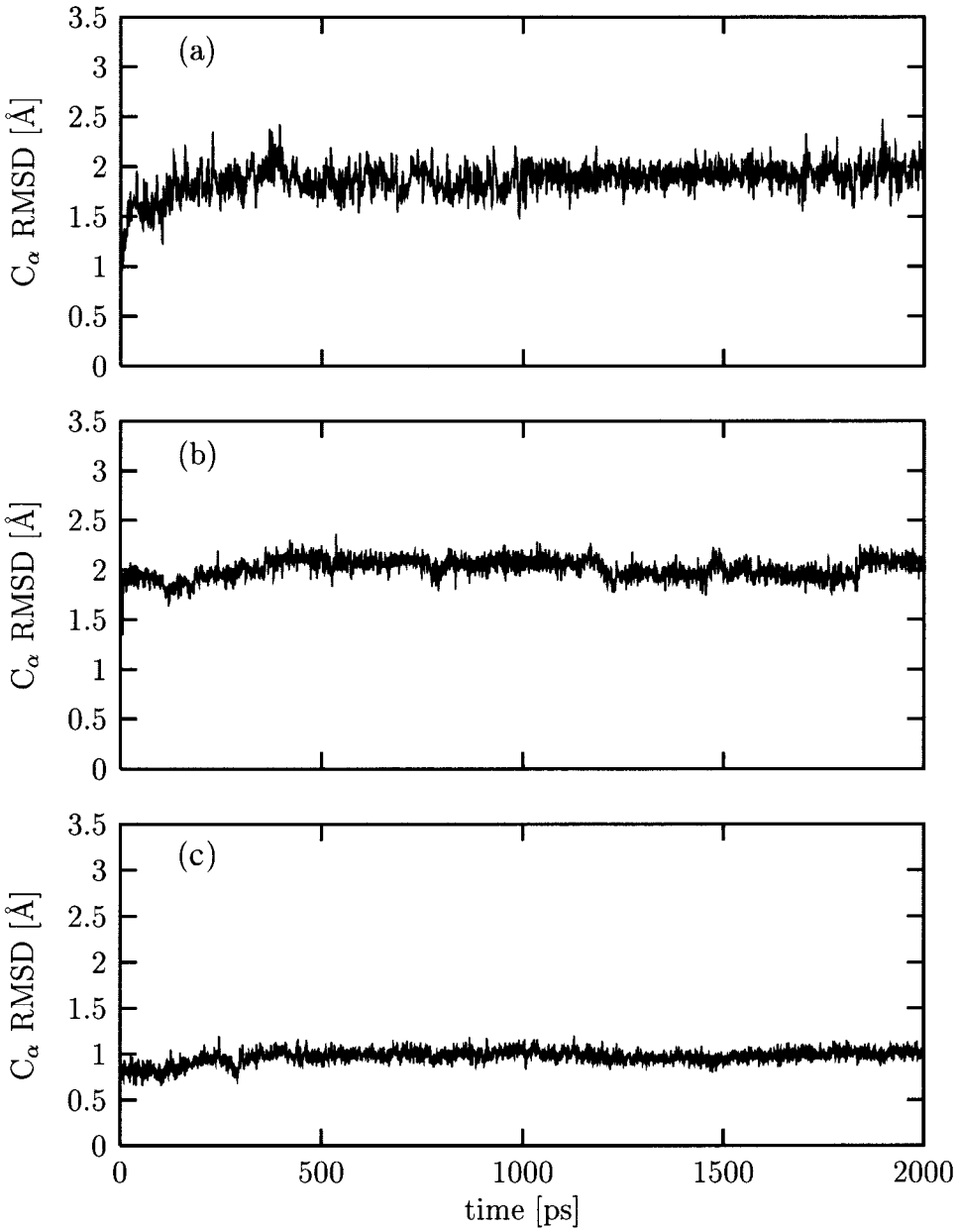


Figure 6. Root mean square deviations (RMSD) from the native conformation of the C_α atoms of the proteins (a) 1ctf, (b) 1gpt, (c) 2cro, during constant temperature molecular dynamics trajectories using the AGBNP implicit solvent model.

decoy sets contain a large variety of high quality decoys, including low energy conformations structurally very different from the native as well as near-native conformations. Each decoy loop conformation was energy minimized using the OPLS/AGBNP energy function and the energy of each loop conformation recorded.

In Table 3 we report the RMSD from the native loop conformation of the decoy loop conformation that scores the best in terms of the OPLS/AGBNP energy function (lowest energy decoy, LED). The OPLS/AGBNP energy function is very successful in discriminating near-native conformations from non-native ones. As shown in Figure 7, in over 90% of the cases the RMSD from the native of the lowest energy decoy is less than 2 Å. In these cases we find that the loop backbone conformation and the side-chain conformations are indeed in good agreement with the native. In five cases the C_α RMSD from the native of the best energy decoy is larger than 2 Å. In one case [1rro(90–96)] the deviation of the predicted conformation from the native is probably caused by the presence of calcium ions in the crystal structure, which are not included in the model; in two more cases [1aba(56–66) and 5fx2(59–68)] the loop in the crystal structure is in close proximity to a ligand not included in the model.

Table 3. Results of Scoring the Protein Loop Decoy Sets Using the OPLS/AGBNP Force Field.

