JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

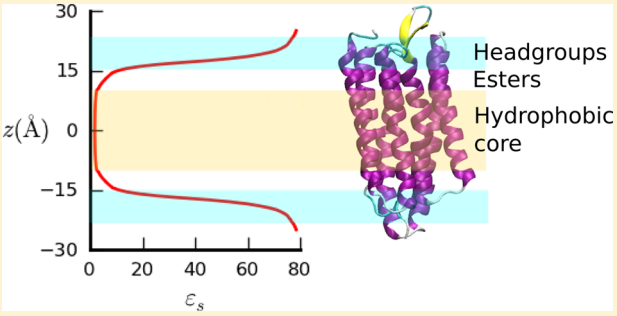
Extension of the FACTS Implicit Solvation Model to Membranes

Martín Carballo-Pacheco,[†] Ioan Vancea,[†,§] and Birgit Strodel*[,†,‡]

[†]Forschungszentrum Jülich GmbH, Institute of Complex Systems: Structural Biochemistry (ICS-6), 52425 Jülich, Germany

[‡]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, Universitätstrasse 1, 40225 Düsseldorf, Germany

ABSTRACT: The generalized Born (GB) formalism can be used to model water as a dielectric continuum. Among the different implicit solvent models using the GB formalism, FACTS is one of the fastest. Here, we extend FACTS so that it can represent a membrane environment. This extension is accomplished by considering a position dependent dielectric constant and empirical surface tension parameter. For the calculation of the effective Born radii in different dielectric environments we present a parameter-free approximation to Kirkwood's equation, which uses the Born radii obtained with FACTS for the water environment as input. This approximation is tested for the calculation of self-free energies, pairwise interaction energies in solution and solvation free energies of complete protein conformations. The results compare well to those from the finite difference Poisson method. The new implicit membrane model is applied to estimate free energy insertion profiles of amino acid analogues and in molecular dynamics simulations of melittin, WALP23 and KALP23, glycophorin A, bacteriorhodopsin, and a Clc channel dimer. In all cases, the results agree qualitatively with experiments and explicit solvent simulations. Moreover, the implicit membrane model is only six times slower than a vacuum simulation.



INTRODUCTION

Membrane proteins account for around 25% of the eurkaryotic genome[1] and are the target of around 50% of drugs. An increasing number of membrane proteins are being resolved experimentally (an up to date count can be checked at http://blanco.biomol.uci.edu/mpstruc/).[2] The dynamics of these recently resolved proteins can be understood with the help of molecular simulations.[3,4] Moreover, molecular simulations can also predict the unknown structure of peptide assemblies inside membranes.[5] In these simulations, the most accurate description of the water and membrane environments is undoubtedly obtained representing both of them explicitly. However, this can become computationally too demanding. To study both longer time scales and larger proteins, there is a need for coarser representations.

One of the most common approximations is the use of implicit solvent models.[6−9] These models, in which the solvent is represented in a mean field approach, estimate the interaction between the protein and the solvent by calculating the solvation free energy ΔG^{solv}. ΔG^{solv} is commonly divided into electrostatic and nonpolar contributions. The electrostatic contribution could be calculated with a numerical solution to Poisson's equation, such as the finite difference Poisson (fdP) method.[10,11] However, fdP calculations are too demanding for molecular dynamics (MD) simulations and most implicit solvents today use the approximate generalized Born (GB) formalism.[6,12−21] Within this formalism, the electrostatic solvation free energy is calculated as

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

where $\tau = 1/\varepsilon_p - 1/\varepsilon_s$, ε_s is the solvent dielectric constant, ε_p is the protein dielectric constant, q_i is charge of atom i, r_{ij} is the distance between atoms i and j, κ is a constant, usually equal to 4, and N is the number of atoms. R_i is the so-called effective Born radius of atom i and can be calculated as

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

where ΔG_i^{el} is the self-electrostatic solvation free energy. The effective Born radius is a measure of the degree of burial of each atom in the protein and the main difference between different GB models is the way in which the effective Born radii are estimated.[13−22] A correct calculation of the effective Born radii leads to a correct estimation of the electrostatic solvation free energy when compared to the fdP method.[23]

The nonpolar contribution to ΔG^{solv} is normally calculated using the solvent accessible surface area (SASA) formalism. In this formalism, the nonpolar contribution is calculated as

$$\Delta G^{np} = \gamma\sum_{i=1}^{N}\mathrm{SASA}_i \quad (3)$$

Received: February 3, 2014
Published: May 27, 2014

dx.doi.org/10.1021/ct500084y | J. Chem. Theory Comput. 2014, 10, 3163−3176

where γ is the effective surface tension parameter and $SASA_i$ is the contribution of atom i to the SASA. The family of methods that combine the GB and SASA formalisms are referred to as GBSA models.

One of the fastest models that follows the GBSA formalism is FACTS (fast analytical continuum treatment of solvation);[19] a simulation with FACTS is only 4 times slower than the same simulation in vacuum.[19,24] This model estimates the effective Born radii and the atomic SASA based on geometric properties. Differing from most GBSA models,[13,14,16,18,22] FACTS does not assume the Coulomb field approximation for the calculation of ΔG^{el} in eq 2. In FACTS, all energy terms are calculated analytically, allowing for a fast calculation of the gradient of ΔG^{solv}. Thus, it can be efficiently applied to MD simulations. FACTS has been used for the study of different proteins, such as rhodopsin[25] or an ankrin repeat protein,[26] and to study protein−ligand docking[24] and protein folding.[27]

Initially, implicit solvent models could only represent the water environment. However, there has been an effort to extend these models to membranes.[28−34] The first extension of a GBSA model to membranes was done by Spassov et al.,[28] who considered the membrane as a planar dielectric slab with dielectric constant equal to ε_p. For the calculation of the effective Born radii, they used an asymptotic pairwise summation.[13,35] To approximate the results obtained by fdP for such two-dielectric model, they added a new empirical term. Following a similar two-dielectric model, Im et al.[14] extended the GBSW (generalized Born with a simple switching function) model[29] to membranes. This model was used, for example, for the study of the folding of the fd coat protein,[36] the interfacial folding and membrane insertion of WALP and TMX peptides,[37] and the membrane oligomerization of glycophorin A, the M2 proton channel and phospholamban.[38]

Similarly, Tanizaki and Feig[31] extended the GBMV (generalized Born using molecular volume) model[15] to membranes. In this new model, which they called HDGB (heterogeneous dielectric generalized Born), a variable solvent dielectric constant was considered. Such profile was motivated by comparison to explicit solvent simulations.[39] For the nonpolar contribution, a variable surface tension parameter was used to reproduce the results of the membrane insertion of O_2.[40] Feig et al.[41] showed that the effective Born radii also depend on the dielectric environment and such dependence was introduced in HDGB.[31] The HDGB model has been applied, for example, to study bacteriorhodopsin and the BtuCD protein,[42] phospholamban,[43,44] the ebola fusion peptide,[45] and the influenza fusion peptide.[46] Even though the application of the different GBSA implicit membrane models has been successful, it has been argued that new contributions should aim to obtain a more accurate picture of the membrane environment.[33,34] To this end, Parisio and Ferrarini[33] divided the nonpolar contribution into dispersion interaction and cavity formation terms and included a term to represent the anisotropic interaction of the protein with the acyl chains. Other groups have included a new term to represent the variation in membrane thickness.[34,47] In this way, the HDGB model has been extended into the DHDGB (dynamic HDGB) model.[34]

Another implicit membrane model worth mentioning is IMM1.[30] This model does not follow the GBSA formalism and is an extension of the EEF1 (effective energy function) model,[48] an experimentally parametrized Gaussian exclusion solvation model, to membranes. The extension was done using experimental data of transfer of amino acids from water to cyclohexane. This fast implicit membrane model has been applied, for example, to the study of peptide aggregation.[5,49]

Despite the fact that most implicit membrane models are quite approximate given the heterogeneity of a membrane environment,[33,34] they are commonly applied, in particular when effective conformational sampling is required.[43−45] Moreover, some implicit solvent models, such as GBSW, have shown instabilities when used in combination with energy minimization methods such as basin hopping.[50] These effects have not been observed with the FACTS implicit solvent model.[27] Therefore, the goal of this work is to extend the FACTS implicit solvent model to membranes. We call this extension FACTSMEM. To achieve this goal, the dielectric constant and surface tension are varied inside the membrane inspired by the HDGB model.[31] The effective Born radii at different dielectric environments are calculated, using an approximation of Kirkwood's equation, as a function of the Born radii calculated at $\varepsilon_s = 78.5$ with FACTS. This approximation is tested for the calculation of self-electrostatic solvation free energies, interaction electrostatic energies, and complete protein conformation solvation free energies. The free energy insertion profiles of amino acid analogues are calculated with FACTSMEM and compared with explicit solvent simulations. Furthermore, FACTSMEM is tested for MD simulations of various proteins: melittin, glycophorin A, WALP23 and KALP23, bacteriorhodopsin, and a Clc channel.

THEORY

FACTS Implicit Solvation Model. In this section, a brief introduction to the FACTS solvation model is given. A more thorough description can be found in the original FACTS paper.[19]

FACTS estimates the effective Born radii and the atomic SASA using geometric properties. It should be recalled that the effective Born radii measure the degree of burial of the atoms in the protein, and the atomic SASAs measure how accessible the atoms are to the solvent. Two geometric measures are used. The first one is a measure of the volume occupied by atoms surrounding a given atom

$$A_i = \sum_{j=1, j\neq i}^{N} V_j \Theta_{ij} \tag{4}$$

while the second one is a measure of symmetry of the surrounding atoms

$$B_i = \left| \frac{\sum_{j=1, j\neq i}^{N} \frac{V_j}{r_{ij}} \Theta_{ij} \hat{\mathbf{x}}_{ij}}{1 + \sum_{j=1, j\neq i}^{N} \frac{V_j}{r_{ij}} \Theta_{ij}} \right| \tag{5}$$

where V_j is the van der Waals (vdW) volume of atom j. The smoothing function Θ_{ij} is calculated as

$$\Theta_{ij} = \begin{cases} \left(1 - \left(\frac{r_{ij}}{R_i^{\text{sphere}}}\right)^2\right)^2 & r_{ij} \leq R_i^{\text{sphere}} \\ 0 & r_{ij} > R_i^{\text{sphere}} \end{cases} \tag{6}$$

where $\hat{\mathbf{x}}_{ij} = (\mathbf{x}_{ij}/r_{ij})$, $r_{ij} = |\mathbf{x}_{ij}|$, $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, and $\mathbf{x}_i$ are the coordinates of atom i. R_i^{sphere} is a cutoff defining the surroundings of atom i. A_i is the sum of the van der Waals

dx.doi.org/10.1021/ct500084y | J. Chem. Theory Comput. 2014, 10, 3163−3176

volumes of all surrounding atoms j weighted by Θ_{ij}. B_i is the normalized Euclidean norm of the sum of the unit vectors pointing from atom i to atoms j, weighted by Θ_{ij} and by the volumes V_j. For a perfectly symmetric distribution of atoms surrounding i, $B_i = 0$, and for a perfectly unsymmetric distribution, $B_i = 1$. The factor Θ_{ij} is used for both weighting and smoothing.

