REGULAR ARTICLE

Michael Feig · Jana Chocholoušová · Seiichiro Tanizaki

Extending the horizon: towards the efficient modeling of large biomolecular complexes in atomic detail

Abstract The application of evaluation of implicit solvent methods for the simulation of biomolecules is described. Detailed comparisons with explicit solvent are described for the modeling of peptide and proteins in continuum aqueous solvent. In addition, we are presenting new data on the simulation of DNA with implicit solvent and describe the development of a heterogeneous dielectric model for the simulation of integral membranes. The performance of implicit solvent simulations based on the GBMV generalized Born method is compared with explicit solvent simulations, and implications for the simulation of very large biomolecular complexes is discussed. We are anticipating that the work described herein will lead to new, efficient modeling tools that will allow the simulation of longer timescales and larger system sizes in order to meet current and future challenges by the experimental community.

Keywords Implicit solvent · Generalized born · Poisson equation · Molecular dynamics

1 Introduction

Classical simulations of biological macromolecules have moved from mere exercises in statistical mechanics to realistic descriptions of molecular dynamics in atomic detail, over the last decades. The progress in this field is a result of methodological advances, such as the efficient treatment of long-range electrostatics [1–4], new enhanced sampling

M. Feig (✉) · J. Chocholoušová · S. Tanizaki
Department of Biochemistry and Molecular Biology,
Michigan State University, East Lansing, MI 48824-1319, USA
E-mail: feig@msu.edu
Tel.: +1-517-4327439
Fax: +1-517-3539334

M. Feig
Department of Chemistry, Michigan State University,
East Lansing, MI 48824-1319, USA

M. Feig
Department of Computer Science and Engineering
Michigan State University, East Lansing, MI 48824-1319, USA

methods [5–9], continuing improvements in force fields [10–19] and advances in incorporating quantum mechanics into simulations [20–22], and also follows an ever increasing speed of computers. It has become possible not only to study folding of peptides and very small proteins in atomic detail [23–32] but also study fundamental mechanisms in large biomolecules such as DNA polymerase [33–35] and the ribosome [36,37]. Nevertheless, the dynamics of many biological processes remain largely inaccessible to atomic-level simulations when relevant timescales reach microseconds and system sizes exceed more than a few hundred residues. It is especially challenging to study the dynamics of protein–protein and protein–nucleic acid interactions in large complexes at the core of fundamental biological functions where much of the current experimental research is focused today. Two of the most impressive recent successes in experimental structure determination have resulted in atomic-level structures of the ribosome [38,39] and eukaryotic RNA polymerase II [40–43]. Both structures offer intriguing insight into gene transcription and translation, but the static structures have posed as many questions as they provide answers about the detailed, inherently dynamic mechanisms. With tremendous effort it has been possible to generate a 10 ns trajectory of the complete ribosome [36] in a first attempt to study its dynamics in atomic detail. However, while such simulations are very impressive, their length remains far from the millisecond timescale of ribosomal function. The great challenge in the years to come will be how to efficiently and accurately study long-time dynamics in large biomolecular complexes for which structures are already available and will likely become available from experiment in the near future.

While rapid increases in computer speed have driven much of the past progress in biomolecular simulations, it appears that the gains in computer speed have begun to slow down recently [44] emphasizing the need for further methodological advances. The computational cost of biomolecular simulations is limited by the finite integration step size at which a simulation progresses, by the level of model detail (i.e. the number of particles or interaction sites) that determines the cost to calculate energies and forces for a single

conformation, and by the time it takes to cross kinetic barriers compared to the biological system. The integration time step is limited fundamentally by the smoothness of the underlying energy landscape; in particular it needs to be short enough to sample the highest frequency components, typically from bond or angle vibrations. As a result, a time step of 1–2 fs is required for atomic-level simulations. While some attempts have been successful to extend the time step to 3–4 fs under certain assumptions [45–47], substantially longer time steps are unlikely to be practical if an atomic level of detail is maintained.