Loop set[a]	N_{res}	N_{decoys}	$RMSD_{max}$[b]	$RMSD_{LED}$[b,c]	Loop set[a]	N_{res}	N_{decoys}	$RMSD_{max}$[b]	$RMSD_{LED}$[b,c]
1a62(89–95)	7	519	6.20	0.48	1aac(58–66)	9	650	8.60	0.67
1aac(69–75)	7	514	1.91	0.17	1aba(36–44)	9	644	7.43	0.38
1g3p(127–133)	7	510	2.97	1.51	1aba(69–77)	9	628	10.58	0.40
1lif(64–70)	7	418	3.65	0.15	1lif(73–81)	9	646	9.46	0.45
1lit(68–74)	7	514	4.53	1.29	1noa(9–17)	9	651	4.34	0.30
1opd(53–59)	7	517	3.77	0.62	1noa(99–107)	9	646	5.87	2.02
1paz(36–42)	7	530	2.03	0.07	1noa(76–84)	9	642	10.46	0.50
1plc(87–93)	7	326	3.90	0.13	1onc(70–78)	9	566	4.67	0.39
1pmy(25–31)	7	485	4.63	0.14	1ptf(10–18)	9	630	6.30	0.08
1ptf(65–71)	7	512	6.91	0.11	1rro(70–78)	9	648	4.14	0.23
1rro(90–96)	7	508	3.50	2.37	3chy(57–65)	9	269	7.49	0.00
1vcc(34–40)	7	517	3.41	0.18	5fx2(8–16)	9	441	6.96	1.74
2a0b(737–743)	7	537	4.24	0.23					
2sns(134–140)	7	321	9.03	0.32	1knt(35–44)	10	967	4.72	0.25
4fgf(58–64)	7	492	2.30	0.31	1plc(42–51)	10	924	6.98	0.00
5fx2(27–33)	7	501	7.24	0.19	1rie(137–146)	10	500	5.70	0.05
7rsa(20–26)	7	493	5.32	0.19	2phy(58–67)	10	926	7.36	0.13
					3lzt(15–24)	10	683	8.18	0.19
135l(84–91)	8	574	6.00	0.16	5fx2(59–68)	10	904	7.63	2.54
1aac(48–55)	8	611	1.99	0.43	5pti(23–32)	10	502	4.39	0.55
1aba(7–14)	8	572	5.88	0.00	7rsa(33–42)	10	925	5.11	0.19
1alc(34–41)	8	407	5.82	0.16	7rsa(87–96)	10	925	8.17	0.25
1cbn(18–25)	8	524	3.48	0.00					
1cbs(55–62)	8	573	6.30	0.39	1aba(56–66)	11	1477	8.79	7.09
1lit(82–89)	8	579	3.70	0.22	1msi(8–18)	11	967	14.10	0.16
1msi(26–33)	8	570	2.84	0.13	1rie(174–184)	11	1060	4.90	0.00
1opd(8–15)	8	575	4.12	0.73	5pti(7–17)	11	1382	3.78	0.09
1plc(6–13)	8	572	7.01	2.79					
1plc(32–39)	8	580	2.51	0.19	7rsa(13–24)	12	1152	9.31	0.17
1poa(71–78)	8	574	4.93	1.24	5nul(54–65)	12	1102	10.12	0.88
1rro(18–25)	8	581	6.85	0.20					
2sns(17–24)	8	408	3.42	1.21					
5icb(15–22)	8	576	2.81	1.78					
7rsa(64–71)	8	577	3.93	0.17					

For each set we report the number of residues in the loop, N_{res}, the number of decoy conformations in the set, N_{decoys}, the largest C_α root mean square deviation (RMSD) from the native conformation in the set, $RMSD_{max}$, and the C_α RMSD from the native of the loop conformation with lowest OPLS/AGBNP energy, $RMSD_{LED}$. The RMSD is calculated for the loop C_α atoms after superimposing the protein frames.

[a]PDB id of protein and loop residues indicated in parentheses.

[b]In Å.

[c]RMSD from the native of the lowest energy decoy (LED).

The protein-ligand decoy complexes, listed in Table 4, were constructed starting from their X-ray experimental structures. The program GLIDE[71] was then used to dock several low energy conformations of the ligands on various candidate binding sites on their corresponding receptor protein, including the native binding site. The complexes were then energy minimized using the OPLS/AGBNP energy function, and the binding energy of each complex, defined as the difference between the energy of the complex and the sum of the energies of the receptor and free ligand, was computed. The results were analyzed in terms of the RMSD from the native ligand coordinates of the ligand in the protein-ligand complex with lowest predicted binding energy (LED), and the rank of the native complex with respect to the OPLS/AGBNP binding energy. The results are shown in Table 4. The OPLS/AGBNP energy is very successful in both predicting the native complex conformation and in giving the native complex a high rank compared to the decoy complexes. In all cases the complex with lowest predicted binding energy is structurally very similar to the native, and the native complexes are consistently ranked among the best 5%. This indicates that the OPLS/AGBNP energy function is very successful in determining the location of the native receptor site and the ligand binding pose among a large number of alternative plausible candidates.

Conclusions

We have developed an implicit solvent effective potential (AGBNP) that is suitable for molecular dynamics simulations and

