

9 Fast Analytical Continuum Treatments of Solvation

François Marchand and Amedeo Caflisch

9.1 Introduction

Successful applications of molecular dynamics (MD) to study the structure and function of a biomolecule depend on the quality of the underlying force field and the sampling efficiency of the simulation protocol. In particular, an accurate representation of the aqueous solvent environment is important to reproduce the structural, functional, and dynamic behavior of soluble biomolecules. The most realistic and physically rigorous way to treat solvation effects is to include explicitly the solvent molecules in the simulation system, at the price of high computational cost. In fact, the solvent molecules greatly increase the number of degrees of freedom and interaction centers. Even with today's computational infrastructure, simulations of single-domain proteins (about 100 residues) cannot sample more than 0.1–1 μ s. Such a short time scale prohibits the study of long-time processes like protein folding, large-scale structural transitions, multimeric assembly processes like complex formation and protein aggregation, as well as the derivation of accurate thermodynamic quantities. This computational drawback has motivated the development of fast implicit solvent models [1–3], where the mean influence of solvent molecules around the solute is described by a potential of mean force that depends only on the atom coordinates of the solute [2, 4]. An implicit solvent model not only considerably reduces the system size, but also avoids the need to average over the extremely large number of solvent configurations, and reduces the viscosity of the solvent environment by eliminating the friction from the solvent molecules, thus accelerating molecular motions [5]. Furthermore, such a model directly yields the so-called effective energy, which is the sum of the solute potential energy *in vacuo* and the solvation free energy. In contrast, explicit water simulations have to be post-processed, for example by finite-difference Poisson–Boltzmann calculations, to obtain the effective energy.

The overall free energy cost of solvating a solute molecule (ΔG_{solv}) is decomposed into a non-polar component and a polar component in most implicit solvent models [2]: $\Delta G_{\text{solv}} = \Delta G_{\text{pol}} + \Delta G_{\text{nonpol}}$. The term ΔG_{pol} is the free-energy change in

the system resulting from the electrostatic interactions. The reorientation and polarization of the individual molecules in the medium cause the solvent to act on the polar contribution in two ways: first, it interacts directly with the individual charges, giving rise to the so-called self-energy contribution to the total free energy of the system; and second, it screens the strength of the Coulomb interactions between charges in the macromolecule. The term ΔG_{nonpol} is the free energy of introducing the solute into the solvent when the electrostatic interactions between the solute and solvent are turned off. It can be further decomposed into a cavity formation term (ΔG_{cav}) and a solute–solvent van der Waals dispersion term (ΔG_{VDW}) [6].

Implicit solvent models can be classified into three main families: surface area models [7–9] (which are the simplest and were developed first), Gaussian solvent-exclusion models [10, 11], and dielectric continuum electrostatics models. The latter can be further classified into finite-difference Poisson–Boltzmann (PB) [12] and generalized Born (GB) [13, 14] models. While PB models are more accurate than GB models, they suffer from high computational cost and difficulties in the derivation of forces. The GB model is related to the PB model but contains several approximations that increase the speed of calculation. There exist models that combine the different approximations, like “generalized Born surface area” (GBSA) [15–17], where the polar part is treated through the GB formalism and the non-polar part through a surface area term.

Here, two fully analytical implicit solvent models are reviewed, the SASA (“solvent-accessible surface area”) model [18], a surface area model, and the FACTS (“fast analytical continuum treatment of solvation”) model [19], a recently introduced GBSA method. SASA and FACTS are both very efficient (only about 1.5 and 4 times slower, respectively, than *in vacuo*) and have been implemented in CHARMM [20]. Because they are fully analytical energy functions, analytical force vectors and the Hessian matrix of second derivatives, which is used in techniques like normal-mode analysis (NMA) [21, 22], can be derived.

9.2

The SASA Implicit Solvent Model: A Fast Surface Area Model

It is assumed that the main contributions to the solvation energy are proportional to the solvent-accessible surface area (SASA) [7] or solvent-accessible volume [10]. Several parameterizations have been proposed in the past [7–9, 23]. The SASA model implemented in CHARMM makes use of a very efficient analytical evaluation of the SASA [24] and was parameterized for the polar hydrogen force field (param19). In SASA, electrostatic screening effects are approximated by a distance-dependent dielectric function and ionic groups are neutralized [11]. The surface area approximation is used for the direct solvation (both polar and non-polar) as introduced by Eisenberg and McLachlan [7]. Because an exact analytical or numerical computation of the SASA is too slow to compete with simulations in explicit solvent, an approximate analytical expression [24] was used. This drastically

reduces the computational cost with respect to an explicit solvent simulation. The model discussed here is based on the assumptions that most of the solvation energy arises from the first water shell around the protein [7], and that two atomic solvation parameters are sufficient to describe the solvation of polar groups (negative, that is, favorable, surface tension-like parameter) and non-polar groups (positive surface tension-like parameter).

Here a description of the SASA implicit solvent model and its calibration is given. This is followed by a discussion on the limitations of the model and a review of its application in studies of conformational transitions of structured peptide, aggregation of peptidic systems, and ligand–receptor interactions.

9.2.1

Description of the Model

In most empirical force fields, the Hamiltonian of the solute–solvent system is additive and consists of the sum of solute–solute, solute–solvent, and solvent–solvent terms. After integration over the solvent coordinates, the potential of mean force $W(\mathbf{r})$, or effective energy, can be divided into two contributions,

$$W(\mathbf{r}) = E_{\text{solute}}(\mathbf{r}) + G_{\text{solv}}(\mathbf{r}) \quad (9.1)$$

for a solute having N atoms with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$. The term “effective energy” for $W(\mathbf{r})$ is used here as in Ref. [11]; it is the sum of intra-solute and mean solvation terms. In the present study, we assume that the mean solvation energy is linearly related to the SASA of the solute:

$$G_{\text{solv}}(\mathbf{r}) = \sum_{i=1}^N \sigma_i S_i(\mathbf{r}) \quad (9.2)$$

where σ_i and $S_i(\mathbf{r})$ are the atomic solvation parameter and SASA of atom i , respectively. The SASA $S_i(\mathbf{r})$ is computed by an approximate analytical expression [24]:

$$S_i(\mathbf{r}) = A_i \prod_{j \neq i}^N [1 - p_i p_j b_{ij}(r_{ij}) / A_i] \quad (9.3)$$

where A_i denotes the SASA of an isolated atom i of radius R_i ,

$$A_i = 4\pi(R_i + R_{\text{probe}})^2 \quad (9.4)$$

and R_{probe} is the radius of the solvent probe. In Equation 9.3, $b_{ij}(r_{ij})$ represents the SASA removed from A_i due to the overlap between atoms i and j separated by a distance $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and is given by

$$b_{ij}(r_{ij}) = \begin{cases} 0 & \text{if } r_{ij} > R_i + R_j + 2R_{\text{probe}} \\ \pi(R_i + R_{\text{probe}})(R_i + R_j + 2R_{\text{probe}} - r_{ij})[1 + (R_j - R_i)r_{ij}^{-1}] & \text{otherwise} \end{cases} \quad (9.5)$$

Using 270 small molecules, the atom type parameters p_i and connectivity parameters p_{ij} have been optimized to reproduce the exact SASA with $R_{\text{probe}} = 1.4 \text{ \AA}$

[24]. The complete list of parameters can be found in the original publication [18].

The SASA model includes the free-energy cost of burying a charged residue in the interior of a protein. However, it does not take into account the solvent screening on the interactions between solute charges. This effect is approximated here using a distance-dependent dielectric function, $\epsilon(r) = 2r$ and was chosen instead of $\epsilon(r) = r$ mainly to reduce the strength of the hydrogen bonds. Larger values of the dielectric constant are expected to lead to partial unfolding of proteins in simulations at room temperature. A cutoff for long-range interactions is used (see below), so that a linear distance-dependent dielectric function does not differ significantly from a more sophisticated one, such as a sigmoidal function [25–27], because the deviation from linearity is negligible for distances smaller than 10 Å [27]. A distance-dependent dielectric function is a very simplified way of accounting for the solvent polarization effects on the solute. In particular, the screening of the electrostatic interactions between charged groups is insufficient, as shown by the formation of too stable salt bridges in MD simulations of the RGDW (Arg–Gly–Asp–Trp) peptide [28]. The limitations of this approximation can be partly overcome by using a set of partial charges with a zero total charge for every residue. In the current SASA implementation, the ionizable amino acids are neutralized [11].

The solvation model has been implemented in CHARMM and is used with a polar hydrogen CHARMM force field (param19), where the only modified parameters are the partial charges of the ionic side chains [11]. The CHARMM param19 default cutoff for long-range interactions is used (that is, a shift function [20] is used with a cutoff at 7.5 Å for both the electrostatic and van der Waals terms). This cutoff length was chosen to be consistent with the original parameterization of CHARMM param19 [20]. Even though the SASA solvation term is calculated at every dynamics step, the CPU time required for simulations with SASA is only about 50% larger than that for a simulation *in vacuo* with the same cutoff (7.5 Å).

