FACTS: Fast Analytical Continuum Treatment of Solvation

URS HABERTHÜR, AMEDEO CAFLISCH
Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, Zurich CH-8057, Switzerland

Received 16 October 2006; Revised 26 June 2007; Accepted 30 July 2007
DOI 10.1002/jcc.20832
Published online 5 October 2007 in Wiley InterScience (www.interscience.wiley.com).

Abstract: An efficient method for calculating the free energy of solvation of a (macro)molecule embedded in a continuum solvent is presented. It is based on the fully analytical evaluation of the volume and spatial symmetry of the solvent that is displaced from around a solute atom by its neighboring atoms. The two measures of solvent displacement are combined in empirical equations to approximate the atomic (or self) electrostatic solvation energy and the solvent accessible surface area. The former directly yields the effective Born radius, which is used in the generalized Born (GB) formula to calculate the solvent-screened electrostatic interaction energy. A comparison with finite-difference Poisson data shows that atomic solvation energies, pair interaction energies, and their sums are evaluated with a precision comparable to the most accurate GB implementations. Furthermore, solvation energies of a large set of protein conformations have an error of only 1.5%. The solvent accessible surface area is used to approximate the nonpolar contribution to solvation. The empirical approach, called FACTS (Fast Analytical Continuum Treatment of Solvation), is only four times slower than using the vacuum energy in molecular dynamics simulations of proteins. Notably, the folded state of structured peptides and proteins is stable at room temperature in 100-ns molecular dynamics simulations using FACTS and the CHARMM force field.

 J Comput Chem 29: 701–715, 2008

Key words: implicit solvent; continuum dielectric; generalized Born model; CHARMM; molecular dynamics simulations

Introduction

An accurate treatment of the effects of aqueous solvent in molecular dynamics (MD) simulations of biological (macro)molecules is of key importance because cells and physiological fluids consist mainly of water. The exact calculation of the electrostatic energy of a protein in solution requires the evaluation of the interactions among all solute–solute, solute–solvent, and solvent–solvent pairs of charges. However, this is computationally expensive for fully hydrated macromolecules. Despite continuous advances in both the development of parallel MD simulation code and the performance of ordinary low cost computer processors, explicit solvent MD simulations of large proteins lasting longer than 100 ns are still almost prohibitive. A simplified treatment that does not require the solvent degrees of freedom and interaction centers explicitly can be very useful, and for large systems implicit models represent the only affordable description of the solvent. For instance, sampling a statistically significant number of folding and unfolding transitions of structured peptides at equilibrium require simulations in the 1–10 μs timescale.[1]

Despite the significant variability of the dielectric constant in the interior of a protein molecule,[2,3] 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.[4–15] In this way, the solvent degrees of freedom and interaction centers are not taken into account explicitly. The Poisson equation provides an exact description of such a solute/solvent system. The numerical solution of the finite-difference Poisson (fdP) equation[16–19] is more efficient than the explicit treatment of the solvent but still not fast enough for effective utilization in computer simulations of macromolecules.

The generalized Born (GB) model was introduced for an efficient evaluation of continuum electrostatic energies.[20] The most critical aspect of the GB model is the calculation of the effective Born radii which measure the degree of burial of individual

This article contains supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/0192-8651/suppmat

Correspondence to: A. Caflisch; e-mail: caflisch@bioc.uzh.ch

Contract/grant sponsor: Swiss National Science Foundation

solute charges. This measure is combined in a heuristic way to obtain a correction to the Coulomb law for each atom pair.[20] In contrast to implicit solvation models which use a distance–dependent screening function,[21–23] the GB equation takes into account the effect of not only the charge–charge distance but also the degree of solvent exposure of the interacting charges. Accurate GB implementations published as of today are between 20 and 40 times slower than simulations *in vacuo*.[24] 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.[25]

Water molecules in the liquid state influence the electrostatic energy of a macromolecule in two ways. They solvate each individual charge of the solute (atomic solvation energy), and they screen the interaction between charge pairs.[2,3,26] In a previous work, we introduced a geometric measure of the degree of burial of pairs of interacting solute charges for quantifying the screening effect.[27] The aim of the present article is to adopt similar steric concepts for the efficient evaluation of the effective Born radius (which is inversely proportional to the atomic solvation energy) using the local environment of each solute atom and empirical formulas. The empirical treatment of electrostatics in FACTS requires five parameters for each atom type (more precisely, for each value of the van der Waals radius), and the parameters are optimized by fitting to fdP data. The same geometric formalism is also proposed for the calculation of the solvent accessible surface area (SASA) of individual atoms of the solute, which is used for approximating the nonpolar contribution to solvation. The resulting continuum model, called FACTS, is a fully analytical and comprehensive treatment of solvation effects. A comparison is given with one of the most accurate GB methods,[24] i.e., GB using molecular volume (GBMV[25]). The extensive validation provides evidence that FACTS is as accurate as the best available GB implementations, and MD simulations with FACTS are only four times slower than using the energy in vacuo.

Methods

Generalized Born Model

The GB approach is an empirical formula originally proposed to approximate the electrostatic contribution to the hydration free energy of small organic compounds[20]

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

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, $\tau = \frac{1}{\varepsilon_m} - \frac{1}{\varepsilon_s}$, 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 is defined by

$$R_i = -\frac{\tau q_i^2}{2\Delta G_i^{\mathrm{el}}} \qquad (2)$$

where ΔG_i^{el} is the electrostatic solvation free energy of atom i. For the evaluation of ΔG_i^{el} in the first generation of GB models, the Coulomb field approximation was used, where the electric displacement $\vec{D}_i$ for each atom i is calculated by supposing that the dielectric boundary is spherical and that atom i lies at the center of this sphere. (Note that this spherical symmetry is only assumed to calculate $\vec{D}_i$.) A large variety of procedures for calculating effective Born radii within the Coulomb field approximation have been presented. These include numerical surface or volume integration,[20,25,28–30] analytical integral expression,[26] and pairwise summation approximations.[31–33]

An important observation is that eq. (1) yields very accurate results if ΔG_i^{el} (or equivalently R_i) is a good approximation of the value obtained by solving the fdP equation.[34] Therefore, the most recent developments of GB models have tried to improve the accuracy of the ΔG_i^{el} evaluation.[12] In particular, corrections to the Coulomb field approximation have been suggested and shown to greatly increase the accuracy of the effective Born radii.[25,30,35] In a different approach it was demonstrated that the quantity $\sqrt{R_i R_j}$ can be interpreted as a measure of enclosure of the (i, j) atom pair and be calculated very efficiently to yield accurate screened interaction energies.[27] The development of FACTS (see next subsection) was inspired by this measure of enclosure.

Fast Analytical Continuum Treatment of Solvation

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. A linear combination with cross-term of these two measures is used as independent variable of a sigmoidal function (see later). 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 fdP equation. The GB formula (1) is used to obtain the electrostatic solvation free energy of the macromolecule. The FACTS model does not assume the Coulomb field approximation (see earlier) and does not require to define a dielectric discontinuity surface. (Such dielectric boundary is only required to calculate the fdP 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.[36] Finally, the nonpolar contribution to the solvation free energy is assumed to be proportional to the sum of the atomic SASA values.[37,38]

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.

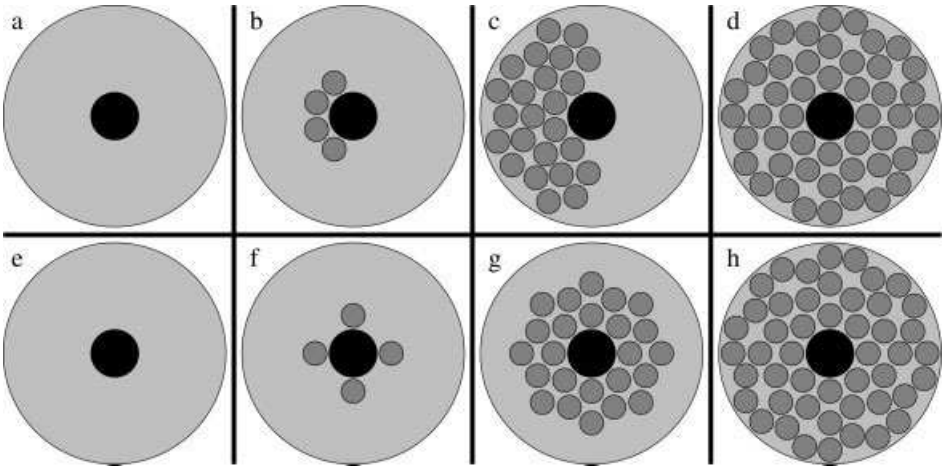

Figure 1. Schematic illustration of the essential concept of the FACTS evaluation of atomic solvation energy. The large circle in light gray represents the sphere of radius R_i^{sphere} that is considered to quantify the atomic solvation energy in the FACTS approach (see text). The small circles in dark gray represent solute atoms that displace the solvent from around the central atom, which is in black. Both pathways (a → b → c → d and e → f → g → h) proceed from a fully solvated to a fully desolvated atom. In the top pathway atoms are added such as to break spatial symmetry as much as possible. In the bottom pathway atoms are added such as to preserve spatial symmetry as much as possible. Crossing from the asymmetric (top) to the symmetric (bottom) pathway in the two intermediate steps, i.e., going from b to f or c to g, the number of neighboring atoms remains constant but the solvation energy of the central atom changes significantly due to the increase in symmetry.