The two measures are combined into

$$C_i = A_i + b_1 B_i + b_2 A_i B_i \tag{7}$$

The self-electrostatic solvation free energy is then calculated as

$$\Delta G_i^{\mathrm{el}} = a_0 + \frac{a_1}{1 + e^{-a_2(C_i - a_3)}} \tag{8}$$

where two of the parameters can be calculated with limiting cases ($a_0 = -a_1 = -(\tau/2r_i^{\mathrm{vdW}})(1 + e^{-a_2 a_3})$), while the other two, a_2 and a_3, were optimized for each given van der Waals radius r_i^{vdW}. With eq 8 and the definition of the effective Born radius (eq 2), the electrostatic contribution to the free energy of solvation (eq 1) can be calculated. In the implementation of FACTS, a shifting truncation scheme[51] is used for the interaction term of the electrostatic contribution.

The atomic SASA is calculated using the same parameters A_i and B_i used for the electrostatic contribution. They are again combined using

$$D_i = A_i + d_1 B_i + d_2 A_i B_i \tag{9}$$

The atomic SASA is then calculated as

$$\mathrm{SASA}_i = c_0 + \frac{c_1}{1 + e^{-c_2(D_i - c_3)}} \tag{10}$$

where again two parameters are calculated with limiting cases ($c_0 = -c_1 = 4\pi(r_i^{\mathrm{vdW}} + 1.4)^2(1 + e^{-c_2 c_3})$), and the other two, c_2 and c_3, were optimized for each given r_i^{vdW}.

The parameters a_2, a_3, b_1, and b_2 were optimized by minimizing the difference between the calculated ΔG_i^{el} with FACTS and fdP.[19] The parameters c_2, c_3, d_1, and d_2 were optimized by minimizing the calculated SASA obtained with FACTS and the exact SASA.[19] The radii R_i^{sphere} were optimized for the calculation of the ΔG_i^{el} with a maximum value of 10 Å.[19] There is a set of parameters for each van der Waals radius r_i^{vdW}.

Extension of FACTS to Membranes. *Dielectric Constant and Nonpolar Profiles.* The extension of a GBSA implicit solvent model to membranes implies estimating the variation in dielectric constant, effective surface tension and effective Born radii. For the extension of FACTS, we decided to use the dielectric constant and effective surface tension profiles proposed by Tanizaki and Feig[31] and later refined by Sayadi et al.[43]

The dielectric constant profile was obtained by calculating ΔG^{el} with fdP in a three-dielectric model, allowing to determine an effective dielectric constant ε_s' for the membrane using Born's equation (eq 2).[31] This calculation was done with a probe of $R = 2$ Å, $q = 1$, and $\varepsilon_p = 1$. This three-dielectric model is symmetric, with z representing the normal of the membrane and $z = 0$ being the center of the membrane. In this model, $\varepsilon_s = 2$ for $|z|$ from 0 to 10 Å, $\varepsilon_s = 7$ from 10 to 15 Å, and $\varepsilon_s = 80$ from 15 Å onward. The model was proposed based on the comparison with a 20 ns explicit solvent simulation[39] and represents the width of a DPPC membrane. The dielectric constant profile was later refined[43] to reproduce the insertion profiles of amino acids analogues from explicit solvent simulations.[52] The effective dielectric constant (ε_s') can be used to calculate ΔG^{el} in a slightly modified generalized Born equation

$$\Delta G^{\mathrm{el}}(z) = -\sum_{i,j=1}^{N} \tau(z) \frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i(z_i) R_j(z_j) \exp(-r_{ij}^2/\kappa R_i(z_i) R_j(z_j))}} \tag{11}$$

where τ now depends on z

$$\tau(z) = \frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_{ij}(\varepsilon_s'(z_i), \varepsilon_s'(z_j))} \tag{12}$$

with $\varepsilon_{ij} = (\varepsilon_s'(z_i) + \varepsilon_s'(z_j))/2$.

The nonpolar contribution is also modified to represent the membrane. In this case, the surface tension coefficient of water γ is multiplied by a prefactor S, which depends on z. The profile was obtained by fitting to an explicit solvent simulation of the insertion of O_2 into a membrane, thereby assuming that the SASA is not modified by the membrane.[40] The profile was also later refined[43] to reproduce the amino acid analogues insertion profiles from explicit solvent simulations.[52] The solvation free energy can then be calculated as

$$\Delta G^{\mathrm{solv}}(z) = \Delta G^{\mathrm{el}}(z) + \gamma \sum_{i=1}^{N} S(z_i) \mathrm{SASA}_i \tag{13}$$

Effective Born Radius at Different ε_s. Feig et al. showed that assuming a constant effective Born radius, when varying the solvent dielectric constant, introduces an error in GB models.[41] For the extension of FACTS to membranes, we decided to estimate R_i at different ε_s as a function of R_i at $\varepsilon_s = 78.5$ (i.e., water) calculated with the original FACTS.

This estimation is done using a parameter-free approximation to Kirkwood's equation.[53] Kirkwoods's equation estimates ΔG^{el} for a protein under two approximations: (i) the protein is spherical and (ii) the protein can be represented as a finite number of point charges inside this sphere.[53,54] Kirkwood's equation has already inspired others in the calculation of the dependence of R_i with ε_s.[41,55−58] When considering a zero salt concentration system, Kirkwood's equation can be written as

$$\Delta G^{\mathrm{el}} = \sum_{i=1}^{N} \Delta G_i^{\mathrm{el}} + \sum_{i=1,j=1,i\neq j}^{N,N} \Delta G_{ij}^{\mathrm{el}} \tag{14}$$

with

$$\Delta G_i^{\mathrm{el}} = -\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_s}\right) \frac{q_i^2}{2A} \sum_{n=0}^{\infty} \frac{t_{ii}^{\,n}}{1 + \frac{n}{n+1}\beta} \tag{15}$$

and

$$\Delta G_{ij}^{\mathrm{el}} = -\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_s}\right) \frac{q_i q_j}{2A} \sum_{n=0}^{\infty} \frac{t_{ij}^{\,n} P_n(\cos\theta)}{1 + \frac{n}{n+1}\beta} \tag{16}$$

where $\beta = \varepsilon_p/\varepsilon_s$, A is the protein radius, $t_{ij} = g_i g_j / A^2$, $g_i = |\mathbf{g}_i|$ is the position of atom i relative to the center of the protein, θ is the angle between $\mathbf{g}_i$ and $\mathbf{g}_j$, and P_n are the ordinary Legendre polynomials. t_{ii} is a measure of the distance of atom i from the center of the protein; if $t_{ii} = 0$ the atom is in the center, and if $t_{ii} = 1$ the atom is at the surface.

Equation 14 has been conveniently separated in self (eq 15) and cross terms (eq 16) in analogy to what can be done with the generalized Born equation (eq 1). In our approach, we

calculate the self-terms in the generalized Born formalism as a function of the self-terms in Kirkwood's equation. In the Results and Discussion section, we show that, under this assumption, the cross terms of the generalized Born equation are calculated correctly. If the self-terms of Kirkwood's equation (eq 15) are set equal to the self-solvation free energies (eq 2), the effective Born radii can be calculated as

$$R_i = \frac{1}{\frac{1}{A}\sum_{n=0}^{\infty}\frac{t_{ii}^{\,n}}{1+\frac{n}{n+1}\beta}} \tag{17}$$

Our intention is to calculate R_i (effective Born radius at an arbitrary ε_s) as a function of $R_{i,w}$ (effective Born radius at ε_s = 78.5). It should be noted that A, t_{ii}, and q_i do not depend on ε_s. Thus, the following equation can be written

$$R_i = R_{i,w}\frac{\sum_{n=0}^{\infty}\frac{t_{ii}^{\,n}}{1+\frac{n}{n+1}\beta_w}}{\sum_{n=0}^{\infty}\frac{t_{ii}^{\,n}}{1+\frac{n}{n+1}\beta}} \tag{18}$$

with $\beta_w = \varepsilon_p/78.5$.

The use of eq 18 in an MD simulation is deemed impossible, as it requires the calculation of many terms in the sum to reach convergence (for example, for t_{ii} = 0.9 around 100 terms are needed). Moreover, the calculation of A and t_{ii} may be difficult, specially for simulations of processes such as peptide aggregation, in which the size of the protein is not constant. It would also create a higher computational cost because the aggregation state would need to be estimated at each time step for a correct estimation of A. For these reasons, we decided to use an approximation to eq 18, that does not depend on A or t_{ii}. The equation we propose is

$$R_i = R_{i,w}\frac{\left(1+\frac{\beta}{2}\right)}{\left(1+\frac{\beta_w}{2}\right)} \tag{19}$$

By comparison of eq 18 with the approximate eq 19, the following relationship is obtained

$$\frac{1+\frac{\beta_w}{2}+t_{ii}+\left(1+\frac{\beta_w}{2}\right)\sum_{n=2}^{\infty}\frac{t_{ii}^{\,n}}{1+\frac{n}{n+1}\beta_w}}{1+\frac{\beta}{2}+t_{ii}+\left(1+\frac{\beta}{2}\right)\sum_{n=2}^{\infty}\frac{t_{ii}^{\,n}}{1+\frac{n}{n+1}\beta}} \approx 1 \tag{20}$$

where the left side is approximated to 1 when using eq 19. To test how good our approximation is, we calculated the left side of eq 20 for different t_{ii} until convergence (see Table 1). Around 100 terms in the sum are needed to reach convergence. Table 1 shows that the error increases the farther ε_s is from the reference value ε_s = 78.5. The largest error is obtained for ε_s = 2 and t_{ii} = 0.3, which is 0.164. In the Results and Discussions section, it will be shown that this error is smaller than the one introduced in the estimation of ΔG_i^{el} by FACTS. The strength of eq 19 comes from the fact that Kirkwood's equation is a power series in t_{ii} and eq 19 eliminates the error introduced by the first power of t_{ii}. Considering that all t_{ii} must be smaller than 1, as all protein atoms are inside the protein, this minimizes the error in eq 20. Moreover, eq 19 is parameter free, which means that it should have a high transferability, as it does not depend on any training set.

Table 1. Calculation of the Left Side of Eq 20 for Different ε_s and t_{ii}[a]

ε_s	t_{ii} = 0.3	t_{ii} = 0.6	t_{ii} = 0.9
40	1.004	1.001	0.997
20	1.012	1.004	0.991
10	1.029	1.010	0.981
8	1.037	1.013	0.976
4	1.079	1.030	0.955
2	1.164	1.066	0.926

[a]All these values are approximated as 1 when using eq 19

It is worth mentioning that we use eq 19 to extend FACTS to membranes because FACTS is one of the fastest GBSA implicit solvent models. However, eq 19 is independent from FACTS and could thus be used to extend other implicit solvent models to membranes. A similar, but highly parametrized, equation was used by Tjong and Zhou to calculate the dielectric constant dependence of solvation free energies of complete proteins.[59] Their equation was later used in the development of a new implicit solvent model.[60] Even though both equations are similar, it should be noted that their equation estimates the solvation free energies of entire proteins as a function of the charge and number of atoms in the protein. On the other hand, our equation is used for the estimation of self-solvation free energies or effective Born radii. We tested the equation from Tjong and Zhou for the estimation of self-solvation free energies using average parameters and found that better results are obtained using eq 19, particularly for small ε_s.