Another avenue is the acceleration of dynamics through enhanced sampling methods which facilitate the crossing of kinetic barriers. Many such methods have been proposed and successfully applied to challenging problems like protein folding [27,48–56]. Most enhanced sampling methods preserve the relative energies of conformational minima, but the extraction of accurate kinetic rates is often not straightforward. In principle, enhanced sampling methods can accelerate dynamics by many orders of magnitude if the dominant kinetic barriers are targeted specifically, e.g. through umbrella sampling [57–60]. However, this requires a priori knowledge of the transitional pathway(s) and barriers. On the other hand, non-specific enhanced sampling methods may facilitate the crossing of all barriers, for example through the use of elevated temperatures in replica exchange simulations [6,50–52,54–56,61] or with multicanonical sampling [6,62]. The result is substantial sampling of less relevant regions of conformational space, which limits the effectiveness in speeding up the most relevant kinetic processes. Nevertheless, even such non-specific enhanced sampling methods can accelerate molecular dynamics by up to several orders of magnitudes [7,63,64] providing access to microsecond timescales with nanosecond timescale simulations. As it appears that fundamental limits have not been reached, future methods are expected to enhance sampling of biological processes even further.

The third option for reaching longer timescales and larger system sizes in biomolecular simulations is the reduction of the level of model detail. An extreme example of this strategy is the use of lattice-based H-P models during early studies of protein folding [65–71]. A variety of more sophisticated low-resolution representations of both proteins and nucleic acids have been proposed over time [72–80]. While such models often lack the level of accuracy needed to understand specific biological processes in detail, it is possible to recover atomic-level detail through multi-scale modeling approaches [81–85]. One can imagine, for example, the use of low-resolution models for the propagation of conformations over kinetic barriers, while reconstructed atomic-level models are used to obtain a more accurate energetic description [85]. However, much of the current efforts in reducing the complexity of biomolecular simulations are focused on a reduced representation of the environment. A commonly employed explicit representation of the environment increases the computational cost substantially over the simulation cost of the biomolecule itself. Typical explicit solvent simulations usually contain more solvent atoms than solute atoms so that most of the cost is actually spent on calculating solvent-solvent interactions rather than solute-solvent or solute-solute interactions. Therefore, numerous attempts have been made to develop implicit descriptions of the environment that do not require an explicit representation while maintaining a comparable level of realism [86,87].

Implicit descriptions of the environment can be distinguished into knowledge-based and physically motivated approaches. Knowledge-based approaches take advantage of the vast amount of information about the structure of proteins and nucleic acids that has become available [88–90]. A typical example are implicit solvation models based on atomic solvation potentials [91–96] that have been used for a long time for water-soluble proteins and more recently also for membrane-bound proteins [97]. Such empirical models provide a good description based on statistical averages and are particularly well suited in scoring functions [89,98–100].

Physically motivated implicit solvent models generally begin by decomposing the solvation free energy into electrostatic and non-polar contributions [101]. In an implicit model, the electrostatic component of the solvation free energy, or the free energy of charging a given biomolecule from zero charge to full charge in the presence of the solvent environment, can be estimated with a continuum dielectric model of the solvent [86,87,102–104]. Such models have been very successful in including the effect of solvent in models of proteins and nucleic acids [105–108]. The non-polar contribution to the solvation free energy contains two components: (1) the entropic cost of creating a cavity for accommodating the biomolecule in the solvent and (2) the biomolecule–solvent van der Waals interactions. Both components are often combined into a simple term that is proportional to the solvent-accessible surface area of the biomolecule [101]. However, some studies suggest that only the cost of cavity formation is reasonably well represented by such an approach, while a different formalism appears to be more appropriate in accounting for the van der Waals interactions with the solvent [109–112].

Implicit solvent models have been successful in many applications, in particular as a component of conformational scoring functions [100,113–118], for estimating ligand–protein and protein–protein binding free energies (P. Ferrara et al. 2004, submitted) [119–124], but also to enable long-time molecular dynamics simulations of peptides and small proteins [24,53,110,126–129]. However, despite such successes, the degree to which implicit solvent models can approximate explicit solvent and, ultimately, reality is not yet fully established [130,131]. Furthermore, it remains challenging to take full advantage of such methods in the modeling of large complexes and biological systems in complex biological environments. In the following we will describe recent progress and ongoing efforts in this respect. In particular we will compare recently developed implicit solvent methods with explicit solvent methods in terms of accuracy and efficiency and we will describe new methods for the implicit modeling of membrane systems.

