JCTC Journal of Chemical Theory and Computation

The AGBNP2 Implicit Solvation Model

Emilio Gallicchio,* Kristina Paris, and Ronald M. Levy

Department of Chemistry and Chemical Biology and BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, New Jersey 08854

Received May 11, 2009

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

Abstract: The AGBNP2 implicit solvent model, an evolution of the Analytical Generalized Born plus NonPolar (AGBNP) model we have previously reported, is presented with the aim of modeling hydration effects beyond those described by conventional continuum dielectric representations. A new empirical hydration free energy component based on a procedure to locate and score hydration sites on the solute surface is introduced to model first solvation shell effects, such as hydrogen bonding, which are poorly described by continuum dielectric models. This new component is added to the generalized Born and nonpolar AGBNP terms. Also newly introduced is an analytical Solvent Excluded Volume (SEV) model which improves the solute volume description by reducing the effect of spurious high dielectric interstitial spaces present in conventional van der Waals representations. The AGBNP2 model is parametrized and tested with respect to experimental hydration free energies of small molecules and the results of explicit solvent simulations. Modeling the granularity of water is one of the main design principles employed for the first shell solvation function and the SEV model, by requiring that water locations have a minimum available volume based on the size of a water molecule. It is shown that the new volumetric model produces Born radii and surface areas in good agreement with accurate numerical evaluations of these quantities. The results of molecular dynamics simulations of a series of miniproteins show that the new model produces conformational ensembles in substantially better agreement with reference explicit solvent ensembles than the original AGBNP model with respect to both structural and energetics measures.

1. Introduction

Water plays a fundamental role in virtually all biological processes. The accurate modeling of hydration thermodynamics is therefore essential for studying protein conformational equilibria, aggregation, and binding. Explicit solvent models arguably provide the most detailed and complete description of hydration phenomena.[1] 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,[2] which are based on the statistical mechanics concept of the solvent potential of mean force,[3] have been shown to be useful alternatives to explicit solvation for applications including protein folding and binding,[4] and small molecule hydration free energy prediction.[5]

Modern implicit solvent models[6,7] include distinct estimators for the nonpolar and electrostatic components of the hydration free energy. The nonpolar component corresponds to the free energy of hydration of the uncharged solute, while the electrostatic component corresponds to the free energy of turning on the solute partial charges. The latter is typically modeled treating the water solvent as a uniform high dielectric continuum.[8] Methods based on the numerical solution of the Poisson−Boltzmann (PB) equation[9] provide a virtually exact representation of the response of the solvent within the dielectric continuum approximation. Recent advances extending dielectric continuum approaches have focused on the development of Generalized Born (GB) models,[10] which have been shown to reproduce with good accuracy PB and explicit solvent[7,11] results at a fraction of

* Corresponding author e-mail emilio@biomaps.rutgers.edu.

10.1021/ct900234u CCC: $40.75 © 2009 American Chemical Society
Published on Web 07/31/2009

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

the computational expense. The development of computationally efficient analytical and differentiable GB methods based on pairwise descreening schemes[6,12,13] has made possible the integration of GB models in molecular dynamics packages for biological simulations.[14−16]

The nonpolar hydration free energy component accounts for all nonelectrostatic solute−solvent interactions as well as hydrophobic interactions,[17] which are essential driving forces in biological processes such as protein folding[18−21] and binding.[22−25] Historically the nonpolar hydration free energy has been modeled by empirical surface area models[26] which are still widely employed.[10,27−35] Surface area models are useful as a first approximation; however, qualitative deficiencies have been observed.[29,36−41]

A few years ago we presented the Analytical Generalized Born plus NonPolar (AGBNP) implicit solvent model,[42] which introduced two key innovations with respect to both the electrostatic and nonpolar components. Unlike most implicit solvent models, the AGBNP nonpolar hydration free energy model includes distinct estimators for the solute−solvent van der Waals dispersion energy and cavity formation work components. The main advantages of a model based on the cavity/dispersion decomposition of the nonpolar solvation free energy stem from its ability to describe both medium-range solute−solvent dispersion interactions, which depend on solute composition, as well as effects dominated by short-range hydrophobic interactions, which can be modeled by an accessible surface area term.[40] A series of studies highlight the importance of the balance between hydrophobicity and dispersion interactions in regulating the structure of the hydration shell and the strength of interactions between macromolecules.[43−45] In AGBNP the work of cavity formation is described by a surface area dependent model,[37,46−48] while the dispersion estimator is based on the integral of van der Waals solute−solvent interactions over the solvent, modeled as a uniform continuum.[38] This form of the nonpolar estimator had been motivated by a series of earlier studies[5,37,49−52] and has since been shown by us[38,53−55] and others[39−41,56] to be qualitatively superior to models based only on the surface area in reproducing explicit solvent results as well as rationalizing structural and thermodynamic experimental observations.

The electrostatic solvation model in AGBNP is based on the pairwise descreening GB scheme[13] whereby the Born radius of each atom is obtained by summing an appropriate descreening function over its neighbors. The main distinction between the AGBNP GB model and conventional pairwise descreening implementations is that in AGBNP the volume scaling factors, which offset the overcounting of regions of space occupied by more than one atom, are computed from the geometry of the molecule rather than being introduced as geometry-independent parameters fit to either experiments or to numerical Poisson−Boltzmann results.[14,57−59] The reduction of the number of parameters achieved with this strategy improves the transferability of the model to unusual functional groups often found in ligand molecules, which would otherwise require extensive parametrization.[60]

Given its characteristics, the AGBNP model has been mainly targeted to applications involving molecular dynamics canonical conformational sampling, and to the study of protein−ligand complexes. Since its inception the model has been employed in the investigation of a wide variety of biomolecular problems ranging from peptide conformational propensity prediction and folding,[54,61−63] ensemble-based interpretation of NMR experiments,[64,65] protein loop homology modeling,[55] ligand-induced conformational changes in proteins,[66,67] conformational equilibria of protein−ligand complexes,[68,69] protein−ligand binding affinity prediction,[70] and structure-based vaccine design.[71] The AGBNP model has been reimplemented and adopted with minor modifications by other investigators.[72,73] The main elements of the AGBNP nonpolar and electrostatic models have been independently validated,[39,40,74,75] and have been incorporated in recently proposed hydration free energy models.[76,77]

In this work we present a new implicit solvent model named AGBNP2 which builds upon the original AGBNP implementation (hereafter referred to as AGBNP1) and improves it with respect to the description of the solute volume and the treatment of short-range solute−water electrostatic interactions.

Continuum dielectric models assume that the solvent can be described by a linear and uniform dielectric medium.[78] This assumption is generally valid at the macroscopic level; however, at the molecular level water exhibits significant deviations from this behavior.[1] Nonlinear dielectric response, the nonuniform distribution of water molecules, charge asymmetry, and electrostriction effects[79] are all phenomena originating from the finite size and internal structure of water molecules as well as their specific interactions which are not taken into account by continuum dielectric models. Some of these effects are qualitatively captured by standard classical fixed-charge explicit water models; however others, such as polarization and hydrogen bonding interactions, can be fully modeled only by adopting more complex physical models.[80] GB models make further simplifications in addition to the dielectric continuum approximation, thereby compounding the challenge of achieving with GB-based implicit solvent models the level of realism required to reliably model phenomena, such as protein folding and binding, characterized by relatively small free energy changes.

In the face of these challenges a reasonable approach is to adopt an empirical hydration free energy model motivated by physical arguments[81] parametrized with respect to experimental hydration free energy data.[20] Models of this kind typically score conformations on the basis of the degree of solvent exposure of solute atoms. Historically[82] the solvent accessible surface area of the solute has been used for this purpose, while modern implementations suitable for conformational sampling applications often employ computationally convenient volumetric measures.[83,84] In this work we take this general approach but we retain the GB model component which we believe is a useful baseline to describe the long-range influence of the water medium. The empirical parametrized component of the model then takes the form of an empirical first solvation shell correction function designed so as to absorb hydration effects not accurately described by the GB model. Specifically, as described below, we employ a short-range analytical hydrogen bonding correction

function based on the degree of water occupancy (taking into account the finite size of water molecules) of appropriately chosen hydration sites on the solute surface. The aim of this model is to effectively introduce some explicit solvation features without actually adding water molecules to the system as for example done in hybrid explicit/implicit models.[85,86]

In this work we also improve the description of the solute volume, which in AGBNP1 is modeled by means of atomic spheres of radius equal to the atomic van der Waals radius. The deficiencies of the van der Waals solute volume model have been recognized.[87,88] They stem from the presence of high dielectric interstitial spaces in the solute interior which are too small to contain discrete water molecules. These spurious high dielectric spaces contribute to the hydration of buried or partially buried atoms causing underestimation of desolvation effects. The volume enclosed by the molecular surface (MS), defined as the surface produced by a solvent spherical probe rolling on the van der Waals surface of the solute,[89] represents the region which is inaccessible to water molecules and is often referred as the Solvent Excluded Volume (SEV).[90] The SEV, lacking the spurious high dielectric interstitial spaces, provides a better representation of the low dielectric region associated with the solute. For this reason accurate Poisson−Boltzmann solvers[9,91,92] have employed the SEV description of the solute region.

Despite its clear advantages, the lack of analytical and computationally efficient representation of the SEV have hampered its deployment in conjunction with GB models for molecular dynamics applications. The Generalized Born Molecular Volume (GBMV) series of models[87,93,94] achieve high accuracy relative to numerical Poisson calculations in part by employing the SEV description of the solute volume. The analytical versions of GBMV[93,94] describe the SEV volume by means of a continuous and differentiable solute density function which is integrated on a grid to yield atomic Born radii. In this work we present a model for the SEV that preserves the analytical pairwise atomic descreening approach employed in the AGBNP1 model,[42] which avoids computations on a grid. We show that this approximate model reproduces some of the key features of the SEV while yielding the same favorable algorithmic scaling of pairwise descreening approaches.

This paper focuses primarily on the description and parametrization of the SEV model and the short-range hydrogen bonding function of AGBNP2. In section 2 we present a brief review of the AGBNP1 model, including the electrostatic and nonpolar models, followed by the derivation of the analytical SEV pairwise descreening model and the short-range hydrogen bonding function which are new for AGBNP2. In section 3 we validate the AGBNP2 analytical estimates for the Born radii and atomic surface areas using as a reference accurate numerical evaluations of these quantities. This is followed by the parametrization of the hydrogen bonding function against experimental hydration free energies of small molecules. This section concludes with a comparison between the structural and energetic properties of a series of structured peptides (miniproteins) predicted with the AGBNP2 model and those obtained with explicit solvation. The paper then concludes with a discussion and implications of the results, and with a perspective on future improvements and validation of the AGBNP2 model.

2. Methods

2.1. The Analytical Generalized Born plus Nonpolar Implicit Solvent Model (AGBNP). In this section we briefly review the formulation of the AGBNP1 implicit solvent model, which forms the basis for the new AGBNP2 model. A full account can be found in the original reference.[42] The AGBNP1 hydration free energy $\Delta G_h(1)$ is defined as

$$\begin{aligned}\Delta G_h(1) &= \Delta G_{elec} + \Delta G_{np} \\ &= \Delta G_{elec} + \Delta G_{cav} + \Delta G_{vdW}\end{aligned} \quad (1)$$

where ΔG_{elec} is the electrostatic contribution to the solvation free energy and ΔG_{np} includes nonelectrostatic contributions. ΔG_{np} is further decomposed into a cavity hydration free energy ΔG_{cav} and a solute−solvent van der Waals dispersion interaction component ΔG_{vdW}.

2.1.1. Geometrical Estimators. Each free energy component in eq 1 is ultimately based on an analytical geometrical description of the solute volume modeled as a set of overlapping atomic spheres of radii R_i centered on the atomic positions $\mathbf{r}_i$. Hydrogen atoms do not contribute to the solute volume. The solute volume is modeled using the Gaussian overlap approach first proposed by Grant and Pickup.[95] In this model the solute volume is computed using the Poincaré formula (also known as the inclusion−exclusion formula) for the volume of the union of a set of intersecting elements

$$V = \sum_i V_i - \sum_{i<j} V_{ij} + \sum_{i<j<k} V_{ijk} - \ldots \quad (2)$$

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. The overlap volumes are approximated by the overlap integral, $V^g_{12\ldots n}$, available in analytical form (see for example eq 10 of ref 42), between n Gaussian density functions each corresponding to a solute atom:

$$V^g_{12\ldots n} \simeq \int d^3\mathbf{r}\, \rho_1(\mathbf{r})\, \rho_2(\mathbf{r}) \ldots \rho_n(\mathbf{r}) \quad (3)$$

where the Gaussian density function for atom i is

$$\rho_i(\mathbf{r}) = p \exp[-c_i(\mathbf{r} - \mathbf{r}_i)^2] \quad (4)$$

where

$$c_i = \frac{\kappa}{R_i^2} \quad (5)$$

and

$$p = \frac{4\pi}{3}\left(\frac{\kappa}{\pi}\right)^{3/2} \quad (6)$$

and κ is a dimensionless parameter that regulates the diffuseness of the atomic Gaussian function. In the AGBNP1 formulation $\kappa = 2.227$.

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

Gaussian integrals are in principle nonzero for any finite distances between the Gaussian densities. Although not mentioned in ref 42, to reduce computational cost AGBNP1 includes a switching function that reduces to zero the overlap volume between two or more Gaussians when the overlap volume is smaller than a certain value. If $V^{g}_{12\ldots}$ is the value of the Gaussian overlap volume between a set of atoms, the corresponding overlap volume $V_{12\ldots}$ used in eq 2 is set as

$$V_{12\ldots n} = \begin{cases} 0 & V^{g}_{12\ldots} \le v_1 \\ V^{g}_{12\ldots n} f_w(u) & v_1 < V^{g}_{12\ldots n} < v_2 \\ V^{g}_{12\ldots n} & V^{g}_{12\ldots} \ge v_2 \end{cases} \quad (7)$$

where

$$u = \frac{V^{g}_{12\ldots} - v_1}{v_2 - v_1} \quad (8)$$

$$f_w(x) = x^3(10 - 15x + 6x^2) \quad (9)$$

where, when using van der Waals atomic radii, $v_1 = 0.1$ and $v_2 = 1$ Å^3, and for the augmented radii used in the surface area model (see below), $v_1 = 0.2$ and $v_2 = 2$ Å^3. This scheme sets to zero Gaussian overlap volumes smaller than v_1, leaves volumes above v_2 unchanged, and smoothly reduces volumes between these two limits. It drastically reduces the number of overlap volumes that need to be calculated since the fact that an n-body overlap volume $V_{12\ldots n}$ between n atoms is zero guarantees that all of the $(n + 1)$-body overlap volumes corresponding to the same set of atoms plus one additional atom are also zero. (Note below that the formulation of AGBNP2 employs modified values of v_1 and v_2 to improve the accuracy of surface areas.)