The calculation of R_i in different dielectric environments is the last ingredient necessary for FACTSMEM. In summary, FACTSMEM will first calculate $R_{i,w}$ and $SASA_i$ with the original FACTS for the water environment (eqs 8, 10, and 2). Then, the effective dielectric constant ε_s' and the normalized effective surface tension $S_i\gamma$ are calculated for each atom as a function of their position z_i. The effective Born radii R_i at ε_s' are then calculated with eq 19. Finally, the solvation free energy is calculated with eqs 11 and 13. All equations and their derivatives are analytical, which means that FACTSMEM can be efficiently applied to MD simulations.

COMPUTATIONAL METHODOLOGY

FACTSMEM has been implemented as a new module in the simulation program CHARMM,[51,61] version c35b4. The effective dielectric constant and normalized effective surface tension profiles from Sayadi et al.[43] were made default. These profiles represent the width of a DPPC membrane. The effective dielectric constant profile was adapted so that ε_s' = 78.5 for water, as this is the value used in FACTS. z represents the normal to the membrane with z = 0 being the center. All calculations were done with κ = 4 and γ = 0.015 kcal/(mol Å^2).[19,31] The dielectric constant of the protein ε_p was set to 1.[19]

Energy Calculations. To test our new approximation to Kirkwood's equation (eq 19) different energy calculations were performed at ε_s equal to 78.5, 40, 20, 10, 8, 4, and 2. Results from FACTSMEM were compared to the results from fdP with a grid space of 0.2 Å. All fdP calculations were accomplished using the PBEQ module[11] in CHARMM, with help of the MMTSB script.[62]

Self-electrostatic solvation free energies were calculated for the atoms of a group of 61 different protein conformations. These included the folded states of single-chain proteins (PDB codes 1cb3, 1l2y, 1a2p, 1abz, 1dvd, 1bpi, 1crn, 1cus, 1enh, 1f8a, 1fmk, 1inc, 1pgb, 1lz1, 1hdn, 1hel, 1pht, 1shg, 1ubq, 2a3d, 2ci2, 2ptl, 3app, and 1vii, and a β-sheet from 1pgb, an α-helix from

1pgb, Beta3s, a 5 ankyrin repeat domain, and the human prion protein) and multichain proteins (PDB codes 1kvd, 1ycq, 1ycr, all chains of 2ins, 5hvp, and chains A and B of 2ins). From now on the β-sheet from 1pgb is called bet1. All but 1inc, 1kvd, 1vii, 1ycq, 1ycr, 3app, 5hvp, and both versions of 2ins were also included in an unfolded state obtained from 50 ns MD simulations starting from the folded conformation at 450 K using a simple implicit solvent model.[63] The fdP calculations of ΔG_i^{el} were done by considering all other charges equal to zero and $q_i = 1$, and subtracting the energies in vacuum ($\varepsilon_s = \varepsilon_p = 1$) from the energies in solution.

Solution pairwise interaction electrostatic energies (from now on called interaction energies) were calculated for atoms in the peptides bet1 and 1cb3. The fdP calculations of $G_{i,j}^{el}$ were conducted by calculating the energies in solution considering all other charges equal to zero and $q_i = q_j = 1$, and then subtracting the self-energies in solution for i and j.

Solvation electrostatic free energies of 2800 different protein conformations were calculated, comprising 100 different conformations of 28 different proteins (all the ones named before but 1vii, 1ycq, 1ycr, 3app, 5hvp, and both versions of 2ins). The 100 different conformations are snapshots of the same high temperature 50 ns unfolding MD simulations mentioned before. In this case, charges from the CHARMM22 force field[64] were used.

Finally, the electrostatic solvation free energy were calculated for 31 different orientations and positions of WALP23 using FACTSMEM and HDGB.[31] The conformations were obtained using a perfect helix as a starting point and rotating and translating it to different regions of the membrane. The charges from the CHARMM22 force field were used. For HDGB, the dielectric constant profile from Sayadi et al.[43] was used. As a benchmark, the electrostatic solvation free energy for the same conformations were calculated in a fdP membrane. The dielectric constant profile for the fdP calculation was defined as $\varepsilon_s = 1.8$ for $0 < z < 12$ Å, $\varepsilon_s = 7$ for 12 Å $< z <$ 17 Å, and $\varepsilon_s =$ 78.5 for $z >$ 17 Å.

Both interaction energies and protein conformation solvation electrostatic free energies were calculated without cutoffs.

Molecular Dynamics Simulations. The free energy insertion profiles of the amino acid analogues were calculated with FACTSMEM. Moreover, MD simulations of melittin, WALP23 and KALP23, glycophorin A, bacteriorhodopsin, and a Clc channel were performed. In all cases, the CHARMM22 force field[64] with the CMAP correction for GBSW[65−67] was used. All bonds involving hydrogen atoms were fixed with the SHAKE algorithm.[68] The integration step was set to 2 fs, nonbonded lists were updated every 20 fs and coordinates saved every 0.2 ps. A cutoff of 12 Å was used for nonbonded interactions with a shifting function for electrostatic interactions and a switching function for van der Waals interactions. The sigmoidal switching function started at 10 Å and nonbonded lists were created for atoms closer than 14 Å. For production runs, Langevin dynamics with a friction coefficient of 10 ps^{-1} was used.[69] The temperature was set to 300 K.

First, we calculated the insertion profiles of the different amino acid analogues. Amino acid analogues were built by eliminating the bond between the C_α and C_β atoms, deleting the backbone and adding a H atom to fill the valence of the C_β atom. All amino acids but histidine, proline, and glycine were studied. Charged amino acids were studied in their neutral state. In the case of arginine, the neutral state was simulated by deleting the hydrogen bonded to the N_ε atom[70] and using the force field parameters developed by Li et al.[71] All analogues were subjected to a 500-step steepest descent minimization, a 500-step conjugate gradient minimization and a 200 ps heating MD simulation with FACTS. Production runs were started with the analogues positioned from the center of the membrane ($z =$ 0 Å) up to the water environment ($z =$ 30 Å) with a 1 Å stride, for a total of 31 runs per amino acid. The center of mass was harmonically restrained with a force constant of 50 kcal/(mol $Å^2$). All simulations were run for 1 ns and the last 0.5 ns were used for analysis.

Next, MD simulations of different proteins were performed to test FACTSMEM. In all cases, hydrogen atoms were built with the HBUILD module in CHARMM.[51,61] All systems were subjected to a 500-step steepest descent and a 500-step conjugate gradient minimization, followed by a 200 ps heating MD simulation. The production runs were performed for 20 ns. Standard protonation states, charged N- and C-termini, and an initial position of the center of mass equal to $z = 0$ Å were used except otherwise noted. The melittin simulation was started from PDB file 2mlt.[72] It was modeled with a charged C-terminus and an uncharged N-terminus. To start the simulation, the peptide was placed at $z =$ 17 Å in a conformation parallel to the membrane surface. WALP23 and KALP23 simulations were started from a perfect α-helical structure as defined by Swiss-Pdb Viewer.[73] The main axis was set parallel to the normal of the membrane. The N-terminus was acetylated and the C-terminus was amidated. The simulation of the glycophorin A dimer was started from PDB file 1afo (model 4).[74] The main axis was set parallel to the normal of the membrane. The simulation of a bacteriorhodopsin monomer was started from PDB file 1qhj.[75] All residues have standard protonation states but Asp96, Asp115, and Glu204 which were protonated.[42,76] The charges[77,78] and force field parameters[79−83] used to model the retinal linked to Lys216 are summarized by Saam et al.[84] Finally, the simulation of the Clc channel dimer was started from the PDB file 1kpk.[85]

CPU Time Measurements. The CPU time needed by FACTSMEM, FACTS and other implicit membrane models (IMM1,[30] GBSAIM,[28] GBSW,[14] and HDGB[31]) for a 1 ns simulation of melittin was measured. The system was set up as the previous melittin simulation. The CHARMM19 force field[86] was used with IMM1 and GBSAIM and the CHARMM22 force field[64] was used with FACTS, IMM1, GBSW, and HDGB. Standard nonbonded options for each model were used.[87] For membrane models which require the selection of the membrane hydrophobic width, a value of 27 Å was chosen. For the IMM1 simulations, the special IMM1 topology and parameter files were used. For GBSAIM, the empirical surface tension parameter was set to 0.025 kcal/(mol $Å^2$). For the GBSW membrane simulation, the half width of the membrane switching function was set to 2.5 Å and the empirical surface tension parameter was set to 0.030 kcal/(mol $Å^2$). For the HDGB simulation, the original nonpolar and dielectric profiles were used and the empirical surface tension parameter was set to 0.015 kcal/(mol $Å^2$). All timing measurements were done on a single 2.93 GHz Intel Xeon X5570 processor.

RESULTS AND DISCUSSION

Self-Electrostatic Solvation Free Energies. First, we tested how well eq 19 estimates effective Born radii in different dielectric environments. For this purpose, the self-electrostatic solvation free energies for the atoms of 61 different protein

conformations at ε_s equal to 78.5, 40, 20, 10, 8, 4, and 2 were calculated with FACTSMEM and compared to the results from fdP calculations. Plots of the results for ε_s = 78.5 and 4 can be seen in Figure 1. It should be noted that FACTSMEM at ε_s =

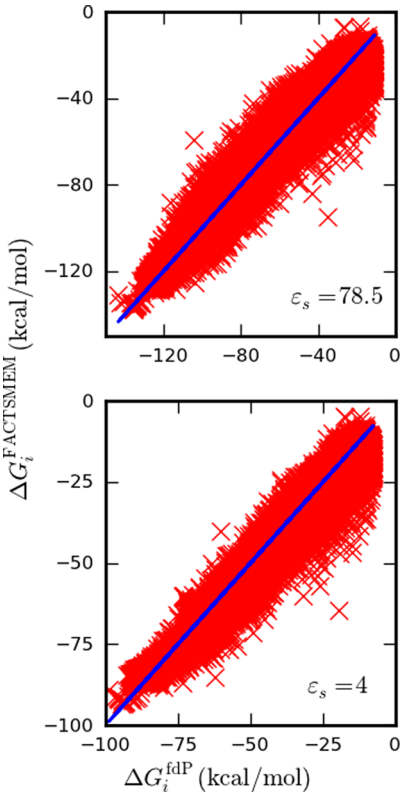


Figure 1. Calculation of ΔG_i^{el} with FACTSMEM and fdp for the atoms of 61 different protein conformations for ε_s = 78.5 and 4. For ε_s = 78.5, FACTSMEM is equivalent to FACTS. In total, there are 78 013 atoms.