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. The radius is large enough to neglect effects on ΔG_i^{el} due to conformational changes outside the sphere. (More precisely, let $\Delta G_i^{\text{el,m}}$ denote ΔG_i^{el} being calculated with the region outside of R_i^{sphere} occupied exclusively by atoms of the macromolecule, assuming an infinitely large protein. Similarly, let $\Delta G_i^{\text{el,s}}$ denote ΔG_i^{el} being calculated with the region outside of R_i^{sphere} occupied exclusively by solvent. Then the value of R_i^{sphere} is chosen large enough so that $\Delta G_i^{\text{el,s}} - \Delta G_i^{\text{el,m}} \cong 0$ holds for any conformation within the sphere.) If only atom i of the macromolecule were present within the sphere of radius R_i^{sphere}, solving the Poisson equation would result in $\Delta G_i^{\text{el}} \cong -\frac{\tau q_i^2}{2r_i^{\text{vdW}}}$. As more and more atoms are gradually added (see Fig. 1) ΔG_i^{el} becomes less favorable depending in a complex way on where the additional atoms are placed. When all the solvent has finally been flushed out from within the sphere, solving the Poisson equation would result in $\Delta G_i^{\text{el}} \cong 0$.

To quantify the atomic solvation energy, it is useful to investigate the change in solvation energy upon sequential addition of solute atoms to the interior of the sphere. Two desolvation pathways are shown in Figure 1. In the leftmost column the atom at the center is completely solvated. In the rightmost column it is completely desolvated. In proceeding from left to right on the top or bottom row in Figure 1, more and more atoms surrounding the atom at the center are added. Thus, the central atom becomes more and more desolvated and its solvation energy ΔG_i^{el} becomes less and less favorable. The difference between the two pathways is that on the top pathway, atoms are added so as to disrupt the spatial symmetry within the sphere, whereas on the bottom pathway atoms are added so as to preserve the spatial symmetry. Crossing from the asymmetric to the symmetric pathway in the two intermediate steps in Figure 1, i.e., going from b to f and c to g, respectively, the number of atoms surrounding the central atom remains constant but they are rearranged such that more solvent close to the central atom is displaced. Thus, the solvation energy of the central atom becomes less favorable. The following two observations are the core of the FACTS model. The increase in solvation energy induced by adding solute atoms (from left to right in Fig. 1) can be accounted for by the change of a suitably defined measure of volume. It quantifies the volume occupied by solute atoms within the sphere of radius R_i^{sphere}. The increase in solvation energy originating from a rearrangement of solute atoms (from top to bottom in Fig. 1) can be approximated by the change of a measure of symmetry, which quantifies the symmetry of the spatial distribution of the atoms surrounding atom i.

From the previous description it is clear that a measure of volume or symmetry alone is not appropriate to calculate the solvation energy of atom i. A fully buried atom (Fig. 1d) and a fully exposed atom (Fig. 1a) are only marginally discriminated by the spatial symmetry within the sphere. (However, the latter situation never arises in proteins since each atom always has neighbors.) Hence, the number of neighbors is the key difference. Analogously, the volume occupied by solute atoms within the sphere is constant in, for instance, snapshots b and f in Figure 1. Nevertheless, the solvation energy becomes less favorable by crossing from b to f. In this case the key difference is the

symmetry. Either of the two measures provides a partial description, but a synergistic combination of the two measures yields a powerful means to calculate the atomic solvation energy.

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

$$A_i = \sum_{j=1, j\neq i}^{N} V_j \Theta_{ij} \tag{3}$$

and the measure of symmetry by

$$B_i = \left| \frac{\sum_{j=1, j\neq i}^{N} \frac{V_j}{r_{ij}} \Theta_{ij} \hat{x}_{ij}}{1 + \sum_{j=1, j\neq i}^{N} \frac{V_j}{r_{ij}} \Theta_{ij}} \right| \tag{4}$$

where

$$\Theta_{ij} := \begin{cases} \left(1 - \left(\frac{r_{ij}}{R_i^{\text{sphere}}}\right)^2\right)^2 & r_{ij} \leq R_i^{\text{sphere}} \\ 0 & r_{ij} > R_i^{\text{sphere}} \end{cases} \tag{5}$$

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 Å^3 and 2000 Å^3 in a sphere of radius $R_i^{\text{sphere}} \cong 10$ Å.

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 it points to, 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 fdP. The value of B_i is normalized to range between 0 and 1. For a fully symmetric distribution $B_i = 0$, whereas for a totally asymmetric distribution (e.g., only one neighboring atom) B_i is close to 1. The additive constant of 1 in the denominator of eq. (4) prevents that the denominator becomes zero for a completely isolated ion.

The purpose of the function Θ_{ij} is twofold: weighting and smoothing. Θ_{ij} is equal to 1 for $r_{ij} = 0$ and drops continuously until $\Theta_{ij} = 0$ 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. Note that due to the function Θ_{ij} the measure of volume includes a small contribution originating from the symmetry. As an example, configurations c and g in Figure 1 yield different values for the measure of volume because of Θ_{ij}.

Having defined the measure of volume and symmetry, the next step is to obtain a functional relationship between atomic solvation energies and the quantities A_i and B_i. The aim is to find a function prototype with some parameters that can be optimized to reproduce accurately the fdP reference values. At this point it is important to note that once a function prototype is found, its parameters have to be optimized separately for each van der Waals radius of the solute atoms. To explain the importance of the van der Waals radius one can consider two fully solvated atoms with differing van der Waals radii. The two atoms have the same values for the measure of volume (zero) and symmetry (zero), but their solvation energies are different and depend on the van der Waals radii according to the Born formula.

To obtain the desired relationship it is helpful to plot fdP derived atomic solvation energies $\Delta G_i^{\text{el,fdP}}$ for unit charges against A_i and B_i in a three-dimensional graph (Fig. 2), classified in sets according to the van der Waals radii of the corresponding atoms. For each set a sigmoidal distribution of data is observed. Therefore, the measures of volume and symmetry are combined linearly and by a mixed term into a single measure of solvent displacement

$$C_i = A_i + b_1 B_i + b_2 A_i B_i \tag{6}$$

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:

$$\Delta G_i^{\text{el,FACTS}} = a_0 + \frac{a_1}{1 + e^{-a_2(C_i - a_3)}} \tag{7}$$

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 (i.e., $C_i \to +\infty$) the value of ΔG_i^{el} should vanish which implies that $a_0 = -a_1$ and $a_2 > 0$. For a fully exposed atom (i.e., $C_i \to 0$) the Born formula applies so that $a_0 = -\frac{\tau}{2r_i^{\text{vdW}}}(1 + e^{-a_2 a_3})$ 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 [eq. (7)] gives an accurate fit to $\Delta G_i^{\text{el,fdP}}$ (Fig. 2). Intuitively, C_i measures the solvent displacement around atom i, and the solvation energy of atom i is a sigmoidal function of this measure.

Total Electrostatic Solvation Energy

Using the definition of effective Born radius [eq. (2)] and the GB formula for the interaction term, the total electrostatic solvation energy in the FACTS model is written as

$$\Delta G^{\text{el,FACTS}} = \sum_{i=1}^{N} \Delta G_i^{\text{el,FACTS}} - \tau \sum_{1 \leq i < j \leq N} \frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i^{\text{FACTS}} R_j^{\text{FACTS}} \exp(-r_{ij}^2 / \kappa R_i^{\text{FACTS}} R_j^{\text{FACTS}})}} \tag{8}$$

$$= -\frac{\tau}{2} \sum_{i,j=1}^{N} \frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i^{\text{FACTS}} R_j^{\text{FACTS}} \exp(-r_{ij}^2 / \kappa R_i^{\text{FACTS}} R_j^{\text{FACTS}})}} \tag{9}$$

where $r_{ii} = 0$ and N is the number of atoms in the macromolecule. Note that the second sum in eq. (8) implies an infinite cutoff while a truncation scheme (shifting[39]) is used in the MD

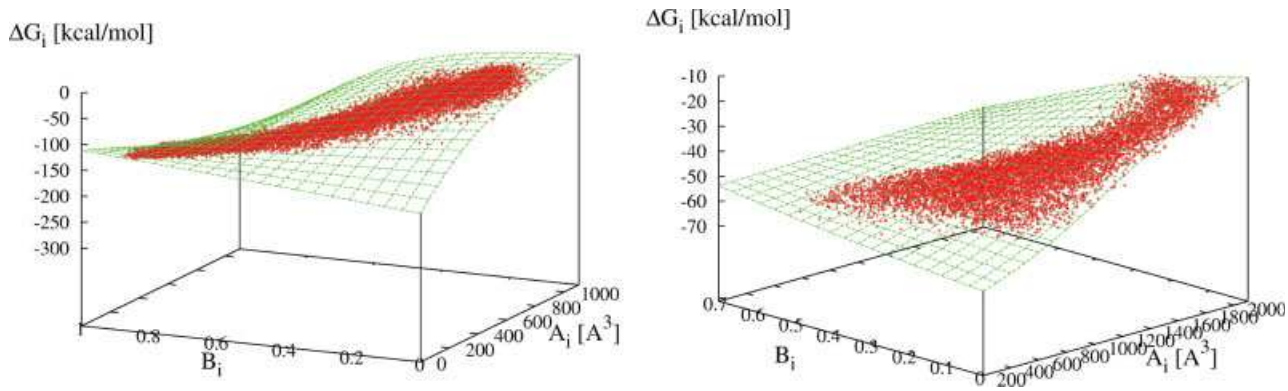

Figure 2. The green surface represents eq. (7), i.e., FACTS atomic electrostatic solvation energy as a function of A_i and B_i for PARAM19 and a van der Waals radius of 1.0 Å (left) and 2.365 Å (right). The red data points are atomic solvation energy values calculated by fdP using unit charges and $\varepsilon_m = 1$. The dependence on the symmetry is more pronounced for the polar hydrogen atoms (left) than the carbon atoms (right) because the latter are almost always buried whereas the former show significant amount of both buried and exposed. Note that a fully symmetric distribution yields $B_i = 0$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