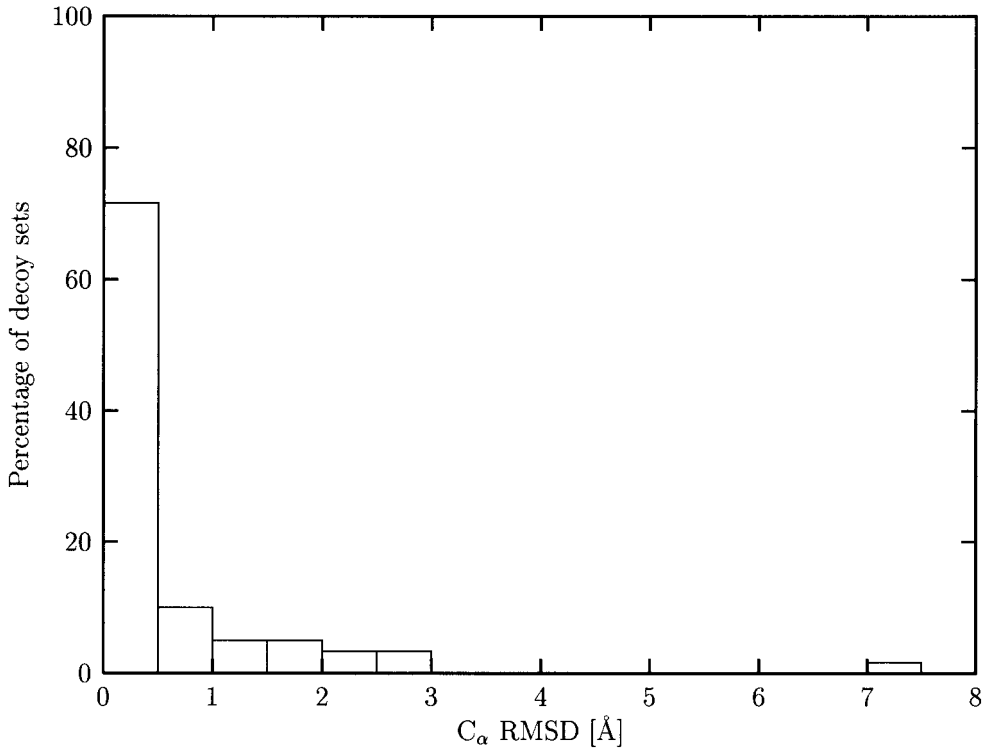


Figure 7. Distribution of the C_α atoms' RMSD from the native of the lowest OPLS/AGBNP energy decoy ($RMSD_{LED}$ from Table 3).

high-resolution modeling. It is based on a novel implementation of the pairwise descreening GB model for the electrostatic component and a new nonpolar hydration free energy estimator. AGBNP has been designed with several goals in mind. The model has been designed to be applicable to a wide range of molecules, from small molecules to large biological macromolecules, and to a wide range of functional group topologies and types (hydrophobic, polar, and ionic). The model has also been designed to be applicable to study absolute hydration free energies as well as conformational equilibria. Both large conformational changes and small conformational rearrangements can be modeled. AGBNP is fully analytical with first derivatives and is computationally efficient, which facilitates incorporation in molecular mechanics simulation packages. The ability to use AGBNP in conjunction with advanced sampling techniques can extend further the predictive power of the model.

These properties make the AGBNP model particularly suitable for studying ligand binding and for high-resolution protein modeling. Ligand binding studies often involve the study of large numbers of ligand candidates with a large variety of topologies and combinations of functional groups. The correct representation of the geometrical parameters of these ligands, which ultimately determine their thermodynamic properties, is therefore crucial.

Table 4. Results of Scoring the Protein-Ligand Complex Sets Using the OPLS/AGBNP Force Field.

Complex set	n_{at}	N_{decoys}	$RMSD_{max}$[a]	$RMSD_{LED}$[a,b]	Rank of native
1aha	15	350	30.39	0.09	6
1bkm	80	340	30.17	2.37	14
1dwc	84	400	36.37	1.01	9
1odw	71	340	30.27	1.62	3

For each set we report the number of atoms in the ligand, n_{at}, the number of decoy conformations in the set, N_{decoys}, the largest root mean square deviation (RMSD) of the ligand from the native conformation in the set, $RMSD_{max}$, the RMSD from the native of the ligand conformation and lowest OPLS/AGBNP binding energy, $RMSD_{LED}$, and the rank of the native conformation with respect to the OPLS/AGBNP binding energy.

[a]In Å.

[b]RMSD from the native of the lowest energy decoy (LED).

The sensitivity of the AGBNP model to conformational changes makes it well suited for homology modeling of protein loops and the prediction of protein-ligand binding modes.

Two key developments set the AGBNP model apart from related implicit solvent models based on the pairwise descreening GB model. The first is the novel parameter-free and conformational-dependent algorithm used to estimate the pairwise descreening scaling coefficients in the evaluation of Born radii. The same algorithm is also used to evaluate atomic surface areas. We show that these methods provide good approximations to the Born radii and atomic surface areas calculated using accurate, but much more expensive, numerical techniques. We have shown that the model is able to reproduce accurately the effect of conformational changes on Born radii and surface areas. This is achieved without the use of adjustable parameters, making the model more easily transferable to a large variety of molecules without the need of extra parameterization steps.

The second key development is the introduction of a nonpolar estimator that does not depend exclusively on the solute surface area. This development has been motivated by studies[11,49,50,53,54,56] that show that surface area models, although valid on a coarse-grained energy scale, do not properly capture nonpolar hydration free energy differences on the fine-grained energy scale necessary for high-resolution modeling. We have developed a nonpolar hydration free energy estimator inspired by the decomposition of the nonpolar hydration free energy into a cavity term, proportional to surface area, and an attractive dispersion energy term, which reproduces the continuum solvent solute-solvent van der Waals interaction energy using a functional form based on the Born radius of each atom. The nonpolar model depends linearly on adjustable parameters that measure the effective surface tension and effective strength of solute-solvent van der Waals interactions. In this article we have set these parameters to reference values from literature sources.

We show that the native conformations of a series of small proteins are stable during long molecular dynamics trajectories conducted using the AGBNP implicit solvent model. The applicability of the AGBNP implicit solvent model to high resolution protein modeling and binding is demonstrated by showing that the AGBNP implicit solvent model in conjunction with the OPLS force field is able to discriminate the native conformation among a large number of non-native conformations of protein loops and protein-ligand complexes. In a following article[77] we analyze the thermodynamic stability of β-hairpin and α-helical peptides with the AGBNP effective potential.

Acknowledgments

We thank Linda Yu Zhang for providing the protein-ligand complex decoys, Matthew Jacobson for providing the database of protein loop decoys, and Anthony Felts for helpful discussions.