78.5 is equivalent to FACTS. The error in the estimation of ΔG_i^{el} comes from two different sources: the error introduced by FACTS for the calculation at ε_s = 78.5 and the error introduced by eq 19 for the estimation of R_i at $\varepsilon_s \neq 78.5$. In Figure 1, it can be seen that the results for both dielectric environments look very similar, which shows that most of the error is introduced by FACTS and the error produced by its extension to different ε_s is negligible. The relative error in the calculation of ΔG_i^{el} for different ε_s can be seen in Table 2. There is a slight increase in the relative error from 9.8% for ε_s = 78.5% to 11.4% for ε_s = 2. This increase, which is expected from Table 1, stays in a reasonable range.

Interaction Energies. The main assumption of the GB models is that an accurate estimation of the N ΔG_i^{el} terms can be used to calculate accurate values of the $\sim N^2$ screening terms,[55] where N is the number of atoms. Moreover, it has been argued by Scarsi and Caflisch[88] that interaction energies in solution are better than solvation energies to validate continuum electrostatic models. Therefore, we calculated interaction energies in solution,

$$G_{ij}^{el} = \frac{q_i q_j}{\varepsilon_p r_{ij}} - \frac{\tau}{2}\frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/\kappa R_i R_j)}} \quad (21)$$

for all atom interactions of the peptides 1cb3 and bet1. Calculations with FACTSMEM and fdP were done for ε_s equal

Table 2. Relative and Absolute Errors (Averages and Standard Deviations (SD)) in the Calculation of the Self Electrostatic Solvation Free Energy ΔG_i^{el} with FACTSMEM with Respect to fdP[a]

ε_s	average (%)	SD (%)	average (kcal/mol)	SD (kcal/mol)
78.5	9.8	16.1	3.5	3.6
40	9.8	16.3	3.4	3.5
20	9.9	16.5	3.2	3.3
10	10.1	16.8	3.0	3.0
8	10.2	16.9	2.9	2.9
4	10.8	16.9	2.6	2.5
2	11.4	15.9	1.8	1.6

[a]Calculation done for the atoms of 61 different protein conformations. For ε_s = 78.5, FACTSMEM is equivalent to FACTS. In total there are 78 013 atoms.

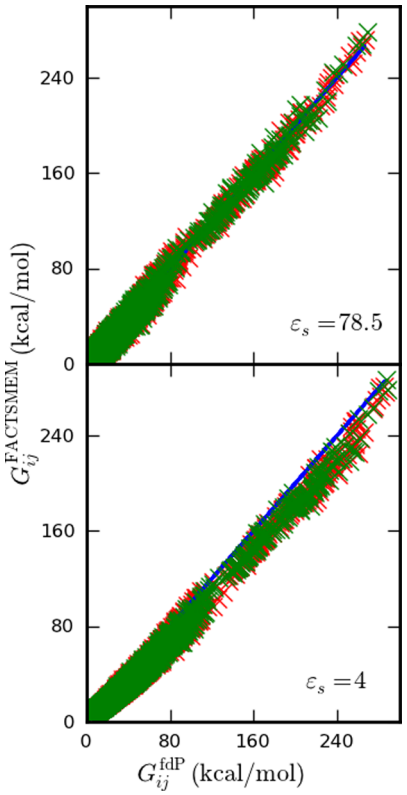


Figure 2. Interaction energies G_{ij}^{el} for all atom interactions in peptides 1cb3 (green) and bet1 (red) for ε_s = 78.5 and 4 calculated with FACTSMEM and fdP. For ε_s = 78.5, FACTSMEM is equivalent to FACTS. In total, there are 42 405 interactions.

to 78.5, 40, 20, 10, 8, 4, and 2. Plots for the results for ε_s = 78.5 and 4 can be seen in Figure 2 and absolute errors for all ε_s are given in Table 3. In this case, absolute errors are calculated as many of the interaction energies G_{ij}^{el} are close to 0, and thus, relative errors are misleading. In Figure 2, it can be observed that interaction energies are slightly underestimated with FACTSMEM for very small ε_s. This result correlates well with the data from Table 3, where it can be seen that the average error increases from 0.86 kcal/mol for ε_s = 78.5 to 2.10 kcal/mol for ε_s = 2. However, the error is still reasonable and does not translate to the results of electrostatic solvation free energies of complete proteins as will be seen in the next section.

Protein Conformation Free Energies. Even though self- and interaction energies are estimated with acceptable accuracy,

dx.doi.org/10.1021/ct500084y | J. Chem. Theory Comput. 2014, 10, 3163−3176

Table 3. Absolute Error in kcal/mol in the Calculation of the Interaction Energies G_{ij}^{el} with FACTSMEM Compared to fdP[a]

ε_s	average (kcal/mol)	SD (kcal/mol)
78.5	0.86	1.56
40	0.91	1.64
20	1.03	1.83
10	1.32	2.26
8	1.46	2.47
4	1.97	3.25
2	2.10	3.49

[a]Calculation done for all atom interactions in peptides bet1 and 1cb3. For ε_s = 78.5, FACTSMEM is equivalent to FACTS. In total, there are 42 405 interactions.

errors could build up and produce an incorrect value of the electrostatic solvation free energy of a protein. Thus, the electrostatic solvation free energies of 2800 different protein conformations were calculated with FACTSMEM and compared to fdP. The results for ε_s = 78.5 and 4 are plotted in Figure 3A. Again, the results are very similar for both ε_s which shows that FACTSMEM does not increase the error introduced by FACTS. The relative errors for ΔG^{el} can be found in Table 4, where it can be seen that the error increases from 3.4% for ε_s = 78.5 to 7.5% for ε_s = 2. It should be noted that the absolute values of ΔG^{el} decreases with decreasing ε_s. Thus, even though the relative error increases when ε_s decreases, the absolute error decreases (data not shown).

Probably, more relevant than the correct calculation of ΔG^{el}, it is to test if FACTSMEM can discriminate correctly between different protein conformations and dielectric environments. Therefore, we calculated the difference in ΔG^{el} for the same protein conformation in $\varepsilon_s \neq 78.5$ and ε_s = 78.5. The results of the differences between ε_s = 40 and 4 with respect to the reference of ε_s = 78.5 are plotted in Figure 3B. The absolute errors for the differences in ΔG^{el} are given in Table 4. In this case, absolute errors are calculated as many of the $\Delta\Delta G^{el}$ are close to 0 and relative errors are thus misleading. The average error in $\Delta\Delta G^{el}$ increases from 2.7 kcal/mol for ε_s = 40 to 17.1 kcal/mol for ε_s = 2. However, it should be noted, that $\Delta\Delta G^{el}$ for the same conformation in different dielectric environment increases drastically when decreasing ε_s. Thus, even though the absolute error in Table 4 increases, the results for small ε_s are better than for large ε_s as it can be observed in Figure 3B. The largest deviation is observed in the calculation of $\Delta\Delta G^{el}$ between ε_s = 40 and ε_s = 78.5. Even though there is an undercalculation of $\Delta\Delta G^{el}$ compared to fdP calculations, both FACTSMEM and fdP always predict the same sign for $\Delta\Delta G^{el}$ for a given conformation in different dielectric environments. Considering the small values of $\Delta\Delta G^{el}$ for ε_s = 40 compared to ΔG^{el}, we consider these results acceptable.

Moreover, we also calculated the difference in ΔG^{el} between different conformations of the same protein in the same dielectric environment. The results for ε_s = 78.5 and 4 are plotted in Figure 3C and the absolute errors are summarized in Table 4. The differences in ΔG^{el} between different conformations in the same dielectric environment are estimated accurately for all ε_s, with the average error decreasing from 17.3 kcal/mol for ε_s = 78.5 to 14.3 kcal/mol for ε_s = 2.

The results are encouraging as, once again, the loss of accuracy between FACTS at ε_s = 78.5 and FACTSMEM at $\varepsilon_s \neq$ 78.5 is small. Notably, the slight underestimation of interaction energies (Figure 2) does not have a great influence in the calculation of complete protein electrostatic solvation free energies (Figure 3A). These results show that FACTSMEM can accurately calculate free energies of solvation and correctly

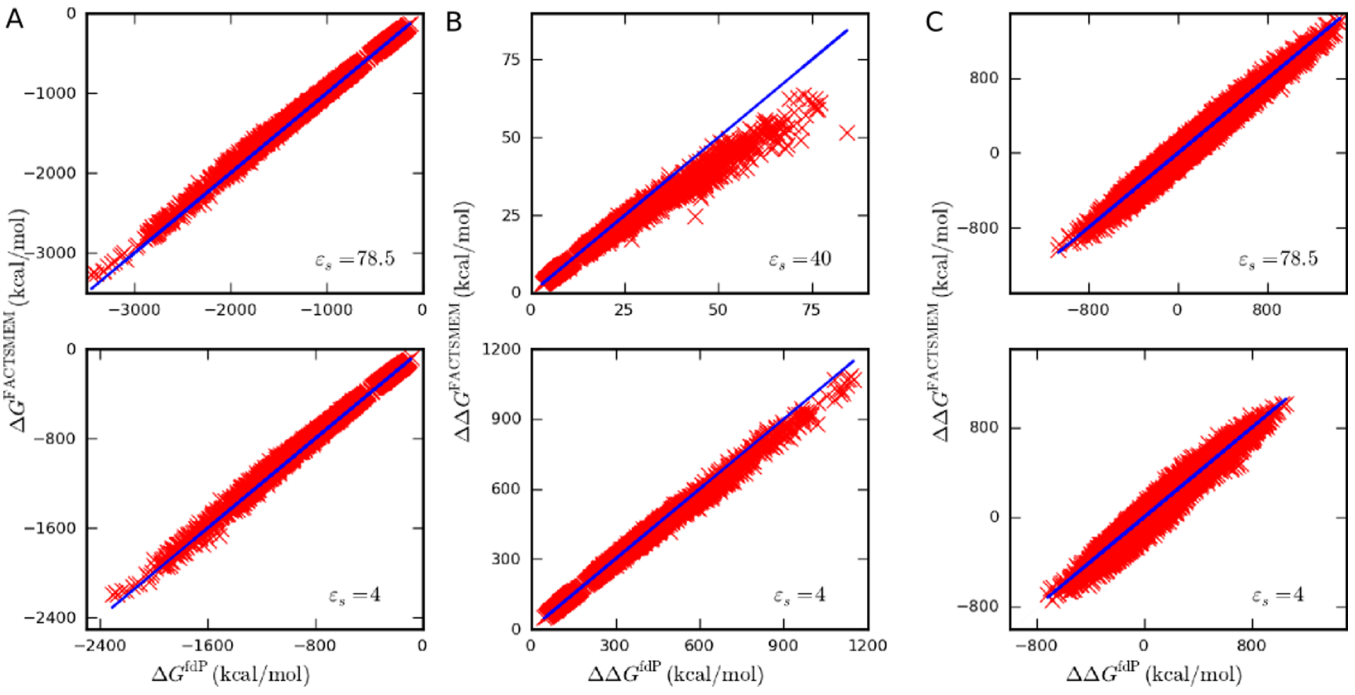


Figure 3. Comparison of electrostatic solvation free energies ΔG^{el} for 2800 different protein conformations calculated with FACTSMEM and fdP. (A) Protein conformation electrostatic solvation free energies, (B) difference in electrostatic solvation free energies between the same conformation at different ε_s ($\varepsilon_s \neq 78.5$ relative to ε_s = 78.5), and (C) difference in electrostatic solvation free energies between different conformations at the same ε_s. For ε_s = 78.5 FACTSMEM is equivalent to FACTS.

dx.doi.org/10.1021/ct500084y | J. Chem. Theory Comput. 2014, 10, 3163−3176

Table 4. Errors in the Calculation of ΔG^{el} and $\Delta\Delta G^{el}$ for Protein Conformations Using FACTSMEM with Respect to fdP[a]

	ΔG^{el}		$\Delta\Delta G^{el}$ for different ε_s[b]		$\Delta\Delta G^{el}$ for different conformations[c]	
ε_s	average (%)	SD (%)	average (kcal/mol)	SD (kcal/mol)	average (kcal/mol)	SD (kcal/mol)
78.5	3.4	1.9			17.3	18.3
40	3.3	1.9	2.7	3.3	17.2	18.4
20	3.2	1.9	6.6	7.9	17.2	18.7
10	3.3	2.0	11.0	13.0	17.4	19.2
8	3.5	2.0	12.3	14.5	17.5	19.4
4	4.6	2.5	15.3	17.0	17.4	19.3
2	7.5	3.7	17.1	18.1	14.3	15.4

[a]2800 protein conformations were considered. For ε_s = 78.5, FACTSMEM is equivalent to FACTS. [b]Error in $\Delta\Delta G^{el}$ between the same conformation at different ε_s (with ε_s = 78.5 as reference). [c]Error in $\Delta\Delta G^{el}$ between different conformations and the same ε_s.

discriminate between different dielectric environments and proteins conformations.