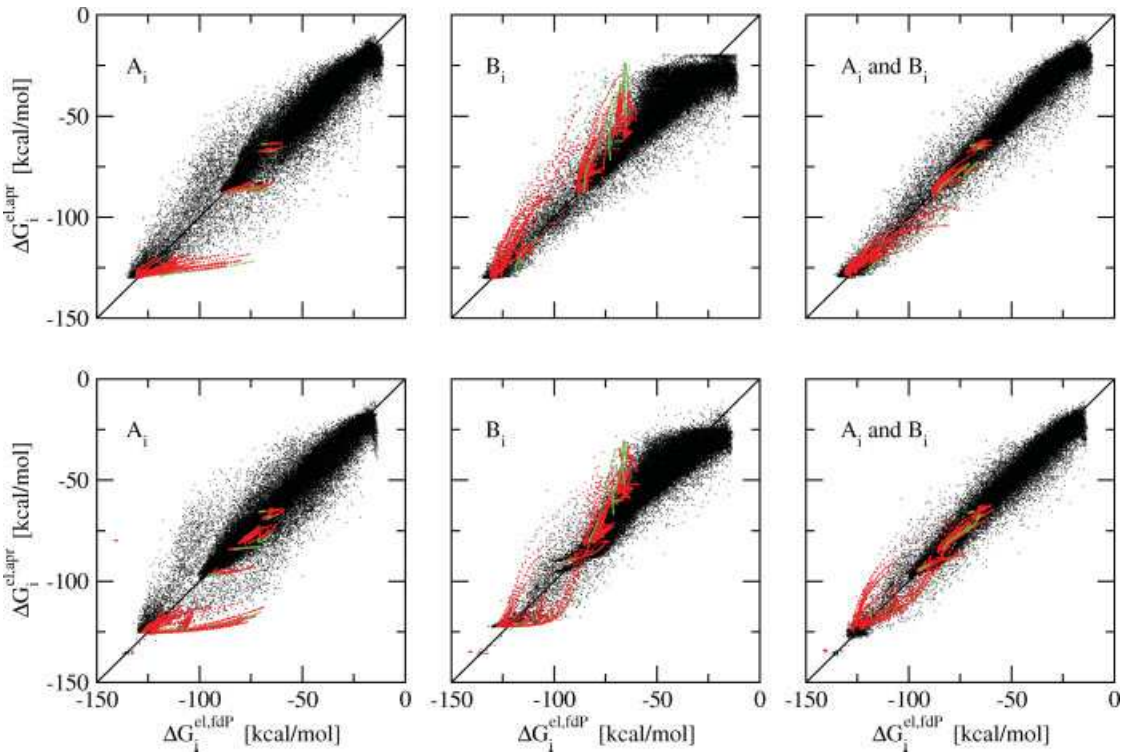

Figure 3. Synergistic effect of volume and symmetry terms in FACTS. In the left column only the measure of volume was used for the FACTS calculations, in the middle column only the measure of symmetry, and in the right column both measures were combined. Unit charges were used because they allow for a more stringent comparison that is not affected by the charge parameter set. The benchmark fdP calculations were performed with $\varepsilon_m = 1$. The data points of the protein conformations are in black, while those of the pairs of charged side chains and the *N*-methyl-acetamide dimer in red and green, respectively. (Top) Atomic electrostatic solvation energy values calculated by FACTS [eq. (7)] versus the fdP values for 77,609 atoms from 1082 molecular structures with the van der Waals radii of PARAM19. (Bottom) Atomic electrostatic solvation energy values calculated by FACTS [eq. (7)] versus the fdP values for 90,747 atoms from 1073 molecular structures with the van der Waals radii of PARAM22. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

simulations reported below. Also, a multiplicative factor of 332.0716 is used in front of τ for obtaining energy values in kcal/mol with interatomic distances in Å and partial charges in electronic units.

Atomic Solvent Accessible Surface Area

Estimating 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 earlier for approximating the atomic electrostatic solvation energy can also be used to calculate the SASA. Several efficient methods that accomplish this task have been suggested in the past. They mainly use interatomic distances only and do not take into account symmetry. It has been suggested that angles between atom triplets could be used,[40] but such an approach is too time consuming. The FACTS approach offers 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 eqs. (6) and (7) one can define

$$D_i = A_i + d_1 B_i + d_2 A_i B_i \tag{10}$$

and

$$S_i^{\mathrm{FACTS}} = c_0 + \frac{c_1}{1 + e^{-c_2(D_i - c_3)}} \tag{11}$$

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 (i.e., $D_i \rightarrow +\infty$) the value of S_i should vanish which implies that $c_0 = -c_1$ and $c_2 > 0$. For a fully exposed atom (i.e., $D_i \rightarrow 0$) the analytical formula applies so that $c_0 = 4\pi(r_i^{\mathrm{vdW}} + 1.4)^2(1 + e^{-c_2 c_3})$ 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.[36]

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 nonpolar term

$$\Delta G^{\mathrm{FACTS}} = \Delta G^{\mathrm{el,FACTS}} + \gamma \sum_{i=1}^{N} S_i^{\mathrm{FACTS}} \tag{12}$$

where $\Delta G^{\mathrm{el,FACTS}}$ is detailed in eq. (9), and γ denotes the empirical surface tension parameter. Values of $\gamma = 0.015$ and $\gamma = 0.025$ kcal mol^{-1} Å^{-2} were used for the MD simulations presented in the Results section.

Comparison between FACTS and Other Analytical Solvation Models

Among the previously published implicit solvation models, those that can be efficiently employed for MD simulations make use of a (linear[21,23] or sigmoidal[41]) distance–dependent screening function instead of a constant dielectric in the denominator of the Coulomb formula. On the other hand, methods based on the GB equation (including FACTS) approximate the screening effects by taking into account not only the charge–charge distance but also the degree of solvent exposure of individual charges. Furthermore, ionic groups are neutralized in two of the three previous models[21,23] which are therefore not appropriate for (poly)peptides with several charged side chains.[42] It is important to note that a direct comparison of the reliability of GB approaches (including FACTS) and simple models based on distance-dependent screening function is not possible because of the different level of physical information and different number of parameters.[43]

There are also important differences in the evaluation of atomic (or self) solvation energy values between FACTS and other efficient implicit models. The Gaussian solvent-exclusion model of Lazaridis and Karplus (EEF1)[21] and the screened Coulomb potential (SCP) model of Hassan et al.[41] do not take into account the spatial symmetry of the displaced solvent whereas a symmetry term is explicitely used in FACTS, i.e., eq. (4). Furthermore, EEF1 assumes that the solvation free energy of a protein is a sum of group contributions and is parameterized with experimental data of small model compounds whereas atomic fdP values are used in the parametrization of FACTS. Hence, the EEF1 solvation energy cannot be decomposed into polar and nonpolar contributions.

Compared to most GB models, a common advantage of FACTS, EEF1,[21] and SCP[41] is that they do not require the definition of a boundary between solute and solvent. On the other hand, values of the dielectric constant of solute and solvent have to be specified in FACTS (and GB models) but not in EEF1 and SCP. Atomic solvation energies strongly depend on the dielectric constant of the solute ε_m.[43] Yet, the possibility of defining a solute dielectric constant increases the range of applicability of FACTS because $\varepsilon_m = 1$ is more appropriate for MD simulations while for structure prediction or docking values of $\varepsilon_m = 2$ or $\varepsilon_m = 4$ better approximate the effects of fluctuating dipoles in single-point energy calculations (see also Results and Discussion).

Parameterization of FACTS

Peptides and Proteins

A composite set of structured peptides (1cb3, a β-sheet from 1pgb, an α-helix from 1pgb, 1ly2, and Beta3s[44]), 18 single-chain proteins (1a2p, 1bpi, 1crn, 1dvd, 1f8a, 1fmk, 1hdn, 1h0l, 1inc, 1lz1, 1pgb, 1pht, 1shg, 1ubq, 2ci2, 2ptl, 3app, and 3pte), and 6 multi-chain proteins (1kvd, 1ycq, 1ycr, 2ins (chains A and B), 2ins (all chains), and 5hvp) of very different sizes, shapes, and secondary structure content was used. The number of residues ranges from 11 in 1cb3 to 347 in 3pte. The set includes almost spherical geometries with no cavities as well as structures with internal cavities. For instance, 5hvp is the HIV-1 aspartic proteinase in a complex with a peptidic ligand that was removed from the active site to obtain an internal cavity. To further diversify the set of structures with many different kinds of irregular shapes (cavities, open loops, etc.) the single-chain proteins were subjected to high temperature unfolding simulations at 450 K for 20 ns with an implicit solvent model.[23] From each trajectory a molten globule-like structure and a significantly extended conformation were selected and added to the initial set of structures. (For 1bpi only a molten globule-like structure was chosen

as it is strongly stabilized by three disulfide bridges and did not unfold sufficiently in the simulation. Similarly, for the very large complexes 1inc, 3app, and 3pte only a molten globule-like structure was added as a significantly extended conformation is too memory demanding for the fdP calculations.) The average increase in the radius of gyration is 26.1% and 92.2% for the molten globule like and significantly unfolded structures, respectively. Their average C_α-RMSD (root mean square deviation) is 13.1 Å and 17.7 Å, respectively. Furthermore, almost completely extended conformations of the structured peptides were included. The final training set consists of 81 (PARAM19, see later) and 72 (PARAM22, see later) conformations from 29 peptides and proteins.