As in a previous work [9], only two σ parameters are considered in SASA: one for carbon and sulfur atoms ($\sigma_{C,S} > 0$), and one for nitrogen and oxygen atoms ($\sigma_{N,O} < 0$). The solvation parameter of the hydrogen atoms is set to zero. The two σ parameters were optimized from 1 ns MD simulations of six small proteins at 300 K by a trial-and-error approach. The two resulting σ values that gave the minimal C_α root-mean-square deviation (RMSD) from the native state are 0.012 kcal mol⁻¹ Å⁻² for carbon and sulfur atoms, and -0.060 kcal mol⁻¹ Å⁻² for nitrogen and oxygen atoms, and correspond to those determined previously [9]. With this parameter set and the CHARMM param19 force field, SASA seems to correctly model the strength of hydrogen bonds; the MD simulations of the six small proteins showed that SASA closely reproduces the number of hydrogen bonds present in the respective X-ray structure, while other solvent models or electrostatic treatment (often) overestimate it [18, 29]. A correct treatment of the strength of hydrogen bonds is important to obtain meaningful energetics in folding–unfolding studies.

9.2.2

Applications of the SASA Implicit Solvent Model

9.2.2.1 Reversible Folding of Structured Peptides

The combination of the fast implicit solvent SASA with the united-atom force field param19, which results in a 40% reduction of the number of atoms compared to all-atom force fields and makes use of short cutoffs, allows the fast and extensive sampling of the conformational space of small to medium-sized systems. One system that was thoroughly studied is β 3s, a designed 20-residue peptide whose solution conformation has been investigated by nuclear magnetic resonance (NMR) spectroscopy [30]. The NMR data indicate that β 3s in aqueous solution forms a monomeric triple-stranded antiparallel β -sheet, in equilibrium with the denatured state. In Rao and Caffisch [31] the conformations sampled during long equilibrium folding–unfolding MD simulations ($>10\mu\text{s}$ in total) were mapped onto a network, with nodes representing clusters of similar conformations and links representing the observed transitions between nodes. With this representation, free-energy minima and their connectivity emerge without requiring projections onto arbitrarily chosen reaction coordinates (Figure 9.1). As previously observed for a variety of networks as diverse as the Internet and the protein interactions within a cell, the conformational space network of polypeptide chains is a scale-free network, that is, the distribution of the number of possible connections of a conformation follows a power law. Interestingly, a correlation was found between the statistical weight (size of the node) and connectivity (number of links to a node) – the most connected nodes are also low-lying minima on the free-energy landscape.

Another observation was that the native basin of the structured peptide shows a hierarchical organization of conformations. Such an organization was not observed for a random heteropolymer that lacks a native state (that is, a predominant free-energy minimum) [31]. The network projection allows the representation of the complexity of the denatured state ensemble, which is very heterogeneous and includes high-entropy, high-enthalpy conformations as well as low-entropy, low-enthalpy traps. Furthermore, the network properties were used to identify transition-state conformations and two main average folding pathways. Such a complexity of the conformational space and kinetic pathways disappears in conventional projections onto one or two progress variables [32]. Other applications of SASA include the investigation of the folding mechanism of structured peptides [33–36] and small proteins [37], as well as the reversible mechanical unfolding of a helical peptide [38]. MD simulations with SASA were used to interpret the kinetic behavior of a photo-switchable cross-linked α -helical peptide [39] and to test a new method to compute the density of states of proteins [40].

9.2.2.2 Peptide Aggregation

SASA was also successfully applied in aggregation studies of amyloidogenic peptides. Simulations of the early steps of aggregation of amyloid-forming peptides

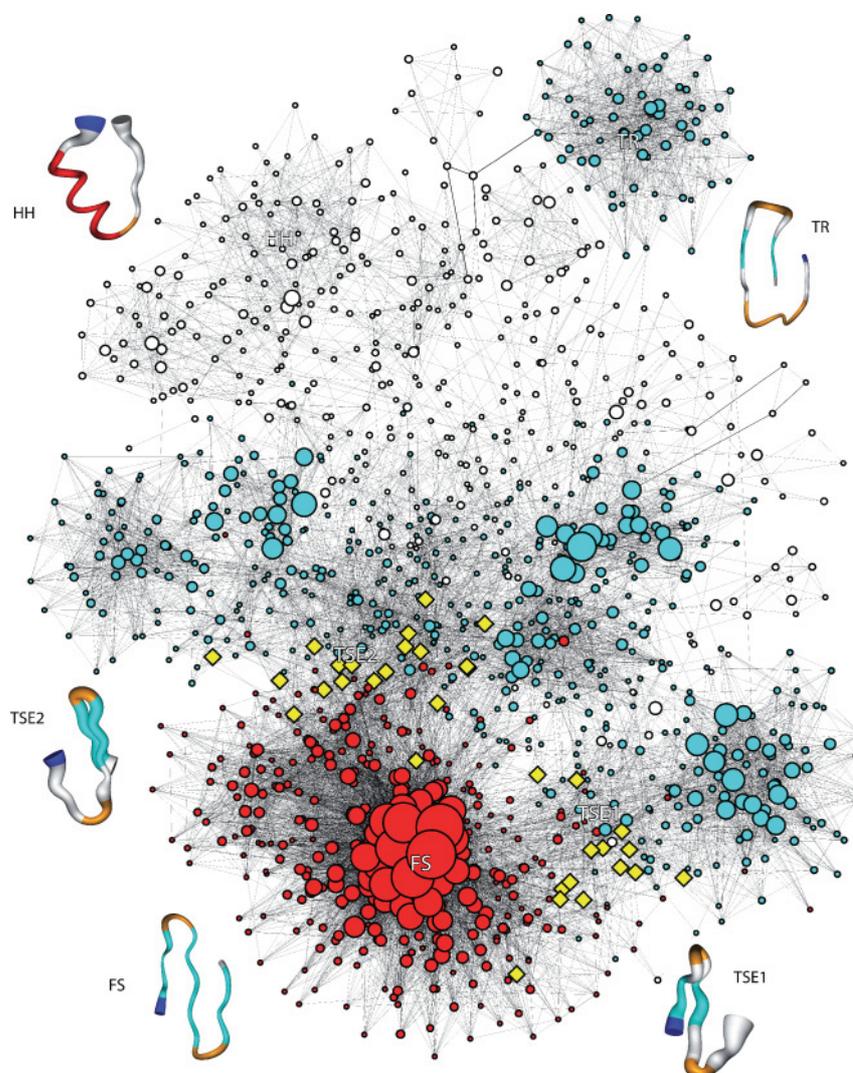


Figure 9.1 Conformational space network of the designed three-stranded antiparallel β -sheet peptide β 3s. Nodes represent conformations and links represent transitions between them, as sampled during 10 μ s implicit solvent molecular dynamics simulations at the melting temperature of 330 K. The size and color of the nodes reflect the statistical weight and average neighbor connectivity, respectively [31]. Representative conformations are shown by a pipe colored

according to secondary structure: white for coil, red for α -helix, orange for turn or bend, cyan for β -strand, and blue for the N-terminus. The variable radius of the pipe reflects the structural variability of the snapshots within a node. The yellow diamonds are folding transition state conformations. HH, TR, TSE, and FS are the helical, trap, transition state ensemble, and folded states, respectively. Reproduced from [31] with permission from Elsevier.

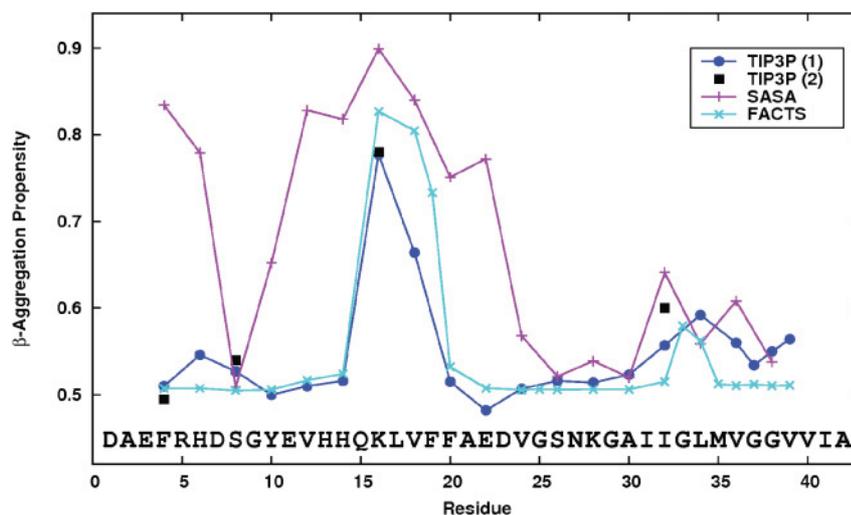


Figure 9.2 The β -aggregation propensity profile of the Alzheimer's amyloid- β peptide ($A\beta_{42}$). The peptide was decomposed into a set of overlapping heptapeptide segments, each shifted by two residues along the sequence, and three copies of each segment were simulated at 310K with param22 and TIP3P explicit water (blue), SASA (magenta), and FACTS param19 with $\epsilon_m = 1$ and a surface tension-like parameter $\gamma = 0.015 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ (cyan). Data from previous explicit water simulations were taken from Cecchini *et al.* [43] (black squares). The simulations were started from a parallel

in-register arrangement in the explicit water simulations and from a conformation in which all peptides were isolated from each other in the implicit solvent simulations. The β -aggregation propensity was calculated as the average of the nematic order parameter \bar{P}_2 [44, 45], which was previously shown to describe the orientational order of the system and discriminate between ordered and disordered conformations [46]. All simulations identify the central hydrophobic cluster $H_{13}HQKLVFFA_{21}$ as a strong aggregation-promoting sequence.

using the SASA model have provided evidence of the importance of side-chain interactions [41, 42]. Cecchini *et al.* devised a strategy where an amyloidogenic sequence is decomposed into overlapping short stretches, and then long MD simulations of multiple copies of each stretch are run in order to sample their tendency to build ordered parallel aggregates [43]. The resulting amyloidogenicity profiles highlight so-called aggregation “hot-spots”, short stretches that promote aggregation of full-length sequences (Figure 9.2). For one such system, the yeast prion Ure2p, this method was used to predict a double-point mutant with lower β -aggregation propensity that was later confirmed by an experimental test [43].