Amino Acid Analogues Insertion Profiles. The free energy insertion profiles ΔG for the different amino acids analogues are calculated with FACTSMEM and the results are plotted in Figure 4. The results are compared with the profiles

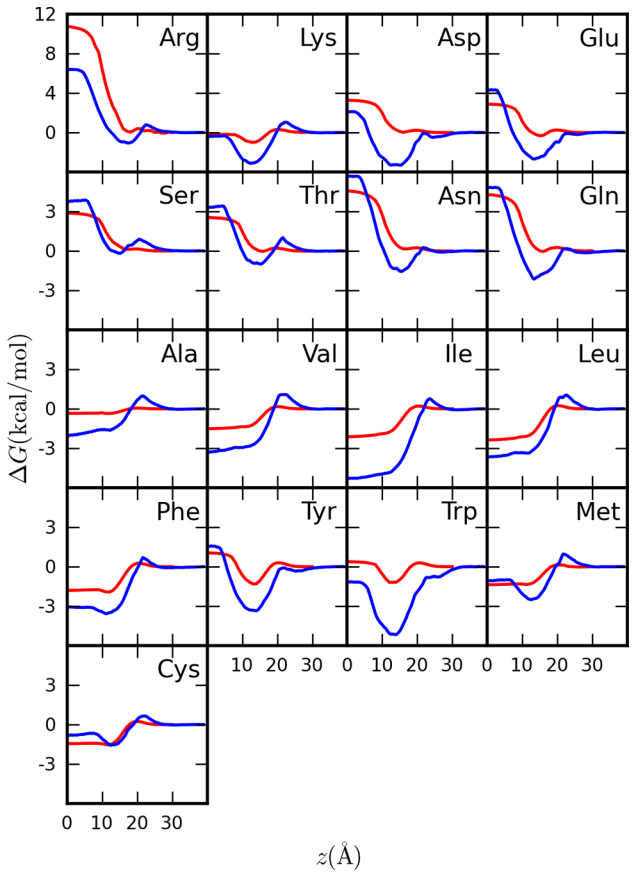


Figure 4. Free energy insertion profiles of the different amino acids calculated with FACTSMEM (red) and with umbrella sampling in an explicitly represented membrane (blue).[52]

obtained from a potential of mean force (PMF) calculation using umbrella sampling MD in an explicitly represented DOPC membrane.[52] It should be noted that the explicit solvent calculation was done using the OPLS force field,[89] which uses different partial charges compared to the CHARMM22 force field.[64] For ionizable side chains, only the uncharged residues were studied. The reason for this is that a dielectric slab of lower ε_s is not good enough for the study of the insertion of charged amino acids into the membrane as discussed by Panahi and Feig.[34] When these amino acids are inserted into the membrane they carry water molecules with them to lower their insertion free energy and this is not represented in FACTSMEM. Other implicit membrane models suffer from the same limitation but nonetheless have success in studying membrane proteins.[5,43−45]

As expected, there are certain differences between explicit and implicit solvent simulations. However, all main trends from the explicit solvent simulations are reproduced with FACTSMEM. For example, all hydrophobic residues are most stable in the center of the membrane and all hydrophilic residues are most stable outside the membrane. Moreover, amphiphilic residues, which are known to anchor at the membrane interface such as tryptophane and lysine,[90] have their ΔG minima at the same position in both implicit and explicit solvent.

When compared to the amino acid insertion profiles obtained with other implicit membrane models, such as HDGB,[31] the largest errors with FACTSMEM are obtained for small hydrophobic residues such as alanine, for which HDGB obtains results much closer to the ones from explicit solvent simulations (compare Figure 11 from ref 31 with Figure 4). This error is probably caused by the problems that FACTS has with calculating the SASA contribution for small molecules (see Figure 5 from ref 19). This error is not relevant for hydrophilic residues, which have a much larger electrostatic component to ΔG. Moreover, it should not be a problem for the simulation of complete proteins, as the SASA is better calculated by FACTS for large molecules.

Melittin. Melittin, the main component of the venom of the European honey bee, is a peptide consisting of 26 amino acids. It has an amphiphilic nature, which allows it to disrupt membranes.[91] It is one of the most studied membrane proteins and has become a standard test for implicit membrane models.[14,30,31] The peptide folds into two α-helices with a small-angled kink at residue Pro14 between the helices.[72,92,93] Melittin can be found either in a transbilayer orientation or parallel to the membrane close to its interface, depending on experimental conditions.[91] We performed a 20 ns Langevin dynamics simulation of melittin starting from the orientation parallel to the membrane with its geometric center at 17 Å from the center of the membrane.

Melittin has been studied experimentally using polarized attenuated total reflection infrared spectroscopy.[94] Using this method, it was measured that the peptide is located on average 17.5 Å away from the center of a DOPC membrane. Moreover, melittin has also been studied theoretically with a short 600 ps explicit solvent simulation in a DOPC membrane.[93] A distance from the center of the membrane between 11 and 13 Å was

obtained. Considering that FACTSMEM represents a DPPC membrane, the aforementioned values must be corrected to compare them faithfully to the results from FACTSMEM. Correcting for the width of the hydrophobic section,[95] we obtain an experimental result of 18.4 Å and an explicit solvent result between 12 and 14 Å. Melittin was also studied using the HDGB implicit membrane model, which also represents a DPPC membrane. In this case, the average distance was equal to 19.1 Å.[31]

The distance from the center of the membrane in the simulation with FACTSMEM is plotted in Figure 5. The

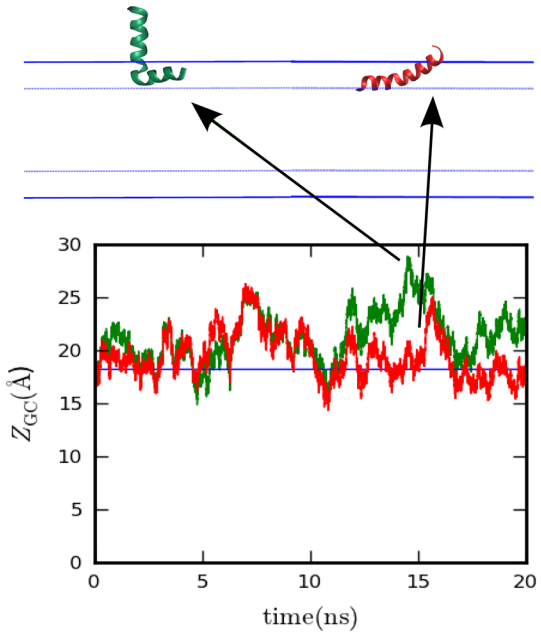


Figure 5. Distance from the center of the membrane of the geometric center z_{GC} of melittin for the FACTSMEM simulation started from z_{GC} = 17.0 Å (red), the FACTSMEM simulation started from z_{GC} = 15.5 Å (green) and for the experimental structure adapted to the DPPC width (blue).[94] On the top, two structures inside the implicit membrane are shown. The membrane is represented by two pairs of lines: solid lines at ±23.3 Å representing the membrane-water interface and dashed lines at ±14.2 Å representing the beginning of the head groups. These lines correspond to the width of a DPPC membrane.[95] The conformation with the pronounced kink (left) and a more typical structure with both helices in the membrane interface (right) are shown.

average value is 21.3 Å, which is larger than the results from experiments and other simulations. The maximum discrepancy between FACTSMEM and the experiment is found around 14 ns. By visually inspecting the simulation with VMD,[96] it can be observed that at this point the peptide has a much more pronounced kink between the two α-helices than what is found experimentally[72,92] or during the rest of this simulation. However, a similar conformation, which was named U-shaped, was found for coarse-grained MD simulations of melittin.[97] The conformation is different though, because the α-helix perpendicular to the membrane is pointing inward in the coarse-grained simulation and outward in our case. Considering that the explicit solvent simulation is only 600 ps[93] and that the simulation with HDGB is 10 ns,[31] this may indicate that the simulation with FACTSMEM is exploring a larger section of conformational space because of the coarse-graining of the membrane system and the longer time scale. A second simulation of melittin with FACTSMEM was started with its geometrical center initially positioned at 15.5 Å. The distance from the center of the membrane is also plotted in Figure 5. In this case, the large deviation around 14 ns is not observed and the average distance is equal to 19.5 Å, which is much closer to the experimental and HDGB results. The helices in melittin are stable with the helicity being 92% at the beginning and 89% at the end of the simulation.

WALP23 and KALP23. WALPs and KALPs are synthetic peptides used to study how α-helical proteins adapt to membranes. These peptides are composed of alanine and leucine residues in the interior and tryptophane (for WALPs) or lysine (KALPs) at the termini. For example, the amino acid sequence of WALP23 is Ac-GWWLALALALALALALALAL-WWA-amide. The length of the peptide can be changed (by adding or deleting LA units) to create different mismatch situations between the hydrophobic length of the peptide and the hydrophobic core width of the membrane. In the case of peptides with positive mismatch, it has been shown that WALPs and KALPs adapt to this mismatch by tilting.[98−100] Changing tryptophane for lysine at the termini can have an effect on the tilting, as these two amino acids anchor in different interface positions.[90] For a review on WALPs and KALPs, the paper from Holt and Killian is recommended.[98]

WALP23 has been studied using ^{2}H solid-state (SS) NMR and analyzed according to the geometric analysis of labeled alanine (GALA).[99] For a DMPC membrane a tilt angle of 5.5° was obtained. This result was later challenged by explicit solvent MD simulations, in which individual runs gave tilt angles over 30°.[101,102] Moreover, when these MD runs are combined and interpreted according to the GALA method, results much closer to the experimental ones are obtained, which lead to the conclusion that care should be taken with results obtained with the GALA method. Later, WALP23 was studied in a DOPC membrane with fluorescence spectroscopy giving a tilt angle of 24.8°.[100] Finally, a tilt angle of 21° was obtained for WALP23 in a DMPC membrane using anisotropic constraints with NMR.[98] WALP23 was also studied in a POPC membrane using an explicit solvent umbrella sampling MD simulation. The PMF calculation leads to a minimum of free energy at an angle of 14.9°.[103] The same system was studied with the MARTINI coarse-grained force field and the DHDGB implicit membrane model with results of 12.3 ± 6.5°[104] and 17.6 ± 1.2°,[34] respectively.