2 Continuum dielectric implicit solvent methodology

Figure 1 illustrates the basic continuum solvent approximation for biomolecules in aqueous solvent, where the partial charges of a given biomolecule are embedded in a low-dielectric cavity and surrounded by a high-dielectric environment representative of the solvent. This kind of model, which applies macroscopic concepts on a microscopic level, is described rigorously by the Poisson equation (or the Poisson–Boltzmann equation if ionic concentrations in the solvent are included):

$$\nabla[\varepsilon(r)\nabla\phi(r)] = -4\pi\rho(r) \qquad (1)$$

The Poisson equation can be solved to yield the electrostatic potential $\phi(r)$ as a function of the biomolecular charge density $\rho(r)$ and the dielectric function $\epsilon(r)$. The electrostatic solvation energy is then readily calculated from the electrostatic potential [103]. While direct solutions of the Poisson equation tend to be costly, more efficient approximations of continuum dielectric descriptions have been proposed [86]. Of these methods, the generalized Born (GB) formalism has become the most popular as it can approximate solutions from Poisson theory accurately and efficiently [132–134]. Furthermore, the GB method provides an analytical expression for the calculation of the electrostatic solvation energy that is advantageous for application in molecular dynamics simulations.

Generalized Born formalisms represent the same dielectric continuum model as Poisson or Poisson-Boltzmann theory. As an extension to the fundamental Born expression for the solvation energy of a single ion in a dielectric medium [135], the solvent-induced reaction field energy of a set of (partial) charges is described with the following empirical expression in GB theory [134,136]:

$$\Delta G^{\text{elec}}_{\varepsilon_p \to \varepsilon_w} = -\frac{1}{2}\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right)\sum_{i,j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i\alpha_j\exp(-r_{ij}^2/F\alpha_i\alpha_j)}}$$
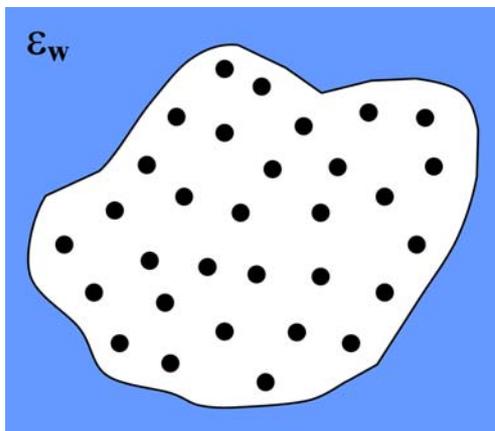
$$\qquad (2)$$

Fig. 1 Schematic illustration of a biomolecule embedded in a dielectric continuum

where ϵ_p, ϵ_w are the interior and exterior dielectric constants, r_{ij} is the distance between atoms i and j, α_i is the so-called GB radius of atom i. The factor F may range from 2 to 10, with 4 being the most commonly used value [134]. While the Born radius of a single, spherical ion corresponds to its size, GB radii reflect the distance of a charge location from the solvent boundary in a given molecule. It has been shown that Eq. (2) is a very good approximation to the electrostatic solvation energy from Poisson theory as long as "perfect" GB radii are used [137]. Perfect (according to Poisson theory) Born radii are obtained from Eq. (3):

$$\alpha_i = -\frac{1}{2}\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right)\frac{1}{G^i_{\text{pol}}} \qquad (3)$$

where the solvation energy G^i_{pol} is obtained from Poisson theory for the given two-dielectric system with a unit charge at the location of the atom i while all the other charges are set to zero. Thus, the key for a successful application of the GB formalism lies in an efficient and accurate calculation of the GB radii, α_i. Following the so-called Coulomb field approximation, it is possible to obtain the GB radius of a given atom from the following expression [133]:

$$\frac{1}{\alpha_i} = \frac{1}{R_i} - \frac{1}{\pi 4}\int_{\text{solute},r>R_i}\frac{1}{r^4}dV \qquad (4)$$