The van der Waals surface area A_i of atom i, which is another geometrical descriptor of the model, is based on the derivative $\partial V/\partial R_i$ of the solute volume with respect to the radius R_i[96]

$$A_i = f_a\left(\frac{\partial V}{\partial R_i}\right) \quad (10)$$

where V is given by eq 2 and

$$f_a(x) = \begin{cases} \dfrac{x^3}{a^2 + x^2} & x > 0 \\ 0 & x \le 0 \end{cases} \quad (11)$$

with $a = 5$ Å^2, is a filter function which prevents negative values for the surface areas for buried atoms while inducing negligible changes to the surface areas of solvent-exposed atoms.

The model further defines the so-called self-volume V'_i of atom i as

$$V'_i = V_i - \frac{1}{2}\sum_j V_{ij} + \frac{1}{3}\sum_{j<k} V_{ijk} + \ldots \quad (12)$$

which is computed similarly to the solute volume and measures the solute volume that is considered to belong exclusively to this atom. Due to the overlaps with other atoms, the self-volume V'_i of an atom is smaller than the van der Waals volume V_i of the atom. The ratio

$$s_i = \frac{V'_i}{V_i} \le 1 \quad (13)$$

is a volume scaling factor used below in the evaluation of the Born radii.

2.1.2. Electrostatic Model. The electrostatic hydration free energy is modeled using a continuous dielectric representation of the water solvent using the Generalized Born (GB) approximation

$$\Delta G_{\mathrm{elec}} = u_\epsilon \sum_i \frac{q_i^2}{B_i} + 2u_\epsilon \sum_{i<j} \frac{q_i q_j}{f_{ij}} \quad (14)$$

where

$$u_\epsilon = -\frac{1}{2}\left(\frac{1}{\epsilon_{\mathrm{in}}} - \frac{1}{\epsilon_{\mathrm{w}}}\right) \quad (15)$$

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

$$f_{ij} = \sqrt{r_{ij}^2 + B_i B_j \exp(-r_{ij}^2/4B_iB_j)} \quad (16)$$

In eqs 14−16 B_i denotes the Born radius of atom i which, under the Coulomb field approximation,[57] is given by the inverse of the integral over the solvent region of the negative fourth power of the distance function centered on atom i

$$\beta_i = \frac{1}{B_i} = \frac{1}{4\pi}\int_{\mathrm{solvent}} \mathrm{d}^3\mathbf{r}\, \frac{1}{(\mathbf{r} - \mathbf{r}_i)^4} \quad (17)$$

In the AGBNP1 model this integral is approximated by a so-called pairwise descreening formula

$$\beta_i = \frac{1}{R_i} - \frac{1}{4\pi}\sum_{j\neq i} s_{ji} Q_{ji} \quad (18)$$

where R_i is the van der Waals radius of atom i, s_{ji} is the volume scaling factor for atom j (eq 13) when atom i is removed from the solute, and Q_{ji} is the integral (available in analytic form; see Appendix B of ref 42) of the function $(\mathbf{r} - \mathbf{r}_i)^{-4}$ over the volume of the sphere corresponding to solute atom j that lies outside the sphere corresponding to atom i. Equation 18 applies to all of the atoms i of the solute (hydrogen atoms and heavy atoms), whereas the sum over j includes only heavy atoms. The AGBNP1 estimates for the Born radii B_i are finally computed from the inverse Born radii β_i from eq 18 after processing them through the function

$$B_i^{-1} = f_b(\beta_i) = \begin{cases} \sqrt{b^2 + \beta_i^2} & \beta_i > 0 \\ b & \beta_i \le 0 \end{cases} \quad (19)$$

where $b^{-1} = 50$ Å. The filter function eq 19 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.

In the AGBNP1 model the scaling factors s_{ji} are approximated by the expression

$$s_{ji} \simeq s_j + \frac{1}{2}\frac{V_{ij}}{V_j} \tag{20}$$

where s_j is given by eq 13 and V_{ij} is the two-body overlap volume between atoms i and j. Also, in the original AGBNP formulation the computation of the scaling factors and the descreening function in eq 18 employed the van der Waals radii for the atoms of the solute and the associated Gaussian densities. These two aspects have been modified in the new formulation (AGBNP2) as described below.

2.1.3. Nonpolar Model. The nonpolar hydration free energy is decomposed into the cavity hydration free energy ΔG_{cav} and the solute−solvent van der Waals dispersion interaction component ΔG_{vdW}:

$$\Delta G_{\mathrm{np}} = \Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{vdW}} \tag{21}$$

The cavity component is described by a surface area model[37,46−48]

$$\Delta G_{\mathrm{cav}} = \sum_i \gamma_i A_i \tag{22}$$

where the summation runs over the solute heavy atoms, A_i is the van der Waals surface area of atom i from eq 10, and γ_i is the surface tension parameter assigned to atom i (see Table 1 of ref 42). Surface areas are computed using augmented radii R_i^c for the atoms of the solute and the associated Gaussian densities. Augmented radii are defined as the van der Waals radii (Table 1 of ref 42) plus a 0.5 Å offset. The computation of the atomic surface areas in AGBNP2 is mostly unchanged from the original implementation,[42] with the exception of the values of the switching function cutoff parameters υ_1 and υ_2 of eq 7, which in the new model are set as $\upsilon_1 = 0.01$ Å^3 and $\upsilon_2 = 0.1$ Å^3. This change was deemed necessary to improve the accuracy of the surface areas which in the new model also affect the Born radii estimates through eq 31 below.

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

$$\Delta G_{\mathrm{vdW}} = \sum_i \alpha_i \frac{a_i}{(B_i + R_{\mathrm{w}})^3} \tag{23}$$

where α_i is an adjustable dimensionless parameter on the order of 1 (see Table 1 of ref 42) and

$$a_i = -\frac{16}{3}\pi\rho_{\mathrm{w}}\epsilon_{i\mathrm{w}}\sigma_{i\mathrm{w}}{}^6 \tag{24}$$

where $\rho_{\mathrm{w}} = 0.033\,28$ Å^{-3} is the number density of water at standard conditions, and $\sigma_{i\mathrm{w}}$ and $\epsilon_{i\mathrm{w}}$ are the OPLS force field[97] Lennard-Jones interaction parameters for the interaction of solute atom i with the oxygen atom of the TIP4P water model.[98] If σ_i and ϵ_i are the OPLS Lennard-Jones parameters for atom i

$$\sigma_{i\mathrm{w}} = \sqrt{\sigma_i \sigma_{\mathrm{w}}} \tag{25}$$

$$\epsilon_{i\mathrm{w}} = \sqrt{\epsilon_i \epsilon_{\mathrm{w}}} \tag{26}$$

where $\sigma_{\mathrm{w}} = 3.153\,65$ Å and $\epsilon_{\mathrm{w}} = 0.155$ kcal/mol are the Lennard-Jones parameters of the TIP4P water oxygen. In eq 23 B_i is the Born radius of atom i from eqs 18 and 19 and $R_{\mathrm{w}} = 1.4$ Å is a parameter corresponding to the radius of a water molecule.

2.2. The AGBNP2 Implicit Solvent Model. The AGBNP2 hydration free energy $\Delta G_{\mathrm{h}}(2)$ is defined as

$$\Delta G_{\mathrm{h}}(2) = \Delta G_{\mathrm{elec}} + \Delta G_{\mathrm{np}} + \Delta G_{\mathrm{hb}} \tag{27}$$

where ΔG_{elec} and ΔG_{np} have the same form as in the AGBNP1 model (eqs 14 and 21−23, respectively). The only major difference is the pairwise descreening model for the Born radii that in AGBNP2 is based on the solvent excluded volume described below rather than the van der Waals volume as in AGBNP1. ΔG_{hb}, described in section 2.2.2, is a novel term for AGBNP2 which represents a first solvation shell correction corresponding to the portion of the hydration free energy not completely accounted for by the uniform continuum model for the solvent. We think of this term as mainly incorporating the effect of solute−solvent hydrogen bonding. As described in detail below, the analytical model for ΔG_{hb} is based on measuring and scoring the volume of suitable hydration sites on the solute surface.

2.2.1. Pairwise Descreening Model Using the Solvent Excluded Volume. When using van der Waals radii to describe the solute volume, small crevices between atoms (Figure 1, panel A) are incorrectly considered as high dielectric solvent regions,[93,99,100] leading to underestimation of the Born radii, particularly for buried atoms. The van der Waals volume description implicitly ignores the fact that the finite size of water molecules prevents them from occupying sites that, even though they are not within solute atoms, are too small to be occupied by water molecules. Ideally a model for the Born radii would include in the descreening calculation all of the volume excluded from water either because it is occupied by a solute atom or because it is located in an interstitial region inaccessible to water molecules. We denote this volume as the solvent excluded volume (SEV). A realistic description of the SEV is the volume enclosed within the molecular surface[89] of the solute obtained by tracing the surface of contact of a sphere with a radius characteristic of a water molecule (typically 1.4 Å) rolling over the van der Waals surface of the solute. The main characteristic of this definition of the SEV (see Figure 2) is that, unlike the van der Waals volume, it lacks small interstitial spaces while it closely resembles the van der Waals volume near the solute−solvent interface. The molecular surface description of the SEV cannot be easily implemented into an analytical formulation. In this section we will present an analytical description of the SEV for the purpose of the pairwise descreening computation of the Born radii, as implemented in AGBNP2, that preserves the main characteristics of the molecular surface description of the SEV.

The main ideas underpinning the SEV model presented here are illustrated in Figure 1. We start with the van der Waals representation of the solute (model A) which presents an

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

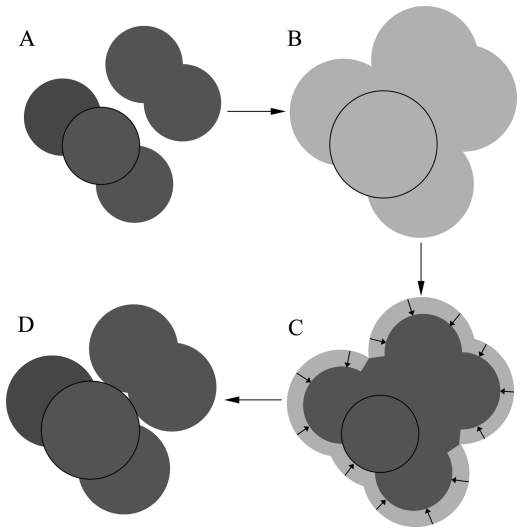


Figure 1. Schematic diagram illustrating the ideas underpinning the model for the solvent excluded volume descreening. Circles represent atoms of two idealized solutes placed in proximity of each other. The van der Waals description of the molecular volume (panel A) leaves high dielectric interstitial spaces that are too small to fit water molecules. The adoption of enlarged van der Waals radii (B) removes the interstitial spaces but incorrectly excludes solvent from the surface of solvent-exposed atoms. The solvent volume subtended by the solvent-exposed surface area is subtracted from the enlarged volume of each atom (C) such that larger atomic descreening volumes are assigned to buried atoms (circled) than exposed atoms (D), leading to the reduction of interstitial spaces while not overly excluding solvent from the surface of solvent-exposed atoms.

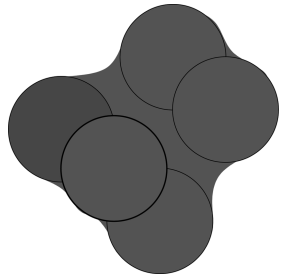


Figure 2. Illustration of the relationship between the van der Waals volume and the solvent excluded volume enclosed by the molecular surface.

undesirable high dielectric interstitial space between the two groups of atoms. Increasing the atomic radii leads to a representation (model B) in which the interstitial space is removed but that also incorrectly excludes solvent from the surface of solvent-exposed atoms. This representation is therefore replaced with one in which the effective volume of each atom in B is reduced by the volume subtended between the solvent-exposed surface of each atom and its van der Waals radius (Figure 1C). This process yields model D in which the effective volume of the most buried atom is larger than those of the solvent-exposed atoms. This SEV model covers the interstitial high dielectric spaces present in a van der Waals description of the solute volume, while approximately maintaining the correct van der Waals volume description of atoms at the solute surface as in the molecular surface description of the SEV (Figure 2).

These ideas have been implemented in the AGBNP2 model as follows. The main modification consists of adopting for the pairwise descreening generalized Born formulation the same augmented van der Waals radii as in the computation of the atomic surface areas. As in the previous model the augmented atomic radii, R_i^c, are set as

$$R_i^c = R_i + \Delta R \tag{28}$$

where R_i is the van der Waals radius of the atom and $\Delta R = 0.5$ Å is the offset. The augmented radii are used in the same way as in the AGBNP1 formulation to define the atomic spheres and the associated Gaussian densities (eqs 3−6). Henceforth in this work all of the quantities (atomic volumes, self-volumes, etc.) are understood to be computed with the augmented atomic radii, unless otherwise specified. In AGBNP2 the form of the expression for the inverse Born radii (eq 18) is unchanged; however, the expressions for the volume scaling factors s_{ji} and the evaluation of the descreening function Q_{ji} are modified as follows to introduce the augmented atomic radii and the reduction of the atomic self-volumes in proportion to the solvent accessible surface areas as discussed above.