Small Molecules

Recently, the potentials of mean force between pairs of charged side chains have been calculated in explicit water.[45] From this study a total of 12 arrangements originating from 7 molecular systems were selected: Glu-Glu head to head and orthogonal, His-Glu orthogonal, Lys-Glu head to head and orthogonal, Lys-Lys head to head, Arg-Glu head to head, Arg-Lys head to head and orthogonal, Arg-Arg head to head, orthogonal, and stacked. The distance was varied from 2.4 to 10 Å resulting in 77 conformations for each pair. Detailed descriptions of the structures and definitions of the distances are given in.[45] Furthermore, 77 conformations of the *N*-methyl-acetamide dimer in a planar arrangement were also considered. Again, the distance between the hydrogen bond donor and acceptor atoms was varied from 2.4 to 10 Å. The union set of all peptides, proteins, and small molecules consists of 1082 structures (81 protein conformations, 77 $\times$ 12 arrangements of pairs of charged side chains, and 77 *N*-methyl-acetamide dimer arrangements) derived from 37 molecular systems (29 proteins, 7 pairs of charged side chains, and the *N*-methyl-acetamide dimer). This constitutes a sound basis for a thorough fitting and assessment of the FACTS model.

Force Field Parameter Set

All calculations were performed using the CHARMM program (version c29b1) with the CHARMM polar hydrogen parameter set (PARAM19[39]) and the CHARMM all-hydrogen parameter set (PARAM22[46]). For the PARAM19 set the van de Waals radii of all hydrogen atoms are set to 1 Å in the fdP, FACTS, and GBMV calculations. For some computations (e.g., atomic solvation energies) all atoms are assigned unit charges to allow for a comparison that is unbiased by the charge parameter set.

Finite difference Poisson (fdP)

The benchmark commonly used to assess the accuracy of continuum electrostatics models are the energy values calculated by fdP. Atomic solvation energies $\Delta G_i^{\mathrm{el,fdP}}$ and pair interaction energies were calculated by numerical solution of the fdP equation with the PBEQ module[47] in CHARMM. All atoms were assigned unit charges for the fdP calculations. A grid spacing of 0.2 Å was used for all fdP calculations with proteins. For the pairs of charged side chains and the *N*-methyl-acetamide dimer a grid spacing of 0.1 Å was used. The van der Waals radii of all hydrogen atoms were set to 1 Å for PARAM19. No adjustments were applied to the van der Waals radii of PARAM22. The dielectric discontinuity boundary was defined by the molecular surface. The atomic solvation energy $\Delta G_i^{\mathrm{el,fdP}}$ of atom i is the solvation energy of the macromolecule when deleting the charges of all atoms except the one of atom i. Solvation energies were calculated by subtracting the self energy *in vacuo* ($\varepsilon_m = 1$, $\varepsilon_s = 1$) from the self energy in solution ($\varepsilon_m = 1$, $\varepsilon_s = 78.5$). The interaction energy of an (i,j) atom pair was obtained by calculating the electrostatic energy of a unit charge at the position of atom j in the electric field generated by a single unit charge at the position of atom i in the presence of solvent ($\varepsilon_m = 1$, $\varepsilon_s = 78.5$).

Parameter Optimization

For each van der Waals radius two sets of parameters have 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. Note that 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 are used to increase efficiency in MD simulations as the same list of atom pairs can be used for the electrostatic solvation energy and SASA. Optimal parameters were obtained by minimizing the deviations of $\Delta G_i^{\mathrm{el,FACTS}}$ from $\Delta G_i^{\mathrm{el,fdP}}$ and of S_i^{FACTS} from S_i^{exact}. A particle swarm algorithm[48] was used for parameter optimization. The fdP data from the 81 conformations (of the 29 peptides and proteins listed above) are included in the training set. The data for the charged side chain pairs and the *N*-methyl-acetamide dimer are only used for tests. All parameters are listed in the Suppl. Mat.

It is interesting and useful to assess the dependency of the FACTS parameters on the training set. The dependency is marginal because fitting all parameters on a single medium sized and globular protein (e.g., the native state of barnase (PDB code 1a2p)) yields a parameter set that performs almost as good as using all 81 protein conformations (Table 1 and Suppl. Mat).

Table 1. Cross Validation of FACTS.

Training set	All	Native 1a2p
Test set	All	All but 1a2p
PARAM19		
Average	3.4	3.5
SD	3.5	3.5
Max	44.7	47.7
PARAM22		
Average	3.3	3.4
SD	3.4	3.5
Max	44.2	43.8

The values are in kcal/mol and represent atomic electrostatic solvation energy deviations from fdP data calculated with unit charges and $\varepsilon_m = 1$. In the second column, the training set for FACTS parameter optimization is identical to the test set and consists of 81 and 72 protein structures for PARAM19 and PARAM22, respectively. In the third column, training and test sets are disjoint; the training set consists of only the native structure of barnase (1a2p) while the test set consists of all the remaining structures.

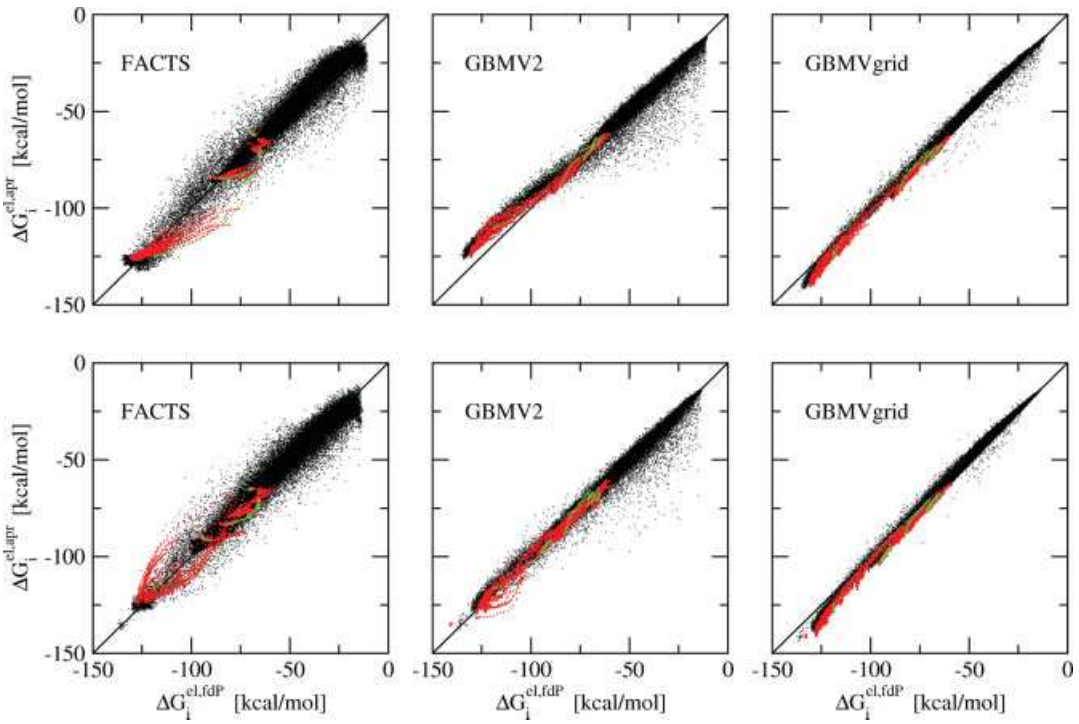

Figure 4. Comparison between FACTS and GBMV. The plots show values of the atomic electrostatic solvation energy evaluated with unit charges and $\varepsilon_m = 1$. The color coding for different molecular systems is the same as in Figure 3. (Top) PARAM19: Slope, correlation, and maximal absolute error for the 60,977 atoms in 81 protein structures are 0.963, 0.980, and 44.7 kcal/mol for FACTS; 0.910, 0.990, and 49.2 kcal/mol for GBMV2[25]; 1.028, 0.998, and 23.0 kcal/mol for GBMVgrid.[30] (Bottom) PARAM22: Slope, correlation, and maximal absolute error for the 62,873 atoms in 72 protein structures are 0.963, 0.981, and 44.2 kcal/mol for FACTS; 0.967, 0.993, and 56.7 kcal/mol for GBMV2; 1.045, 0.998, and 19.2 kcal/mol for GBMVgrid. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The only protocol that fails to produce reliable parameters is to fit only on small or very extended conformations. In both these cases the radius R^{sphere} is estimated too small, resulting in a significant loss of accuracy for large and compact conformations. On the other hand, a larger radius does not compromise the accuracy for small or very open structures (but has a negative effect on the efficiency). In retrospect these findings show that the data set used in this study to obtain the FACTS parameters is redundant. Yet, these findings are useful for additional parameterizations of the FACTS model (e.g., for CHARMM and $\varepsilon_m = 4$ or for another force field), which can be done with much less fdP data and therefore much faster.

Additional evidence for the robustness of the parameters is provided by a leave-one-out cross-validation procedure using 10 structures. The variation of the values of individual parameters is much smaller than their value averaged over the 10 models (Suppl. Mat.) which indicates that the fitting is statistically significant.

Molecular Dynamics Simulations

All MD simulations were performed with CHARMM[39] starting from the native structure downloaded from the PDB database.[49] Constant temperature MD simulations were carried out using weak coupling to a Berendsen's bath with a coupling constant of 5 ps. The CHARMM default truncation schemes of long-range electrostatics were used, i.e., a shift to zero energy at 7.5 Å and 12 Å for PARAM19 and PARAM22, respectively. The same cutoff values were employed for the van der Waals energy with a shifting and polynomial switching function for PARAM19 and PARAM22, respectively. The SHAKE algorithm was used to fix the length of the covalent bonds involving hydrogen atoms, which allows an integration time step of 2 fs. The nonbonding interactions were updated every 20 fs or using a heuristic update algorithm for simulations with PARAM19 or PARAM22, respectively, and coordinate frames were saved every 10 ps for analysis.