Appendix A: Calculation of the Gradient of ΔG_{cav}

Taking into account the switching function in eq. (40), the cavity free energy is given by $\Delta G_{\mathrm{cav}} = \sum_j \gamma_j f_a(A_j) A_j$, where the surface areas, A_j, of heavy atoms are given by eq. (7). The surface areas of hydrogen atoms are set to zero. The derivative of ΔG_{cav} with respect to the position $\mathbf{r}_i$ of a nonhydrogen atom is

$$\frac{\partial \Delta G_{\mathrm{cav}}}{\partial \mathbf{r}_i} = \sum_j \gamma_j [f_a(A_j) + A_j f'_a(A_j)] \frac{\partial A_j}{\partial \mathbf{r}_i} \tag{43}$$

where $f'_a(A_j)$ is the first derivative of the switching function. By defining

$$\tilde{\gamma}_j = \gamma_j [f_a(A_j) + A_j f'_a(A_j)] \tag{44}$$

eq. (43) is written in a form formally equivalent to what would be obtained in the absence of the switching function

$$\frac{\partial \Delta G_{\mathrm{cav}}}{\partial \mathbf{r}_i} = \sum_j \tilde{\gamma}_j \frac{\partial A_j}{\partial \mathbf{r}_i} \tag{45}$$

By inserting eq. (7) in eq. (45) we obtain

$$\frac{\partial \Delta G_{\mathrm{cav}}}{\partial \mathbf{r}_i} = -\left(\frac{\partial \Delta G_{\mathrm{cav}}}{\partial \mathbf{r}_i}\right)_{2-\mathrm{body}} + \left(\frac{\partial \Delta G_{\mathrm{cav}}}{\partial \mathbf{r}_i}\right)_{3-\mathrm{body}} - \cdots \tag{46}$$

where

$$\left(\frac{\partial \Delta G_{\mathrm{cav}}}{\partial \mathbf{r}_i}\right)_{2-\mathrm{body}} = \sum_j \left(\tilde{\gamma}_i \frac{\partial}{\partial R_i} + \tilde{\gamma}_j \frac{\partial}{\partial R_j}\right) \frac{\partial V^g_{ij}}{\partial \mathbf{r}_i} \tag{47}$$

$$\left(\frac{\partial \Delta G_{\mathrm{cav}}}{\partial \mathbf{r}_i}\right)_{3-\mathrm{body}} = \sum_{j<k} \left(\tilde{\gamma}_i \frac{\partial}{\partial R_i} + \tilde{\gamma}_j \frac{\partial}{\partial R_j} + \tilde{\gamma}_k \frac{\partial}{\partial R_k}\right) \frac{\partial V^g_{ijk}}{\partial \mathbf{r}_i} \tag{48}$$

and similarly for intersections of higher order.

The derivative of the Gaussian overlap volume, $V^g_{12\cdots n}$, of n atoms with respect to the position, $\mathbf{r}_i$, of one of the atoms is, from eq. (10)

$$\frac{\partial V^g_{12\cdots n}}{\partial \mathbf{r}_i} = -2c_i(\mathbf{r}_i - \mathbf{r}^c_{12\cdots n}) V^g_{12\cdots n} \tag{49}$$

where c_i is the Gaussian exponent coefficient of atom i and the coalescence center $\mathbf{r}^c_{12\cdots n}$ is defined by eq. (39).

The second derivatives of the Gaussian overlap volumes that appear in eqs. (47) and (48) have the form

$$\frac{\partial}{\partial R_j}\frac{\partial V^g_{12\cdots n}}{\partial \mathbf{r}_i} = \frac{2\kappa}{R_j^3}\left\{\left[2\left(\delta_{ij} - \frac{c_i}{\Delta_{12\cdots n}}\right)(\mathbf{r}_j - \mathbf{r}^c_{12\cdots n})\right] V^g_{12\cdots n} + \left[\frac{3}{2\Delta_{12\cdots n}} + |\mathbf{r}_j - \mathbf{r}^c_{12\cdots n}|^2\right] \frac{\partial V^g_{12\cdots n}}{\partial \mathbf{r}_i}\right\} \tag{50}$$

where $\partial V^g_{12\cdots n}/\partial \mathbf{r}_i$ is given by eq. (49).

It should be noted that, because the modified surface tension parameters $\tilde{\gamma}_i$ depend on the surface area [eq. (44)], the derivatives of the cavity free energy can be collected only after the atomic

surface areas are known. It is therefore required to loop twice over the intersection volumes, as described in the section Volume and Surface Area, once to collect surface areas and again to collect the gradients of the cavity free energy.

Appendix B: Derivation of the Pair Descreening Function

In reference to Figure 2, the cosine of the angle σ_{ji} subtended by the portion of the surface of atom i inside atom j is given by

$$\cos \sigma_{ji} = \frac{R_i^2 - R_j^2 + r_{ij}^2}{2 r_{ij} R_i} \tag{51}$$

where R_i and R_j are the radii of atoms i and j, respectively, and r_{ij} is the distance between the two atoms.

To calculate the integral Q_{ji} of the $1/r^4$ function centered on atom i over the crescent moon-shaped portion of atom j not overlapping with atom i [eq. (23)], four possible cases are identified[29]:

1. $r_{ij}^2 \geq (R_i + Rj)^2$
In this case the spheres of atoms i and j do not overlap and we express Q_{ji} in spherical polar coordinates

$$Q_{ji} = 2\pi \int_{r_{ij}-R_j}^{r_{ij}+R_j} dr r^2 \frac{1}{r^4} \int_{\cos \sigma_{ji}(r)}^{1} d(\cos \theta) \tag{52}$$

where $\cos \sigma_{ji}(r)$ is given by eq. (51) with R_i replaced by r

$$\cos \sigma_{ji}(r) = \frac{r^2 - R_j^2 + r_{ij}^2}{2 r_{ij} r} \tag{53}$$

After performing the integral in eq. (52) we obtain

$$Q_{ji} = 2\pi \left[\frac{R_j}{r_{ij}^2 - R_j^2} - \frac{1}{2 r_{ij}} \log \frac{r_{ij} + R_j}{r_{ij} - R_j} \right] \tag{54}$$