where R_i is usually the atomic van der Waals radius of atom i and the volume integral is carried out over the entire solute cavity. The integral can be approximated further by discrete sums of overlapping spheres [138,139] or Gaussians [140]. However, some methods carry out the integration directly [141,142] or replace the volume integral with a surface integral [143] in order to evaluate Born radii according to Eq. (4) more accurately. The Coulomb field approximation is exact for a charge at the center of a spherical cavity, but becomes inaccurate for off-center charges. For small molecules, where charge sites are relatively close to the center, the errors remain small; however, significant deviations are found in larger molecules [141,144,145]. Corrections to the Coulomb field approximation have been introduced recently [145,146] with the result of substantially more accurate electrostatic solvation energies for larger molecules. The Coulomb field approximation correction terms have also opened an avenue to extend the GB formalism to the accurate implicit treatment of low-dielectric environments. According to Kirkwood [147], the reaction field energy for a single off-center charge q at a distance r from the center of a spherical cavity with radius R is given as:

$$\Delta G^{\text{elec}}_{\varepsilon_p/\varepsilon_p \to \varepsilon_p\varepsilon_w} = q\phi_{RF}(\mathbf{r})$$

$$= -\frac{q^2}{2}\left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}\right)$$

$$\times\left(\frac{1}{R} + \frac{2\varepsilon_w}{2\varepsilon_w + \varepsilon_p}\frac{r^2}{R^3} + \frac{3\varepsilon_w}{3\varepsilon_w 2\varepsilon_p}\frac{r^4}{R^5} + \cdots\right) \qquad (5)$$

where ϵ_w and ϵ_p are the external and internal dielectric constants, respectively. The Coulomb field approximation corresponds to the zeroth-order term, in which case the Born

radii [calculated from Eq. (3)] do not depend on r or the external dielectric constant. Corrections to the Coulomb field approximation effectively approximate the higher order terms in the Kirkwood expression and introduce a dependence on ϵ_w and ϵ_p [144]. This has led to the following formulation for the calculation of GB radii in arbitrary two-dielectric environments:

$$\alpha_i = \frac{1}{C_0 A_4 + C_1 \left(\frac{3\varepsilon_w}{3\varepsilon_w + 2\varepsilon_p}\right) A_7} + D + \frac{E}{\varepsilon_w + 1} \quad (6)$$

In this expression, A_4 is calculated as $1/\alpha_i$ according to the Coulomb field approximation [Eq. (4)], A_7 is a higher-order correction described elsewhere [146], and C_0, C_1, D, and E are constants that are fitted to reproduce solutions from Poisson theory.

A critical issue of continuum dielectric models is the exact definition of the dielectric boundary. The canonical approach is to construct the so-called molecular surface [148]. This surface is defined by rolling a sphere with the approximate size of a water molecule around the molecular object constructed from van der Waals spheres for all atomic sites. This definition leads to the exclusion of small cavities that are essentially inaccessible to water molecules and is presumed to result in the surface that best describes true solvent accessibility. Such a surface definition is commonly used when electrostatic energies of biomolecules are calculated from Poisson theory [103, 149]. Existing methods for the exact calculation of the molecular surface [150] have been too expensive, however, to be effective in GB methods. Therefore, most GB methods represent the dielectric boundary as an overlap of atomic van der Waals spheres that may be smoothed or replaced by Gaussians in order to approximate the molecular surface [140, 142, 151]. As a consequence, the agreement of such GB methods with Poisson theory, where the exact molecular surface is used to define the dielectric boundary, is limited because different dielectric functions are compared [132]. Better agreement has become possible, however, with new formulations that approximate the additional solvent-excluded volume created by the molecular surface rather than the actual surface. One such method is the GBMV formalism [146], which also carries out the integral in Eq. (4) directly and includes a correction to the Coulomb field approximation. As a result, highly accurate Born radii are obtained with the GBMV method when compared to the corresponding radii calculated from Poisson theory according to Eq. (3) while total electrostatic solvation energies agree very well with relative errors of $\leq 1\%$ between GBMV and Poisson theory [132, 146].