The pairwise volume scaling factors s_{ji}, that is the volume scaling factor for atom j when atom i is removed from the solute, are set as

$$s_{ji} = s_j + \frac{V'_{ji}}{V_j} \tag{29}$$

where s_j (defined below) is the volume scaling factor for atom j analogous to eq 13 computed with all the atoms present, and the quantity

$$V'_{ji} = V'_{ij} = \frac{1}{2}V_{ij} - \frac{1}{3}\sum_k V_{ijk} + \frac{1}{4}\sum_{k<l} V_{ijkl} - \ldots \tag{30}$$

subtracts from the expression for the self-volume of atom j all those overlap volumes involving both atoms i and j.

Two differences with respect to the original AGBNP1 formulation are introduced. The first is that s_j is computed from the self-volume after subtracting from it the volume of the region subtended by the solvent-exposed surface between the enlarged and van der Waals atomic spheres of atom j, according to the expression

$$s_j = \frac{V'_j - d_j A_j}{V_j} \tag{31}$$

where A_j is the surface area of atom j from eq 10. Referring to Figure 3, the volume of the subtended region is $d_j A_j$ as in eq 31 with

$$d_j = \frac{1}{3}R'_j\left[1 - \left(\frac{R_j}{R'_j}\right)^3\right] \tag{32}$$

The other difference concerns the V'_{ji} term which in the AGBNP1 formulation is approximated by the two-body overlap volume V_{ij} (see eq 13), the first term in the right-hand side of eq 30. This approximation is found to lack

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

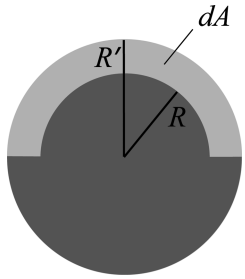


Figure 3. Graphical construction showing the volume subtracted from the atomic self-volume to obtain the surface area corrected atomic self-volume. R is the van der Waals radius of the atom; $R' = R + \Delta R$ is the enlarged atomic radius. dA is the volume of the region (light gray) subtended by the solvent-exposed surface area between the enlarged and van der Waals atomic spheres.

sufficient accuracy for the present formulation given the relative increase in size of all overlap volumes. Therefore in AGBNP2 V'_{ji} is computed including in eq 30 all nonzero overlap volumes after the application of the switching function from eq 7.

In the AGBNP2 formulation the functional form for the pair descreening function Q_{ji} is the same as in the original formulation (see Appendix of ref 42); however, in the new formulation this function is evaluated using the van der Waals radius R_i for atom i (the atom being "descreened") and the augmented radius R_j^c for atom j (the atom that provides the solvent descreening), rather than using the van der Waals radius for both atoms. Thus if the pair descreening function is denoted by $Q(r,R_1,R_2)$, where r is the interatomic distance, R_1 the radius of the atom being descreened, and R_2 the radius that provides descreening, we set in eq 18

$$Q_{ji} = Q(r_{ij}, R_i, R_j^c) \tag{33}$$

The alternative of using enlarged atoms for both atoms and the inclusion of a properly weighted self-descreening term (to take into account the SEV of the atom being descreened) was also tried and judged to be less accurate than eq 33 relative to numerical integration.

2.2.2. Short-Range Hydrogen Bonding Correction Function. In this section we present the analytical model that implements the short-range hydrogen bonding correction function for AGBNP2. The model is based on a geometrical procedure, described below, to measure the degree to which a solute atom can interact with hydration sites on the solute surface. The procedure is as follows. A sphere of radius R_s representing a water molecule is placed in a position that provides near-optimal interaction with a hydrogen bonding donor or acceptor atom of the solute. The position $\mathbf{r}_s$ of this water sphere s is function of the positions of two or more parent atoms that compose the functional group including the acceptor/donor atom:

$$\mathbf{r}_s = \mathbf{r}_s(\{\mathbf{r}_{ps}\}) \tag{34}$$

where $\{\mathbf{r}_{ps}\}$ represents the positions of the set of parent atoms of the water site s. For instance, the water site position in correspondence with a polar hydrogen is

$$\mathbf{r}_s = \mathbf{r}_D + \frac{\mathbf{r}_H - \mathbf{r}_D}{|\mathbf{r}_H - \mathbf{r}_D|} d_{HB}$$

where $\mathbf{r}_D$ is the position of the heavy atom donor, $\mathbf{r}_H$ is the position of the polar hydrogen, and d_{HB} is the distance between the heavy atom donor and the center of the water sphere (see Figure 4). Similar relationships (see the Appendix) are employed to place candidate water spheres in correspondence with hydrogen bonding acceptor atoms of the solute. These relationships are based on the local topology of the hydrogen bonding acceptor group (linear, trigonal, and tetrahedral). This scheme places one or two water spheres in correspondence with each hydrogen bonding acceptor atom (see Table 1).

The magnitude of the hydrogen bonding correction corresponding to each water sphere is a function of the predicted water occupancy of the location corresponding to the water sphere. In this work the water occupancy is measured by the fraction w_s of the volume of the water site sphere that is accessible to water molecules without causing steric clashes with solute atoms (see Figure 4)

$$w_s = \frac{V_s^{free}}{V_s} \tag{35}$$

where $V_s = (4/3)\pi R_s^3$ is the volume of the water sphere and

$$V_s^{free} = V_s - \sum_i V_{si} + \sum_{i<j} V_{sij} - \sum_{i<j<k} V_{sijk} \tag{36}$$

is the "free" volume of water site s, obtained by summing over the two-body, three-body, etc. overlap volumes of the water sphere with the solute atoms. Note that the expression of the free volume is the same as the expression for the self-volume (eq 12) except that it lacks the fractional coefficients 1/2, 1/3, etc. The overlap volumes in eq 36 are computed using radius R_s for the water site sphere (here set to 1.4 Å) and augmented radii R_i^c for the solute atoms. Equation 36 is derived similarly to the expression for the self-volumes by removing overlap volumes from the volume of the water sphere rather than evenly distributing them across the atoms participating in the overlap.

Given the water occupancy w_s of each water sphere, the expression for the hydrogen bonding correction for the solute is

$$\Delta G_{hb} = \sum_s h_s S(w_s; w_a, w_b) \tag{37}$$

where h_s is the maximum correction energy which depends on the type of solute−water hydrogen bond (see Table 1), and $S(w;w_a,w_b)$ is a polynomial switching function which is 0 for $w < w_a$, 1 for $w > w_b$, and smoothly (with continuous first derivatives) interpolates from 0 to 1 between w_a and w_b (see Figure 5). The expression of $S(w;w_a,w_b)$ is

$$S(w;w_a,w_b) = \begin{cases} 0 & w \le w_a \\ f_w\left(\dfrac{w - w_a}{w_b - w_a}\right) & w_a < w < w_b \\ 1 & w \ge w_b \end{cases} \tag{38}$$

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

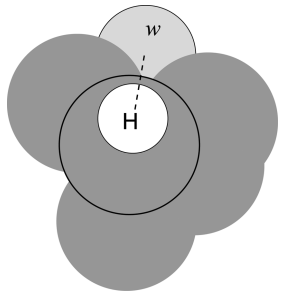


Figure 4. Schematic diagram for the placement of the water sphere (*w*, light gray) corresponding to the hydrogen bonding position relative to the a polar hydrogen (white sphere) of the solute (dark gray). The dashed line traces the direction of the hydrogen−parent heavy atom (circled) bond along which the water sphere is placed. The magnitude of hydrogen bonding correction grows as a function of the volume (light gray) of the water site sphere not occupied by solute atoms.

Table 1. Optimized Surface Tension Parameters and Hydrogen Bonding Correction Parameters for the Atom Types Present in Protein Molecules[a]

atom type	γ (cal/mol/Å^2)	geometry	N_w	h (kcal/mol)
C (aliphatic)	129			
C (aromatic)	120			
H on N		linear	1	−0.25
H on N (Arg)		linear	1	−2.50
H on O		linear	1	−0.40
H on S		linear	1	−0.50
O (alcohol)	117	tetrahedral	2	−0.40
O (carbonyl)	117	trigonal	2	−1.25
O (carboxylate)	40	trigonal	2	−1.80
N (amine)	117	tetrahedral	1	−2.00
N (aromatic)	117	trigonal	1	−2.00
S	117	tetrahedral	2	−0.50

[a] γ is the surface tension parameter, N_w is the number of water spheres, and h is the maximum correction corresponding to each atom type (eq 37). Atom types not listed do not have hydrogen bonding corrections and are assigned γ = 117 cal/mol/Å^2.

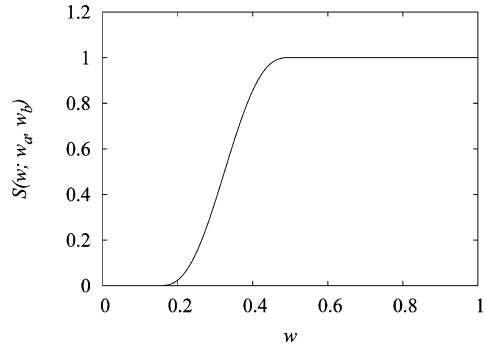


Figure 5. Switching function $S(w;w_a,w_b)$ from eqs 38 and 9 with w_a = 0.15 and w_b = 0.5.

where $f_w(x)$ is a switching function given by eq 9. In this work we set w_a = 0.15 and w_b = 0.5. This scheme establishes (see Figure 5) that no correction is applied if more than 85% of the water sphere volume is not water accessible, whereas maximum correction is applied if 50% or more of the water sphere volume is accessible.

2.3. Molecular Dynamics of Miniproteins. We conducted molecular dynamics simulations of what we will refer to as miniproteins (Figure 6), that is, peptides that have been shown to form stable secondary structures in solution: the 23-residue trp-cage peptide of sequence ALQELLGQWLKDGGPSSGRPPPS [Protein Data Bank (PDB) ID 1RIJ],[101] the 28-residue cdp-1 peptide of sequence KPYTARIKGRTFSNEKELRDFLETFTGR (PDB ID 1PSV),[102] and the 28-residue fsd-1 peptide of sequence QQYTAKIKGRTFRNEKELRDFIEKFKGR (PDB ID 1FSD).[103] The structure of trp-cage (see Figure 6) is characterized by a tryptophan side chain enclosed in a cage formed by an α-helix on one side and a proline-rich loop on the other. The cdp-1 and fsd-1 miniproteins (Figure 6) adopt a mixed $\alpha\beta$ conformation and are particularly rich in charged residues. The trp-cage miniprotein was chosen because it has been the target of several computational studies.[104−107] The cdp-1 and fsd-1 peptides were of interest because they showed in preliminary tests with AGBNP1 solvation a significant tendency to deviate from the experimental structures.

Molecular dynamics simulations were conducted for 12 ns starting with the first NMR model deposited in the PDB. The temperature was set to 300 K with the Nosé−Hoover thermostat,[108,109] a molecular dynamics (MD) time step of 2 fs was employed, and covalent bond lengths involving hydrogen atoms were fixed at their equilibrium positions. Backbone motion was restricted by imposing a positional harmonic restraint potential with a force constant of 0.3 kcal/mol/Å^2 on the positions of the Cα atoms, which allows for a range of motion of about 5 Å at the simulation temperature. These restraints are sufficiently weak to allow substantial backbone and side chain motion while preserving overall topology.

Molecular dynamics simulations were conducted with the OPLS-AA potential[97,110] with explicit solvation (SPC water model with 2450, 3110, and 3250 water molecules for trp-cage, cdp-1, and fsd-1, respectively) and with both AGBNP1 and AGBNP2 implicit solvation. The DESMOND program[111] was used for the explicit solvent simulations, and the IMPACT program[15] was used for those with implicit solvation. Identical force field settings were employed in these two programs. The explicit solvent simulations were conducted in the NPT ensemble using the Martyna−Tobias−Klein barostat[112] at 1 atm pressure and employed the smooth Particle Mesh Ewald (PME) method[113] for the treatment of long-range electrostatic interactions with a real-space cutoff of 9 Å. Equilibrium averages and energy distributions were obtained by analysis of the latter 10 ns of saved trajectories. Convergence was tested by comparing averages obtained using the first and second halves of simulation data. Hydrogen bonds were detected using a minimum hydrogen-acceptor distance of 2.5 Å and a minimum donor angle of 120°.

3. Results

3.1. Accuracy of Born Radii and Surface Areas. The quality of any implicit solvent model depends primarily on the reliability of the physical model on which it is based. The accuracy of the implementation, however, is also a critical aspect for the success of the model in practice. This is true in particular for models, such as AGBNP, based on

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

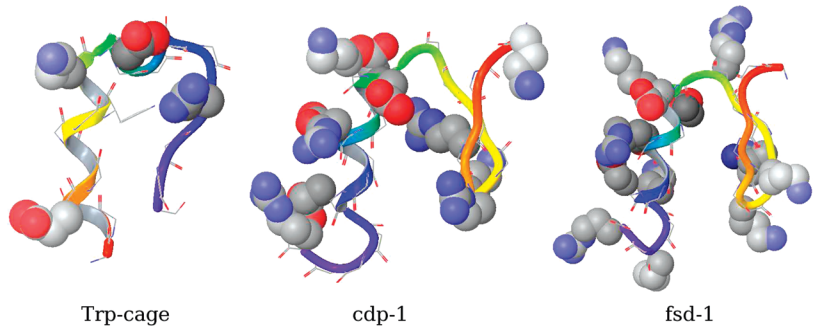


Figure 6. Graphical representations of the NMR structures of the three miniproteins investigated in this work: trp-cage (PDB ID 1RIJ), cdp-1 (PDB ID 1PSV), and fsd-1 (PDB ID 1FSD). In each case the first deposited NMR model is shown. Backbone ribbon is colored from the N-terminal (red) to the C-terminal (blue). Charged side chains are shown in space-filling representation.