2. $(R_i + R_j)^2 > r_{ij}^2 \geq (R_i - R_j)^2$
In this case the two spheres overlap but neither one is totally engulfed inside the other. Q_{ji} is given by eq. (52) in which the lower limit of integration for the radial integral is R_i rather than $r_{ij} - R_j$. The result of the integration is

$$Q_{ji} = 2\pi \left[\left(\frac{1}{R_i} - \frac{1}{r_{ij} + R_j} \right) - \frac{r_{ij}^2 - R_j^2}{4 r_{ij}} \left(\frac{1}{R_i^2} - \frac{1}{(r_{ij} + R_j)^2} \right) - \frac{1}{2 r_{ij}} \log \frac{r_{ij} + R_j}{R_i} \right] \tag{55}$$

3. $r_{ij}^2 < (R_i - R_j)^2$ and $R_j \geq R_i$
In this case sphere i is totally engulfed by sphere j. Q_{ji} is obtained by the integral of $1/r^4$ centered on i over a spherical shell surrounding sphere i of inner radius R_i and outer radius $R_j - r_{ij}$, plus the value of Q_{ji} from eq. (55) when $R_i = R_j - r_{ij}$. The result is

$$Q_{ji} = 2\pi \left[\frac{2}{R_i} + \frac{R_j}{r_{ij}^2 - R_j^2} - \frac{1}{2 r_{ij}} \log \frac{r_{ij} + R_j}{R_j - r_{ij}} \right] \tag{56}$$

4. $r_{ij}^2 < (R_i - R_j)^2$ and $R_j < R_i$
In this case sphere j is totally engulfed by sphere i therefore

$$Q_{ji} = 0 \tag{57}$$

The derivative of Q_{ji} with respect to the interatomic distance r_{ij} is:

1. $r_{ij}^2 \geq (R_i + R_j)^2$

$$\frac{\partial Q_{ji}}{\partial r_{ij}} = 2\pi \left[\frac{R_j}{r_{ij}(r_{ij}^2 - R_j^2)} \left(1 - \frac{2 r_{ij}^2}{r_{ij}^2 - R_j^2} \right) + \frac{1}{2 r_{ij}^2} \log \frac{r_{ij} + R_j}{r_{ij} - R_j} \right] \tag{58}$$

2. $(R_i + R_j)^2 > r_{ij}^2 \geq (R_i - R_j)^2$

$$\frac{\partial Q_{ji}}{\partial r_{ij}} = 2\pi \left[\frac{1}{2} \left(1 - \frac{r_{ij}^2 - R_j^2}{2 r_{ij}^2} \right) \left(\frac{1}{(r_{ij} + R_j)^2} - \frac{1}{R_i^2} \right) + \frac{1}{2 r_{ij}^2} \log \frac{r_{ij} + R_j}{R_i} \right] \tag{59}$$

3. $r_{ij}^2 < (R_i - R_j)^2$ and $R_j \geq R_i$

$$\frac{\partial Q_{ji}}{\partial r_{ij}} = 2\pi \left[\frac{R_j}{r_{ij}(r_{ij}^2 - R_j^2)} \left(1 - \frac{2 r_{ij}^2}{r_{ij}^2 - R_j^2} \right) + \frac{1}{2 r_{ij}^2} \log \frac{r_{ij} + R_j}{R_j - r_{ij}} \right] \tag{60}$$

4. $r_{ij}^2 < (R_i - R_j)^2$ and $R_j < R_i$

$$\frac{\partial Q_{ji}}{\partial r_{ij}} = 0 \tag{61}$$

The derivatives of Q_{ji} with respect to $\mathbf{r}_i$ and $\mathbf{r}_j$ are

$$\frac{\partial Q_{ji}}{\partial \mathbf{r}_i} = -\frac{\partial Q_{ji}}{\partial \mathbf{r}_j} = \frac{\mathbf{r}_j - \mathbf{r}_i}{r_{ij}} \frac{\partial Q_{ji}}{\partial r_{ij}} \tag{62}$$

Appendix C: Calculation of the Gradient of ΔG_{vdW}

From eqs. (30) and (41) we obtain

$$\frac{\partial \Delta G_{\mathrm{vdW}}}{\partial \mathbf{r}_i} = \sum_j \frac{3 \alpha_j B_j^2}{(B_j + R_w)^4} f_b'(\beta_j) \frac{\partial \beta_j}{\partial \mathbf{r}_i} \tag{63}$$

where β_j is the inverse Born radius returned by the pair descreening formula (24), and $f'_b(\beta_j)$ is the first derivative of the filter function.

A change in position of atom i affects the Born radius of atom j in two ways. Not only does the direct descreening of atom i on atom j change, but also the descreening of the atoms around atom i changes due to the change of their self-volumes. These effects are born out in the expression for the gradient $\partial\beta_j/\partial\mathbf{r}_i$ of β_j obtained from eq. (24):

$$\frac{\partial\beta_j}{\partial\mathbf{r}_i} = \left(\frac{\partial\beta_j}{\partial\mathbf{r}_i}\right)_s + \left(\frac{\partial\beta_j}{\partial\mathbf{r}_i}\right)_Q \tag{64}$$

where

$$\left(\frac{\partial\beta_j}{\partial\mathbf{r}_i}\right)_s = -\frac{1}{4\pi}\sum_k s_{kj}\frac{\partial Q_{kj}}{\partial\mathbf{r}_i} \tag{65}$$

is the component of the gradient keeping the self-volume scaling factors constant and

$$\left(\frac{\partial\beta_j}{\partial\mathbf{r}_i}\right)_Q = -\frac{1}{4\pi}\sum_k \frac{\partial s_{kj}}{\partial\mathbf{r}_i} Q_{kj} \tag{66}$$

is the component corresponding to the changes of self-volumes.