The result from a 20 ns MD simulation with FACTSMEM is 14.1 ± 6.4°, which is very close to the 14.9° obtained from the PMF explicit solvent simulation. The result is also qualitatively in agreement with the latest experimental results, specially considering that DPPC (the membrane represented by FACTSMEM) has a thicker hydrophobic core than DOPC or DMPC. The result from FACTSMEM agrees with experiments much better than that obtained with the GBSW implicit membrane, which gives a tilt angle of 32.7 ± 8.3 for DMPC.[37] It should be noted though, that individual runs from explicit solvent simulations.[101,102] give larger tilt angles than FACTSMEM or the PMF explicit solvent simulation. The reason for this is that the tilt angle is a highly dynamical feature, as it does not equilibrate to a stable value but varies constantly. This feature is seen both with FACTSMEM (see Figure 6) and in explicit solvent simulations (see Figure 1 from ref 101 and Figure 2 from ref 102). Thus, very long explicit solvent simulations are needed to obtain relevant statistics. However, because of the instantaneous relaxation and limited friction of

dx.doi.org/10.1021/ct500084y | J. Chem. Theory Comput. 2014, 10, 3163−3176

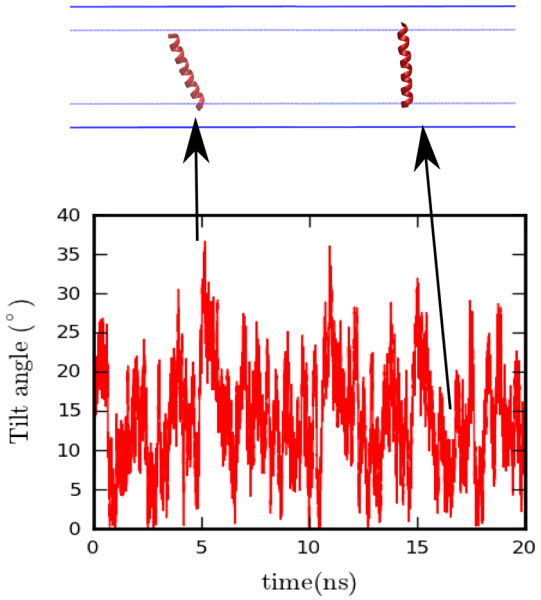


Figure 6. Instantaneous tilt angle for a WALP23 simulation with FACTSMEM. On the top, two typical structures from the simulation are shown. The implicit membrane is represented as in Figure 5. The structure on the left is more tilted, while the structure on the right shows an almost vertical WALP23 orientation.

the implicit membrane, relevant statistics can be obtained from shorter implicit membrane simulations. Using FACTSMEM, WALP23 is stable inside the membrane and the difference in ΔG^{solv} between WALP23 inside and outside the membrane is equal to −13.0 kcal/mol.

As a further test, we calculated the solvation free energy of WALP23 in 31 different orientations and positions using FACTSMEM and the HDGB implicit membrane model.[31] This is a good test to check if FACTSMEM obtains results comparable to those from one of the most accurate yet computationally more intensive implicit membrane models.[105] The benchmark to compare both models is a fdP membrane as explained in the Computational Methodology section. Figure 7 shows the results for both models. The results are similar, with FACTSMEM having an average error of 7.1% and HDGB an error of 12.9%. A similar test was done for the M2 channel lining segment (PDB code 1eq8, model 4[106]). In this case the

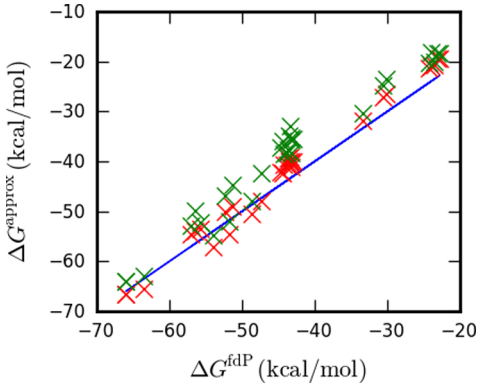


Figure 7. ΔG^{solv} for 31 different orientations and positions of WALP23 calculated with FACTSMEM (red) and HDGB (green) compared to the results obtained with a fdP membrane.

average error is 2.9 and 2.6% for FACTSMEM and HDGB, respectively.

Finally, a 20 ns MD simulation of KALP23 in FACTSMEM was performed. In this case a tilt angle of 10.3 ± 4.9° was obtained. This result correlates with the trend that lysine anchors farther away from the membrane than tryptophane and therefore smaller tilt angles are to be expected for KALPs.[90] Similar trends were obtained by Kim and Im[103] using PMF explicit solvent simulations. For example, in DMPC, a tilt angle of 20.7° was obtained for KALP23 and a tilt angle of 28.1° for WALP23.

Glycophorin A. Glycophorin A is a membrane protein consisting of 131 amino acids that can be found in human erythrocyte cells. Its transmembrane domain forms dimers and it is used for the study of protein helix−helix interactions inside membranes. Each monomer folds as an α-helix and the two monomer cross at an angle of −40°.[74] The dimer forms a packed structure with specific intermonomer hydrogen bonds to stabilize it. The dimer has been studied by solution NMR,[74] SS-NMR,[107] explicit solvent simulations,[108] coarse-grained simulations[109] and with the GBSW implicit membrane model.[14] Distances between different specific atoms in each monomer can be used to describe the packing of the dimer. The results for different methods are summarized in Table 5.

A 20 ns MD simulation of the dimer was performed with FACTSMEM starting from the solution NMR structure. The dimer is stable with an average C_α-RMSD of the transmembrane domain of 1.3 Å. The distances between the relevant atoms for the simulation with FACTSMEM can also be found in Table 5. The results obtained with FACTSMEM are very similar to the ones obtained with GBSW and explicit solvent simulations, and partially different from the experimental ones. Even though all three simulations (explicit solvent, GBSW, FACTSMEM) were started from the solution NMR structure, the final results are closer to the ones from SS-NMR. It can be concluded that FACTSMEM can be used to study intermonomer interactions of transmembrane domains with results similar to the ones from explicit solvent simulations.

Bacteriorhodopsin and Clc Channel. Because implicit membrane models only represent protein atoms explicitly, large systems can be studied with them. Thus, it is relevant to test if FACTSMEM can reproduce the dynamics of large membrane proteins. In this study, we performed 20 ns MD simulations of a bacteriorhodopsin monomer and a Clc channel dimer.

Bacteriorhodopsin, a membrane protein from *Halobacterium salinarum*, is used to convert light into proton gradients. The structure used for the simulation contains residues 5−232 from PDB file 1qhj. The C_α-RMSD for the 228 amino acids is plotted in Figure 8 and shows that the simulation is stable. The value of C_α-RMSD compares very well with the results from a 5 ns explicit solvent simulation of the bacteriorhodopsin trimer, in which the C_α-RMSD is between 1.1 and 1.6 Å.[110] The fact that the C_α-RMSD is similar for FACTSMEM and explicit solvent simulations is very relevant as one of the main criticism of implicit solvent models is that they tend to overstabilize protein−protein interactions, in particular salt bridges,[111,112] which would lead to lower C_α-RMSDs. Moreover, the protein does not turn to unphysical positions as it happened with the HDGB implicit membrane model when used with the standard cutoff scheme.[42] We also tested introducing bacteriorhodopsin in a horizontal orientation. This orientations is not stable and the protein leaves the membrane. The difference in ΔG^{solv} between the horizontal and vertical conformations is 720 kcal/

Table 5. Interhelical Distances in the Glycophorin A Dimer Measured in Å

helix 1	helix 2	solution NMR[74]	SS-NMR[107]	explicit solvent[108]	GBSW[14]	FACTSMEM
Gly79 C_α	Gly79 C	5.3	4.1	4.0 ± 0.3	4.0 ± 0.2	4.0 ± 0.2
Ile76 C	Gly79 C_α	5.4	4.8	4.5 ± 0.3	4.5 ± 0.3	4.5 ± 0.3
Gly83 C_α	Gly83 C	4.9	4.3	4.3 ± 0.4	4.4 ± 0.4	4.6 ± 0.4
Gly83 C_α	Val80 C	4.6	4.2	4.7 ± 0.3	4.9 ± 0.3	5.0 ± 0.3
Gly79 C	Val80 C_γ	2.8	4.0	4.1 ± 0.3	4.2 ± 0.3	4.2 ± 0.3
Gly83 C	Val84 C_γ	3.4	4.0	4.5 ± 0.4	4.5 ± 0.5	4.9 ± 0.8

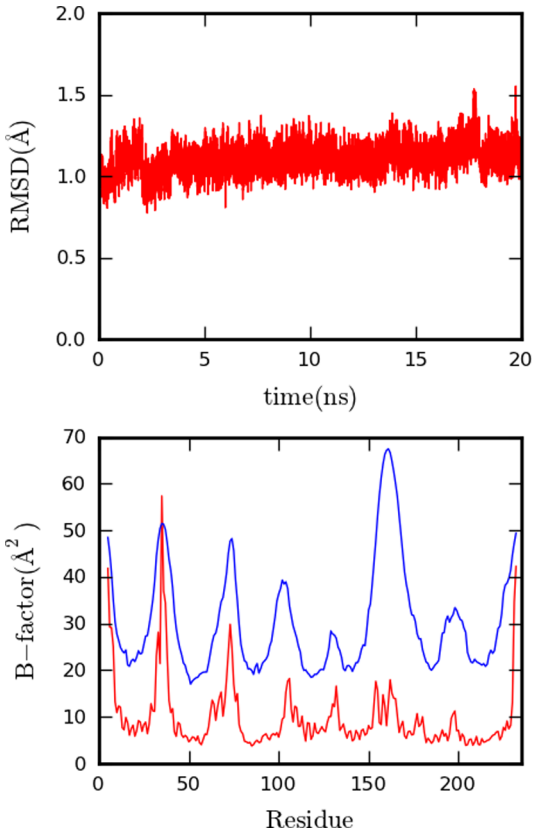


Figure 8. Results for the 20 ns MD simulation of bacteriorhodopsin. Top: C_α-RMSD. Bottom: Simulated (red) and experimental (blue)[75] B-factor.