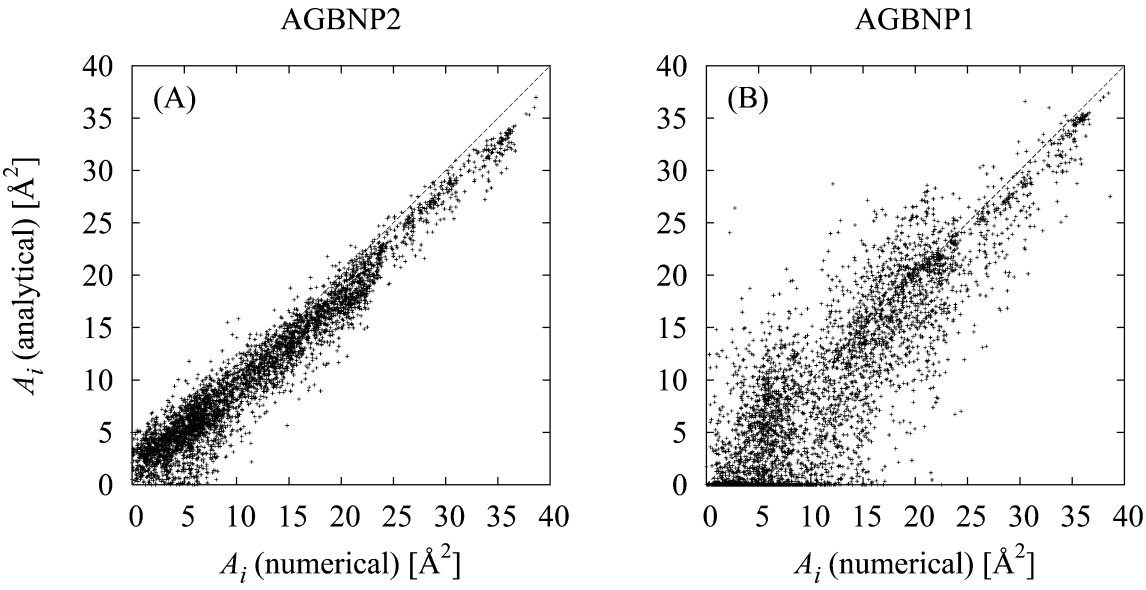


Figure 7. Comparisons between numerical and analytical molecular surface areas of the heavy atoms of the crystal structures (1ctf and 1lz1, respectively) of the C-terminal domain of the ribosomal protein L7/L12 (74 aa) and human lysozyme (130 aa), and of four conformations each of the trp-cage, cdp-1, and fsd-1 miniproteins extracted from the corresponding explicit solvent MD trajectories. (A) Analytical molecular surface areas computed using the present model and (B), for comparison, analytical surface areas computed using the original model from ref 42.

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

the generalized Born formula. It has been pointed out, for instance, that approximations in the integration procedure to obtain the Born radii may actually be of more significance than the physical approximations on which the GB model is based.[114] It is therefore important to test that the conformational-dependent quantities employed by AGBNP2 are a good approximation to the geometrical parameters that they are supposed to represent. The present AGBNP2 formulation relies mainly on three types of conformational-dependent quantities: Born radii (eq 18), solvent accessible surface areas (eq 10), and solvent accessibilities of hydration sites (eq 38). In this section we analyze the validity of the AGBNP2 analytical estimates for the Born radii and surface areas against accurate numerical results for the same quantities.

We employ the GEPOL program[90] to compute numerically atomic solvent accessible surface areas with a solvent probe diameter of 1 Å, the same probe diameter that defines the solute−solvent boundary in the AGBNP model. Figure 7A shows the comparison between the surface area estimates given by the present formulation of AGBNP and the numerical surface areas produced by GEPOL for a series of native and modeled protein conformations. In Figure 7B we show the same comparison for the surface areas of the original AGBNP1 model. These representative results show that the present analytical surface area implementation, which as described above employs a weaker switching function for the overlap volumes, produces significantly more accurate atomic surface areas than the original model. These improvements in the computation of the surface areas, introduced mainly to obtain more accurate Born radii through eq 31, are also expected to yield more reliable cavity hydration free energy differences.

Figure 8 illustrates on the same set of protein conformations the accuracy of the inverse Born radii, B_i^{-1}, obtained using the AGBNP2 pairwise descreening method using the SEV model for the solute volume described above (eq 18), by comparing them to accurate estimates obtained by evaluating the integral in eq 17 numerically on a grid. The comparison is performed for the inverse Born radii because these, being proportional to GB self-energies, are more reliable accuracy indicators than the Born radii themselves. The grid for the numerical integration was prepared as

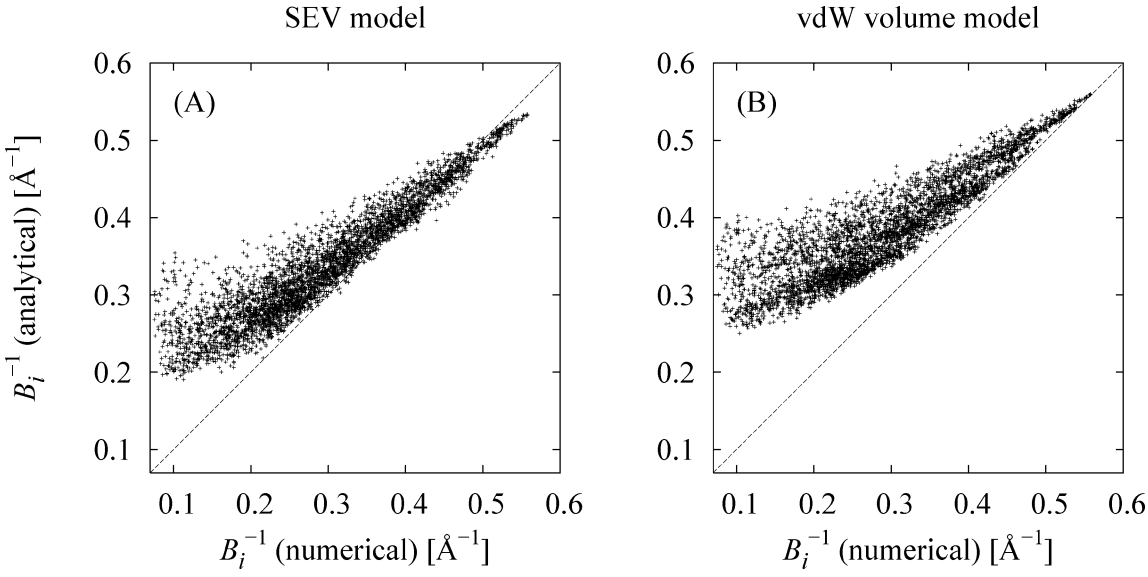


Figure 8. Comparisons between numerical and analytical inverse Born radii for the heavy atoms of the same protein conformations as in Figure 7. (A) Analytical Born radii computed using the present SEV model. (B) Analytical Born radii computed using the van der Waals volume model (ref 42).

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

previously reported,[42] except that the solvent excluded volume (SEV) of the solute was employed here rather than the van der Waals volume. The integration grid over the SEV was obtained by taking advantage of the particular way that the GEPOL algorithm describes the SEV of the solute; GEPOL iteratively places auxiliary spheres of various dimensions in the interstitial spaces between solute atoms in such a way that the van der Waals volume of the solute plus the auxiliary spheres accurately reproduces the SEV of the solute. Therefore in the present application a grid point was considered part of the SEV of the solute if it was contained within any solute atom or any one of the auxiliary spheres placed by GEPOL. The default 1.4 Å solvent probe radius was chosen for the numerical computation of the SEV with the GEPOL program to assess the accuracy of the model with respect to a full representation of the solute solvent excluded volume as in the GBMV series of models.[93,94] The results of this validation (Figure 8) show that the analytical SEV pairwise descreening model described above is able to yield Born radii which are not as affected by the spurious high dielectric interstitial spaces present in the van der Waals volume description of the solute. With the van der Waals volume model (Figure 8B) the Born radii of the majority of solute atoms start to significantly deviate from the reference values for Born radii larger than about 2.5 Å ($B^{-1} = 0.4$ Å^{-1}). Born radii computed with the analytical SEV model instead (Figure 8A) track the reference values reasonably well further up to about 4 Å ($B^{-1} = 0.25$ Å^{-1}). Despite this significant improvement most Born radii are still underestimated by the improved model (and, consequently, the inverse Born radii are overestimated—see Figure 8), particularly those of nonpolar atoms near the hydrophobic core of the larger proteins. These regions tend to be loosely packed and tend to contain interstitial spaces too large to be correctly handled by the present model. Because it mainly involves groups of low polarity, this limitation has a small effect on the GB solvation energies. It has however a more significant effect on the van der Waals solute−solvent interaction energy estimates through eq 23, which systematically overestimate the magnitude of the interaction for atoms buried in hydrophobic protein core. While the present model in general ameliorates in all respects the original AGBNP model, we are currently exploring ways to address this residual source of inaccuracy.

3.2. Small Molecule Hydration Free Energies. The validation and parametrization of the hydrogen bonding and cavity correction parameters have been performed based on the agreement between experimental and predicted AGBNP2 hydration free energies of a selected set of small molecules, listed in Table 2, containing the main functional groups present in proteins. This set of molecules includes only small and relatively rigid molecules whose hydration free energies can be reliably estimated using a single low energy representative conformation[115] as was done here. Table 2 lists for each molecule the experimental hydration free energy, the AGBNP2 hydration free energy computed without hydrogen bonding (HB) corrections and the default $\gamma = 117$ cal/mol/Å^2 surface tension parameter, denoted by AGBNP2/SEV, as well as the hydration free energy from the AGBNP2 model including the HB correction term and the parameters listed in Table 1. For comparison, the corresponding predictions with the original AGBNP1[42] model are reported in the Supporting Information.

Going down the results in Table 2, we notice a number of issues addressed by the new implementation. With the new surface area implementation and without corrections (third column in Table 2), the hydration free energies of the normal alkanes are too small compared to experiments; furthermore, in contrast with the experimental behavior, the predicted hydration free energies incorrectly become more favorable with increasing chain length. A similar behavior is observed for the aromatic hydrocarbons. Clearly this is due to the rate of increase of the positive cavity term with increasing alkane size which is insufficient to offset the solute−solvent van der Waals interaction energy term, which becomes more negative with increasing solute size. We have chosen to

Table 2. Experimental and Predicted Hydration Free Energies of a Set of Small Molecules

molecule	exptl[a,b]	AGBNP2/SEV[a,c]	AGBNP2[a,d]
n-ethane	1.83	0.98	1.80
n-propane	1.96	0.92	1.97
n-butane	2.08	0.88	2.14
n-pentane	2.33	0.78	2.26
n-hexane	2.50	0.70	2.40
cyclopentane	1.20	0.34	1.63
cyclohexane	1.23	0.05	1.50
benzene	−0.87	−1.50	−1.14
toluene	−0.89	−1.66	−0.94
acetone	−3.85	−1.09	−3.83
acetophenone	−4.58	−2.74	−5.07
ethanol	−5.01	−4.77	−5.30
phenol	−6.62	−4.51	−5.38
ethanediol	−9.60	−7.99	−9.87
acetic acid	−6.70	−2.73	−7.05
propionic acid	−6.48	−2.58	−6.38
methyl acetate	−3.32	−0.10	−3.92
ethyl acetate	−3.10	−0.02	−3.60
methyl amine	−4.56	−2.39	−4.37
ethyl amine	−4.50	−2.24	−3.95
dimethyl amine	−4.29	−1.95	−3.21
trimethyl amine	−3.24	−1.78	−2.39
acetamide	−9.71	−6.81	−10.45
N-methylacetamide	−10.08	−4.75	−7.51
pyridine	−4.70	−3.62	−5.30
2-methylpyridine	−4.63	−2.94	−4.22
3-methylpyridine	−4.77	−2.82	−4.13
methanethiol	−1.24	−0.61	−1.46
ethanethiol	−1.30	−0.57	−1.22
neutral AUE[a,e]		1.90	0.45
acetate ion	−79.90	−77.32	−87.70
propionate ion	−79.10	−76.29	−86.29
methylammonium ion	−71.30	−73.21	−73.54
ethylammonium ion	−68.40	−70.63	−70.75
methyl guanidinium	−62.02[f]	−57.30	−69.81
ions AUE[a,g]		2.85	5.47

[a] In kcal/mol. [b] Experimental hydration free energy from ref 116 except where indicated. [c] AGBNP predicted hydration free energies with the default γ parameter for all atoms types ($\gamma = 117$ cal/mol/Å^2) and without HB corrections. [d] AGBNP predicted hydration free energies with optimized parameters listed in Table 1. [e] Average unsigned error of the AGBNP predictions for the neutral compounds relative to the experiments. [f] From ref 117. [g] Average unsigned error of the AGBNP predictions for the ionic compounds relative to the experiments.

address this shortcoming by increasing by 10.2% and 2.5% respectively the effective surface tensions for aliphatic and aromatic carbon atoms rather than decreasing the corresponding α parameters of the van der Waals term since the latter had been previously validated against explicit solvent simulations. We have chosen to limit the increases of the surface tension parameters to aliphatic and aromatic carbon atoms since the results for polar functional groups did not indicate that this change was necessary for the remaining atom types. With this new parametrization we achieve (compare the second and fourth columns in Table 2) excellent agreement between the experimental and predicted hydration free energies of the alkanes and aromatic compounds. Note that the AGBNP2 model, regardless of the parametrization, correctly predicts the more favorable hydration free energies of the cyclic alkanes relative to their linear analogues. AGBNP2, thanks to its unique decomposition of the nonpolar solvation free energy into an unfavorable cavity term and an opposing favorable term, is, to our knowledge, the only analytic implicit solvent implementation capable of describing correctly this feature of the thermodynamics of hydration of hydrophobic solutes.