3 Implicit versus explicit solvent

A connection between implicit and explicit solvent is possible in multiple ways. The most direct quantitative comparison between explicit and implicit solvent is the calculation of solvation free energies. Especially attractive is the calculation of charging free energies for a fixed conformation that corresponds directly to the electrostatic solvation energy based on a continuum dielectric model. Such a comparison can be used to adjust the dielectric boundary, e.g. by modifying the atomic radii that are used to define the molecular surface. This strategy has been followed for proteins and nucleic acids based on explicit charging free energies for amino acid dipeptides and single nucleotides [152–155]. A modification of the atomic radii from the default van der Waals radii, or more precisely the σ values from the force field Lennard–Jones potential, can be justified by comparing the molecular surface with the surface of the solvent-excluded volume from explicit solvent simulations. The example in Fig. 2a shows the default molecular surface and explicit solvent-excluded volume around a β-hairpin structure. Many areas are clearly visible where the molecular surface either over- or underestimates the solvent-accessible surface. The surface can be brought into much better agreement when the atomic radii are adjusted appropriately (Fig. 2b). In addition, Table 1 shows that the agreement with explicit solvent charging free energies of amino acid dipeptides also improves consistently as a result of the adjusted dielectric boundary. It should be stressed that the actual Lennard–Jones parameters remain unaffected. Only the radii that define the dielectric boundary are changed in order to reflect that the effective point of closest contact between the solvent and atoms on the molecular surface is not at the minimum of the Lennard–Jones potential but substantially closer or further as a result of electrostatic interactions.

Comparisons of conformational energy landscapes between explicit and implicit solvent descriptions are most relevant for actual applications of implicit solvent methods. Molecular dynamics simulations with implicit solvent that are easily performed with GB methods [110, 126, 127, 156–158] and to a more limited extent also with Poisson-based methods [129, 159–162] can serve as a means to carry out such tests. Figure 3 shows the results of implicit solvent simulations with a recent GB method for the three proteins, protein G, ubiquitin, and cyclophilin A. These three proteins were selected as good test cases where structures from crystallography and NMR are available and in close agreement. The average Cα root mean square deviations over the last 5 ns are 0.77 Å (protein G), 1.74 Å (cyclophilin A), and 1.59 Å (ubiquitin) with respect to the experimental structures 3GB1, 1OCA, and 1D3Z. For comparison, typical deviations of 1–3 Å are found in explicit solvent simulations [162]. As another example, we have carried out implicit solvent simulations of the DNA dodecamer d(CGCGAATTCGCG)$_2$. Figure 4 shows that the structure is well maintained over the course of a 10 ns simulation whereas more detailed comparisons of structural parameters with explicit solvent simulations of the same sequence show surprisingly good agreement (see Table 2), although we note larger structural fluctuations in the implicit simulations. This finding is especially remarkable given difficulties in obtaining stable simulations of nucleic acids with both implicit [158] and explicit solvent [1, 163, 164].
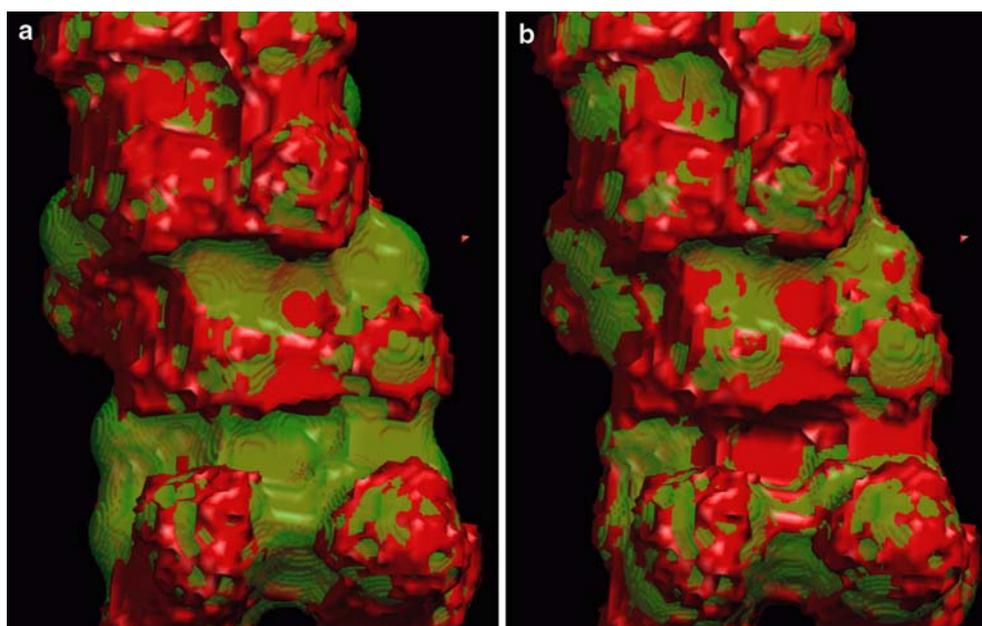
Fig. 2 Comparison of molecular surface used in continuum solvent (*green*) and true solvent-accessible surface (*red*) from explicit solvent simulations for β-hairpin structure. The molecular surface on the left (**a**) is based on radii taken from default CHARMM22 force field Lennard–Jones parameters [192]. Atomic radii were adjusted as follows for the molecular surface on the right (**b**): Cα: 2.107/2.4 Å (Gly), backbone C: 2.239 Å, backbone O: 1.444 Å, backbone N: 2.140 Å, backbone Hα: 1.207 Å, backbone HN: 0.225 Å, alanine Cβ; 2.142 Å, Hβ;1.240 Å, phenylalanine/tyrosine ring C: 1.8 Å, tryptophane ring C: 2.0 Å, tyrosine/serine/threonine hydroxyl O: 1.83 Å, glutamine/asparigine side chain O: 1.6 Å, glutamine/asparigine side chain N: 2.04 Å, histidine ring N: 1.92 Å, aspartic acid/glutamic acid side chain O: 1.4 Å, lysine/arginine side chain N: 2.05 Å. In addition a reduced water probe radius of 1 Å was used