The gradient of ΔG_{vdW} is similarly decomposed

$$\frac{\partial\Delta G_{\mathrm{vdW}}}{\partial\mathbf{r}_i} = \left(\frac{\partial\Delta G_{\mathrm{vdW}}}{\partial\mathbf{r}_i}\right)_s + \left(\frac{\partial\Delta G_{\mathrm{vdW}}}{\partial\mathbf{r}_i}\right)_Q \tag{67}$$

Inserting eq. (65) into eq. (63) yields

$$\left(\frac{\partial\Delta G_{\mathrm{vdw}}}{\partial\mathbf{r}_i}\right)_s = -\frac{1}{4\pi}\sum_j\left[\frac{3\alpha_i B_i^2}{(B_i+R_{\mathrm{w}})^4}f'_b(\beta_i)s_{ji}\frac{\partial Q_{ji}}{\partial\mathbf{r}_i} + \frac{3\alpha_j B_j^2}{(B_j+R_{\mathrm{w}})^4}f'_b(\beta_j)s_{ij}\frac{\partial Q_{ij}}{\partial\mathbf{r}_i}\right] \tag{68}$$

Inserting eq. (66) into eq. (63), using the fact that $s_{kj} = V'_k/V_k + V^g_{kj}/2V_k$ and setting

$$W_{ij} = \frac{3\alpha_j B_j^2}{(B_j+R_{\mathrm{w}})^4}f'_b(\beta_j)\frac{Q_{ij}}{V_i} \tag{69}$$

yields

$$\left(\frac{\partial\Delta G_{\mathrm{vdw}}}{\partial\mathbf{r}_i}\right)_Q = -\frac{1}{4\pi}\sum_j W_j\frac{\partial V'_j}{\partial\mathbf{r}_i} - \frac{1}{4\pi}\sum_j\frac{1}{2}(W_{ij}+W_{ji})\frac{\partial V^g_{ij}}{\partial\mathbf{r}_i} \tag{70}$$

where

$$W_i = \sum_j W_{ij} \tag{71}$$

By inserting eq. (3) in eq. (70) we finally obtain

$$\left(\frac{\partial\Delta G_{\mathrm{vdw}}}{\partial\mathbf{r}_i}\right)_Q = \frac{1}{4\pi}\sum_j\frac{1}{2}(W_i - W_{ij} + W_j - W_{ji})\frac{\partial V^g_{ij}}{\partial\mathbf{r}_i} - \frac{1}{4\pi}\sum_{j<k}\frac{1}{3}(W_i+W_j+W_k)\frac{\partial V^g_{ijk}}{\partial\mathbf{r}_i} + \frac{1}{4\pi}\sum_{j<k<l}\frac{1}{4}(W_i+W_j+W_k+W_l)\frac{\partial V^g_{ijkl}}{\partial\mathbf{r}_i} - \cdots \tag{72}$$

Notice that if i is a hydrogen atom $Q_{ji} = 0$ in eq. (68) and $(\partial\Delta G_{\mathrm{vdw}}/\partial\mathbf{r}_i)_Q = 0$ because $V^g_{ijk\cdots} = 0$. Also notice that the calculation of W_i for each particle requires prior knowledge of the Born radius of each atom of the system. Furthermore, the self-volume scaling factors s_i must be known before eq. (68) is implemented. The calculation of the gradient of ΔG_{vdW} proceeds therefore in the following steps:

1. The self-volumes and volume scaling factors are computed as described in the section Numerical Implementation.
2. The Born radii are computed using the pairwise descreening formula, and the values Q_{ij} and $\partial Q_{ij}/\partial r_{ij}$ are stored in a memory cache.
3. The gradients $(\partial\Delta G_{\mathrm{vdW}}/\partial\mathbf{r}_i)_s$ are computed [eq. (68)].
4. The W_i values are computed [eqs. (69) and (71)] and the gradients $(\partial\Delta G_{\mathrm{vdW}}/\partial\mathbf{r}_i)_Q$ are calculated [eq. (73)] by looping a second time over the intersection volumes in order to compute the gradients of the intersection volumes.

The use of a memory cache to store the values of Q_{ij} and its radial derivatives avoids looping multiple times over atom pairs. The size of the memory cache increases linearly with system size if a distance cutoff is applied to the pair descreening function. For the calculation of the Born radii we employ the same cutoff scheme used for the nonbonded electrostatic (direct Coulomb and generalized Born pair interactions) and Lennard-Jones interactions. Table 2 lists the amount of memory used by the memory cache for a representative sample of molecular systems.

Appendix D: Calculation of the Gradient of ΔG_{GB}

The derivation of the formulas for the gradient of ΔG_{GB} [eq. (16)] is similar to the derivation for ΔG_{vdW} (Appendix C), with the exception that an additional term is present due to the explicit dependence of f_{ij} on the interatomic distance. We obtain

$$\frac{\partial\Delta G_{\mathrm{GB}}}{\partial\mathbf{r}_i} = \left(\frac{\partial\Delta G_{\mathrm{GB}}}{\partial\mathbf{r}_i}\right)_{s,Q} + \left(\frac{\partial\Delta G_{\mathrm{GB}}}{\partial\mathbf{r}_i}\right)_s + \left(\frac{\partial\Delta G_{\mathrm{GB}}}{\partial\mathbf{r}_i}\right)_Q \tag{73}$$

where

$$\left(\frac{\partial \Delta G_{\mathrm{GB}}}{\partial \mathbf{r}_i}\right)_{s,Q} = 2u_\varepsilon \sum_j \frac{q_i q_j}{f_{ij}^3}\left[1 - \frac{\exp(-r_{ij}^2/4B_iB_j)}{4}\right]\mathbf{r}_{ji} \quad (74)$$

$$\left(\frac{\partial \Delta G_{\mathrm{GB}}}{\partial \mathbf{r}_i}\right)_{s} = -\frac{u_\varepsilon}{4\pi}\sum_j \left[(q_i^2 + Y_iB_i)\frac{\partial Q_{ji}}{\partial \mathbf{r}_i} + (q_j^2 + Y_jB_j)\frac{\partial Q_{ij}}{\partial \mathbf{r}_i}\right] \quad (75)$$