The AGBNP2 model without corrections markedly underpredicts the magnitudes of the experimental hydration free energies of the compounds containing carbonyl groups (ketones, organic acids, and esters). The hydration free energies of alcohols are also underpredicted but by smaller amounts. Much better agreement with the experimental hydration free energies of these oxygen-containing compounds (see Table 2) is achieved after applying hydrogen bonding corrections with $h = -1.25$ kcal/mol for the carbonyl oxygen and $h = -0.4$ kcal/mol for both the hydrogen and oxygen atoms of the hydroxy group (Table 1). Note that the same parameters employed individually for carbonyl and hydroxy groups in ketones and alcohols are applied to the more complex carboxylic groups of acids and esters as well as amides and carboxylate ions. The thiol groups of organic sulfides required similar corrections as the hydroxy groups (Tables 1 and 2). The AGBNP2 model without corrections also markedly underpredicted the magnitude of the experimental hydration free energies of amines and amides and, to a smaller extent, of compounds with nitrogen-containing heterocyclic aromatic rings. The addition of HB corrections of −0.25 kcal/mol for amine hydrogens and $h = -2.0$ kcal/mol for both amine and aromatic nitrogen atoms yields improved agreement (Table 2), although the effect of N-methylation is still overemphasized.

3.3. Miniprotein Results. As described in section 2.3, we have performed restricted MD simulations of a series of so-called miniproteins (trp-cage, cdp-1, and fsd-1) to study the extent of the agreement between the conformational ensembles generated with the original AGBNP implementation (AGBNP1) and the present implementation (AGBNP2) with respect to explicit solvent generated ensembles. The results of earlier studies[4,54,55] suggest that the AGBNP/OPLS-AA model correctly reproduces for the most part the backbone secondary structure features of protein and peptides. The tests in the present study are therefore focused on side chain conformations. The backbone atoms were harmonically restricted to remain within approximately 3 Å Cα root-mean-square deviation of the corresponding NMR experimental models. We structurally analyzed the ensembles in terms of the extent of occurrence of intramolecular interactions.

As shown in Table 3, we measured a significantly higher average number of intramolecular hydrogen bonds and ion pairing in the AGBNP1 ensembles relative to the explicit solvent ensembles for all miniproteins studied. The largest deviations are observed for cdp-1 and fsd-1, two miniproteins particularly rich in charged side chains, with on average nearly twice as many intramolecular hydrogen bonds compared to explicit solvent. Many of the excess intramolecular hydrogen bonds with AGBNP1 involve interactions between polar groups (polar side chains or the peptide backbone) and the side chains of charged residues. For example, for cdp-1 we observe (see Table 3) approximately eight hydrogen bonds between polar and charged groups on average compared to nearly none with explicit solvation.

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

Table 3. Average Number of Some Types of Intramolecular Electrostatic Interactions in the Explicit Solvent Conformational Ensembles, and the Ensembles Generated from Simulations Using the AGBNP1 and AGBNP2 Effective Potentials for the trp-cage, cdp-1, and fsd-1 Miniproteins

miniprotein	explicit	AGBNP1	AGBNP2
	Intramolecular Hydrogen Bonds		
trp-cage	13.5	18.3	15.3
cdp-1	12.6	24.5	15.4
fsd-1	14.1	24.6	14.3
all	40.2	67.4	45.0
	Polar−Polar Hydrogen Bonds		
trp-cage	12.9	17.1	13.9
cdp-1	12.5	16.4	14.1
fsd-1	12.0	15.0	12.9
all	37.4	48.5	40.9
	Polar−Charged Hydrogen Bonds		
trp-cage	0.6	1.2	1.4
cdp-1	0.1	8.1	1.3
fsd-1	2.1	9.6	1.4
all	2.8	18.9	4.1
	Ion Pairs		
trp-cage	0.3	1.0	1.0
cdp-1	2.5	2.9	2.7
fsd-1	1.4	4.6	4.0
all	4.2	8.5	7.7

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

Despite the introduction of empirical surface tension correction to penalize ion pairs,[55] AGBNP1 overpredicts ion pair formation. We found that ion pairing involving arginine was particularly overstabilized by AGBNP1 as we observed stable ion pairing between arginine and either glutamate or aspartate residues during almost the entire duration of the simulation in virtually all cases in which this was topologically feasible given the imposed backbone restrains. In contrast, with explicit solvation some of the same ion pairs were seen to form and break numerous times, indicating a balanced equilibrium between contact and solvent-separated conformations. This balance is not reproduced with implicit solvation, which instead strongly favors ion pairing. In any case, the relative stability of ion pairs appeared to depend in subtle ways on the protein environment as, for example, the two ion pairs between arginine and glutamate of cdp-1 were found to be stable with either explicit solvation or AGBNP1 implicit solvation whereas other Arg−Glu ion pairs in trp-cage and fsd-1 were found to be stable only with implicit solvation.

This analysis generally confirms quantitatively a series of past observations made in our laboratory indicating that the original AGBNP implementation tends to be biased toward conformations with excessive intramolecular electrostatic interactions, at the expense of more hydrated conformations in which polar groups are oriented so as to interact with the water solvent. During the process of development of the modifications to address these problems, we found it useful to rescore with varying AGBNP formulations and parametrizations the miniprotein conformational ensembles obtained with AGBNP1 and explicit solvation, rather than performing simulations with each new parametrizations. An example of this analysis is shown in the first row of plots of Figure 9, which compare the probability distributions of the

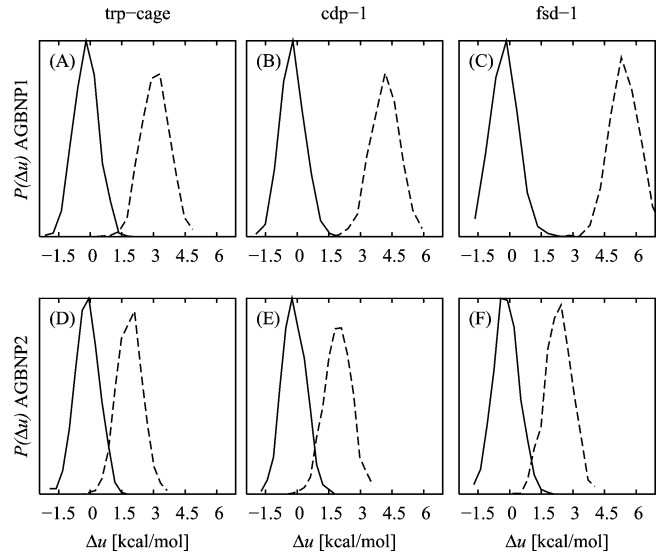


Figure 9. Potential energy distributions of the conformational ensembles for the trp-cage (first column, panels A, D), cdp-1 (second column, panels B, E), and fsd-1 (third column, panels C, F) miniproteins obtained using the AGBNP1/OPLS-AA (first row, panels A−C; full line) and AGBNP2/OPLS-AA (second row, panels D−F; full line) effective potentials and explicit solvation (dashed line). The distributions are shown as a function of the energy gap per residue (Δu) relative to the mean effective potential energy of the implicit solvent ensemble distribution.

AGBNP1 effective potential energies over the conformational ensembles generated with AGBNP1 implicit solvation and with explicit solvation. These results clearly show that the AGBNP1/OPLS-AA effective potential disfavors conformations from the explicit solvent ensemble relative to those generated with implicit solvation. The AGBNP1 energy scores of the explicit solvent ensembles of all miniproteins are shifted toward higher energies than those of the AGBNP1 ensemble, indicating that conformations present in the explicit solvent ensemble would be rarely visited when performing conformational sampling with the AGBNP1/OPLS-AA potential. AGBNP1/OPLS-AA assigns a substantial energetic penalty (see Figure 9A−C) to the explicit solvent ensemble relative to the AGBNP1 ensemble (on average 3.3, 4.4, and 5.7 kcal/mol per residue for, respectively, the trp-cage, cdp-1, and fsd-1 miniproteins). This energetic penalty, being significantly larger than thermal energy, rules out the possibility that conformational entropy effects could offset it to such an extent so as to equalize the relative free energies of the two ensembles. Detailed analysis of the energy scores shows that, as expected, the AGBNP1 implicit solvent ensemble is mainly favored by more favorable electrostatic Coulomb interaction energies due to its greater number of intramolecular electrostatic contacts relative to the explicit solvent ensemble (see above). Conversely, the AGBNP1 solvation model does not assign sufficiently favorable hydration free energy to the more solvent-exposed conformations obtained in explicit solvation so as to make them competitive with the more compact conformations of the AGBNP1 ensemble.

Similar energetic scoring analysis with the AGBNP2 model (see Figure 1 of the Supporting Information) with and

without hydrogen bonding to solvent corrections showed that the introduction of the SEV model for the solute volume significantly reduced the energetic gap between the explicit solvent and AGBNP1 conformational ensembles, and that the introduction of the hydrogen bonding corrections further favors the explicit solvent ensemble. We proceeded to vary the AGBNP2 parameters to achieve the best possible scoring of the explicit solvent ensembles relative to the AGBNP1 ensembles while maintaining an acceptable level of agreement with the small molecule experimental hydration free energies. This procedure eventually yielded the parameters listed in Table 1, which produce small molecule hydration free energies in good agreement with the experiments (Table 2), as well as energy distributions for the three miniproteins that, while still favoring the AGBNP1 ensembles, displayed energy gaps between the explicit and AGBNP1 implicit solvation ensembles comparable to thermal energy and smaller than the spread of the energy distributions.

The energy scoring experiments on the explicit solvent and AGBNP1 ensembles described above were very useful for tuning the formulation of the AGBNP2 model without requiring running lengthy MD simulations. They do not, however, guarantee that the conformational ensembles generated with the AGBNP2 solvation model will more closely match the explicit solvent ensembles than those generated with AGBNP1. This is because the new solvation model could introduce new energy minima not encountered with AGBNP1 or explicit solvation that would be visited only by performing conformational sampling with AGBNP2. To validate the model in this respect, we obtained MD trajectories with the AGBNP2 implicit solvent model and compared the corresponding probability distributions of the effective energy with those of the explicit solvent ensembles similarly as above. The results for the three miniproteins, shown in Figure 9D−F, indicate that the AGBNP2-generated ensembles display significantly smaller bias (mean energy gaps per residue of 2.0, 2.1, and 2.5 kcal/mol for, respectively, the trp-cage, cdp-1, and fsd-1 miniproteins) than AGBNP1 (Figure 9A−C), which yielded energy gaps of 3.3, 4.4, and 5.7 kcal/mol per residue, respectively. This observation shows that AGBNP2 produces conformational ensembles with energy distributions that more closely match on average that of the explicit solvent ensemble without producing unphysical minima that deviate significantly from it.

We have analyzed structural features of the conformational ensembles obtained with the AGBNP1 and AGBNP2 models to establish the degree of improvement achieved with the new model with respect to intramolecular interactions. The salient results of this analysis are shown in Table 3. This table reports for each miniprotein the average number of intramolecular hydrogen bonds and ion pairs. The number of hydrogen bonds is further decomposed into those involving only polar groups (including the backbone) and those involving a polar group and the side chain of a charged residue (arginine, lysine, aspartate, and glutamate). As noted above, it is apparent from these data that the AGBNP1 model produces conformations with too many hydrogen bonds and ion pairs. The majority of the excess hydrogen bonds with AGBNP1 involve residue side chains. Similarly, too many ion pairs are observed in the AGBNP1 ensemble particularly for the fsd-1 miniprotein (4.6 ion pairs on average with AGBNP1 compared to only 1.4 in explicit solvent). The AGBNP2 ensembles, in comparison, yield considerably fewer intramolecular hydrogen bonds. For instance, the average number of hydrogen bonds for cdp-1 is reduced from 24.5 with AGBNP1 to 15.4 with AGBNP2, which is to be compared with 12.6 in explicit solvent. With AGBNP2 the number of polar−polar hydrogen bonds is generally in good agreement with explicit solvation. However, the greatest improvement is observed with polar−charged interactions. For example, the number of polar−charged hydrogen bonds of fsd-1 is reduced by almost 10-fold in going from AGBNP1 to AGBNP2 to reach good agreement with explicit solvation. Importantly, a significant fraction of the excess polar−charged interactions observed with AGBNP1 corrected by AGBNP2 are interactions between the peptide backbone and charged side chains that would otherwise interfere with the formation of secondary structures.

With AGBNP2 we observe small but promising improvements in terms of ion pair formation. The average number of ion pairs of cdp-1 consistently agrees between all three solvation models, and the only possible ion pair in trp-cage is more stable in both implicit solvent formulations than in explicit solvent (it occurs in virtually all implicit solvent conformations compared to only 30% of the conformations in explicit solvent). However, the average number of ion pairs for fsd-1 is reduced from 4.6 with AGBNP1 to 4.0 with AGBNP2. We observe good agreement between the number of ion pairs involving lysine with either AGBNP1 or AGBNP2 and explicit solvation. However, ion pairs involving arginine are generally more stable with implicit solvation than with explicit solvation. The agreement in the number of ion pairs with cdp-1 is due to the fact that for this miniprotein the two possible ion pairs involving arginine result stable with explicit solvation as well as with implicit solvation. For the other two miniproteins, however, ion pairs involving arginine that are marginally stable with explicit solvation are found to be significantly more stable with implicit solvation, although less so with AGBNP2 solvation.

4. Discussion

Modern implicit solvent models for biomolecular simulations are generally based on the uniform dielectric continuum representation of the solvent which is accurately modeled by the Poisson−Boltzmann (PB) equation.[9] Generalized Born (GB) models,[10] which approximate the PB formalism, are applicable to molecular dynamics thanks to their low computational complexity. GB models have reached a high level of accuracy compared to PB following the introduction of more realistic solute volume descriptions[87,100] and of higher order corrections to the Coulomb field approximation.[118−120] However, at the molecular level water is sometimes poorly described by uniform continuum models. Even the best GB models have been found to deviate considerably from, for example, explicit solvent benchmarks.[121,127] The nonlinear and asymmetric dielectric response of water stems primarily from the finite extent and

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

internal structure of water molecules.[1] The modeling of effects due to water granularity is important for the proper description of molecular association equilibria. Integral equation methods[122] provide an accurate implicit solvation description from first principles; however, despite recent progress,[123] they are not yet applicable to molecular dynamics of biomolecules. The primary aim of the present study has been to formulate an analytical and computational efficient implicit solvent model incorporating solvation effects beyond those inherent in standard continuum dielectric models and, by so doing, achieving an improved description of solute conformational equilibria.