Table 1 Comparison of charging free energies for blocked dipeptides

System	Explicit solvent	Implicit solvent molecular surface	Implicit solvent modified surface
Ala	−11.9	−11.5	−12.8
Val	−10.4	−9.7	−11.2
Leu	−11.4	−11.3	−12.6
Ile	−10.8	−9.7	−11.1
Ser	−19.7	−19.3	−19.8
Thr	−19.0	−17.8	−18.4
Phe	−14.6	−12.7	−14.6
Tyr	−19.0	−18.7	−19.7
Cys	−15.5	−14.8	−16.0
Met	−11.2	−10.8	−11.7
Asn	−20.8	−19.7	−20.5
Gln	−16.7	−14.8	−16.1
Trp	−17.9	−16.4	−17.5
Hsd	−25.7	−24.3	−24.9
Asp	−91.9	−78.6	−91.9
Glu	−87.8	−74.9	−88.3
Arg	−66.0	−77.5	−70.3
Lys	−72.9	−82.1	−77.3

The dipeptide conformations were obtained from M. Nina and are the same as the structures used in a previous study [152]. Explicit solvent charging free energy results were taken from the same study. Implicit solvent electrostatic solvation energies were calculated from Poisson theory with the PBEQ module in CHARMM [188,189] both with default CHARMM radii and the standard molecular surface modified as described in the caption for Fig. 2. A grid spacing of 0.1 Å and an external dielectric constant of 80 were used for the Poisson calculations. All energies are in kcal/mol