and

$$\begin{aligned}\left(\frac{\partial \Delta G_{\mathrm{GB}}}{\partial \mathbf{r}_i}\right)_{Q} = {} & \frac{u_\varepsilon}{4\pi}\sum_j \frac{1}{2}(U_i - U_{ij} + U_j - U_{ji})\frac{\partial V_{ij}^g}{\partial \mathbf{r}_i} \\ & - \frac{1}{4\pi}\sum_{j<k}\frac{1}{3}(U_i + U_j + U_k)\frac{\partial V_{ijk}^g}{\partial \mathbf{r}_i} \\ & + \frac{1}{4\pi}\sum_{j<k<l}\frac{1}{4}(U_i + U_j + U_k + U_l)\frac{\partial V_{ijkl}^g}{\partial \mathbf{r}_i} - \cdots \end{aligned} \quad (76)$$

where

$$Y_i = \sum_j \frac{q_i q_j \exp(-r_{ij}^2/4B_iB_j)}{f_{ij}^3}\left(\frac{r_{ij}^2}{4} + B_iB_j\right) \quad (77)$$

$$U_{ij} = (q_j^2 + Y_jB_j)\, f_b'(\beta_j)\frac{Q_{ij}}{V_i} \quad (78)$$

and

$$U_i = \sum_j U_{ij} \quad (79)$$

The algorithm for the calculation of the gradient of ΔG_{GB} is similar to the one presented for ΔG_{vdW} (Appendix C); the same memory cache used in the calculation of the gradient of ΔG_{vdW} is used.

Appendix E: Numerical Calculation of Born Radii

By expressing eq. (21) in spherical polar coordinates centered on $\mathbf{r}_i$ we obtain

$$\beta_i = \frac{1}{R_i} - \frac{1}{4\pi}\int_{\Omega_i} d\phi d(\cos\theta) dr \frac{1}{r^2} \quad (80)$$

The integral in eq. (80) is calculated on a grid of points. Because the points nearest to atom i contribute the most to the integral, the density of radial points should be chosen to be greater near R_i. This is achieved by performing the change of variable $z = 1/r$, which transforms the integral of $1/r^2$ in eq. (80) into the integration of the unit function:

$$\beta_i = \frac{1}{R_i} - \frac{1}{4\pi}\int_{\Omega_i} d\phi d(\cos\theta) dz \quad (81)$$

The integral above is then evaluated using a uniform grid of points $\{z_k, \cos\theta_l, \phi_m\}$ in $0 < z_k \le 1/R_i$, $-1 \le \cos\theta_l \le 1$, and $0 \le \phi_m \le 2\pi$, such that the corresponding point $\mathbf{r}_{klm}$ in Cartesian space is inside the solute volume ($|\mathbf{r} - \mathbf{r}_j|^2 < R_j^2$ for at least one atom j) but outside atom i ($|\mathbf{r} - \mathbf{r}_i|^2 > R_i^2$)

$$\beta_i = \frac{1}{R_i} - \frac{1}{4\pi}\sum_{\substack{klm \\ \mathbf{r}_{klm}\in\Omega_i}} \Delta\phi\Delta(\cos\theta)(z_{k+1} - z_k) \quad (82)$$

where $\Delta(\cos\theta)$ and $\Delta\phi$ are the grid spacings for the angular variables and we have used the fact that within the kth radial integration interval $\int_{z_k}^{z_{k+1}} dz = z_{k+1} - z_k$. In the calculations presented in this article we have used 80 integration points in each angular and radial direction (for a total of 512,000 grid points) for each solute atom.

References

1. Honig, B.; Yang, A.-S. Adv Protein Chem 1995, 46, 27.
2. Dill, K. A. Biochemistry 1990, 29, 7133.
3. Apostolakis, J.; Ferrara, P.; Calfish, A. J Chem Phys 1999, 110, 2099.
4. Lazaridis, T.; Karplus, M. Curr Opin Struct Biol 2000, 10, 139.
5. Felts, A. K.; Gallicchio, E.; Wallqvist, A.; Levy, R. M. Proteins: Struct, Funct, Genet 2002, 48, 404.
6. Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. J Phys Chem 1996, 100, 11775.
7. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, C. W. J Phys Chem A 1997, 101, 3005.
8. Froloff, N.; Windemuth, A.; Honig, B. Protein Sci 1997, 6, 1293.
9. Cramer, C. J.; Truhlar, D. Chem Rev 1999, 99, 2161.
10. Bashford, D.; Case, D. A. Annu Rev Phys Chem 2000, 51, 129.
11. Gallicchio, E.; Zhang, L. Y.; Levy, R. M. J Comput Chem 2002, 23, 517.
12. Lee, M. S.; Feig, M.; Salsbury Jr., F. R.; Brooks III, C. L. J Comput Chem 2003, 24, 1348.
13. Levy, R. M.; Gallicchio, E. Annu Rev Phys Chem 1998, 49, 531.
14. Roux, B.; Simonson, T. Biophys Chem 1999, 78, 1.
15. Dinner, A. R.; Lazaridis, T.; Karplus, M. Proc Natl Acad Sci USA 1999, 96, 9068.
16. Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. Nature 2002, 420, 102.
17. Lazaridis, T.; Karplus, M. J Mol Biol 1999, 288, 477.
18. Wallqvist, A.; Gallicchio, E.; Felts, A. K.; Levy, R. M. In Computational Methods for Protein Folding: A Special Volume of Advances in Chemical Physics, Vol. 120; Friesner, R., Ed.; John Wiley & Sons: New York, 2002, 459.
19. Sitkoff, D.; Sharp, K. A.; Honig, B. J Phys Chem 1994, 98, 1978.
20. Zhou, R.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. J Phys Chem 2001, 105, 10388.
21. Schwarzl, S. M.; Tschopp, T. B.; Smith, J. C.; Fisher, S. J Comput Chem 2002, 23, 1143.
22. Tomasi, J.; Persico, M. Chem Rev 1994, 94, 2027.
23. Cortis, C. M.; Friesner, R. A. J Comput Chem 1997, 18, 1591.

24. Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. J Comput Chem 2002, 23, 128.
25. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrikson, T. J Am Chem Soc 1990, 112, 6127.
26. Onufriev, A.; Bashford, D.; Case, D. A. J Phys Chem B 2000, 104, 3712.
27. Ghosh, A.; Rapp, C. S.; Friesner, R. A. J Phys Chem B 1998, 102, 10983.
28. Zhang, L.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. J Comput Chem 2001, 22, 591.
29. Schaefer, M.; Froemmel, C. J Mol Biol 1990, 216, 1045.
30. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J Phys Chem 1996, 100, 19824.
31. Schaefer, M.; Karplus, M. J Phys Chem 1996, 100, 1578.
32. Dominy, B. N.; Brooks, C. L. I. J Phys Chem B 1999, 103, 3765.
33. Tsui, V.; Case, D. A. Biopolymers 2000, 56, 275.
34. Ben-naim, A. Hydrophobic Interactions; Plenum Press: New York, 1980.
35. Kauzmann, W. Adv Protein Chem 1959, 14, 1.
36. Privalov, P. L.; Makhatadze, G. I. J Mol Biol 1993, 232, 660.
37. Sturtevant, J. M. Proc Natl Acad Sci USA 1977, 74, 2236.
38. Williams, D. H.; Searle, M. S.; Mackay, J. P.; Gerhard, U.; Maplestone, R. A. Proc Natl Acad Sci USA 1993, 90, 1172.
39. Siebert, X.; Hummer, G. Biochemistry 2002, 41, 2965.
40. Ooi, T.; Oobatake, M.; Nemethy, G.; Sheraga, A. Proc Natl Acad Sci USA 1987, 84, 3086.
41. Lee, M. R.; Duan, Y.; Kollman, P. A. Proteins 2000, 39, 309.
42. Hünenberger, P. H.; Helms, V.; Narayana, N.; Taylor, S. S.; McCammon, J. A. Biochemistry 1999, 38, 2358.
43. Simonson, T.; Brünger, A. T. J Phys Chem 1994, 98, 4683.
44. Rapp, C. S.; Friesner, R. A. Proteins: Struct, Funct, Genet 1999, 35, 173.
45. Fogolari, F.; Esposito, G.; Viglino, P.; Molinari, H. J Comput Chem 2001, 22, 1830.
46. Pellegrini, E.; Field, M. J. J Phys Chem A 2002, 106, 1316.
47. Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Ruiz-Lòpez, M. F.; Rinaldi, D.; Orozco, M.; Luque, F. J. J Comput Chem 2003, 24, 284.
48. Wallqvist, A.; Covell, D. G. J Phys Chem 1995, 99, 13118.
49. Gallicchio, E.; Kubo, M. M.; Levy, R. M. J Phys Chem B 2000, 104, 6271.
50. Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. J Am Chem Soc 2003, 25, 9523.
51. Nina, M.; Beglov, D.; Roux, B. J Phys Chem B 1997, 101, 5239.
52. Tsui, V.; Case, D. A. J Am Chem Soc 2000, 122, 2489.
53. Pitarch, J.; Moliner, V.; Pascual-Ahuir, J.-L.; Silla, A.; Tuñón, I. J Phys Chem 1996, 100, 9955.
54. Pitera, J. W.; van Gunsteren, W. F. J Am Chem Soc 2001, 123, 3163.
55. Zhang, L.; Gallicchio, E.; Levy, R. M. In AIP Conference Proceedings (Simulation and Theory of Electrostatic Interactions in Solutions), Vol. 492; AIP Publishing Center: New York, 1999.
56. Su, Y.; Gallicchio, E.; Levy, R. M. 2003, in preparation.
57. McCammon, J. A.; Straatsma, T. P. Annu Rev Phys Chem 1992, 43, 407.
58. Petitjean, M. J Comput Chem 1994, 15, 507.
59. Kratky, K. W. J Statist Phys 1981, 25, 619.
60. Grant, J. A.; Pickup, B. T. J Phys Chem 1995, 99, 3503.
61. Onufriev, A.; Case, D. A.; Bashford, D. J Comput Chem 2002, 23, 1297.
62. Lee, M. S.; Salsbury, F. F.; Brooks III, C. L. J Chem Phys 2002, 116, 10606.
63. Pierotti, R. A. Chem Rev 1976, 76, 717.
64. Hummer, G.; Garde, S.; García, A. E.; Paulaitis, M. E.; Pratt, L. R. Phys Chem B 1998, 102, 10469.
65. Lum, K.; Chandler, D.; Weeks, J. D. J Phys Chem B 1999, 103, 4570.
66. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. J Am Chem Soc 1996, 118, 11225.
67. Jorgensen, W. L.; Madura, J. D. Mol Phys 1985, 56, 1381.
68. Pratt, L. R.; Chandler, D. J Chem Phys 1977, 67, 3683.
69. Huang, D. M.; Chandler, D. J Phys Chem B 2002, 106, 2047.
70. Kitchen, D. B.; Hirata, F.; Kofke, D. A.; Westbrook, J. D.; Yormush, M.; Levy, R. M. J Comput Chem 1990, 11, 1169.
71. Schrödinger, Inc., Portland, OR.
72. Weiser, J.; Shenkin, P. S.; Still, C. J Comput Chem 1999, 20, 688.
73. Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. J Am Chem Soc 1999, 121, 9243.
74. Pascual-Ahuir, J. L.; Silla, E. J Comput Chem 1990, 11, 1047.
75. Figueirido, F.; Levy, R. M.; Zhou, R.; Berne, B. J. J Chem Phys 1997, 106, 9835.
76. Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. Proteins 2003, in press.
77. Harano, Y.; Felts, A. K.; Gallicchio, E.; Levy, R. M. 2003, in preparation.
78. Park, B.; Levitt, M. J Mol Biol 1996, 258, 367.
79. Simons, K. T.; Bonneau, R.; Ruczinski, I.; Baker, D. Proteins: Struct, Funct, Genet 1999, S3, 171.
80. Rizzo, R.; Tirado-Rives, J.; Jorgensen, W. L. J Med Chem 2001, 44, 145.