9.2.2.3 Other Applications

The SASA implicit solvent model was also used in the characterization of the unbinding mechanism of odorant molecules from the odorant binding protein

(OBP) [47]. MD simulations with SASA allowed the identification of a consensus pathway for thymol unbinding from a rat OBP and the description of the associated conformational changes in the receptor. The binding affinity calculated from potentials of mean force (PMFs) was comparable to the value measured by isothermal titration calorimetry.

9.2.3

Limitations of the SASA Implicit Solvent Model

The SASA model is not expected to describe correctly the stability of large proteins: first, because the screening between partial charges does not depend on the local environment; and second, because it is unaffected by atoms that are near the surface but remain completely inaccessible to solvent. SASA should work best for small systems where most or all atoms are at least partially exposed. Limitations in modeling correctly the behavior of large systems is exemplified in Ferrara *et al.* [18], where MD simulations of three proteins—barnase (1a2p, 110 residues), hen egg-white lysozyme (1hel, 129 residues), and cutinase (1cus, 197 residues)—result in fast unfolding even at 300 K, with a C_α RMSD from their respective native conformation above 3.5 Å. The fact that SASA is not appropriate for large systems is also reflected in its evaluation as a scoring function for the CASP4 protein structure prediction competition, where its performance was lower than that of more sophisticated approaches like Poisson–Boltzmann or generalized Born methods [48].

Limitations for small systems were also apparent in the simulation of two designed mini-protein motifs BBA5 and $\alpha\alpha$. BBA5 is a 23-residue peptide with a $\beta\beta\alpha$ architecture [49], whereas $\alpha\alpha$ is a 38-residue peptide designed to adopt a helical hairpin conformation in aqueous solution [50]. During simulations of the two systems at 280 K starting from their respective native conformation, the C_α RMSD relative to the native state rapidly increased above 4 Å. Detailed analysis of the structural deviation showed that, while the distinct secondary structure elements were mostly preserved (the β -hairpin and the helix in BBA5, and helix 1 and helix 2 in $\alpha\alpha$), most of the deviation arises from the loss of their respective native tertiary structures (the packing of the hairpin and helix in BBA5, and the packing of helix 1 and helix 2 in $\alpha\alpha$). Another source of deviation was an increased content of π -helicity. It is likely that the largest error in implicit solvent models originates from the treatment of the charged groups. The use of a distance-dependent dielectric function and the EEF1 modifications of the CHARMM param19 force field, where ionic side chains are neutralized, lead to a rather crude approximation of the electrostatic contribution. This can be a major source of error for BBA5 and $\alpha\alpha$, since both have a high charge density (7 and 15, respectively). On the other hand, it was experimentally observed that, at pH 10.5, $\alpha\alpha$, which has four lysine residues, is largely disordered even at a temperature of 5 °C [51]. Finally, BBA5 was not stable in a 300 K MD simulation with explicit water and the AMBER force field with an 8 Å cutoff, whereas it was stable with a 10 Å cutoff [52].

9.3

The FACTS Implicit Solvent Model: A Fast Generalized Born Approach

The limitations of the SASA model motivated the development of the FACTS implicit model [19]. FACTS implements a more rigorous treatment of electrostatic solvation, which does not require the neutralization of ionic groups, and also takes into account the degree of solvent exposure for the calculation of screening effect.

Despite the significant variability of the dielectric constant in the interior of a protein molecule [53, 54], several implicit solvent models are based on the assumption that the protein is a uniform, low-dielectric region. The essential approximation in such continuum electrostatics models is to represent the solvent as a featureless high-dielectric medium, and the macromolecule as a region with a low-dielectric constant and a spatial charge distribution [1–4, 55–62]. In this way, the solvent degrees of freedom and interaction centers are not taken into account explicitly. The Poisson equation would provide an exact description of such a solute–solvent system, and its numerical solution, obtained either by a finite-difference algorithm [12, 63–65] or by a boundary-element algorithm [66–69], is more efficient than the explicit treatment of the solvent, but still not fast enough for effective utilization in computer simulations of macromolecules. Based on the Born model for ionic solvation [70], the generalized Born (GB) model extends this formalism to treat solutes containing multiple charged particles and an arbitrarily shaped molecular surface. The GB formalism has become an efficient method for the evaluation of continuum electrostatic energies [14] and approximates the PB electrostatic solvation energy as an efficient pairwise summation that allows analytical force calculations:

$$\Delta G^{\text{el,GB}} = -\frac{1}{2} \tau \sum_{i,j=1}^N \frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j} \exp(-r_{ij}^2 / \kappa R_i R_j)} \quad (9.6)$$

where r_{ij} is the distance between charges q_i and q_j , $r_{ii} = 0$, the constant κ is usually set to 4 or 8, $\tau = 1/\epsilon_m - 1/\epsilon_s$, and N is the number of atoms in the solute. The volume occupied by the solute is assigned a low dielectric constant ϵ_m (typically 1, 2, or 4) and the charge distribution is defined by the partial charges of the solute atoms. The solvent is replaced by a uniform medium with a high dielectric constant ϵ_s (typically 78.5 or 80 in the case of water).

The effective Born radius, R_i , is a key quantity in the GB formalism. It measures the degree of burial of individual solute charges and corresponds to the distance between a particular atom and its hypothetical spherical dielectric boundary, chosen such that the self (or atomic) electrostatic solvation energy, ΔG_i^{el} , satisfies the Born equation [70]:

$$\Delta G_i^{\text{el}} = -\frac{\tau q_i^2}{2R_i} \quad (9.7)$$

In principle, the “exact” effective Born radii can be calculated from Equation 9.7 using the self electrostatic solvation energy obtained through the PB theory, but

this would bring no computational advantage. One key observation was that Equation 9.6 yields very accurate results if ΔG_i^{el} (or equivalently R_i) is a good approximation of the value obtained by solving the PB equation [71].

The first generation of GB models used the Coulomb field approximation for the evaluation of ΔG_i^{el} , where the electric displacement \mathbf{D}_i for each atom i is calculated by supposing that the hypothetical dielectric boundary is spherical and that atom i lies at the center of this sphere. A large variety of procedures for calculating effective Born radii within the Coulomb field approximation have been presented. These include numerical surface or volume integrations [14, 72–75], analytical integral expressions [15], and pairwise summation approximations [76–78]. After recognizing that the Coulomb field approximation was a major source of deviation from PB values [3, 79], corrections to the Coulomb field approximation have been suggested and shown to greatly increase the accuracy of the effective Born radii [16, 74, 75, 80]. Current accurate GB implementations are between 10 and 20 times slower than simulations *in vacuo* [81]. Moreover, for proteins of about 100 residues, the computational cost per MD time step is about the same for accurate GB models and explicit water simulations with periodic boundary conditions [75].

In FACTS, the self electrostatic solvation energy and SASA of individual atoms are calculated using intuitive geometric properties of the solute whose evaluation requires only solute interatomic vectors. For each solute atom, the volume and spatial symmetry of its neighboring atoms (or, equivalently, of the solvent displaced by the neighboring atoms) are approximated (Figure 9.3). A combination of these two measures is used as variable of a sigmoidal function (see below). The parameters of the sigmoidal function, together with those of the linear combination with cross-term, are derived by fitting to atomic electrostatic solvation energy values calculated by numerical solution of the PB equation. The GB formula (Equation 9.6) is used to obtain the electrostatic solvation free energy of the macromolecule. The FACTS model does not assume the Coulomb field approximation and does not require the definition of a dielectric discontinuity surface (such dielectric boundary is only required to calculate the PB reference data to which the parameters of the FACTS model are fitted). The same two measures of solvent displacement are combined and used in another sigmoidal function to estimate the SASA of individual atoms. The parameters of the sigmoidal function are derived by fitting to SASA values calculated by an exact analytical method [82]. Finally, the non-polar contribution to the solvation free energy is assumed to be proportional to the sum of the atomic SASA values [7, 8].