More extensive comparisons between implicit and explicit solvent energy landscapes have been carried out in peptide folding studies [128,130,131,165]. Generally, the agreement between implicit and explicit solvent is quite good, however, some deficiencies have been pointed out, in particular with respect to the enhancement of salt bridges in implicit solvent [130]. Such disagreement between implicit and explicit solvent models may suggest that the dielectric surface requires further modification or that a pure implicit solvent model is simply insufficient. Many examples have been
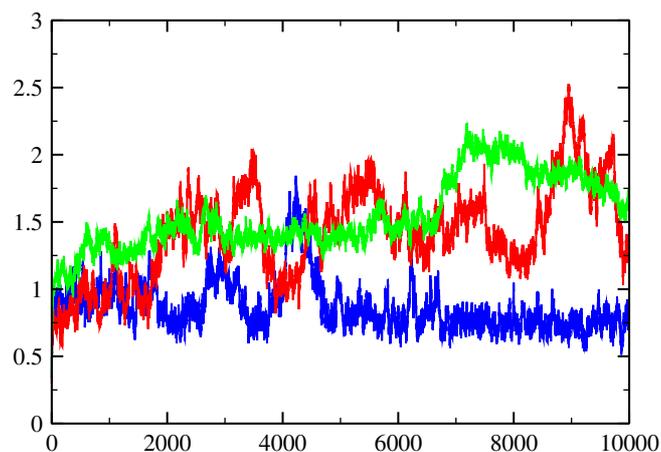
Fig. 3 Cα root mean square deviation in Å as a function of time in ps for molecular dynamics simulations at 300 K of protein G (*blue*), ubiquitin (*red*), and cyclophilin A (*green*) with implicit solvent. The GBMV method was used with the modified surface described in the caption of Fig. 2. Solvent friction was included through Langevin dynamics with a friction coefficient of 50/ps. A coefficient of $\gamma = 15$ cal/mol/Å^2 was used in solvent-accessible surface area based non-polar solvation term
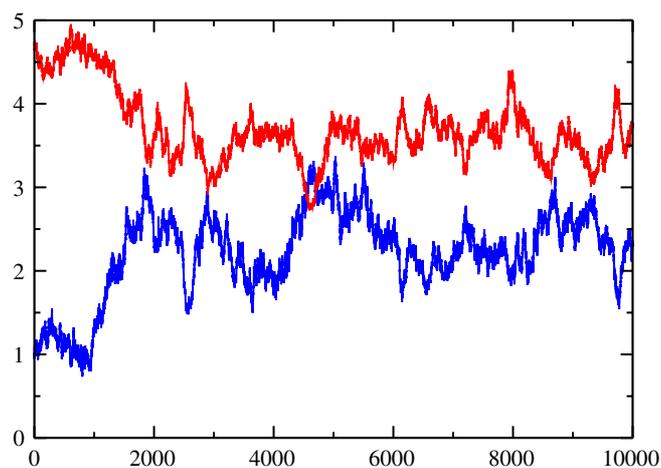

Fig. 4 Heavy-atom root mean square deviation in Å as a function of time in ps for molecular dynamics simulation of DNA d(CGCGAATTCGCG)$_2$ with the CHARMM27 force field for nucleic acids [191,193]. Terminal base pairs are excluded. RMSD values are calculated with respect to the NMR structure 1 GIP (*blue*) and canonical A-DNA (*red*). Implicit solvent simulations were carried out with the GBMV method, Langevin dynamics and hydrophobic contribution as described in Fig. 3, but without modifying the dielectric boundary

Table 2 Comparison of DNA helical parameters

	Explicit solvent	Implicit solvent
Rise	3.4 ± 0.4 Å	3.2 ± 0.6 Å
Inclination	8.7 ± 3.0	11.2 ± 8.8
Roll	5.6 ± 3.0	6.2 ± 7.8
Twist	34.9 ± 2.5	34.3 ± 5.4
x-displacement	-0.9 ± 0.4 Å	-1.3 ± 1.2 Å
Propeller twist	-11.4 ± 5.4	-11.5 ± 8.9
Slide	0.1 ± 0.3 Å	-0.7 ± 0.5 Å

Helical parameters were calculated with FREEHEL98 [190] from the implicit solvent simulation shown in Fig. 4. Averages over the last 9 ns simulation time and over all base pairs except the termini are compared with explicit solvent results for the same sequence from previous work [191]

documented where solvent molecules interact in a highly specific manner with biological macromolecules [166–169]. Specific interactions with solvent often rely on the discrete nature of solvent molecules with dynamic and electric properties that are very different from bulk solvent. One would expect that implicit solvent alone is not sufficient in such cases so that it may be necessary to develop hybrid models that include a small set of explicit solvent molecules. Some first steps have been undertaken in this direction [170,171], but it is clear that many more studies are needed to evaluate these issues in more detail as the agreement with explicit solvent and, ultimately, experiment is crucial in establishing the reliability of implicit solvent methods.