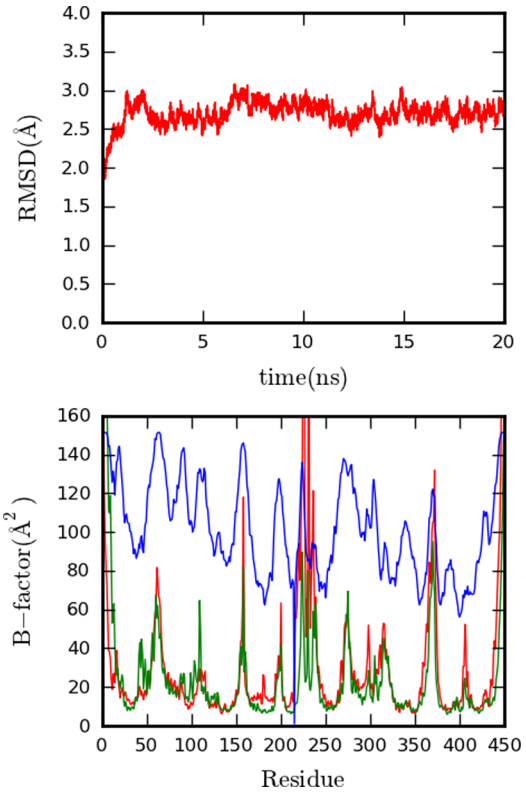


Figure 9. Results for the 20 ns MD simulation of the Clc channel. Top: C_α-RMSD. Bottom: simulated (red and green) and experimental (blue)[85] B-factor.

mol. The B-factor from the FACTSMEM simulation and the X-ray experiment[75] are compared in Figure 8. The flexible regions from the simulations correspond well with the experimental result. The only region in which some difference is observed is in helix E (around residue 154 in Figure 8), where the experimental B-factor is much higher than the simulated one. However, the same discrepancy was observed in explicit solvent simulations of the bacteriorhodopsin trimer.[110] It was argued, that the reason for the large experimental B-factor is not the flexibility of the region but the difference in behavior between the monomers.[110]

We also performed a simulation of a Clc channel dimer, which contains 900 amino acids. Clc channels transport Cl^- selectively across membranes. In humans, they have different functions such as regulating electrical excitation in skeletal muscles. The Clc channel studied in this work is from *Escherichia coli*. The C_α-RMSD of the 20 ns simulation with FACTSMEM is plotted in Figure 9. It takes around 3 ns for the system to equilibrate after which the dimer is stable. The simulated B-factor for each monomer and for the X-ray experiment[85] are plotted in Figure 9. The flexible protein sections simulated with FACTSMEM correspond well with the experimental results.

These results show that FACTSMEM can be applied to study large membrane proteins.

Computational Requirements. One of the strengths of implicit membrane models is the smaller amount of computational resources needed to perform an MD simulation compared to explicit solvent simulations. Thus, it is important to compare FACTSMEM to other implicit membrane models regarding computational requirements. A 1 ns MD simulation of melittin was performed with different models on a single 2.93 GHz Intel Xeon X5570 processor. The CHARMM19 force field was used together with IMM1 and GBSAIM. It should be noted that the CHARMM19 force field does not represent aliphatic hydrogens explicitly, which means that the same simulation with CHARMM19 is faster than with CHARMM22. Simulations with FACTS, IMM1, the GBSW membrane and HDGB were performed with CHARMM22. A simulation in vacuum was performed with both force fields as a benchmark.

The CPU time and relative time as a function of the vacuum simulation for all models are given in Table 6. FACTSMEM is only 6 times slower than a vacuum simulation and only slightly slower than FACTS. Other implicit membrane models that follow the GBSA formalism such as GBSW or HDGB are much slower than FACTSMEM. Notably, FACTSMEM is even faster than the GBSAIM model which uses the CHARMM19 force field. When using the CHARMM22 force field, the only model with a comparable speed to FACTSMEM is IMM1. When IMM1 is used in conjunction with the CHARMM19 force field, it is faster than FACTSMEM.

Table 6. Computational Requirements of Different Implicit Solvent Models[a]

model	CPU time (h)	relative time[b]
CHARMM19		
VACUUM	0.15	1
IMM1	0.53	3.5
GBSAIM	6.42	42.8
CHARMM22		
VACUUM	0.36	1
FACTS	1.95	5.4
IMM1	2.08	5.7
FACTSMEM	2.19	6.0
GBSW	7.76	21.4
HDGB	15.9	43.9

[a]For a 1 ns MD simulation of melittin on a single CPU. [b]As a function of the vacuum simulation using the same force field.

CONCLUSIONS

The generalized Born formalism is a mean-field continuum model used to represent the water environment. One of the fastest models that follow this formalism is the FACTS implicit solvent model. Here, we present an extension of FACTS to membranes. The model, called FACTSMEM, uses a position dependent dielectric constant ε_s and effective surface tension for the estimation of the solvation free energy. The effective Born radii at different ε_s are calculated as a function of the value obtained by FACTS at ε_s = 78.5 using a parameter-free approximation to Kirkwood's equation. The approximation was tested for the calculation of self-electrostatic solvation free energies of all atoms of 61 different protein conformations. When comparing to the results of finite difference Poisson calculations, only a very small error is added to the one originally introduced by FACTS. The approximation was also tested by calculating interatomic interaction energies and the solvation free energy of 2800 different protein conformations. In all cases, errors at different ε_s are similar. Furthermore, FACTSMEM is able to discriminate correctly between different conformations in the same dielectric environment and the same conformation in different dielectric environments. It should be noted, that the avoidance of parameters in the approximation to Kirkwood's equation enables a high transferability of eq 19 for the extension of other implicit solvent models to membranes.

FACTSMEM was used to calculate the free energy profiles for membrane insertion of amino acid analogues. The results qualitatively agree with those obtained with explicit solvent simulations. Notably, the free energy minimum for interface anchoring residues such as tryptophane and lysine are correctly represented by FACTSMEM. Moreover, our new implicit membrane model was tested for MD simulations of various proteins. First, melittin was studied in its orientation parallel to the membrane. The peptide stays at a distance from the membrane center similar to what is obtained experimentally. We also found that the kink between the two helices is more flexible than expected; similar results were obtained from coarse-grained simulations. We also studied WALP23 and KALP23 with FACTSMEM. For WALP23, a similar tilt angle to explicit solvent umbrella sampling MD simulations was obtained. Furthermore, we could also reproduce the trend that KALPs tilt less than WALPs because of the different anchoring positions of their terminal amino acids. FACTSMEM was also tested for the study of protein−protein interactions inside membranes. For this purpose, we studied a glycophorin A dimer. The distances between representative atoms of both helices are close to what is obtained from explicit solvent simulations. Finally, we applied FACTSMEM to larger proteins such as bacteriorhodopsin and a Clc channel dimer. Both proteins are stable and the experimentally determined flexible regions are reproduced correctly in our simulations.

Many of the proteins studied with FACTSMEM have already been studied with other implicit membrane models. This allows us to compare the quality of FACTSMEM and other implicit membrane models. GBSW and FACTSMEM obtain similar results for the helix−helix interactions of glycophorin A. However, the tilt angle of WALP23 calculated with GBSW is much larger than the one obtained with FACTSMEM. The result obtained with FACTSMEM, however, is closer to the one obtained with potential of mean force calculations using an explicitly represented membrane. The implicit solvent model which is most similar to FACTSMEM is HDGB, as FACTSMEM uses the same profiles for the dielectric constant and surface tension parameter. Both models produce similar results for melittin and bacteriorhodopsin. However, FACTSMEM does not stabilize the unphysical horizontal position of bacteriorhodopsin using the standard cutoff scheme. Furthermore, the estimated free energies of solvation of 31 different protein orientations inside a membrane are very similar for HDGB and FACTSMEM. Moreover, the computational requirements of FACTSMEM are lower than the ones of other implicit membrane models, particularly those that also follow the GB formalism. Thus, FACTSMEM can be used to generate results comparable to those from other implicit membrane models yet at a fraction of their computing time.

The results obtained here show that FACTSMEM can be used to obtain quantitative results close to experiments and explicit solvent simulations. However, we know that approximations are made in developing such a model and the simulations using it cannot be considered equal to explicit solvent simulations. For example, FACTSMEM does not consider that the membrane can change its width to adapt to certain proteins. As with all other implicit solvents, which consider the interaction as a mean field approximation, specific interactions between protein and lipid atoms cannot be represented in FACTSMEM. However, a fast implicit membrane model such as FACTSMEM can be used for certain studies which exceed the time scales accessible with explicit solvent simulations, such as structure prediction simulations. It is specially interesting when implicit and explicit solvent simulations are combined for such goal. For example, Zhang et al.[113] used the GBSW implicit membrane, and later refined their results with an explicit solvent simulation, to predict the structure of Erb and Eph dimers. FACTSMEM can play an important role in such structure prediction studies.

AUTHOR INFORMATION

Corresponding Author

*E-mail: b.strodel@fz-juelich.de.

Present Address

[§]Ioan Vancea: European Molecular Biology Laboratory c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Amedeo Caflisch for providing the protein structure files and Michael Feig and Parimal Kar for the profiles of the effective dielectric constant and normalized surface tension parameter. M.C.P. and B.S. gratefully acknowledge the computing time granted by the JARA-HPC Vergabegremium and provided on the JARA-HPC Partition part of the RWTH Compute Cluster in Aachen (grant number JARA0018).