Both electrostatic solvation energy and SASA are determined using the same geometrical properties and analytical framework, which makes FACTS a comprehensive and efficient implicit solvation model. When compared with one of the most accurate GB methods [81], like “generalized Born using molecular volume” (GBMV) [75], it is shown that solvation energies computed by FACTS are of similar accuracy to the best available GB implementations, and MD simulations with FACTS are only four times slower than using the *in vacuo* energy.

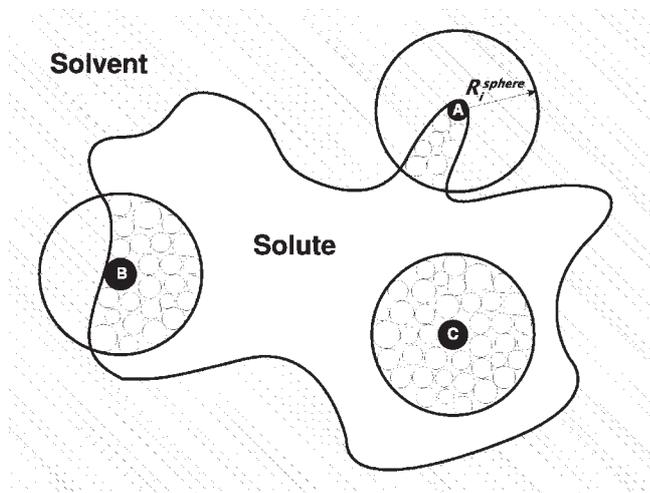


Figure 9.3 Schematic illustration of the measure of degree of burial in FACTS. The solvent exposure of an atom depends on the location of neighboring atoms. The large circle represents the sphere of radius R_i^{sphere} used in FACTS to quantify the atomic solvation energy. FACTS computes the volumes occupied by the neighboring atoms inside the sphere and the symmetry of their spatial distribution. For example, for atom A

(which may be part of a protruding side chain), the volume occupied by neighboring atoms is small and the symmetry is low, so that its atomic solvation energy ΔG_A^{el} is large. The surface atom B has an intermediate measure of volume and a low symmetry value, which result in an intermediate ΔG_B^{el} . Atom C is completely buried, its volume and symmetry values are maximal, and ΔG_C^{el} vanishes to zero.

9.3.1

Description of the Model

9.3.1.1 Atomic (or Self) Electrostatic Solvation Energy

The essential idea in FACTS is that the electrostatic solvation free energy of atom i , ΔG_i^{el} , is evaluated by considering a sphere of radius R_i^{sphere} around atom i [19]. The radius is large enough so that the atom distribution outside the sphere has only a negligible effect on ΔG_i^{el} . If only atom i of the macromolecule were present within the sphere of radius R_i^{sphere} , solving the Poisson equation would result in $\Delta G_i^{\text{el}} \cong -\tau q_i^2 / 2r_i^{\text{VDW}}$. As more and more atoms are gradually added (as for atoms A and B in Figure 9.3), ΔG_i^{el} becomes less favorable depending in a complex way on the spatial distribution of the additional atoms. When all the solvent has finally been flushed out from within the sphere (as for atom C in Figure 9.3), solving the Poisson equation would result in $\Delta G_i^{\text{el}} \cong 0$.

The solvent exposure of a given atom i in a macromolecule depends obviously on the volume occupied by neighboring atoms, a second factor being the distance between the atom i and a neighboring atom j ; atom i will be more desolvated by a close atom j than by a distant one. Then, for a given number of neighboring

atoms inside the sphere of radius R_i^{sphere} with fixed distances from atom i , numerous different atomic distributions can be obtained by rotations of the individual neighboring atoms (especially in three dimensions), and each of these configurations may result in a different electrostatic solvation free energy of atom i . To account for this effect, FACTS introduces a measure of symmetry. In Figure 9.3 atom B lies at the surface of a protein, directly contacting the dielectric discontinuity surface, and is therefore well exposed to solvent. In this atomic configuration, most neighboring atoms are packed in one side of the sphere R_i^{sphere} and the atomic distribution is thus highly asymmetric. One can imagine different configurations with more symmetric atom distributions, where atom i is no longer in direct contact with the solvent and is thus more shielded from it, resulting in a less favorable solvation free energy ΔG_i^{el} . It should be noted that the boundary between the macromolecular and solvent environments in the PB evaluation is defined by the molecular surface [83] and, contrary to the van der Waals surface, it avoids the presence of solvent inside interstitial volumes and microcavities.

To cast the above ideas into mathematical form, the abbreviations $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, $r_{ij} = |\mathbf{x}_{ij}|$, and $\hat{\mathbf{x}}_{ij} = \mathbf{x}_{ij}/r_{ij}$ are introduced. The measure of *volume* occupied by the solute around atom i is defined by

$$A_i = \sum_{j=1, j \neq i}^N V_j \Theta_{ij} \quad (9.8)$$

and the measure of *symmetry* by

$$B_i = \frac{\left| \sum_{j=1, j \neq i}^N (V_j / r_{ij}) \Theta_{ij} \hat{\mathbf{x}}_{ij} \right|}{1 + \sum_{j=1, j \neq i}^N (V_j / r_{ij}) \Theta_{ij}} \quad (9.9)$$

where

$$\Theta_{ij} := \begin{cases} \left[1 - (r_{ij} / R_i^{\text{sphere}})^2 \right]^2 & r_{ij} \leq R_i^{\text{sphere}} \\ 0 & r_{ij} > R_i^{\text{sphere}} \end{cases} \quad (9.10)$$

The measure of volume A_i is simply the sum of the van der Waals volumes V_j of the atoms surrounding atom i within the sphere, weighted by Θ_{ij} . Typically A_i ranges between 100 and 2000 \AA^3 in a sphere of radius $R_i^{\text{sphere}} \cong 10 \text{\AA}$.

The measure of symmetry B_i is a weighted Euclidean norm of the sum of the unit vectors pointing from the central atom i to the neighboring atoms. Thereby each unit vector is weighted by Θ_{ij} , and additionally by the volume of the neighboring atom V_j to which it points, divided by its distance r_{ij} from atom i . There is no other reason for the additional weighting factor V_j/r_{ij} except for the fact that it was found to improve the correlation between the values of B_i and atomic solvation energies calculated by PB. The value of B_i is normalized to range between 0 and 1. For a fully symmetric distribution, B_i equals 0; whereas for a totally asymmetric distribution (for example, only one neighboring atom), B_i is close to 1. The additive constant of 1 in the denominator of Equation 9.9 prevents the denominator becoming zero for a completely isolated ion.

The purpose of the function Θ_{ij} is two-fold: weighting and smoothing. The function Θ_{ij} is equal to 1 for $r_{ij} = 0$ and drops continuously until Θ_{ij} equals 0 at $r_{ij} = R_i^{\text{sphere}}$. Thus, on the one hand, Θ_{ij} accounts for the fact that, the further an atom is placed from atom i , the less it influences its solvation energy. On the other hand, Θ_{ij} ensures the existence of continuous (first and second) derivatives.

When the PB-derived atomic solvation energies $\Delta G_i^{\text{el,PB}}$ for unit charges are plotted against A_i and B_i , a sigmoidal distribution of data is observed (Figure 9.4). Therefore, the measures of volume and symmetry are combined linearly and by a cross-term into a single measure of solvent displacement

$$C_i = A_i + b_1 B_i + b_2 A_i B_i \quad (9.11)$$

and a sigmoidal shaped function of C_i is used to calculate the electrostatic solvation energy $\Delta G_i^{\text{el,FACTS}}$ of atom i for a unit charge:

$$\Delta G_i^{\text{el,FACTS}} = a_0 + \frac{a_1}{1 + e^{-a_2(C_i - a_3)}} \quad (9.12)$$

The parameters a_0 and a_1 are determined using the limiting cases of a fully buried and fully exposed atom. In the case of a fully buried atom (that is, $C_i \rightarrow +\infty$), the value of ΔG_i^{el} should vanish, which implies that $a_0 = -a_1$ and $a_2 > 0$. For a fully exposed atom (that is, $C_i \rightarrow 0$), the Born formula applies, so that $a_0 = -(\tau/2r_i^{\text{VDW}})(1 + e^{-a_2 a_3})$ for a unit charge. Hence, for each van der Waals radius, the five parameters b_1 , b_2 , a_2 , a_3 , and R^{sphere} have to be determined by an optimization procedure. The sigmoidal function (Equation 9.12) gives an accurate fit to $\Delta G_i^{\text{el,PB}}$ (Figure 9.4). Intuitively, C_i measures the solvent displacement around atom i , and the solvation energy of atom i is a sigmoidal function of this measure.