Results and Discussion

This section focuses on the results obtained using an interior (i.e., solute) dielectric $\varepsilon_m = 1$. Note that using $\varepsilon_m = 1$ (instead of $\varepsilon_m = 2$ or $\varepsilon_m = 4$) is the most stringent test of the accuracy of a continuum dielectric model. For single-point energy calculations (e.g., for structure prediction or ranking in ligand binding)

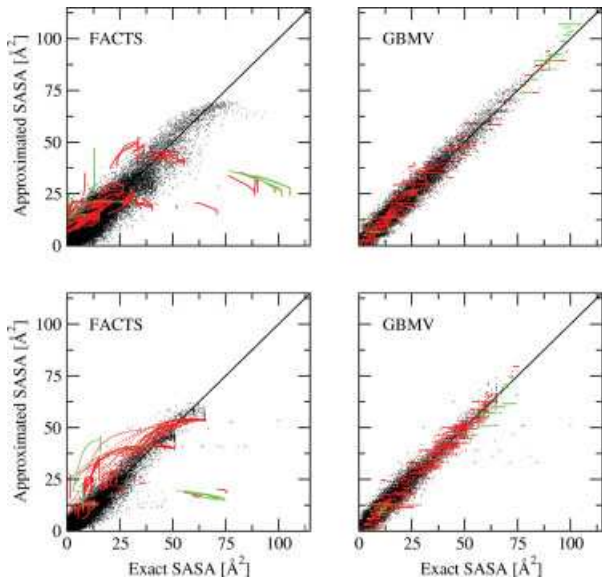

Figure 5. Comparison of atomic SASA evaluation by FACTS (ε_m = 1 parametrization) and GBMV.[25] The benchmark are the exact values of atomic SASA.[36] The color coding for different molecular systems is the same as in Figure 3. (Top) PARAM19: Slope, correlation, and maximal absolute error for the atoms in the protein structures are 0.919, 0.955, and 44.2 Å^2 for FACTS; 1.001, 0.986, and 14.1 Å^2 for GBMV. (Bottom) PARAM22: Slope, correlation, and maximal absolute error for the atoms in the protein structures are 0.942, 0.970, and 33.2 Å^2 for FACTS; 1.001, 0.983, and 14.3 Å^2 for GBMV. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

ε_m = 2 or ε_m = 4 would be more appropriate since values of $\varepsilon_m > 1$ account for thermal fluctuations of protein dipoles. As the FACTS model is primarily aimed to be used in MD simulations, the validation with ε_m = 1 is discussed in detail in the present study. However, parameterizations of the FACTS model for ε_m = 2 have also been performed and results are presented in the Suppl. Mat.

Atomic (or Self) Electrostatic Solvation Energy

It is interesting to assess the gain in accuracy by combining the measures of volume and symmetry instead of using only either of them. For this purpose the optimizing procedure for atomic electrostatic solvation energies was performed three times: by using both measures [as in eq. (6), i.e., $C_i = A_i + b_1B_i + b_2A_iB_i$], by using only the measure of volume ($\tilde{C}_i = A_i + \tilde{b_1}A_i^2 + \tilde{b_2}A_i^3$), and by using only the measure of symmetry ($\hat{C} = B_i + \hat{b_1}B_i^2 + \hat{b_2}B_i^3$). Note that the number of parameters is the same in all three situations. Plots of atomic solvation energy values calculated with the three different C_i's versus fdP values are shown in Figure 3. Interestingly, the measure of volume yields more accurate solvation energies than the measure of symmetry for buried atoms (solvation energy close to zero), whereas the measure of symmetry is better for solvent exposed atoms (favorable solvation energy). This observation provides evidence for the synergistic effect of combining the two measures. [Note that for an objective evaluation of the data plotted in Figure 3 the parameters a_0 and a_1 in eq. (7) were not fixed because the limites to determine a_0 and a_1 cannot be applied for the symmetry measure alone.]

Figure 4 shows atomic electrostatic solvation energy values calculated by FACTS [eq. (7)], GBMV2,[25] and GBMVgrid[30] versus the benchmark fdP values. The numerical approach GBMVgrid is the most accurate method, followed by GBMV2 and FACTS. However, the maximal absolute error is largest for GBMV2 because of some significant outliers. Similar behavior is observed for both PARAM19 and PARAM22.

Atomic Solvent Accessible Surface Area (SASA)

The correlation between SASA values of atoms in protein structures calculated by FACTS and exact values is 0.96 and 0.97 for PARAM19 and PARAM22, respectively (Fig. 5). The accuracy of the GBMV surface algorithm[25] is slightly higher than FACTS, and more so for PARAM19. The largest deviations in FACTS PARAM19 are observed for atoms with little solvent accessibility and originate from the relatively large sphere radii of the carbon atoms, which are close to 10 Å (Suppl. Mat.). It has to be remembered that 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. Large discrepancies are observed mainly for the small molecular systems, i.e., pairs of charged side chains and the *N*-methyl-acetamide dimer, which is also a consequence of the large sphere radii. Interestingly, both GBMV and FACTS yield more accurate atomic SASA values than the approach by Hasel et al.[40] (see Suppl. Mat.).

Pairwise Electrostatic Energies and Their Sums

Screened interaction energies, i.e., pairwise energies in solution,[50] are calculated for FACTS, GBMV2, and GBMVgrid by the formula

$$G_{ij} = \frac{q_iq_j}{\epsilon_m r_{ij}} - \frac{\tau q_iq_j}{\sqrt{r_{ij}^2 + R_iR_j \exp(-r_{ij}^2/\kappa R_iR_j)}} \qquad (13)$$

where the Born radii R_i are evaluated using the respective models. The agreement with fdP values is excellent for all three methods (Suppl. Mat.). For assessment of individual conformations, e.g., in structure prediction or docking, one is not primarily interested in individual pair interaction energies G_{ij}. The relevant quantity is rather the sum over all pairwise energies involving a given solute atom i, i.e., $G_i = \sum_{j \neq i} G_{ij}$. Accurate reproduction of G_{ij} in a given model with respect to the fdP values does not necessarily imply accurate reproduction of G_i since individual errors may not compensate among each other. A good agreement with fdP values of G_i is obtained using FACTS (Suppl. Mat.).

Electrostatic Solvation Energy of Protein Conformations

To assess the accuracy of FACTS in calculating macromolecular solvation energy a large variety of conformations for 29 proteins

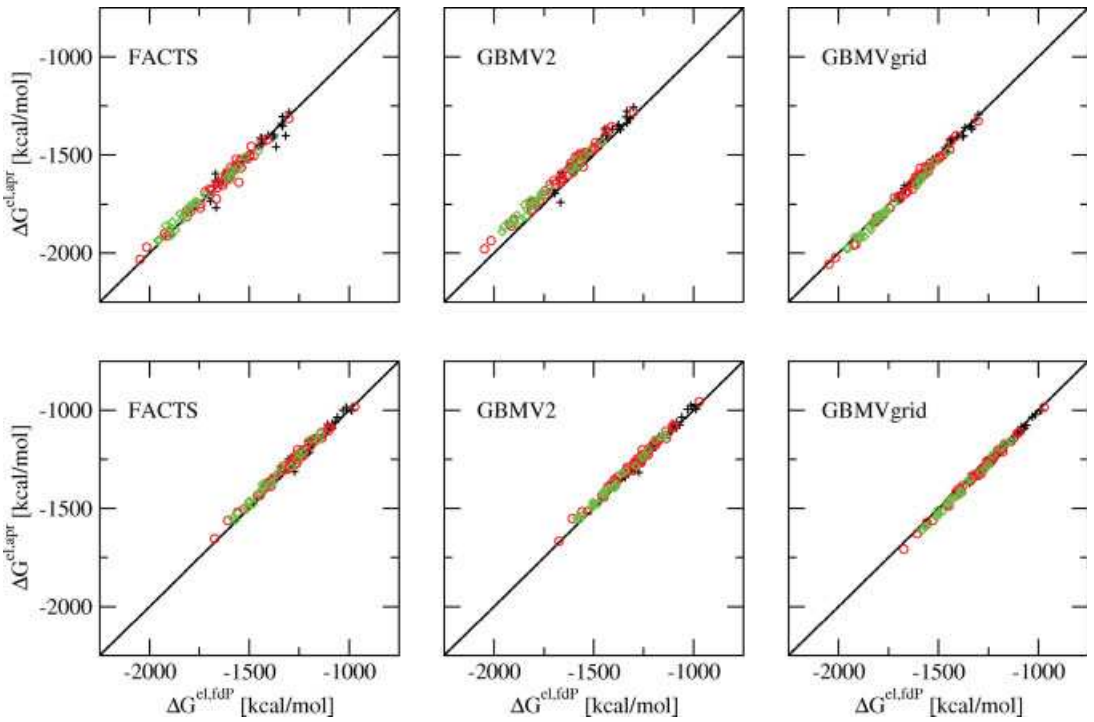

Figure 6. Comparison of protein electrostatic solvation energy values calculated by FACTS [eq. (9), $\kappa = 12$] and GBMV. Each plot shows data for 100 conformations of barnase with PARAM19 (top) and PARAM22 (bottom). The structures were chosen along a high temperature unfolding trajectory started from the 1a2p X-ray structure. Different symbols discriminate between different ranges of the radius of gyration. Plus and diamond symbols represent the 25 conformations with small and large radius of gyration, respectively, while circles the 50 intermediate ones. The benchmark are the fdP values with $\varepsilon_m = 1$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

were generated by 50-ns MD simulations of unfolding at 450 K using an implicit solvent model.[23] Coordinates were saved every 10 ps and all snapshots were sorted according to increasing radius of gyration (R_g). A total of 100 conformations were chosen from each trajectory as follows: every 20th conformation from the 500 snapshots with the lowest R_g (25 conformations), every 20th conformation from the 500 snapshots with the largest R_g (25 conformations), and every 80th conformation from the remaining 4000 snapshots (50 conformations). The 100 conformations of each protein cover a wide range of RMSD and R_g. For the 29×100 conformations the values of the electrostatic solvation energy ΔG^{FACTS} [eq. (8)], ΔG^{GBMV2}, and $\Delta G^{GBMVgrid}$ were calculated and compared to ΔG^{fdP}.