REFERENCES

(1) Wallin, E.; Heijne, G. V. *Protein Sci.* **1998**, *7*, 1029−1038.
(2) White, S. Membrane Proteins of Known 3D structure, 2013. http://blanco.biomol.uci.edu/mpstruc/ (accessed May 12, 2014).
(3) Gumbart, J.; Wang, Y.; Aksimentiev, A.; Tajkhorshid, E.; Schulten, K. *Curr. Opin. Struct. Biol.* **2005**, *15*, 423−431.
(4) Lindahl, E.; Sansom, M. S. *Curr. Opin. Struct. Biol.* **2008**, *18*, 425−431.
(5) Strodel, B.; Lee, J. W. L.; Whittleston, C. S.; Wales, D. J. *J. Am. Chem. Soc.* **2010**, *132*, 13300−13312.
(6) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.
(7) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1−20.
(8) Feig, M.; Brooks, C. L., III *Curr. Opin. Struct. Biol.* **2004**, *14*, 217−224.
(9) Chen, J.; Brooks, C. L., III; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140−148.
(10) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.
(11) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59−75.
(12) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(13) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(14) Im, W.; Feig, M.; Brooks, C. L., III *Biophys. J.* **2003**, *85*, 2900−2918.
(15) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L., III *J. Comput. Chem.* **2003**, *24*, 1348−1356.
(16) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479−499.
(17) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383−394.
(18) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J. Chem. Theory Comput.* **2007**, *3*, 156−169.
(19) Haberthür, U.; Caflisch, A. *J. Comput. Chem.* **2008**, *29*, 701−715.
(20) Aguilar, B.; Shadrach, R.; Onufriev, A. V. *J. Chem. Theory Comput.* **2010**, *6*, 3613−3630.
(21) Brieg, M.; Wenzel, W. *J. Chem. Theory Comput.* **2013**, *9*, 1489−1498.
(22) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(23) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297−1304.
(24) Zoete, V.; Grosdidier, A.; Cuendet, M.; Michielin, O. *J. Mol. Recogn.* **2010**, *23*, 457−461.
(25) Felline, A.; Seeber, M.; Rao, F.; Fanelli, F. *J. Chem. Theory Comput.* **2009**, *5*, 2472−2485.
(26) Serquera, D.; Lee, W.; Settanni, G.; Marszalek, P. E.; Paci, E.; Itzhaki, L. S. *Biophys. J.* **2010**, *98*, 1294−1301.
(27) Hoffmann, F.; Strodel, B. *J. Chem. Phys.* **2013**, *138*, 025102.
(28) Spassov, V. Z.; Yan, L.; Szalma, S. *J. Phys. Chem. B* **2002**, *106*, 8726−8738.
(29) Im, W.; Lee, M. S.; Brooks, C. L., III *J. Comput. Chem.* **2003**, *24*, 1691−1702.
(30) Lazaridis, T. *Proteins: Struct., Funct., Bioinf.* **2003**, *52*, 176−192.
(31) Tanizaki, S.; Feig, M. *J. Chem. Phys.* **2005**, *122*, 124706.
(32) Ulmschneider, M. B.; Ulmschneider, J. P.; Sansom, M. S.; Nola, A. D. *Biophys. J.* **2007**, *92*, 2338−2349.
(33) Parisio, G.; Ferrarini, A. *J. Chem. Theory Comput.* **2010**, *6*, 2267−2280.
(34) Panahi, A.; Feig, M. *J. Chem. Theory Comput.* **2013**, *9*, 1709−1719.
(35) Dominy, B. N.; Brooks, C. L., III *J. Phys. Chem. B* **1999**, *103*, 3765−3773.
(36) Im, W.; Brooks, C. L., III *J. Mol. Biol.* **2004**, *337*, 513−519.
(37) Im, W.; Brooks, C. L., III *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6771−6776.
(38) Bu, L.; Im, W.; Brooks, C. L., III *Biophys. J.* **2007**, *92*, 854−863.
(39) Stern, H. A.; Feller, S. E. *J. Chem. Phys.* **2003**, *118*, 3401−3412.
(40) Marrink, S. J.; Berendsen, H. J. C. *J. Phys. Chem.* **1996**, *100*, 16729−16738.
(41) Feig, M.; Im, W.; Brooks, C. L., III *J. Chem. Phys.* **2004**, *120*, 903−911.
(42) Tanizaki, S.; Feig, M. *J. Phys. Chem. B* **2006**, *110*, 548−556.
(43) Sayadi, M.; Tanizaki, S.; Feig, M. *Biophys. J.* **2010**, *98*, 805−814.
(44) Sayadi, M.; Feig, M. *Biochim. Biophys. Acta, Biomembr.* **2013**, *1828*, 577−585.
(45) Jaskierny, A. J.; Panahi, A.; Feig, M. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1109−1117.
(46) Panahi, A.; Feig, M. *J. Phys. Chem. B* **2010**, *114*, 1407−1416.
(47) Choe, S.; Hecht, K. A.; Grabe, M. *J. Gen. Physiol.* **2008**, *131*, 563−573.
(48) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Bioinf.* **1999**, *35*, 133−152.
(49) Miyashita, N.; Straub, J. E.; Thirumalai, D.; Sugita, Y. *J. Am. Chem. Soc.* **2009**, *131*, 3438−3439.
(50) Strodel, B.; Wales, D. J. *J. Chem. Theory Comput.* **2008**, *4*, 657−672.
(51) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(52) MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. *Biophys. J.* **2008**, *94*, 3393−3404.
(53) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351−361.
(54) Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, *79*, 5333−5339.
(55) Sigalov, G.; Scheffel, P.; Onufriev, A. *J. Chem. Phys.* **2005**, *122*, No. 094511.
(56) Sigalov, G.; Fenley, A.; Onufriev, A. *J. Chem. Phys.* **2006**, *124*, No. 124902.
(57) Grycuk, T. *J. Chem. Phys.* **2003**, *119*, 4817−4826.
(58) Wojciechowski, M.; Lesyng, B. *J. Phys. Chem. B* **2004**, *108*, 18368−18376.
(59) Tjong, H.; Zhou, H.-X. *J. Chem. Phys.* **2006**, *125*, No. 206101.
(60) Tjong, H.; Zhou, H.-X. *J. Phys. Chem. B* **2007**, *111*, 3055−3061.
(61) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(62) Feig, M.; Karanicolas, J.; Brooks, C. L., III *J. Mol. Graphics Modell.* **2004**, *22*, 377−395.
(63) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 24−33.
(64) MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; 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.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(65) MacKerell, A. D.; Feig, M.; Brooks, C. L., III *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(66) MacKerell, A. D.; Feig, M.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(67) Chen, J.; Im, W.; Brooks, C. L., III *J. Am. Chem. Soc.* **2006**, *128*, 3728−3736.
(68) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327−341.
(69) Feig, M. *J. Chem. Theory Comput.* **2007**, *3*, 1734−1748.
(70) Norberg, J.; Foloppe, N.; Nilsson, L. *J. Chem. Theory Comput.* **2005**, *1*, 986−993.
(71) Li, L.; Vorobyov, I.; M, A. D., Jr.; Allen, T. W. *Biophys. J.* **2008**, *94*, L11−L13.
(72) Terwilliger, T. C.; Eisenberg, D. *J. Biol. Chem.* **1982**, *257*, 6010−6015.
(73) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714−2723.
(74) MacKenzie, K. R.; Prestegard, J. H.; Engelman, D. M. *Science* **1997**, *276*, 131−133.
(75) Belrhali, H.; Nollert, P.; Royant, A.; Menzel, C.; Rosenbusch, J. P.; Landau, E. M.; Pebay-Peyroula, E. *Structure* **1999**, *7*, 909−917.
(76) Gerwert, K.; Hess, B.; Soppa, J.; Oesterhelt, D. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4943−4947.
(77) Hayashi, S.; Ohmine, I. *J. Phys. Chem. B* **2000**, *104*, 10678−10691.
(78) Hayashi, S.; Tajkhorshid, E.; Pebay-Peyroula, E.; Royant, A.; Landau, E. M.; Navarro, J.; Schulten, K. *J. Phys. Chem. B* **2001**, *105*, 10124−10131.
(79) Nina, M.; Roux, B.; Smith, J. *Biophys. J.* **1995**, *68*, 25−39.
(80) Baudry, J.; Crouzy, S.; Roux, B.; Smith, J. C. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1018−1024.
(81) Tajkhorshid, E.; Paizs, B.; Suhai, S. *J. Phys. Chem. B* **1997**, *101*, 8021−8028.
(82) Tajkhorshid, E.; Suhai, S. *J. Phys. Chem. B* **1999**, *103*, 5581−5590.
(83) Tajkhorshid, E.; Baudry, J.; Schulten, K.; Suhai, S. *Biophys. J.* **2000**, *78*, 683−693.
(84) Saam, J.; Tajkhorshid, E.; Hayashi, S.; Schulten, K. *Biophys. J.* **2002**, *83*, 3097−3112.
(85) Dutzler, R.; Campbell, E. B.; Cadene, M.; Chait, B. T.; MacKinnon, R. *Nature* **2002**, *415*, 287−294.
(86) Reiher III, W. E. Theoretical Studies of Hydrogen Bonding. Ph.D. Thesis, Harvard University, Cambridge, MA, USA, 1985.
(87) Feig, M. *Methods Mol. Biol.* **2008**, *443*, 181−196.
(88) Scarsi, M.; Caflisch, A. *J. Comput. Chem.* **1999**, *20*, 1533−1536.
(89) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657−1666.
(90) de Planque, M. R. R.; Kruijtzer, J. A. W.; Liskamp, R. M. J.; Marsh, D.; Greathouse, D. V.; Koeppe, R. E.; de Kruijff, B.; Killian, J. A. *J. Biol. Chem.* **1999**, *274*, 20839−20846.
(91) Raghuraman, H.; Chattopadhyay, A. *Biosci. Rep.* **2007**, *27*, 189−223.
(92) Bazzo, R.; Tappin, M. J.; Pastore, A.; Harvey, T. S.; Carver, J. A.; Campbell, I. D. *Eur. J. Biochem.* **1988**, *173*, 139−146.
(93) Bernèche, S.; Nina, M.; Roux, B. *Biophys. J.* **1998**, *75*, 1603−1618.
(94) Frey, S.; Tamm, L. K. *Biophys. J.* **1991**, *60*, 922−930.
(95) Nagle, J. F.; Tristram-Nagle, S. *Biochim. Biophys. Acta, Rev. Biomembr.* **2000**, *1469*, 159−195.
(96) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(97) Santo, K. P.; Berkowitz, M. L. *J. Phys. Chem. B* **2012**, *116*, 3021−3030.
(98) Holt, A.; Killian, J. *Eur. Biophys. J.* **2010**, *39*, 609−621.
(99) Strandberg, E.; Ozdirekcan, S.; Rijkers, D. T.; van der Wel, P. C.; Koeppe, R. E., II; Liskamp, R. M.; Killian, J. A. *Biophys. J.* **2004**, *86*, 3709−3721.
(100) Holt, A.; Koehorst, R. B.; Rutters-Meijneke, T.; Gelb, M. H.; Rijkers, D. T.; Hemminga, M. A.; Killian, J. A. *Biophys. J.* **2009**, *97*, 2258−2266.
(101) Ozdirekcan, S.; Etchebest, C.; Killian, J. A.; Fuchs, P. F. J. *J. Am. Chem. Soc.* **2007**, *129*, 15174−15181.
(102) Esteban-Martín, S.; Salgado, J. *Biophys. J.* **2007**, *93*, 4278−4288.
(103) Kim, T.; Im, W. *Biophys. J.* **2010**, *99*, 175−183.
(104) Monticelli, L.; Tieleman, D. P.; Fuchs, P. F. *Biophys. J.* **2010**, *99*, 1455−1464.
(105) Yuzlenko, O.; Lazaridis, T. *J. Comput. Chem.* **2013**, *34*, 731−738.
(106) Opella, S. J.; Marassi, F. M.; Gesell, J. J.; Valente, A. P.; Kim, Y.; Oblatt-Montal, M.; Montal, M. *Nat. Struct. Mol. Biol.* **1999**, *6*, 374−379.
(107) Smith, S. O.; Song, D.; Shekar, S.; Groesbeek, M.; Ziliox, M.; Aimoto, S. *Biochemistry* **2001**, *40*, 6553−6558.
(108) Cuthbertson, J. M.; Bond, P. J.; Sansom, M. S. P. *Biochemistry* **2006**, *45*, 14298−14310.
(109) Psachoulia, E.; Fowler, P. W.; Bond, P. J.; Sansom, M. S. P. *Biochemistry* **2008**, *47*, 10503−10512.
(110) Kandt, C.; Schlitter, J.; Gerwert, K. *Biophys. J.* **2004**, *86*, 705−717.
(111) Zhou, R.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777−12782.
(112) Zhou, R. *Proteins: Struct., Funct., Bioinf.* **2003**, *53*, 148−161.
(113) Zhang, L.; Sodt, A. J.; Venable, R. M.; Pastor, R. W.; Buck, M. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 365−376.