9.3.1.2 Total Electrostatic Solvation Energy

The total electrostatic solvation energy in the FACTS model is the sum of the atomic self-energies and the GB interaction term:

$$\Delta G^{\text{el,FACTS}} = \sum_{i=1}^N \Delta G_i^{\text{el,FACTS}} - \tau \sum_{1 \leq i < j \leq N} \frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i^{\text{FACTS}} R_j^{\text{FACTS}}} \exp(-r_{ij}^2 / \kappa R_i^{\text{FACTS}} R_j^{\text{FACTS}})} \quad (9.13)$$

Here $\Delta G_i^{\text{el,FACTS}}$ is calculated according to Equation 9.12, and $R_i^{\text{FACTS}} = -\tau q_i^2 / 2\Delta G_i^{\text{el,FACTS}}$, with N being the number of atoms in the macromolecule. Equivalently,

$$\Delta G^{\text{el,FACTS}} = -\frac{1}{2} \tau \sum_{i,j=1}^N \frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i^{\text{FACTS}} R_j^{\text{FACTS}}} \exp(-r_{ij}^2 / \kappa R_i^{\text{FACTS}} R_j^{\text{FACTS}})} \quad (9.14)$$

with $r_{ii} = 0$. Note that the second sum in Equation 9.13 implies an infinite cutoff, while a truncation scheme (shifting [20]) is used in the FACTS implementation. Also, a multiplicative factor of 332.0716 is used in front of τ to obtain energy values in kilocalories per mole (kcal mol^{-1}) with interatomic distances in Ångström (Å) and partial charges in electronic units (e).

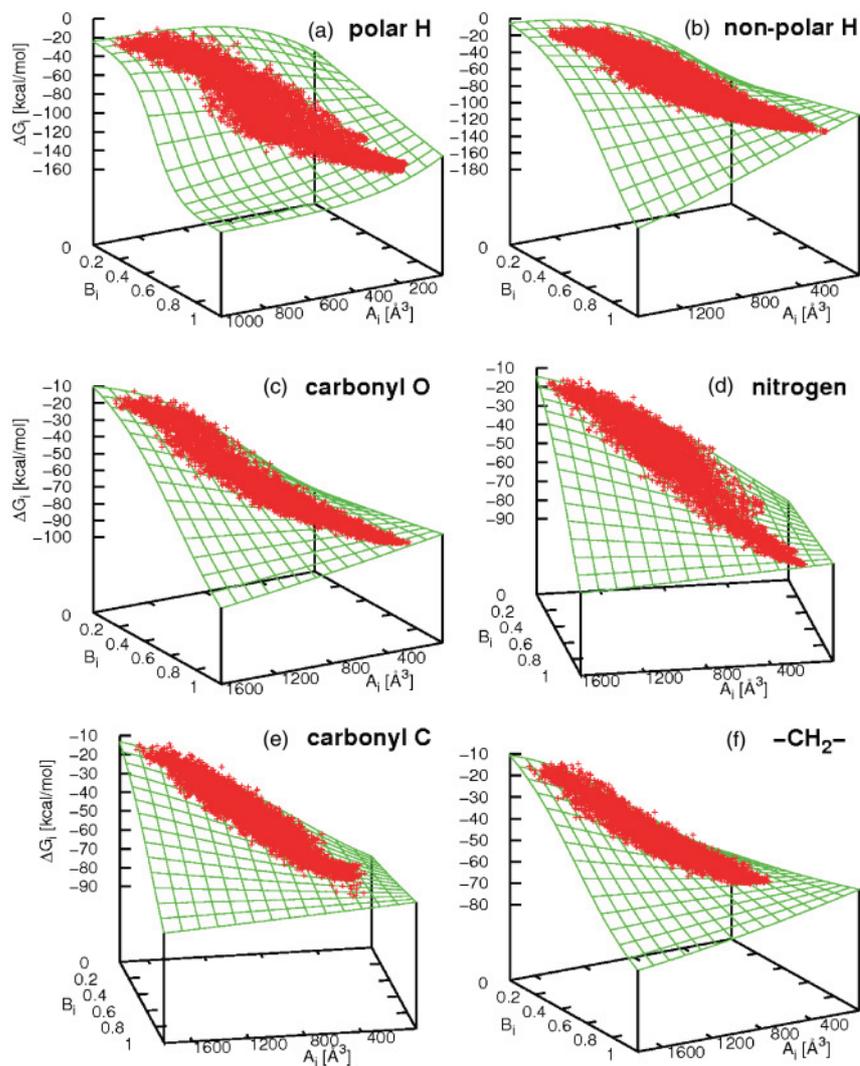


Figure 9.4 The green surface represents Equation 9.12, that is, the FACTS atomic electrostatic solvation energy as a function of A_i and B_i for param22. The red data points are atomic solvation energy values calculated by PB using unit charges and $\epsilon_m = 1$. The dependence on the symmetry is more

pronounced for the polar hydrogen atoms (a) than for the aliphatic carbon atoms (f) because the latter are almost always buried, whereas the former have similar probabilities to be buried or exposed. Note that a fully symmetric distribution yields $B_i = 0$.

9.3.1.3 Atomic Solvent-Accessible Surface Area

Estimating the amount and symmetry of the solvent that is displaced around a given atom provides information on how much the atom is accessible to solvent. Therefore, the geometric concepts described above for approximating the atomic electrostatic solvation energy offer a straightforward way to approximate the SASA of atom i , S_i , by taking into account the relative positions of the surrounding atoms. Analogously to Equations 9.11 and 9.12 one can define

$$D_i = A_i + d_1 B_i + d_2 A_i B_i \quad (9.15)$$

and

$$S_i^{\text{FACTS}} = c_0 + \frac{c_1}{1 + e^{-c_2(D_i - c_3)}} \quad (9.16)$$

for the SASA of atom i . The parameters c_0 and c_1 are determined using the limiting cases of a fully buried and fully exposed atom. In the case of a fully buried atom (that is, $D_i \rightarrow +\infty$), the value of S_i should vanish, which implies that $c_0 = -c_1$ and $c_2 > 0$. For a fully exposed atom (that is, $D_i \rightarrow 0$), the analytical formula applies, so that $c_0 = 4\pi(r_i^{\text{VDW}} + 1.4)^2(1 + e^{-c_2 c_3})$ using a probe sphere of 1.4 Å radius. The parameters d_1 , d_2 , c_2 , and c_3 are derived by fitting to exact values of the SASA [82].

9.3.1.4 Total Solvation Free Energy in the FACTS Model

The solvation free energy of a macromolecule is written as the sum of a polar and a non-polar term,

$$\Delta G^{\text{FACTS}} = \Delta G^{\text{el,FACTS}} + \gamma \sum_{i=1}^N S_i^{\text{FACTS}} \quad (9.17)$$

where γ denotes the empirical surface tension parameter. Values of γ between 0.015 and 0.025 kcal mol⁻¹ Å⁻² have been used to run MD simulations [19].

9.3.2

Parameterization of FACTS

The parameterization was done separately for the polar hydrogen parameter set param19 and the all-atom parameter set param22. Briefly, two training sets, of 81 (param19) and 72 (param22) conformations, from 29 peptides and proteins containing native, molten globule-like and extended conformations and spanning a wide spectrum of secondary structures and irregular shapes (cavities, open loops, and so on), were used for the parameterization. Atomic solvation energies $\Delta G_i^{\text{el,PB}}$ were calculated by numerical solution of the PB equation with the PBEQ module [84] in CHARMM. Reference atomic SASA values (S_i^{exact}) were computed with an analysis module in CHARMM. For each van der Waals radius, two sets of parameters had to be optimized separately: the five parameters b_1 , b_2 , a_2 , a_3 , and R^{sphere} for the atomic solvation energies, and the four parameters d_1 , d_2 , c_2 , and c_3 for the atomic SASA. An upper bound of 10 Å was imposed for the optimization of R^{sphere} . Furthermore, R^{sphere} was optimized only for electrostatic solvation energies. For

atomic SASA values the R^{sphere} parameters determined for the electrostatic solvation were used to increase the efficiency in MD simulations, as the same list of atom pairs can be used for the evaluation of electrostatic solvation energy and SASA. Optimal parameters were obtained by minimizing the deviations of $\Delta G_i^{\text{el,FACTS}}$ from $\Delta G_i^{\text{el,PB}}$ and of S_i^{FACTS} from S_i^{exact} . A particle swarm algorithm [85] was used for parameter optimization.

9.3.3

Validation and Applications of FACTS

The FACTS model was compared with GBMV [81], one of the most accurate GB methods available (we refer the reader to the original FACTS publication [19] for details). Quantities like atomic solvation energies, screened interaction energies and SASA values are slightly better approximated by GBMV, but the correlation between FACTS and reference values is still very high, while FACTS is about 10 times faster than GBMV. The biggest deviations of FACTS from “exact” values occur with the param19 parameter set (a force field with short cutoffs and less hydrogenatoms than param22), and especially in the evaluation of atomic SASA values, mainly because, as mentioned above, the sphere radii were not optimized *ad hoc* for the atomic SASA evaluation, but set equal to those of the electrostatic atomic solvation energy for computational efficiency.