The results for barnase (1a2p) show that the agreement between approximated and exact (i.e., fdP) values is very good for the three models (Fig. 6). As indicated by the percentage

Table 2. Percentage Error of Electrostatic Solvation Energy Values of 2900 Protein Conformations (100 Conformations from Each of 29 Trajectories).

	FACTS, $\kappa = 4$	FACTS, $\kappa = 8$	FACTS, $\kappa = 12$	GBMV2, $\kappa = 8$	GBMVgrid, $\kappa = 8$
PARAM19					
Average [%]	1.96	1.36	1.32	2.37	2.04
SD [%]	1.57	1.25	1.26	1.34	1.31
Max [%]	10.37	11.78	12.21	8.62	8.76
PARAM22					
Average [%]	3.33	1.72	1.42	1.28	2.29
SD [%]	1.88	1.20	1.08	1.03	1.48
Max [%]	11.64	7.40	6.97	6.26	7.21

The parameter κ is in the interaction term of eq. (8). The benchmark are the fdP values with $\varepsilon_m = 1$.

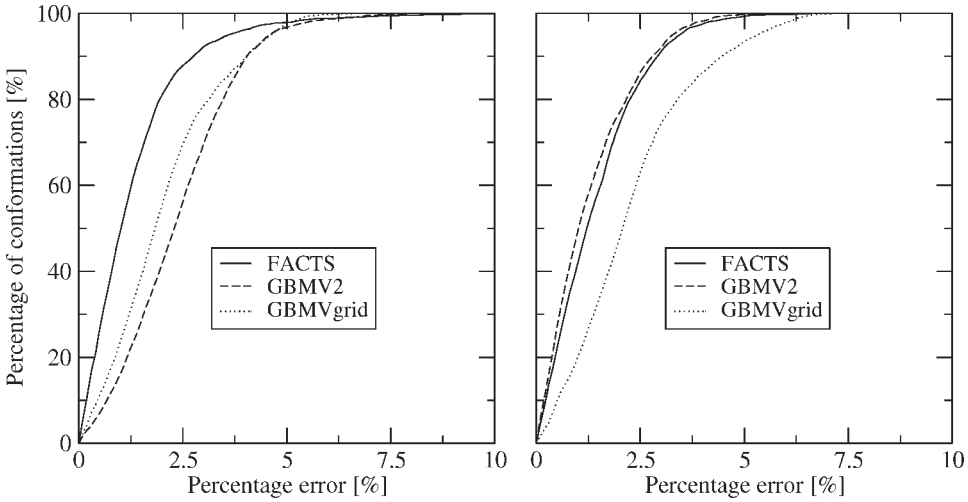

Figure 7. Cumulative histogram of percentage errors of electrostatic solvation energy values. The FACTS eq. (9) with κ=12 was used for 2900 structures of 29 proteins for PARAM19 (left) and PARAM22 (right). The benchmark are the fdP values with $\varepsilon_m = 1$.

error for all proteins, the accuracy of FACTS improves significantly by using $\kappa = 8$ instead of $\kappa = 4$, while only a marginal improvement is observed with $\kappa = 12$ (Table 2). Notably, with $\kappa = 12$ the percentage error of $\Delta G^{\mathrm{FACTS}}$ averaged over all 2900 conformations is only 1.32% and 1.42% with PARAM19 and PARAM22, respectively. Only with PARAM22 is GBMV2 (with its default value of $\kappa = 8$[25]) more accurate than FACTS, which is probably a consequence of the fact that GBMV2 was

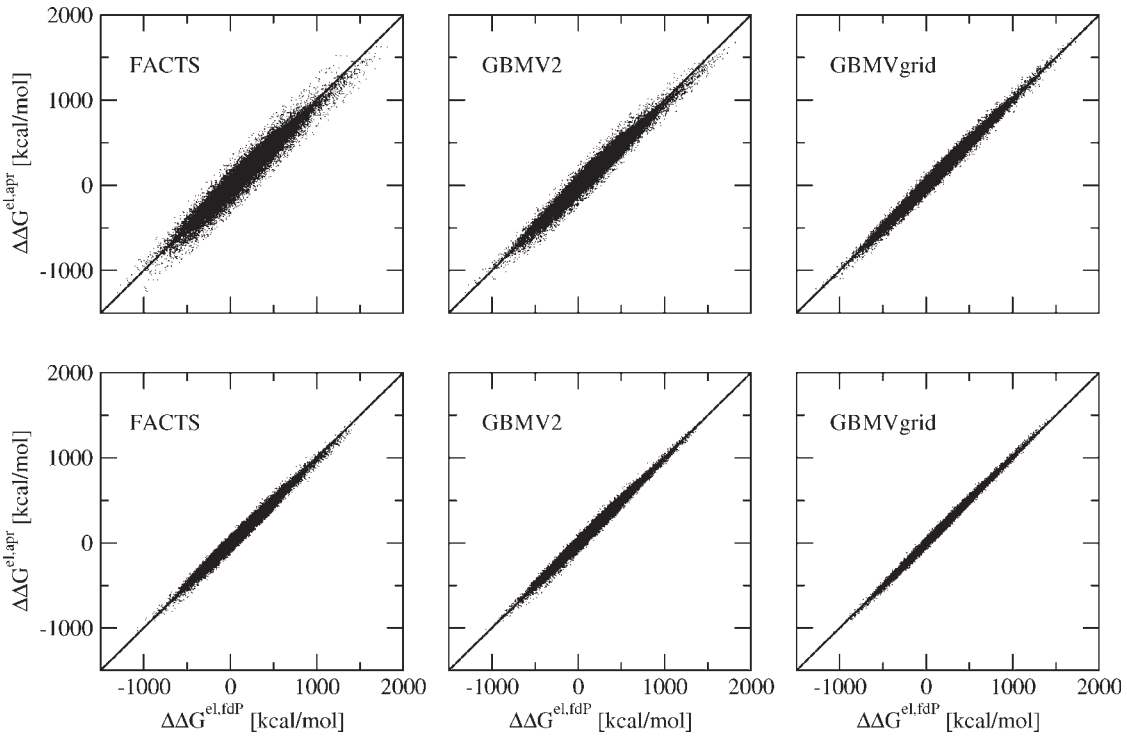

Figure 8. Comparison of relative electrostatic solvation energy values calculated by FACTS ($\kappa = 12$) and GBMV for $\varepsilon_m = 1$. From each of 29 trajectories, 100 conformations were chosen as described in the text and the difference in solvation energy ($\Delta\Delta G$) for all possible pairs of structures was evaluated. In this way, an eventual systematic offset in solvation energy relative to the benchmark fdP values is eliminated. Such offset is irrelevant for MD simulations.

Table 3. Differences in Electrostatic Solvation Energy from Pairs of Protein Conformations ($\Delta\Delta G$).

	FACTS, $\kappa = 4$	FACTS, $\kappa = 8$	FACTS, $\kappa = 12$	GBMV2, $\kappa = 8$	GBMVgrid, $\kappa = 8$
PARAM19					
Average	26.91	25.45	25.18	20.04	14.43
SD	36.70	34.99	34.46	26.10	16.70
Max	393	374	367	269	193
PARAM22					
Average	17.28	15.89	15.77	13.94	10.39
SD	17.92	16.46	16.25	14.98	10.20
Max	209	180	169	177	122

A total of 29 × 4950 values of solvation energy differences were calculated. All values are in kcal/mol. The parameter κ is in the interaction term of eq. (8). The benchmark are the fdP values with $\varepsilon_m = 1$.

optimized mainly for PARAM22. The cumulative histogram (Fig. 7) shows that 95% of the 2900 conformations have an error in the FACTS solvation energy smaller than 3.60% and 3.46% with PARAM19 and PARAM22, respectively.

For most applications of force-field based methods, the crucial quantity is the difference in electrostatic solvation energy between two structures of the same molecular system, i.e., $\Delta\Delta G$. These differences are calculated for all pairs of structures for each trajectory for FACTS, GBMV2, and GBMVgrid and compared to the fdP values. The results are shown in Figure 8 and Table 3. FACTS performs almost as well as GBMV2. The GBMVgrid approach is the most accurate of the three models but cannot be used for MD simulations because it is a numerical method. All the three methods show higher accuracy for PARAM22 than PARAM19. For FACTS, this result is probably a consequence of the larger number of values of van der Waals radii (15 in PARAM22 and 7 in PARAM19), while for GBMV it is due to the aforementioned optimization for PARAM22.

Molecular Dynamics Simulations

FACTS has been implemented into CHARMM version c29b1. Because of the fully analytical treatment of solvation in FACTS the total energy does not drift in NVE simulations even with a time step of 2 fs (Fig. 9), whereas GBMV requires a time step of 1 fs to reduce the energy drift.[51]

The native state of structured peptides and proteins is stable over 100-ns MD runs at 300 K (Table 4). Interestingly, the MD results with FACTS PARAM22 are similar for two different values of the surface tension-like parameter ($\gamma = 0.015$ and $\gamma = 0.025$ kcal mol^{-1} Å^{-2}), which indicates robustness with respect to the relative weighting (i.e., balancing) of polar and nonpolar solvation. Only 1cb3 and 1abz show a C_α-RMSD larger than 3.5 Å after 100 ns with both values of the parameter γ. These findings are consistent with experimental data. The PDB entry 1cb3 is an ensemble of NMR conformers of the segment 101–111 of α-lactalbumin, which is flexible when isolated from the context of the protein. In fact, the five C-terminal residues of this segment were shown by NMR to be essentially unstructured in water at 283 K.[53] The de novo designed 38-residue α-helical hairpin peptide $\alpha t\alpha$ (PDB 1abz) was estimated to be only 60% helical at 298 K by circular dichroism.[54]

A common artifact of MD simulations in vacuo is the very small atomic fluctuations. The RMS fluctuations of the C_α atoms of chymotrypsin inhibitor 2 (PDB 2ci2) along a FACTS PARAM22 300 K MD simulation are in agreement with the corresponding values derived from crystallographic B-factors (Fig. 10). In particular, the N-terminal segment and the loop (residues 38–44) are the most flexible regions according to both MD simulations and X-ray data.[55] As expected, slightly larger fluctuations are observed with the smaller of the two values of the surface tension-like parameter γ used for the nonpolar term in the MD simulations.