In this work we have developed the AGBNP2 implicit solvent model which is based on an empirical (but physically motivated) first solvation shell correction function parametrized against experimental hydration free energies of small molecules and the results of explicit solvent molecular dynamics simulations of a series of miniproteins. The correction function favors conformations of the solute in which polar groups are oriented so as to form hydrogen bonds with the surrounding water solvent, thereby achieving a more balanced equilibrium with respect to the competing intramolecular hydrogen bond interactions. A key ingredient of the model is an analytical prescription to identify and measure the volume of hydration sites on the solute surface. Hydration sites that are deemed too small to contain a water molecule do not contribute to the solute hydration free energy. Conversely, hydration sites of sufficient size form favorable interactions with nearby polar groups. This model thus incorporates the effects of both water granularity and nonlinear first shell hydration effects.

The GB and nonpolar models in the AGBNP2 implicit solvent model provide the linear continuum dielectric model basis of the model as well as a description of nonelectrostatic hydration effects.[42] In this work the GB and solute−solvent dispersion interaction energy models are further enhanced by replacing the original van der Waals solute volume model with a more realistic solvent excluded volume (SEV) model. The new volume description improves the quality of the Born radii of buried atoms and atoms participating in intramolecular interactions which would otherwise be underestimated due to high dielectric interstitial spaces present with the van der Waals volume description.[88] GB models with these characteristics have been previously proposed. The GBMV series of models[87,93,94] represent the SEV on a grid which, although accurate, is computationally costly and lacks frame of reference invariance. The pairwise descreening based GB^{OBC} model[120] introduced an empirical rectifying function to increase the Born radii of buried atoms in an averaged, geometry-independent manner. The GBn model[100] introduced the neck region between pairs of atoms as additional source of descreening, dampened by empirical parameters to account in an average way for overlaps between neck regions and between solute atoms and neck regions. The approach proposed here to represent the SEV consists of computing the atomic self-volumes, used in the pairwise descreening computation, using enlarged atomic radii so as to cover the interatomic interstitial spaces. The self-volume of each atom is then reduced proportionally to its solvent accessible surface area (see eq 31) to subtract the volume in van der Waals contact with the solvent. We show (Figure 8) that this model reproduces well Born radii computed from an accurate numerical representation of the SEV, noting that improvements for the Born radii of atoms in a loosely packed hydrophobic interior, while significant, are still not optimal. Although approximate, this representation of the SEV maintains the simplicity and computational efficiency of pairwise descreening schemes, while accounting for atomic overlaps in a consistent and parameter-free manner.

The new AGBNP2 model has been formulated to be employed in molecular dynamics conformational sampling applications, which require potential models of low computational complexity and favorable scaling characteristics, and with analytical gradients. These requirements have posed stringent constraints on the design of the model and the choice of the implementation algorithms. In the formulation of AGBNP2 we have reused as much as possible well-established and efficient algorithms developed earlier for the AGBNP1 model. For example, the key ingredient of the hydrogen bonding correction function is the free volume of a hydration site, which is computed using a methodology developed for AGBNP1 to compute atomic self-volumes. Similarly, the SEV-based pairwise descreening procedure employs atomic surface areas (see eq 31), computed as previously described.[42] AGBNP2 suffers additional computational cost associated with the SEV-based pairwise descreening procedure and the hydrogen bonding correction function. This handicap, however, is offset by having only one solute volume model in AGBNP2 rather than two in AGBNP1. AGBNP1 requires two separate invocations of the volume overlaps machinery (eq 2) for each of the two volume models it employs, corresponding to the van der Waals atomic radii for the pairwise descreening calculation and enlarged radii for the surface area calculation.[42] AGBNP2 instead employs a single volume model for both the pairwise descreening and surface area calculations. A direct CPU timing comparison between the two models cannot be reported at this time because the preliminary implementation of the AGBNP2 computer code used in this work lacks key data caching optimizations similar to those already employed in AGBNP1. Given the computational advantages of the new model discussed above, we expect to eventually obtain similar or better performance than with AGBNP1.

The AGBNP2 model has been parametrized against experimental hydration free energies of a series of small molecules and with respect to the ability of reproducing energetic and structural signatures of the conformational ensemble of three miniproteins generated with explicit solvation. These data sources are chosen so as to ensure that the resulting model would be applicable to both hydration free energy estimation and conformation equilibria, which are fundamental characteristics for models aimed at protein−ligand binding affinity estimation. On the other hand, experimental hydration free energies and explicit solvent conformational ensembles are to some extent incompatible with one another given the limitations of even the best fixed-charge force fields and explicit solvation models to reproduce experimental hydration free energies of small molecules with

high accuracy.[41,124,125] Mindful of this issue we did not seek a perfect correspondence with the experimental hydration free energy values. We first obtained parameters by fitting against the small molecule experimental hydration free energies and then adjusted the parameters to improve the agreement with the explicit solvent data, making sure that the predicted small molecule hydration free energies remained within a reasonable range relative to the experimental values. In practice this procedure yielded predicted hydration free energies in good agreement with the experimental values with the exception of the carboxylate and guanidinium ions (see Table 2), for which AGBNP2 predicts more favorable hydration free energies than the experiments, a consequence of the large hydrogen bonding corrections necessary to reduce the occurrence of intramolecular electrostatic interactions in the investigated proteins. As discussed further below, limitations in the description of hydration sites adopted for carboxylate and guanidinium ions may be partly the cause of the observed inconsistencies for these functional groups.

The parametrization and quantitative validation of the model, which is the primary focus of this work, has been based on comparing the effective potential energy distributions of implicit solvent conformational ensembles with those of explicit solvent ensembles. We observed that the AGBNP1 solvation model energetically ranked explicit solvent conformations significantly less favorably than implicit solvent conformations. The AGBNP2 model is characterized by smaller energetic bias relative to the explicit solvent ensembles, indicating that conformational sampling with the AGBNP2/OPLS-AA energy function produces conformations that more closely match those obtained with explicit solvation. This result is a direct consequence of employing the more realistic solvent excluded volume description of the solute, which yields larger Born radii for buried groups, as well as the hydrogen bonding to solvent corrections, which favor solvent exposed conformations of polar groups. Furthermore, comparison of the energy distributions of the AGBNP2 and explicit solvent ensembles for the three miniproteins (Figure 9D−F) shows, in contrast to the AGBNP1 results, that the AGBNP2 bias for the two more charge-rich miniproteins (cdp-1 and fsd-1) is similar to that of the least charged one (trp-cage). This suggests that the residual energetic bias of the AGBNP2 model is probably related to the nonpolar model rather than the electrostatic model. Future studies will address this particular aspect of the model.

The energy scoring studies conducted in this work indicate that AGBNP2 is a significant improvement over AGBNP1. They also show, however, that the new model falls short of consistently scoring explicit solvent conformations similarly to implicit solvent conformations. Although an optimal match between energy distributions is a necessary condition for complete agreement between implicit and explicit solvation results, it is unrealistic to expect to reach this ultimate goal at the present level of model simplification. Increasing the magnitude of the hydrogen bonding corrections can improve the agreement between the explicit and implicit solvation energy distributions, albeit at the expense of the quality of the predicted small molecule hydration free energies. It seems likely that the no parametrization of the current model would yield both good relative conformational free energies and hydration free energies. Future work will pursue the modeling of additional physical and geometrical features, such us the use of variable dielectric approaches to model polarization effects,[126] necessary to improve the agreement between implicit and explicit solvation energy distributions. The energy gap between the implicit solvent and explicit solvent energy distributions used here is, we believe, a meaningful measure of model quality and could serve as a useful general validation tool to compare the accuracy of implicit solvent models.

The excessive number of intramolecular electrostatic interactions involving charged groups has been one of the most noticeable shortcomings of GB-based implicit solvent models.[127] To correct this tendency, we have in the past adopted in the AGBNP1 model ad hoc strategies aimed at either destabilizing electrostatic intramolecular interactions[54] or, alternatively, stabilizing the competing solvent-separated conformations.[55] This work follows the latter approach using a more robust and physically motivated framework based on locating and scoring hydration sites on the solute surface as well as adopting a more realistic volume model. Structural characterization of the conformational ensembles has shown that AGBNP2 produces significantly fewer intramolecular interactions than AGBNP1, reaching good agreement with explicit solvent results. The reduction of intramolecular interactions has been the greatest for interactions between polar and charged groups. We believe the excessive tendency toward the formation of intramolecular interactions to be the root cause of the high melting temperatures of structured peptides[64] predicted with AGBNP1. Given the reduction of intramolecular interactions achieved with AGBNP2, we expect the new model to yield more reasonable peptide melting temperatures, a result which we hope to report in future publications.

Less-visible improvements have been obtained for ion pairs involving arginine side chains which remain more stable with implicit solvation than with explicit solvation. However, significantly, with AGBNP2 we observed a more dynamic equilibrium between ion-paired and solvent-separated conformations of arginine side chains that was not observed with AGBNP1. This result is promising because it indicates that the AGBNP2 solvation model, although still favoring ion-paired conformations, produces a more balanced equilibrium, which is instead almost completely shifted toward contact conformations with AGBNP1. Nevertheless it is apparent that the AGBNP treatment of the guanidinium group of arginine is not as good as for other groups. This limitation appears to be shared with other functional groups containing sp^2-hybridized nitrogen atoms as evidenced, for example, by the relatively lower quality of the hydration free energy predictions for amides and nitrogen-containing aromatic compounds (Table 2). Similar implicit solvent overstabilization solvation of arginine-containing ion pairs has been observed by Yu et al.[85] in their comparison of Surface Generalized Born (SGB) and SPC explicit solvation with the OPLS-AA force field. Despite quantitative differences, the explicit solvent studies (with the TIP3P water model) of Masunov and Lazaridis[128] and Hassan,[129] using the CHARMM force field, and that of Mandell at al.,[130] using the OPLS-AA force field, have confirmed that arginine forms the

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

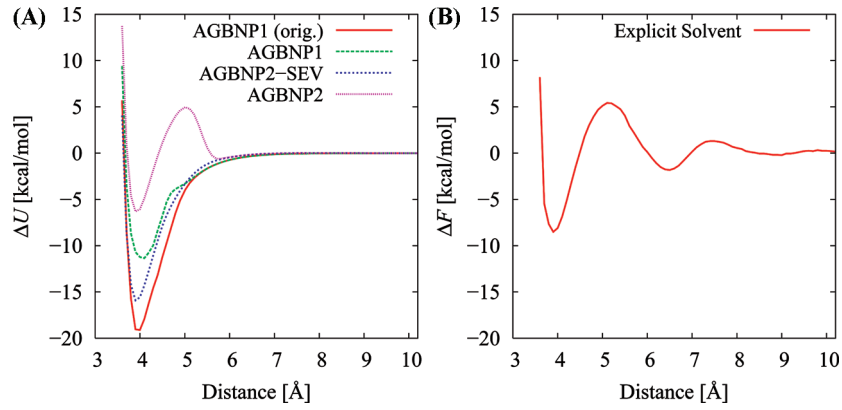


Figure 10. Potential of mean force of ion pair formation between propyl guanidinium and ethyl acetate in the coplanar orientation with AGBNP implicit solvation (A) and explicit solvation (B; ref 130). In (A) "AGBNP1 (orig.)" refers to the original AGBNP1 parametrization,[42] "AGBNP1" refers to the AGBNP1 model used in this work which includes a surface tension parameter correction for the carboxylate group aimed at reducing the occurrence of ion pairs,[55] "AGBNP2" refers to the current model, and "AGBNP2-SEV" refers to the current model without hydrogen bonding and surface tension corrections. The potentials of mean force are plotted with respect to the distance between the atoms of the protein side chain analogues equivalent to the Cζ of arginine and the Cγ of aspartate.

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

strongest ion pairing interactions, especially in the bidentate coplanar conformation. These observations are consistent with the present explicit solvent results using OPLS-AA and the SPC water model, where we find that most of the ion pairs of the miniproteins were found to involve arginine side chains. In contrast to our present implicit solvent results, however, the work of Masunov and Lazaridis[128] indicated that the GB-based implicit solvent model that they analyzed[14] produced potentials of mean force for arginine-containing ion pairs in general agreement with explicit solvation.

To rationalize the present implicit solvent results concerning ion pair formation, it has been instructive to analyze the potentials of mean force (PMFs) of ion pair association with the AGBNP model. As an example, Figure 10 shows the PMF for the association of propyl guanidinium (arginine side chain analogue) and ethyl acetate (aspartate and glutamate analogue) in a bidentate coplanar conformation (similar to the arrangement used previously)[85,128−130] for various AGBNP implementations. The corresponding explicit solvent PMF obtained by Mandell et al.[130] is also shown in Figure 10 for comparison. The original AGBNP1 parametrization[42] clearly leads to an overly stable salt bridge with the contact conformation scored at approximately −19 kcal/mol relative to the separated conformation, compared with −8.5 kcal/mol with explicit solvation. The AGBNP1 parametrization analyzed here, which includes an empirical surface area correction to reduce the occurrence of ion pairs,[55] yields a contact free energy (−11 kcal/mol) in much better agreement with explicit solvation, although the shape of the PMF is not properly reproduced. The present AGBNP2 model without hydrogen bonding corrections (labeled "AGBNP2-SEV" in Figure 10) yields a PMF intermediate between the original and corrected AGBNP1 parametrizations. The AGBNP2 model with hydrogen bonding corrections yields the PMF with the closest similarity to the one obtained in explicit solvent. Not only the contact free energy (−6.5 kcal/mol) is in good agreement with the explicit solvent result, but, importantly, it also reproduces the solvation barrier of the PMF at 5 Å separation, corresponding to the distance below which there is insufficient space for a water layer between the ions.

It is in this range of distances that the greatest discrepancies are observed between PMFs with explicit solvation and some GB-based implicit solvation models[85,128] that do not model effects due to the finite size of water molecules. Both the hydrogen bonding correction and the SEV volume description employed in AGBNP2, which are designed to take into account the granularity of the water solvent—the hydrogen bonding correction through the minimum free volume of water sites (eq 37) and the SEV model through the water radius offset (eq 28)—make it possible to reproduce the solvation barrier typical of molecular association processes in water.