9.3.3.1 Potential of Mean Forces of Side-Chain Dimers

One important test of the quality of an implicit solvent model is the comparison of potentials of mean forces (PMFs) between pairs of ionic or polar side-chain analogs or backbone fragments computed in implicit solvent with PMFs obtained in explicit water. Six PMFs calculated in explicit water and the implicit models GBMV and FACTS, are shown in Figure 9.5. A general trend is that, compared to TIP3P (transferable intermolecular potential, three-point [86]) explicit water, GBMV underestimates the interaction energies whereas FACTS overestimates them. Close agreement of FACTS with the explicit water profile is seen in the backbone–Arg and Ser–Asn systems, and a large deviation is observed in the Lys–Glu system. In most cases FACTS does not reproduce the desolvation barrier, while GBMV tends to overestimate it. It has been observed that the reproduction of the desolvation barrier, which only influences the kinetics, is less important than the correct estimation of the minimum of the interaction energy, which influences the overall thermodynamics of a system [29].

9.3.3.2 Atomic Fluctuations

MD simulations *in vacuo* suffer from too small atomic fluctuations. The RMSD fluctuations of the C_α atoms of ubiquitin (Protein Data Bank (PDB) ID: 1UBQ) and chymotrypsin inhibitor 2 (PDB: 2CI2) computed from param19 300K MD simulations with either EEF1, SASA or FACTS as implicit solvent model are shown in Figure 9.6. In the case of ubiquitin (Figure 9.6a), the fluctuations in FACTS and SASA trajectories are slightly higher than in the EEF1 trajectory, and

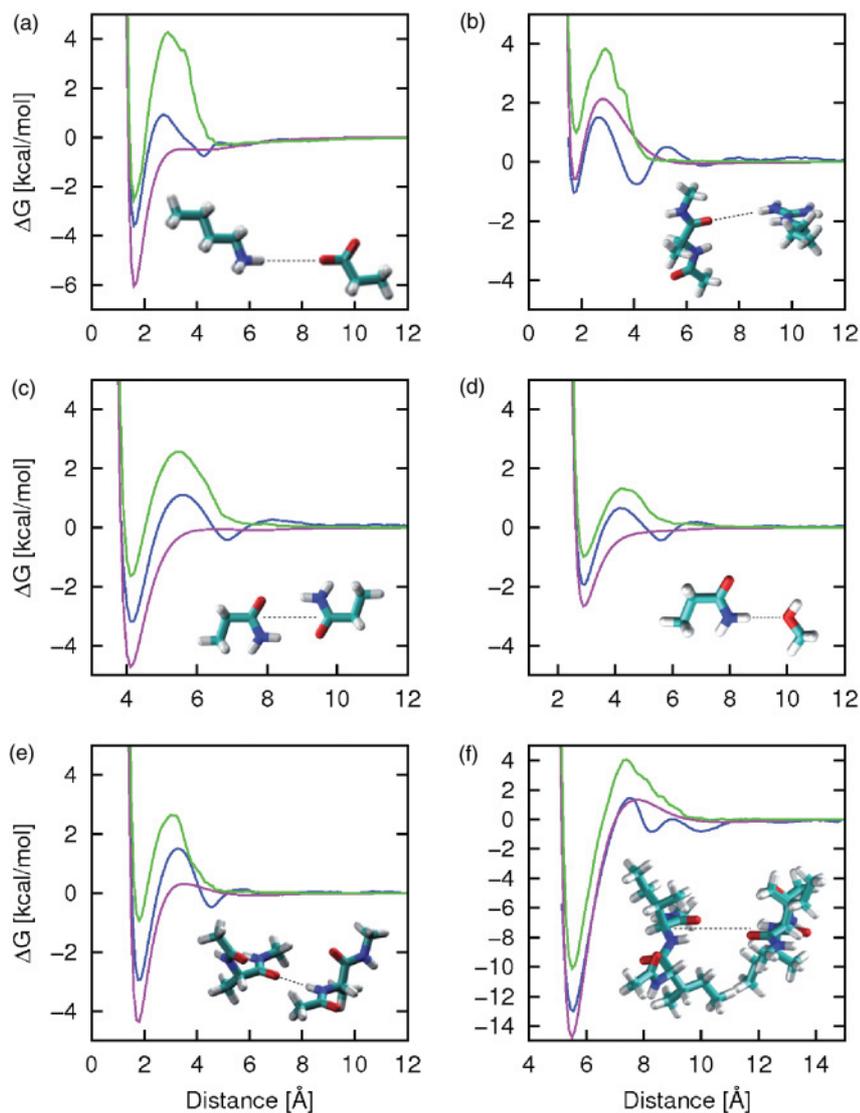


Figure 9.5 Free-energy profiles of six dimers in TIP3P water (blue), GBMV implicit solvent (green), and FACTS param22 with $\epsilon_m = 1$ (magenta). The dimer configurations are shown in the inserts. The reaction coordinates plotted along the x -axis are (a) $r(\text{Lys HZ1-Glu OE1})$, (b) $r(\text{Ala O-Arg HH22})$, (c) $r(\text{Gln$

$\text{CD-Gln CD})$, (d) $r(\text{Ser OG-Gln NE2})$, (e) $r(\text{Ala HN-Ala O})$, and (f) $r(\text{Ile-Ile CA-Ile-Ile CA})$. The molecules were constrained so that their only degree of freedom was translation along the dashed line.

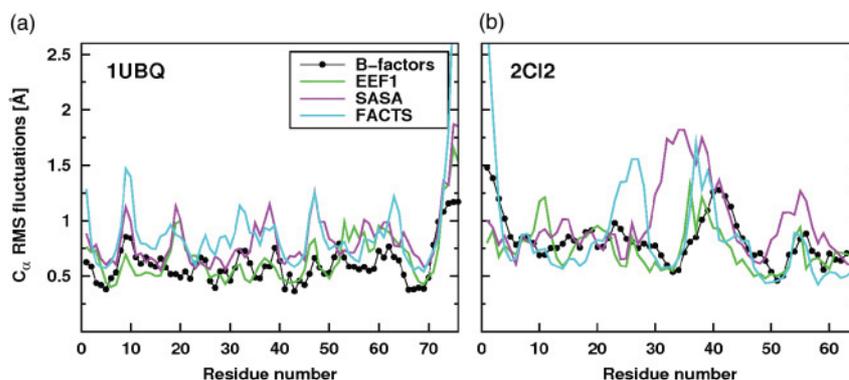


Figure 9.6 RMS fluctuations (Å) of the C_{α} atoms of (a) ubiquitin and (b) chymotrypsin inhibitor 2, extracted from 300K simulations started from their respective native structures. Simulations were performed with param19 and either with EEf1 (green), SASA (magenta), and FACTS with $\epsilon_m = 1$ and a surface tension-like parameter $\gamma = 0.025 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ (cyan). The plotted values are averages of fluctuations of five consecutive 2 ns trajectory segments. The bold line with circles represents the fluctuations derived from the crystallographic B -factors [87] using the formula RMS fluctuation = $[3B/(8\pi^2)]^{0.5}$.

slightly overestimate the crystallographic B -factors [87]. A small increase of the fluctuation may be expected for proteins in solution due to the lack of crystal packing. More importantly, FACTS and SASA reproduce the peaks (that is, the regions of highest mobility) as can be seen in the case of 2CI2 (Figure 9.6b), where the N-terminal segment and the loop (residues 38–44) are the most flexible regions according to both MD simulations and X-ray data.

9.3.3.3 Peptide Aggregation

Following the work of Cecchini *et al.* on the β -aggregation propensity profile of the Alzheimer's amyloid- β peptide ($A\beta_{42}$) [43], simulations were repeated with either TIP3P explicit water or FACTS as solvent model. As shown in Figure 9.2, the profile calculated in explicit water is better reproduced by FACTS. The high charge density in the N-terminal region of $A\beta_{42}$ is responsible for the strong deviation between the β -aggregation propensities calculated by the SASA and explicit water simulations.

9.3.3.4 Scalar and Parallel Performance

The main advantage of FACTS is its speed; it is only four times slower than *in vacuo* and about 10 times faster than GBMV. The CPU time scales linearly with system size, as shown in Figure 9.7a. Furthermore, FACTS has been parallelized and scales well on up to eight CPUs. A scaling factor of 5.6 is obtained in simulations of the 389-residue protein β -secretase (PDB: 1SGZ) on eight Xeon 2.33 GHz cores (70% of ideal scaling; Figure 9.7b). Remarkably, simulations of a much smaller system, a 56-residue designed protein (PDB: 2JWS), scale almost as well as the large 1SGZ system (a scaling factor of 5.3 on eight cores).