Using FACTS, the reversible folding to the NMR conformer has been observed in 1-μs simulations of Beta3s, a designed 20-residue three-stranded antiparallel β-sheet.[44] Moreover, the thermodynamic stability (i.e., free energy difference between folded and denatured state) of Beta3s is lower using FACTS (about 65 and 2% folded population at 300 and 330 K, respectively) than a simple SASA-based solvation model (about 95 and 50% folded population at 300 and 330 K, respectively[56,57]). At 283 K *in vitro*, the β-sheet population is 13–31% according to the Nuclear Overhauser Enhancements data and 30–55% according to the chemical shifts.[52] The FACTS simulations of Beta3s were

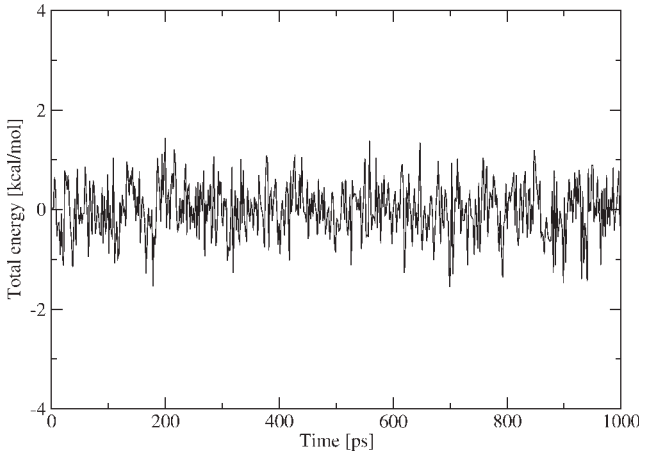

Figure 9. Time series of the total energy relative to the starting conformation of protein G (1pgb) during 1 ns MD simulation in the NVE ensemble using FACTS PARAM19 and a time step of 2 fs.

Table 4. Deviation from the Native Structure During MD Simulations at 300 K.

PDB	Residues	$\langle\rangle_{10}$	$\langle\rangle_{20}$	$\langle\rangle_{30}$	$\langle\rangle_{40}$	$\langle\rangle_{50}$	$\langle\rangle_{60}$	$\langle\rangle_{70}$	$\langle\rangle_{80}$	$\langle\rangle_{90}$	$\langle\rangle_{100}$
1cb3	11	3.5	3.7	4.0	4.1	3.9	4.1	3.8	3.9	4.1	4.1
1l2y	20	1.3	1.1	1.0	1.3	1.1	1.1	1.1	1.1	1.0	1.1
Beta3s[a]	20	2.0	2.0	2.0	2.4	3.2	3.0	2.4	2.4	2.5	2.4
1f8a	33	1.7	3.4	3.4	3.4	3.4	3.4	3.5	3.4	3.4	3.4
1vii	36	2.2	2.0	2.0	1.9	2.0	1.9	2.0	1.9	1.9	1.9
1abz	38	4.2	3.8	3.6	4.0	4.3	4.2	4.5	4.6	4.5	4.5
1crn	46	0.9	0.9	1.0	1.1	1.5	1.4	1.1	0.9	0.9	0.8
1enh	54	1.6	4.2	4.2	4.1	4.2	4.1	4.1	3.6	3.3	3.3
1pgb	56	1.0	1.0	1.1	1.0	1.2	1.0	1.0	1.0	0.9	1.0
1bpi	58	1.8	1.9	2.1	2.1	2.1	2.2	2.2	2.2	2.3	2.2
1fmk	59	1.7	1.8	1.8	2.2	2.3	2.3	2.3	2.3	2.3	2.3
2ci2	65	2.0	2.3	2.3	2.3	2.4	2.3	2.6	2.5	2.5	2.6
2a3d	73	2.9	3.0	3.2	3.1	3.0	2.8	2.9	2.8	2.9	3.1
1ubq	76	1.4	1.4	1.4	1.9	2.0	2.0	2.0	2.0	2.0	2.0
1pht	83	1.8	2.2	2.1	2.2	2.1	2.2	2.3	2.4	2.4	2.4
1hdn	85	1.9	2.0	1.9	2.0	2.0	2.4	2.6	2.7	2.7	2.6
1a2p	108	2.9	3.2	3.2	3.5	3.9	4.2	4.3	4.3	4.3	4.3
1cb3	11	3.5	4.1	4.0	4.0	4.1	4.1	3.4	3.6	3.6	3.7
1l2y	20	3.8	4.3	4.5	5.0	5.0	5.0	5.0	5.0	4.9	5.1
Beta3s[a]	20	2.4	3.1	3.0	2.8	2.7	2.8	2.6	2.7	2.7	2.8
1f8a	33	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.1
1vii	36	2.0	1.9	1.9	1.9	2.0	1.9	1.9	1.9	1.9	1.9
1abz	38	5.3	6.0	6.3	5.5	5.4	5.5	6.2	6.2	5.9	5.9
1crn	46	0.9	1.0	0.8	0.8	0.9	0.8	0.9	1.0	1.0	1.0
1enh	54	1.5	1.8	2.1	2.4	2.4	2.3	2.3	2.1	1.6	1.6
1pgb	56	0.9	0.9	0.9	1.1	1.1	1.1	1.1	1.2	1.0	1.1
1bpi	58	1.8	2.0	2.1	2.0	2.1	2.0	2.0	2.0	2.0	2.0
1fmk	59	1.3	1.4	1.6	1.8	1.7	1.6	1.6	1.7	1.7	1.7
2ci2	65	2.0	1.9	1.7	1.7	1.9	1.8	1.8	2.1	2.1	2.0
2a3d	73	3.3	3.2	3.3	3.4	3.2	3.2	3.3	3.2	3.3	3.3
1ubq	76	1.4	1.8	2.0	2.0	2.0	2.0	1.9	2.0	1.9	2.0
1pht	83	1.9	2.3	2.5	2.7	3.5	3.5	2.9	2.7	2.8	2.8
1hdn	85	1.5	1.5	1.5	1.4	1.4	1.5	1.5	1.5	1.5	1.5
1a2p	108	2.4	2.6	2.5	2.7	2.9	3.0	3.4	3.3	3.3	3.3

Individual columns contain values in Å of the C_α-RMSD from the native structure averaged over 10 ns intervals, e.g., for the last column $<>_{100}$ the C_α-RMSD was averaged over the 90–100 ns interval. The simulations were performed with FACTS PARAM22, $\kappa = 4$, $\varepsilon_m = 1.0$, and using $\gamma = 0.015$ kcal mol^{-1} Å^{-2} (top) and $\gamma = 0.025$ kcal mol^{-1} Å^{-2} (bottom).

[a]Beta3s is a three-stranded antiparallel β-sheet peptide.[44,52]

performed using PARAM19 with a solute dielectric constant $\varepsilon_m = 1$ and $\gamma = 0.015$ kcal mol^{-1} Å^{-2}. An in-depth analysis of reversible folding of structured peptides and small proteins will be presented elsewhere.

FACTS Computational Requirements

Using the same nonbonding cutoff, MD simulations with FACTS are about four times slower than *in vacuo* but about 10 times faster than with GBMV2. Notably, on a single Opteron 1.8 GHz processor, a 100-ns MD run of the 46-residue crambin (1crn) requires 4 and 22 CPU-days with FACTS PARAM19 (396 atoms and 7.5 Å cutoff) and FACTS PARAM22 (642 atoms and 12 Å cutoff), respectively (Table 5). Moreover, the CPU-time scales linearly with protein size (Fig. 11). The extra memory requirements for FACTS with respect to a vacuum calculation are marginal and they originate solely from the atom-pair list, which is used for both electrostatic and SASA calculations. As an example, with the current implementation of FACTS into the c29b1 version of CHARMM only 12 Mb of RAM are needed for FACTS PARAM19 MD simulations of the 389-residue protein β-secretase (PDB 1sgz).

Conclusions

A fully analytical treatment of solvation in the continuum model has been developed using empirical formulas and fitting to atomic solvation energy values calculated by numerical solution of the Poisson equation. The method, called FACTS, is very efficient because it is based on simple measures of solvent dis-

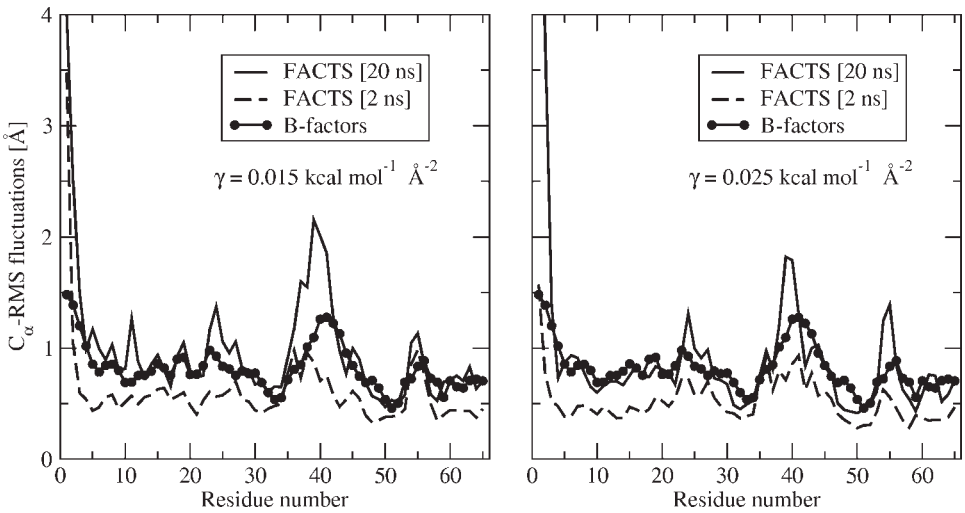

Figure 10. RMS fluctuations in Å of the C_α atoms of CI2. FACTS PARAM22 was used with $\varepsilon_m = 1$, κ=4, and surface tension-like parameter $\gamma = 0.015$ kcal mol^{-1} Å^{-2} (left), and $\gamma = 0.025$ kcal mol^{-1} Å^{-2} (right). The fluctuations were extracted from a 300 K simulation started from the native structure (2ci2) and considering a trajectory segment of 2 ns (dashed line) and 20 ns (solid line). The bold line with circles represents the fluctuations derived from the crystallographic B-factors[55] using the formula RMS fluctuation = $(3B/(8\pi^2))^{0.5}$.