It is notable in the PMF results shown in Figure 10 the lack of a free energy maximum with the AGBNP2/SEV model (AGBNP2 without HB corrections), which would be expected on the basis of results with the GBMV model, indicating that a SEV treatment of the GB model leads to a higher and much broader PMF maximum relative to explicit solvent.[88] There are two possible factors contributing to this discrepancy. The first is that the OPLS-AA force field used in this work seems to consistently produce stronger ionic interactions than the CHARMM force field (on which the GBMV model is based) as suggested by the relatively small free energies of salt bridge formation obtained with CHARMM-based implicit solvent models[88,128] relative to OPLS-AA-based ones (see for example ref 85 and the present results with AGBNP1). Because the shape of the PMF at intermediate separations is determined by a delicate balance between attractive electrostatic interactions and repulsive desolvation forces, stronger electrostatic interactions with OPLS-AA are potentially responsible in part for a missing or smaller PMF maximum. The lack of the PMF maximum with AGBNP2/SEV is most likely also due to the reduced radius offset used in AGBNP[42] used to construct the SEV. The small probe radius leads to a smaller reduction, compared to a full SEV treatment, of the high dielectric volume surrounding the ionic groups as they approach each other. The consequence is

a smaller rate of increase of the desolvation penalty which in turn leads to a smaller or absent PMF maximum.

Increasing the magnitude of the AGBNP radius offset is not feasible as we observed that the Gaussian overlap approximation for the overlap volumes (eq 10 of ref 42) breaks down for atomic radii much larger than the van der Waals radii. On the other hand, as the results in Figure 10 show, the added desolvation provided by the short-range HB function is able to properly correct this deficiency, yielding a PMF maximum in good correspondence with explicit solvation. This shows that the HB function as parametrized is likely taking into account not only short-range nonlinear hydration effects but also inaccuracies in the GB and nonpolar models, as well as approximations in the implementation such as the small probe radius discussed above.

The good correspondence between the AGBNP2 and explicit solvent PMFs for propyl guanidinium and ethyl acetate (Figure 10) stands in contrast with the residual AGBNP2 overprediction of arginine salt bridges compared to explicit solvation (Table 3). We observed that, in the majority of arginine salt bridges occurring with AGBNP2, the guanidinium and carboxylate groups interact at an angle rather than in the coplanar configuration discussed above. We have confirmed that the PMF of ion pair formation for an angled conformation (not shown) indeed shows a significantly more attractive contact free energy than the coplanar one. This result indicates that the in-plane placement of the hydration sites for the carboxylate groups (see the Appendix) does not sufficiently penalize angled ion pair arrangements. This observation is consistent with the need for introducing an isotropic surface area based hydration correction for carboxylate groups (the reduced γ parameter for the carboxylate oxygen atoms in Table 1), which showed some advantage in terms of reducing the occurrence of salt bridges. Future work will focus on developing a more general hydration shell description for carbonyl groups and related planar polar groups to address this issue.

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

5. Conclusions

We have presented the AGBNP2 implicit solvent model, an evolution of the AGBNP1 model we have previously reported, with the aim of incorporating hydration effects beyond the continuum dielectric representation. To this end a new hydration free energy component based on a procedure to locate and score hydration sites on the solute surface is used to model first solvation shell effects, such as hydrogen bonding, which are poorly described by continuum dielectric models. This new component is added to the generalized Born and nonpolar AGBNP models which have been improved with respect to the description of the solute volume description. We have introduced an analytical solvent excluded volume (SEV) model which reduces the effect of artifactual high dielectric interstitial spaces present in conventional van der Waals representations of the solute volume. The new model is parametrized and tested with respect to experimental hydration free energies and the results of explicit solvent simulations. The modeling of the granularity of water is one of the main principles employed in the design of the empirical first shell solvation function and the SEV model, by requiring that hydration sites have a minimum available volume based on the size of a water molecule. We show that the new volumetric model produces Born radii and surface areas in good agreement with accurate numerical evaluations. The results of molecular dynamics simulations of a series of miniproteins show that the new model produces conformational ensembles in much better agreement with reference explicit solvent ensembles than the AGBNP1 model with respect to both structural and energetics measures.

Future development work will focus on improving the modeling of some functional groups, particularly ionic groups involving sp^2 nitrogen, which we think are at the basis of the residual excess occurrence of salt bridges, and on the optimization of the AGBNP2 computer code implementation. Future work will also focus on further validation of the model on a wide variety of benchmarks including protein homology modeling and peptide folding.

Acknowledgment. This work was supported in part by National Institute of Health Grant GM30580. The calculations reported in this work have been performed at the BioMaPS High Performance Computing Center at Rutgers University funded in part by NIH shared instrumentation Grant 1 S10 RR022375.

Appendix A: Hydration Site Locations

Figure 11 shows the location of the hydration sites for the functional groups listed in Table 1. Each hydration site is represented by a sphere of 1.4 Å radius. The distance d_{HB} between the donor or acceptor heavy atom and the center of the hydration site sphere is set to 2.5 Å.

There is a single linear geometry for HB donor groups. The corresponding hydration site is placed at a distance d_{HB} from the heavy atom donor along the heavy atom−hydrogen bond.

Acceptor trigonal geometries have one or two hydration sites depending on whether the acceptor atom is bonded to, respectively, two or one other atom. In the former case the water site is placed along the direction given by the sum of the unit vectors corresponding to the sum of the NR_1 and NR_2 bonds (following the atom labels in Figure 11). In the latter case the W_1 site (see Figure 11) is placed in the R_1CO plane forming an angle of 120° with the CO bond. The W_2 site is placed similarly.

Acceptor tetrahedral geometries have one or two hydration sites depending on whether the acceptor atom is bonded, respectively, to three or two other atoms. In the former case the water site is placed along the direction given by the sum of the unit vectors corresponding to the sum of the NR_1, NR_2, and NR_3 bonds. In the latter case the positions of the W_1 and W_2 sites are given by

$$\mathbf{w}_1 = \mathbf{O} + d_{HB}(\cos\theta\mathbf{u}_1 + \sin\theta\mathbf{u}_2)$$

$$\mathbf{w}_2 = \mathbf{O} + d_{HB}(\cos\theta\mathbf{u}_1 - \sin\theta\mathbf{u}_2)$$

where $\mathbf{O}$ is the position of the acceptor atoms, $\theta = 104.4°$, and $\mathbf{u}_1$ and $\mathbf{u}_2$ are, respectively, the unit vectors corresponding to the OR_1 and OR_2 bonds.

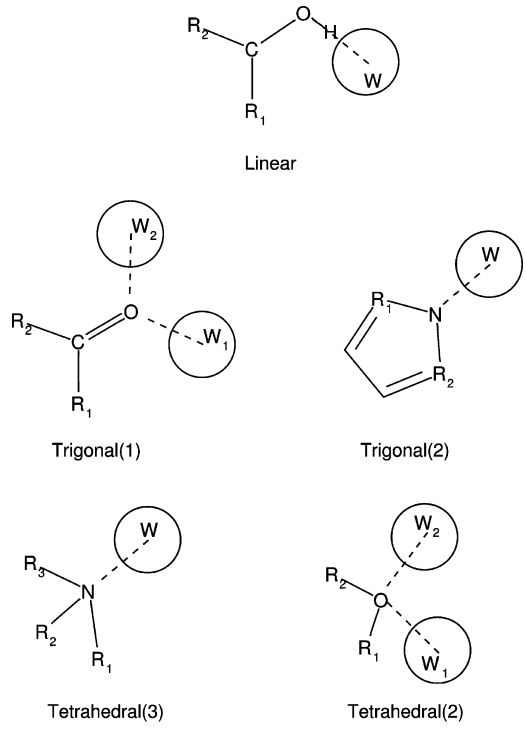


Figure 11. Diagram illustrating the hydration site locations for each of the functional group geometries used in this work. Linear, hydrogen bond donor; trigonal(1) and trigonal(2), trigonal planar geometries with, respectively, one and two covalent bonds on the acceptor atom; tetrahedral(2) and tetrahedral(3), tetrahedral geometries with, respectively, two and three covalent bonds on the acceptor atom. Representative molecular structures are shown for each geometry.

Appendix B: Gradients of GB and van der Waals Energies

The component of the gradient of the AGBNP2 van der Waals energy at constant self-volumes is the same as in the AGBNP1 model (see Appendix C of ref 42). In AGBNP2 the expression for the component of the gradient corresponding to variations in the atomic scaling factors, s_{ij}, includes pair corrections at all overlap levels because of the presence of multibody volumes in V''_{ij}. In addition, a new component corresponding to the change in surface areas appears:

$$\left(\frac{\partial \beta_j}{\partial \mathbf{r}_i}\right)_Q = -\frac{1}{4\pi}\sum_k \frac{\partial s_{kj}}{\partial \mathbf{r}_i} Q_{kj} = -\frac{1}{4\pi}\sum_k \frac{1}{V_k}\frac{\partial V'_k}{\partial \mathbf{r}_i} Q_{kj} \tag{39}$$

$$-\frac{1}{4\pi}\sum_k \frac{1}{V_k}\frac{\partial V'_{kj}}{\partial \mathbf{r}_i} Q_{kj} \tag{40}$$

$$+\frac{1}{4\pi}\sum_k \frac{1}{V_k} p_k \frac{\partial A_k}{\partial \mathbf{r}_i} Q_{kj} \tag{41}$$

Equation 39 leads to the same expression of the derivative component as in the AGBNP1 model (eq 72 in ref 42) (except for the extra elements in the two-body terms due to the inclusion of the $1/2V_{kj}$ correction term). Equation 40 corresponds to the component of the derivative due to variations in V'_{jk}, the volume to be added to the self-volumes of j and k to obtain the s_{jk} and s_{kj} scaling factors. In the AGBNP1 model this component included only two-body overlap volumes; in AGBNP2 this term instead includes all overlap volumes greater than zero. Finally, eq 41, where A_k is the surface area of atom k, leads to the component of the derivatives of the GB and vdW terms due to variations of the exposed surface area. The latter two terms are new for AGBNP2.

B.1. Component of Derivative from eq 40. From eq 63 in ref 42 and eq 40 we have

$$-4\pi\left(\frac{\partial \Delta G_{\mathrm{vdW}}}{\partial \mathbf{r}_i}\right)_{Q2} = \sum_{jk} W_{kj}\frac{\partial V'_{kj}}{\partial \mathbf{r}_i} \tag{42}$$

where W_{kj} has the same expression as in eq 69 in ref 42. In working with eq 42 it is important to note that, whereas V'_{kj} is symmetric with respect to swapping the j and k indices, W_{kj} and W_{jk} are different from each other. Substituting eq 30 into eq 42 and expanding over symmetric terms we obtain

$$-4\pi\left(\frac{\partial \Delta G_{\mathrm{vdW}}}{\partial \mathbf{r}_i}\right)_{Q2} = \frac{1}{2}\sum_{jk} W_{kj}\frac{\partial V_{kj}}{\partial \mathbf{r}_i} - \frac{1}{3}\sum_{jkl} W_{kj}\frac{\partial V_{jkl}}{\partial \mathbf{r}_i} + \frac{1}{2}\frac{1}{4}\sum_{jklp} W_{kj}\frac{\partial V_{jklp}}{\partial \mathbf{r}_i} - \ldots \tag{43}$$

Equation 43 is simplified by noting that

$$\frac{\partial V_{jk\ldots}}{\partial \mathbf{r}_i} = \delta_{ij}\frac{\partial V_{ik\ldots}}{\partial \mathbf{r}_i} + \delta_{ik}\frac{\partial V_{ji\ldots}}{\partial \mathbf{r}_i} + \ldots \tag{44}$$

Equation 44 is inserted in eq 43 and sums are reduced accordingly; then symmetric terms are collected into single sums by reindexing the summations, obtaining

$$-4\pi\left(\frac{\partial \Delta G_{\mathrm{vdW}}}{\partial \mathbf{r}_i}\right)_{Q2} = \frac{1}{2}\sum_j (W_{ij}+W_{ji})\frac{\partial V_{ij}}{\partial \mathbf{r}_i} - \frac{1}{3}\sum_{j<k}[(W_{ij}+W_{ji})+(W_{jk}+W_{kj})+(W_{ik}+W_{ki})]\frac{\partial V_{ijk}}{\partial \mathbf{r}_i} + \frac{1}{4}\sum_{j<k<l}[(W_{ij}+W_{ji})+(W_{ik}+W_{ki})+(W_{il}+W_{li})+(W_{jk}+W_{kj})+(W_{jl}+W_{lj})+(W_{kl}+W_{lk})]\frac{\partial V_{ijkl}}{\partial \mathbf{r}_i} - \ldots \tag{45}$$

The corresponding expression for the gradient of ΔG_{GB} is similar but employs the U_{ij} factors of eq 78 of ref 42 rather than W_{ij}.

B.2. Component of Derivative from eq 41. Inserting eq 41 in eq 63 of ref 42 gives

$$4\pi\left(\frac{\partial \Delta G_{\mathrm{vdW}}}{\partial \mathbf{r}_i}\right)_{Q3} = \sum_{jk} W_{kj}p_k\frac{\partial A_k}{\partial \mathbf{r}_i} = \sum_k W_k p_k \frac{\partial A_k}{\partial \mathbf{r}_i}$$

which is the same expression as that for the gradient of ΔG_{cav} (see Appendix A of ref 42) with the replacement

$$\gamma_k \rightarrow \frac{1}{4\pi}W_k p_k$$

The corresponding expression for the gradient of ΔG_{GB} follows from the substitution:

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

$$\gamma_k \rightarrow \frac{1}{4\pi} U_k p_k$$

B.3. Derivatives of HB Correction Energy. From eq 37 we have

$$\frac{\partial \Delta G_{\mathrm{hb}}}{\partial \mathbf{r}_i} = \sum_s h_s S'(w_s) \frac{\partial w_s}{\partial \mathbf{r}_i} \tag{46}$$

Inserting eqs 35 and 36 in eq 46 gives

$$\frac{\partial \Delta G_{\mathrm{hb}}}{\partial \mathbf{r}_i} = -\sum_{sj} \frac{h_s S'(w_s)}{V_s} \frac{\partial V_{sj}}{\partial \mathbf{r}_i} + \sum_{s,j<k} \frac{h_s S'(w_s)}{V_s} \frac{\partial V_{sjk}}{\partial \mathbf{r}_i} - \ldots \tag{47}$$

where

$$\frac{\partial V_{sjk\ldots}}{\partial \mathbf{r}_i} = \left(\frac{\partial V_{sjk\ldots}}{\partial \mathbf{r}_i}\right)_{\mathbf{r}_s} + \frac{\partial \mathbf{r}_s}{\partial \mathbf{r}_i} \frac{\partial V_{sjk\ldots}}{\partial \mathbf{r}_s} \tag{48}$$

where the first term on the right-hand side represents the derivative of the overlap volume with respect to the position of atom i keeping the position of the water site s fixed, and the second term reflects the change of overlap volume due to a variation of the position of the water site caused by a shift in position of atom i. The latter term is nonzero only if i is one of the parent atoms of the water site.