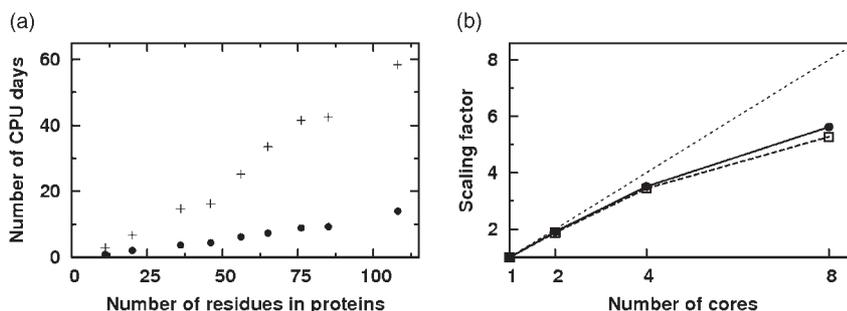


Figure 9.7 FACTS timing analysis. (a) System size scaling of CPU time required for 100ns MD simulations with FACTS. Circles and plus symbols correspond to simulations with param19 and param22, respectively. (b) Scaling factors as a function of number of cores used in parallel MD simulations of the 389-residue protein β -secretase (PDB: 1SGZ;

solid line with circles) and of a 56-residue designed protein (PDB: 2JWS; bold dashed line with squares). The thin dashed line represents an ideal scaling behavior. All simulations were performed on a dual motherboard with Xeon E5410 2.33 GHz quad-core processors.

9.4 Conclusions

The SASA implicit solvent model combines a fast approximation of the surface area with two atomic solvation parameters, and uses a linear distance-dependent dielectric function and neutralized ionic side chains to approximate the electrostatic screening effects. It reasonably describes solvation effects for peptides and small proteins, where most of the charges are exposed to solvent. The typical problems that arise in *in vacuo* MD simulations, that is, large deviations from the native conformation and an excessive number of intra-solute hydrogen bonds, are reduced when including the mean solvation term. The low computational burden of SASA allowed the study of slow processes like aggregation of amyloidogenic peptides [41–43, 46], and reversible folding of structured peptides [33–35, 52], which in turn helped to develop network-based analysis methods of the conformational space [31, 88, 89]. SASA shows its limitations in MD simulations of proteins or highly charged molecules, where the distance-dependent dielectric function and the neutralization of ionic side chains fails to correctly describe the electrostatics in aqueous solution [18].

Attempts to alleviate these limitations led to the development of FACTS [19], an implicit solvent model based on the generalized Born treatment of electrostatics. Notably, FACTS is very efficient because it uses simple measures of solvent displacement and requires only distances between solute atoms that are close in three-dimensional space and are therefore included in standard non-bonding lists. With FACTS the structural integrity of globular proteins in long MD simulations is preserved, and the atomic fluctuations correlate well with values derived from crystallographic *B*-factors. FACTS can be further improved, as the deviations in the PMFs from explicit water curves demonstrate. A source of deviation may be

in the use of van der Waals radii for the definition of the solute–solvent dielectric boundary in the derivation of PB atomic free energy. The van der Waals radii are not optimized to reproduce solvation energy, and a different set of radii may lead to improved agreement between explicit water free energies and PB values [90], which in turn will improve the overall dynamical behavior of biomolecular systems. A more realistic description of solvation effects can also be obtained with an improved treatment of the non-polar component beyond the linear approximation of the solvent-accessible surface area, as it is increasingly recognized that different hydrophobic solvation regimes exist, depending on the length scale of the molecular system [91–93].

Acknowledgments

We thank Jianhan Chen and Charles L. Brooks III for generously sharing their data on potential of mean forces (PMFs). We thank Stefanie Muff and Andrea Prunotto for helpful discussions and comments. This work was supported in part by the Swiss National Competence Center for Research in Neural Plasticity and Repair.

References

- 1 Feig, M. and Brooks, C.L. III (2004) Recent advances in the development and applications of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.*, **14**, 217–224.
- 2 Roux, B. and Simonson, T. (1999) Implicit solvent models. *Biophys. Chem.*, **78**, 1–20.
- 3 Bashford, D. and Case, D.A. (2000) Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.*, **51**, 129–152.
- 4 Cramer, C.J. and Truhlar, D.G. (1999) Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev.*, **99**, 2161–2200.
- 5 Zagrovic, B. and Pande, V. (2003) Solvent viscosity dependence of the folding rate of a small protein: distributed computing study. *J. Comput. Chem.*, **24**, 1432–1436.
- 6 Levy, R.M., Zhang, L.Y., Gallicchio, E., and Felts, A.K. (2003) On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute–solvent interaction energy. *J. Am. Chem. Soc.*, **125**, 9523–9530.
- 7 Eisenberg, D. and McLachlan, A.D. (1986) Solvation energy in protein folding and binding. *Nature*, **319**, 199–203.
- 8 Ooi, T., Oobatake, M., Némethy, M., and Scheraga, H.A. (1987) Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 3086–3090.
- 9 Fraternali, F. and van Gunsteren, W.F. (1996) An efficient mean solvation force model for use molecular dynamics simulations of proteins in aqueous solution. *J. Mol. Biol.*, **256**, 939–948.
- 10 Stouten, P.F.W., Froemmel, C., Nakamura, H., and Sander, C. (1993) An effective solvation term based on atomic occupancies for use in protein simulations. *Mol. Simul.*, **10**, 97–120.
- 11 Lazaridis, T. and Karplus, M. (1999) Effective energy function for proteins in solution. *Proteins*, **35**, 133–152.
- 12 Warwicker, J. and Watson, H.C. (1982) Calculation of the electric potential in the active site cleft due to α -helix dipoles. *J. Mol. Biol.*, **157**, 671–679.

- 13 Constanciel, R. and Contreras, R. (1984) *Theor. Chim. Acta*, **65**, 1.
- 14 Still, W.C., Tempczyk, A., Hawley, R.C., and Hendrickson, T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, **112**, 6127–6129.
- 15 Schaefer, M. and Karplus, M. (1996) A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.*, **100**, 1578–1599.
- 16 Im, W., Lee, M.S., and Brooks, C.L. III (2003) Generalized Born model with a simple smoothing function. *J. Comput. Chem.*, **24**, 1691–1702.
- 17 Mongan, J., Simmerling, C., McCammon, J.A., Case, D.A., and Onufriev, A. (2007) Generalized born model with a simple, robust molecular volume correction. *J. Chem. Theory Comput.*, **3**, 156–169.
- 18 Ferrara, P., Apostolakis, J., and Caffisch, A. (2002) Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins*, **46**, 24–33.
- 19 Haberthür, U. and Caffisch, A. (2008) Fast analytical continuum treatment of solvation. *J. Comput. Chem.*, **29**, 701–715.
- 20 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **4**, 187–217.
- 21 Tama, F. and Sanejouand, Y.H. (2001) Conformational change of proteins arising from normal mode calculations. *Protein Eng. Des. Sel.*, **14**, 1–6.
- 22 Cecchini, M., Houdusse, A., and Karplus, M. (2008) Allosteric communication in myosin V: from small conformational changes to large directed movements. *PLoS Comput. Biol.*, **4**, e1000129.
- 23 Wesson, L. and Eisenberg, D. (1992) Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.*, **1**, 227–235.
- 24 Hasel, W., Hendrickson, T.F., and Still, W.C. (1988) A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput. Methodol.*, **1**, 103–116.
- 25 Mehler, E.L. and Eichele, G. (1984) Electrostatic effects in water-accessible regions of proteins. *Biochemistry*, **23**, 3887–3891.
- 26 Ramstein, J. and Lavery, R. (1988) Energetic coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 7231–7235.
- 27 Mehler, E.L. (1990) Comparison of dielectric response models for simulating electrostatic effects in proteins. *Protein Eng.*, **3**, 415–417.
- 28 Bartels, C., Stote, R., and Karplus, M. (1998) Characterization of flexible molecules in solution: the RGDW peptide. *J. Mol. Biol.*, **284**, 1641–1660.
- 29 Chen, J.H., Im, W.P., and Brooks, C.L. III (2006) Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field. *J. Am. Chem. Soc.*, **128**, 3728–3736.
- 30 De Alba, E., Santoro, J., Rico, M., and Jiménez, M.A. (1999) *De novo* design of a monomeric three-stranded antiparallel β -sheet. *Protein Sci.*, **8**, 854–865.
- 31 Rao, F. and Caffisch, A. (2004) The protein folding network. *J. Mol. Biol.*, **342**, 299–306.
- 32 Caffisch, A. (2006) Network and graph analyses of folding free energy surfaces. *Curr. Opin. Struct. Biol.*, **16**, 71–78.
- 33 Cavalli, A., Haberthür, U., Paci, E., and Caffisch, A. (2003) Fast protein folding on downhill energy landscape. *Protein Sci.*, **12**, 1801–1803.
- 34 Ferrara, P., Apostolakis, J., and Caffisch, A. (2000) Targeted molecular dynamics simulations of protein unfolding. *J. Phys. Chem.*, **104**, 4511–4518.
- 35 Paci, E., Cavalli, A., Vendruscolo, M., and Caffisch, A. (2003) Analysis of the distributed computing approach applied to the folding of a small beta-peptide. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 8217–8222.
- 36 Settanni, G., Rao, F., and Caffisch, A. (2005) Φ -value analysis by molecular dynamics simulations of reversible folding. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 628–633.
- 37 Gsponer, J. and Caffisch, A. (2001) Role of native topology investigated by multiple unfolding simulations of four SH3 domains. *J. Mol. Biol.*, **309**, 285–298.
- 38 Rathore, N., Yan, Q.L., and de Pablo, J.J. (2004) Molecular simulation of the