placement and thus requires only distances between solute atoms which are close in three-dimensional space. These interatomic distances have anyway to be calculated for the nonbonding terms of a force field. FACTS does not use a dielectric boundary nor does it assume the Coulomb field approximation. The agreement between FACTS and energy values calculated by the finite-difference Poisson technique is good and comparable to the one of the most accurate GB methods that use empirical corrections to the Coulomb field approximation. In MD simulations of proteins FACTS is about ten times faster than the most accurate GB implementations. The native state of structured peptides and proteins is stable during 300 K MD runs of 100 ns using FACTS in combination with the CHARMM force field. Moreover, marginally stable peptides and unstructured loops in proteins are flexible under the same conditions. The accuracy and efficiency of FACTS suggest that it could also be used for protein structure prediction and docking.

Table 5. Computation Time Required for 100-ns MD Simulations with FACTS.

		PARAM19		PARAM22	
PDB	Residues	Atoms	CPU-days	Atoms	CPU-days
1cb3	11	112	0.6	186	3.5
Beta3s[a]	20	215	2.2	329	8.4
1crn	46	396	3.8	642	21.9
2ci2	65	636	6.6	1076	43.1
1a2p	108	1073	14.5	1700	71.3

All simulations were performed on a single Opteron 1.8 GHz processor.
[a]Beta3s is a three-stranded antiparallel β-sheet peptide.[44,52]

Acknowledgments

We thank Francois Marchand, Riccardo Pellarin, Andrea Prunotto, Manuel Walker, Nicolas Majeux, Stefanie Muff, and Giovanni Settanni for very interesting and useful discussions. We also thank the anonymous referees for very interesting and useful comments. The calculations were performed on a Beowulf Linux cluster at the Informatikdienste of the University of Zurich and we thank C. Bolliger, T. Steenbock and Dr. A. Godknecht for their help in setting up and maintaining the cluster.

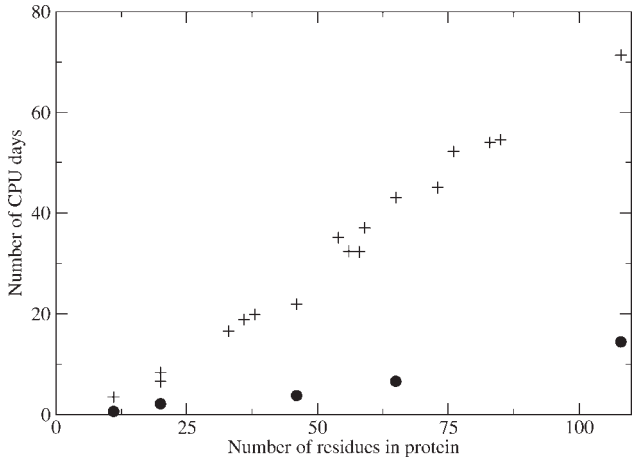

Figure 11. System-size scaling of CPU-time required for 100-ns MD simulations with FACTS. Circles and plus symbols correspond to simulations with PARAM19 and PARAM22, respectively.

References

1. Settanni, G.; Rao, F.; Caflisch, A. Proc Natl Acad Sci USA 2005, 102, 628.
2. Warshel, A.; Russell, S. T. Q Rev Biophys 1984, 17, 283.
3. Warshel, A.; Papazyan, A. Curr Opin Struct Biol 1998, 8, 211.
4. Tomasi, J.; Persico, M. Chem Rev 1994, 94, 2027.
5. Gilson, M. K. Curr Opin Struct Biol 1995, 5, 216.
6. Roux, B.; Simonson, T. Biophys Chem 1999, 78, 1.
7. Cramer, C. J.; Trulhar, D. G. Chem Rev 1999, 99, 2161.
8. Bashford, D.; Case, D. A. Annu Rev Phys Chem 2000, 51, 129.
9. Orozco, M.; Luque, F. J. Chem Rev 2000, 100, 4187.
10. Simonson, T. Curr Opin Struct Biol 2001, 11, 243.
11. Werner, P.; Caflisch, A. J Am Chem Soc 2003, 125, 4600.
12. Feig, M.; Brooks, C. L., III. Curr Opin Struct Biol 2004, 14, 217.
13. Baker, N. A. Curr Opin Struct Biol 2005, 15, 137.
14. Im, W.; Chen, J.; Brooks, C. L., III. Adv Protein Chem 2006, 72, 171.
15. Feig, M.; Chocholousova, J.; Tanizaki, S. Theor Chem Acc 2006, 116, 194.
16. Warwicker, J.; Watson, H. C. J Mol Biol 1982, 157, 671.
17. Gilson, M. K.; Honig, B. H. Proteins: Struct Funct Bioinfo 1988, 4, 7.
18. Bashford, D.; Karplus, M. Biochemistry 1990, 29, 10219.
19. Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. Comput Phys Comm 1991, 62, 187.
20. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127.
21. Lazaridis, T.; Karplus, M. Proteins: Struct Funct Bioinfo 1999, 35, 133.
22. Hassan, S. A.; Guarnieri, F.; Mehler, E. L. J Phys Chem B 2000, 104, 6478.
23. Ferrara, P.; Apostolakis, J.; Caflisch, A. Proteins: Struct Funct Bioinfo 2002, 46, 24.
24. Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III. J Comput Chem 2003, 25, 264.
25. Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L., III. J Comput Chem 2003, 24, 1348.
26. Schaefer, M.; Karplus, M. J Phys Chem 1996, 100, 1578.
27. Haberthür, U.; Majeux, N.; Werner, P.; Caflisch, A. J Comput Chem 2003, 24, 1936.
28. Scarsi, M.; Apostolakis, J.; Caflisch, A. J Phys Chem A 1997, 101, 8098.
29. Ghosh, A.; Rapp, C. S.; Friesner, R. A. J Phys Chem B 1998, 102, 10983.
30. Lee, M. S.; Salsbury, F. R.; Brooks, C. L., III. J Chem Phys 2002, 116, 10606.
31. Hawkins, G. D.; Cramer, C. J.; Trulhar, D. G. J Phys Chem 1996, 100, 19824.
32. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. J Phys Chem A 1997, 101, 3005.
33. Dominy, B. N.; Brooks, C. L., III. J Phys Chem B 1999, 103, 3765.
34. Onufriev, A.; Bashford, D.; Case, D. A. J Comput Chem 2002, 23, 1297.
35. Im, W.; Lee, M. S.; Brooks, C. L., III. J Comput Chem 2003, 24, 1691.
36. Lee, B.; Richards, F. M. J Mol Biol 1971, 55, 379.
37. Eisenberg, D.; McLachlan, A. D. Nature 1986, 319, 199.
38. Ooi, T.; Oobatake, M.; Némethy, M.; Scheraga, H. A. Proc Natl Acad Sci USA 1987, 84, 3086.
39. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
40. Hasel, W.; Hendrickson, T. F.; Still, W. C. Tetrahedron Comput Methodol 1988, 1, 103.
41. Hassan, S. A.; Mehler, E. L.; Zhang, D.; Weinstein, H. Proteins: Struct Funct Bioinfo 2003, 51, 109.
42. Cecchini, M.; Curcio, R.; Pappalardo, M.; Melki, R.; Caflisch, A. J Mol Biol 2006, 357, 1306.
43. Hassan, S. A.; Mehler, E. L. Int J Quant Chem 2005, 102, 986.
44. Ferrara, P.; Caflisch, A. Proc Natl Acad Sci USA 2000, 97, 10780.
45. Masunov, A.; Lazaridis, T. J Am Chem Soc 2003, 125, 1722.
46. MacKerell, A. D. Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L. Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. J Phys Chem B 1998, 102, 3586.
47. Im, W.; Beglov, D.; Roux, B. Comput Phys Commun 1998, 111, 59.
48. Kennedy, J.; Eberhart, R. C. Swarm Intelligence; Morgan Kaufmann Publishers Inc.: San Fransisco; 2001.
49. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. Nucl Acids Res 2000, 28, 235.
50. Scarsi, M.; Caflisch, A. J Comput Chem 1999, 14, 1533.
51. Chocolousova, J.; Feig, M. J Comput Chem 2006, 27, 719.
52. De Alba, E.; Santoro, J.; Rico, M.; Jiménez, M. A. Protein Science 1999, 8, 854.
53. Demarest, S. J.; Hua, Y.; Raleigh, D. P. Biochemistry 1999, 38, 7380.
54. Fezoui, Y.; Weaver, D. L.; Osterhout, J. J. Proc Natl Acad Sci USA 1994, 91, 3675.
55. McPhalen, C. A.; James, M. N. G. Biochemistry 1987, 26, 261.
56. Cavalli, A.; Haberthür, U.; Paci, E.; Caflisch, A. Protein Science 2003, 12, 1801.
57. Muff, S.; Caflisch, A. Proteins: Struct Funct Bioinfo DOI: 10.1002/prot.21565.