Supporting Information Available: Figure showing potential energy distributions of the AGBNP1 and explicit solvent conformational ensembles for the the trp-cage, cdp-1, and fsd-1 miniproteins scored with the AGBNP2-SEV/OPLS-AA and AGBNP2/OPLS-AA effective potentials; table listing experimental and AGBNP1 predicted hydration free energies of the set of small molecules in Table 2. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531–567.

(2) Feig, M.; Brooks, C. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–224.

(3) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1–20.

(4) Felts, A. K.; Andrec, M.; Gallicchio, E.; Levy, R. Protein Folding and Binding: Effective Potentials, Replica Exchange Simulations, and Network Models. In *Water and Biomolecules—Physical Chemistry of Life Phenomena*; Springer Science: New York, 2008.

(5) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517–529.

(6) Onufriev, A. *Annu. Rep. Comput. Chem.* **2008**, *4*, 125–137.

(7) Chen, J.; Brooks, C.; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140–148.

(8) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094.

(9) Baker, N. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137–143.

(10) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrikson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–129.

(11) Zhang, L.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591–607.

(12) Schaefer, M.; Froemmel, C. *J. Mol. Biol.* **1990**, *216*, 1045–1066.

(13) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824–19839.

(14) Dominy, B. N.; Brooks, C. L. I. *J. Phys. Chem. B* **1999**, *103*, 3765–3773.

(15) Banks, J.; et al. *J. Comput. Chem.* **2005**, *26*, 1752–1780.

(16) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(17) Ben-naim, A. *Hydrophobic Interactions*; Plenum Press: New York, 1980.

(18) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1–63.

(19) Dill, K. A. *Biochemistry* **1990**, *29*, 7133–7155.

(20) Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1993**, *232*, 660–679.

(21) Honig, B.; Yang, A.-S. *Adv. Protein Chem.* **1995**, *46*, 27–58.

(22) Sturtevant, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 2236–2240.

(23) Williams, D. H.; Searle, M. S.; Mackay, J. P.; Gerhard, U.; Maplestone, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1172–1178.

(24) Froloff, N.; Windemuth, A.; Honig, B. *Protein Sci.* **1997**, *6*, 1293–1301.

(25) Siebert, X.; Hummer, G. *Biochemistry* **2002**, *41*, 2965–2961.

(26) Ooi, T.; Oobatake, M.; Nemethy, G.; Sheraga, H. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086–3090.

(27) Lee, M. R.; Duan, Y.; Kollman, P. A. *Proteins* **2000**, *39*, 309–316.

(28) Hünenberger, P. H.; Helms, V.; Narayana, N.; Taylor, S. S.; McCammon, J. A. *Biochemistry* **1999**, *38*, 2358–2366.

(29) Simonson, T.; Brünger, A. T. *J. Phys. Chem.* **1994**, *98*, 4683–4694.

(30) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(31) Rapp, C. S.; Friesner, R. A. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 173–183.

(32) Fogolari, F.; Esposito, G.; Viglino, P.; Molinari, H. *J. Comput. Chem.* **2001**, *22*, 1830–1842.

(33) Pellegrini, E.; Field, M. J. *J. Phys. Chem. A* **2002**, *106*, 1316–1326.

(34) Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Ruiz-Lòpez, M. F.; Rinaldi, D.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2003**, *24*, 284–297.

(35) Jorgensen, W.; Ulmschneider, J.; Tirado-Rives, J. *J. Phys. Chem. B* **2004**, *108*, 16264–16270.

(36) Wallqvist, A.; Covell, D. G. *J. Phys. Chem.* **1995**, *99*, 13118–13125.

(37) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem. B* **2000**, *104*, 6271–6285.

(38) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. *J. Am. Chem. Soc.* **2003**, *25*, 9523–9530.

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u

(39) Wagoner, J.; Baker, N. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8331–8336.

(40) Chen, J.; Brooks, C. *Phys. Chem. Chem. Phys.* **2008**, *10*, 471–481.

(41) Mobley, D.; Bayly, C.; Cooper, M.; Shirts, M.; Dill, K. *J. Chem. Theory Comput.* **2009**, *5*, 350–358.

(42) Gallicchio, E.; Levy, R. *J. Comput. Chem.* **2004**, *25*, 479–499.

(43) Wallqvist, A.; Gallicchio, E.; Levy, R. M. *J. Phys. Chem. B* **2001**, *105*, 6745–6753.

(44) Huang, D. M.; Chandler, D. *J. Phys. Chem. B* **2002**, *106*, 2047–2053.

(45) Zhou, R.; Huang, X.; Margulis, C.; Berne, B. *Science* **2004**, *305*, 1605–1609.

(46) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717–726.

(47) Hummer, G.; Garde, S.; García, A. E.; Paulaitis, M. E.; Pratt, L. R. *J. Phys. Chem. B* **1998**, *102*, 10469–10482.

(48) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570–4577.

(49) Pitarch, J.; Moliner, V.; Pascual-Ahuir, J.-L.; Silla, A.; Tuñón, I. *J. Phys. Chem.* **1996**, *100*, 9955–9959.

(50) Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. *J. Am. Chem. Soc.* **1999**, *121*, 9243–9244.

(51) Pitera, J. W.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2001**, *123*, 3163–3164.

(52) Zacharias, M. *J. Phys. Chem. A* **2003**, *107*, 3000–3004.

(53) Su, Y.; Gallicchio, E. *Biophys. Chem.* **2004**, *109*, 251–260.

(54) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 310–321.

(55) Felts, A.; Gallicchio, E.; Chekmarev, D.; Paris, K.; Friesner, R.; Levy, R. *J. Chem. Theory Comput.* **2008**, *4*, 855–858.

(56) Dong, F.; Wagoner, J.; Baker, N. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4889–4902.

(57) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578–1599.

(58) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, C. W. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.

(59) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489–2498.

(60) Schaefer, M.; Bartels, C.; Leclerc, F.; Karplus, M. *J. Comput. Chem.* **2001**, *22*, 1857–1879.

(61) Chekmarev, D.; Ishida, T.; Levy, R. *J. Phys. Chem. B* **2004**, *108*, 19487–19495.

(62) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6801–6806.

(63) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. *J. Phys. Chem. B* **2005**, *109*, 6722–6731.

(64) Weinstock, D.; Narayanan, C.; Felts, A. K.; Andrec, M.; Levy, R.; Wu, K.; Baum, J. *J. Am. Chem. Soc.* **2007**, *129*, 4858–4859.

(65) Weinstock, D.; Narayanan, C.; Baum, J.; Levy, R. *Protein Sci.* **2008**, *17*, 950–954.

(66) Ravindranathan, K.; Gallicchio, E.; Levy, R. *J. Mol. Biol.* **2005**, *353*, 196–210.

(67) Messina, T.; Talaga, D. *Biophys. J.* **2007**, *93*, 579–585.

(68) Ravindranathan, K.; Gallicchio, E.; Friesner, R. A.; McDermott, A. E.; Levy, R. M. *J. Am. Chem. Soc.* **2006**, *128*, 5786–5791.

(69) Ravindranathan, P.; Gallicchio, E.; McDermott, A.; Levy, R. *J. Am. Chem. Soc.* **2007**, *129*, 474–475.

(70) Su, Y.; Gallicchio, E.; Das, K.; Arnold, E.; Levy, R. *J. Chem. Theory Comput.* **2007**, *3*, 256–277.

(71) Lapelosa, M.; Gallicchio, E.; Ferstandig Arnold, G.; Arnold, E.; Levy, R. M. *J. Mol. Biol.* **2009**, *385*, 675–691.

(72) Tjong, H.; Zhou, H. *J. Phys. Chem. B* **2007**, *111*, 3055–3061.

(73) Tjong, H.; Zhou, H. *J. Chem. Phys.* **2007**, *126*, 195102.

(74) Zhu, J.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2005**, *109*, 3008–3022.

(75) Fan, H.; Mark, A. E.; Zhu, J.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6760–6764.

(76) Grant, J. A.; Pickup, B.; Sykes, M. J.; Kitchen, C.; Nicholls, A. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4913–4922.

(77) Labute, P. *J. Comput. Chem.* **2008**, *29*, 1693–1698.

(78) Levy, R.; Belhadj, M.; Kitchen, D. *J. Chem. Phys.* **1991**, *95*, 3627–3633.

(79) Alper, H.; Levy, R. M. *J. Phys. Chem.* **1990**, *94*, 8401–8403.

(80) Morozov, A.; Kortemme, T. *Adv. Protein Chem.* **2005**, *72*, 1–38.

(81) Lazaridis, T.; Karplus, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 139–145.

(82) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199–203.

(83) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133–152.

(84) Vitalis, A.; Pappu, R. *J. Comput. Chem.* **2009**, *30*, 673–699.

(85) Yu, Z.; Jacobson, M.; Josovitz, J.; Rapp, C.; Friesner, R. *J. Phys. Chem. B* **2004**, *108*, 6643–6654.

(86) Okur, A.; Wickstrom, L.; Simmerling, C. *J. Chem. Theory Comput.* **2008**, *4*, 488–498.

(87) Lee, M. S.; Salsbury, F. R.; Brooks, C. L., III. *J. Chem. Phys.* **2002**, *116*, 10606.

(88) Swanson, J. M. J.; Mongan, J.; McCammon, J. A. *J. Phys. Chem. B* **2005**, *109*, 14769–14772.

(89) Lee, B.; Richards, F. *J. Mol. Biol.* **1971**, *55*, 379–400.

(90) Pascual-Ahuir, J. L.; Silla, E. *J. Comput. Chem.* **1990**, *11*, 1047–1060.

(91) Cortis, C. M.; Friesner, R. A. *J. Comput. Chem.* **1997**, *18*, 1591–1608.

(92) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128–137.

(93) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1348–1356.

(94) Chocholousova, J.; Feig, M. *J. Comput. Chem.* **2006**, *27*, 719–729.

(95) Grant, J. A.; Pickup, B. T. *J. Phys. Chem.* **1995**, *99*, 3503–3510.

(96) Kratky, K. W. *J. Stat. Phys.* **1981**, *25*, 619–634.

(97) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(98) Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381–1392.

(99) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

(100) Mongan, J.; Simmerling, C.; McCammon, J.; Case, D.; Onufriev, A. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.

(101) Liu, Y.; Liu, Z.; Androphy, E.; Chen, J.; Baleja, J. *Biochemistry* **2004**, *43*, 7421–7431.

(102) Dahiyat, B.; Sarisky, C.; Mayo, S. *J. Mol. Biol.* **1997**, *273*, 789–796.

(103) Dahiyat, B.; Mayo, S. *Science* **1997**, *278*, 82–87.

(104) Snow, C.; Zagrovic, B.; Pande, V. *J. Am. Chem. Soc.* **2002**, *124*, 14548–14549.

(105) Pitera, J.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587–7592.

(106) Zhou, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.

(107) Paschek, D.; Hempel, S.; García, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17754–17759.

(108) Nosè, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

(109) Hoover, W. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(110) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(111) Bowers, K.; Chow, E.; Xu, H.; Dror, R.; Eastwood, M.; Gregersen, B.; Klepeis, J.; Kolossváry, I.; Moraes, M.; Sacerdoti, F.; Salmon, J.; Shan, Y.; Shaw, D. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*; IEEE: Tampa, FL, 2006.

(112) Martyna, G.; Tobias, D.; Klein, M. *J. Comput. Phys.* **1994**, *101*, 4177–4189.

(113) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(114) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297–1304.

(115) Mobley, D.; Chodera, J.; Dill, K. *J. Phys. Chem. B* **2008**, *112*, 938–946.

(116) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563–595.

(117) Vorobyov, I.; Li, L.; Allen, T. *J. Phys. Chem. B* **2008**, *112*, 9588–9602.

(118) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983–10990.

(119) Im, W.; Lee, M.; Brooks, C. *J. Comput. Chem.* **2003**, *24*, 1691–1702.

(120) Onufriev, A.; Bashford, D.; Case, D. *Proteins* **2004**, *55*, 383–394.

(121) Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 1846–1857.

(122) Yoshida, N.; Imai, T.; Phongphanphanee, S.; Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **2009**, *113*, 873–886.

(123) Miyata, T.; Hirata, F. *J. Comput. Chem.* **2008**, *29*, 871–882.

(124) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2004**, *108*, 16567–16576.

(125) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(126) Zhu, K.; Shirts, M.; Friesner, R. *J. Chem. Theory Comput.* **2007**, *3*, 2108–2119.

(127) Zhou, R.; Berne, B. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777–12782.

(128) Masunov, A.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 1722–1730.

(129) Hassan, S. *J. Phys. Chem. B* **2004**, *108*, 19501–19509.

(130) Mandell, D. J.; Chorny, I.; Groban, E.; Wong, S.; Levine, E.; Rapp, C.; Jacobson, M. *J. Am. Chem. Soc.* **2007**, *129*, 820–827.

CT900234U

Downloaded by UNIV OF TEXAS AUSTIN on October 13, 2009 | http://pubs.acs.org
Publication Date (Web): July 31, 2009 | doi: 10.1021/ct900234u