- reversible mechanical unfolding of proteins. *J. Chem. Phys.*, **120**, 5781–5788.
- 39 Ihalainen, J.A., Bredenbeck, J., Pfister, R., Helbing, J., Woolley, G.A., and Hamm, P. (2007) Folding and unfolding of a photoswitchable peptide from picoseconds to microseconds. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 5383–5388.
- 40 Rathore, N., Knotts, T.A., and de Pablo, J.J. (2003) Configurational temperature density of states simulations of proteins. *Biophys. J.*, **85**, 3963–3968.
- 41 Gsponer, J., Haberthür, U., and Caffisch, A. (2003) The role of sidechain interactions in the early steps of aggregation: molecular dynamics simulations of an amyloid forming peptide from the yeast prion Sup35. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 5154–5159.
- 42 Paci, E., Gsponer, J., Salvatella, X., and Vendruscolo, M. (2004) Molecular dynamics studies of the process of amyloid aggregation of peptide fragments of transthyretin. *J. Mol. Biol.*, **340**, 555–569.
- 43 Cecchini, M., Curcio, R., Pappalardo, M., Melki, R., and Caffisch, A. (2006) A molecular dynamics approach to the structural characterization of amyloid aggregation. *J. Mol. Biol.*, **357**, 1306–1321.
- 44 Chandrasekhar, S. (1992) *Liquid Crystals*, Cambridge University Press, Cambridge.
- 45 Zannoni, C. (2001) Molecular design and computer simulations of novel mesophases. *J. Mater. Chem.*, **11**, 2637–2646.
- 46 Cecchini, M., Rao, F., Seeber, M., and Caffisch, A. (2004) Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.*, **121**, 10748–10756.
- 47 Hajjar, E., Perahia, D., Debat, H., Nespoulous, C., and Robert, C.H. (2006) Odorant binding and conformational dynamics in the odorant binding protein. *J. Biol. Chem.*, **281**, 29929–29937.
- 48 Feig, M. and Brooks, C.L. III (2002) Evaluating casp4 predictions with physical energy functions. *Proteins*, **49**, 232–245.
- 49 Struthers, M.D., Cheng, R.P., and Imperiali, B. (1996) Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science*, **271**, 342–345.
- 50 Fezoui, Y., Connolly, P.J., and Osterhout, J.J. (1997) Solution structure of $\alpha\alpha$, a helical hairpin peptide of *de novo* design. *Protein Sci.*, **6**, 1869–1877.
- 51 Fezoui, Y., Weaver, D.L., and Osterhout, J.J. (1994) *De novo* design and structural characterization of an α -helical hairpin peptide: a model system for the study of protein folding intermediates. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 3675–3697.
- 52 Ferrara, P., Apostolakis, J., and Caffisch, A. (2000) Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. B*, **104**, 5000–5010.
- 53 Warshel, A. and Russell, S.T. (1984) Calculations of electrostatic interactions in biological systems and in solutions. *Q. Rev. Biophys.*, **17**, 283–422.
- 54 Warshel, A. and Papazyan, A. (1998) Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.*, **8**, 211–217.
- 55 Tomasi, J. and Persico, M. (1994) Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. *Chem. Rev.*, **94**, 2027–2094.
- 56 Gilson, M.K. (1995) Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.*, **5**, 216–223.
- 57 Orozco, M. and Luque, F.J. (2000) Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.*, **100**, 4187–4225.
- 58 Simonson, T. (2001) Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.*, **11**, 243–252.
- 59 Werner, P. and Caffisch, A. (2003) A sphere-based model for the electrostatics of globular proteins. *J. Am. Chem. Soc.*, **125**, 4600–4608.
- 60 Baker, N.A. (2005) Improving implicit solvent simulations: a Poisson-centric view. *Curr. Opin. Struct. Biol.*, **15**, 137–143.
- 61 Im, W., Chen, J., and Brooks, C.L. III (2006) Peptide and protein folding and conformational equilibria: theoretical treatment of electrostatics and hydrogen

- bonding with implicit solvent models. *Adv. Protein Chem.*, **72**, 171–195.
- 62 Feig, M., Chocholousova, J., and Tanizaki, S. (2006) Extending the horizon: towards the efficient modeling of large biomolecular complexes in atomic detail. *Theor. Chem. Accts.*, **116**, 194–205.
- 63 Gilson, M.K. and Honig, B.H. (1988) Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins*, **4**, 7–18.
- 64 Bashford, D. and Karplus, M. (1990) pK_s of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry*, **29**, 10219–10225.
- 65 Davis, M.E., Madura, J.D., Luty, B.A., and McCammon, J.A. (1991) Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.*, **62**, 187–197.
- 66 Zauhar, R.J. and Morgan, R.S. (1985) A new method for computing the macromolecular electric potential. *J. Mol. Biol.*, **186**, 815–820.
- 67 Cortis, C.M. and Friesner, R.A. (1997) An automatic three-dimensional finite element mesh generation system for the Poisson–Boltzmann equation. *J. Comput. Chem.*, **18**, 1570–1590.
- 68 Cortis, C.M. and Friesner, R.A. (1997) Numerical solution of the Poisson–Boltzmann equation using tetrahedral finite-element meshes. *J. Comput. Chem.*, **18**, 1591–1608.
- 69 Vorobjev, Y.N. and Scheraga, H.A. (1997) A fast adaptive multigrid boundary element method for macromolecular electrostatic computations in a solvent. *J. Comput. Chem.*, **18**, 569–583.
- 70 Born, M. (1920) Volumen und Hydrationswärme der Ionen. *Z. Phys.*, **1**, 45–48.
- 71 Onufriev, A., Bashford, D., and Case, D.A. (2002) Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.*, **23**, 1297–1304.
- 72 Scarsi, M., Apostolakis, J., and Caffisch, A. (1997) Continuum electrostatic energies of macromolecules in aqueous solutions. *J. Phys. Chem. A*, **101**, 8098–8106.
- 73 Ghosh, A., Rapp, C.S., and Friesner, R.A. (1998) Generalized Born model based on a surface integral formulation. *J. Phys. Chem. B*, **102**, 10983–10990.
- 74 Lee, M.S., Salsbury, F.R., and Brooks, C.L. III (2002) Novel generalized Born methods. *J. Chem. Phys.*, **116**, 10606–10614.
- 75 Lee, M.S., Feig, M., Salsbury, F.R., and Brooks, C.L. III (2003) New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.*, **24**, 1348–1356.
- 76 Hawkins, G.D., Cramer, C.J., and Truhlar, D.G. (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.*, **100**, 19824–19839.
- 77 Qiu, D., Shenkin, P.S., Hollinger, F.P., and Still, W.C. (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A*, **101**, 3005–3014.
- 78 Dominy, B.N. and Brooks, C.L. III (1999) Development of a generalized Born model parametrization for proteins and nucleic acids. *J. Phys. Chem. B*, **103**, 3765–3773.
- 79 Schaefer, M. and Froemmel, C. (1990) A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution. *J. Mol. Biol.*, **216**, 1045–1066.
- 80 Grycuk, T. (2003) Deficiency of the Coulomb-field approximation in the generalized Born model: an improved formula for Born radii evaluation. *J. Chem. Phys.*, **119**, 4817–4826.
- 81 Feig, M., Onufriev, A., Lee, M.S., Im, W., Case, D.A., and Brooks, C.L. III (2003) Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.*, **25**, 264–284.
- 82 Lee, B. and Richards, F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.
- 83 Richards, F.M. (1977) Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.*, **6**, 151–176.

- 84 Im, W., Beglov, D., and Roux, B. (1998) Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson–Boltzmann equation. *Comput. Phys. Commun.*, **111**, 59–75.
- 85 Kennedy, J. and Eberhart, R.C. (2001) *Swarm Intelligence*, Morgan Kaufmann, San Francisco.
- 86 Jorgensen, W.L., Chandrasekhar, J., Madura, J., Impey, R.W., and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
- 87 McPhalen, C.A. and James, M.N.G. (1987) Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley-seeds. *Biochemistry*, **26**, 261–269.
- 88 Krivov, S.V., Muff, S., Caffisch, A., and Karplus, M. (2008) One-dimensional barrier preserving free-energy projections of a beta-sheet miniprotein: new insights into the folding process. *J. Phys. Chem. B*, **112**, 8701–8714.
- 89 Muff, S. and Caffisch, A. (2008) Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β -sheet miniprotein. *Proteins*, **70**, 1185–1195.
- 90 Nina, M., Beglov, D., and Roux, B. (1997) Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B*, **101** (26), 5239–5248.
- 91 Chandler, D. (2005) Interfaces and the driving force of hydrophobic assembly. *Nature*, **437**, 640–647.
- 92 Rajamani, S., Truskett, T.M., and Garde, S. (2005) Hydrophobic hydration from small to large lengthscales: understanding and manipulating the crossover. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 9475–9480.
- 93 Graziano, G. (2007) Cavity thermodynamics and surface tension of water. *Chem. Phys. Lett.*, **442**, 307–310.