Another aspect that has not been addressed in much detail is how kinetic rates obtained with implicit solvent relate to kinetics from explicit solvent simulations. By itself, implicit solvent based on continuum electrostatics does not include any solvent friction or stochastic collisions with the solvent. As a consequence, the dynamics with implicit solvent may be distorted substantially over explicit solvent simulations. In particular, it is not obvious how to extract meaningful kinetic rates from such simulations, and it has recently been noted that the kinetics differ substantially even between different GB models [172]. Langevin dynamics can be used to restore the effects of solvent friction and stochastic collisions [173]. It has been reported that kinetic rates correlate linearly with the friction coefficient up to a factor 10 less than the value appropriate for water [174], but some practical questions remain how to apply Langevin dynamics correctly to simulations of large molecules in implicit solvent [175]. We have begun to address these questions by examining the kinetics of conformational transitions in alanine dipeptide that can be observed in direct molecular dynamics simulation. Figure 5 shows the energy landscapes projected onto the backbone dihedral angles ϕ and ψ from a 200 ns explicit solvent simulation and a $0.5\,\mu$s simulation with implicit solvent and Langevin dynamics. The sampling is very similar with 51% α vs. 43% β in explicit solvent and 55% α vs. 37% β in implicit solvent although some differences in detailed features can be identified. The rates for crossings between the α and β basins (see Table 3) are also very similar between the explicit solvent simulation and the Langevin simulation when a friction coefficient between 10/ps and 50/ps is employed. The rates become accelerated with smaller friction coefficients but not in a uniform way as suggested previously [174]. In fact, the rates change differently for α ↔ β and α$_L$ ↔ β transitions. α ↔ β transitions become more than four times frequent with $f = 5$/ps vs. $f = 50$/ps, while α$_L$ ↔ β transitions only change by a factor of 2. It may be suspected that the simulated dynamics in large biomolecules is affected in a similar way. Further studies will be needed to address these questions in more detail.
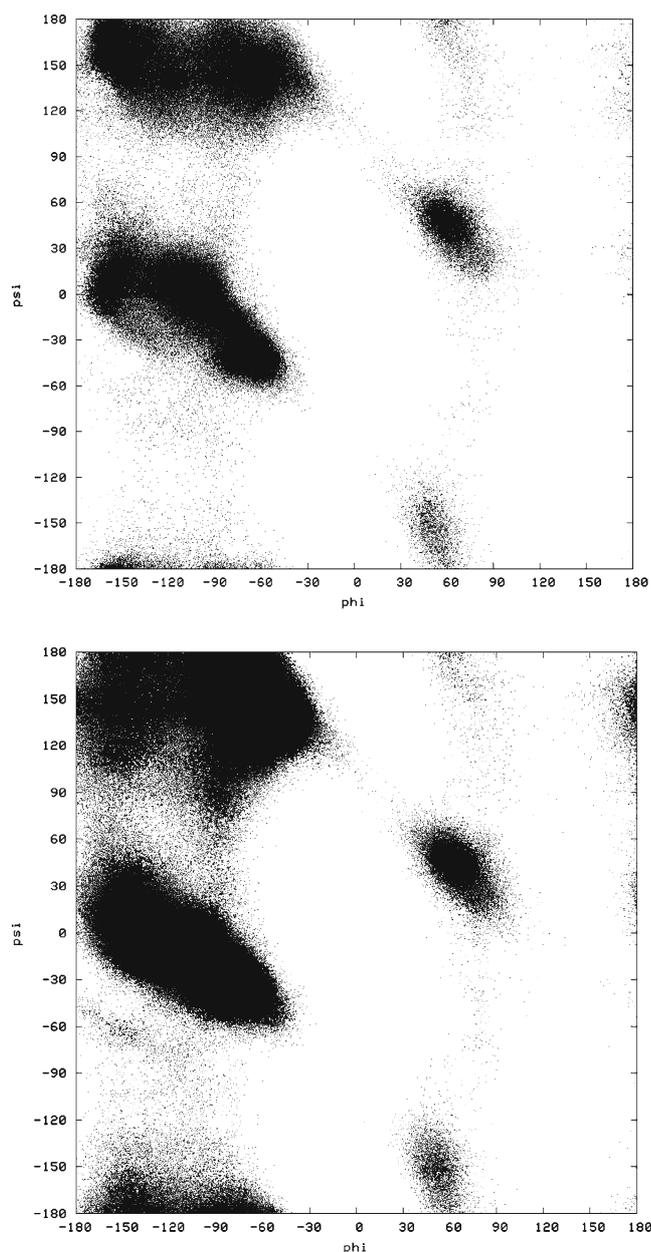
Fig. 5 Conformational sampling of backbone dihedral angles ϕ and ψ in alanine dipeptide from explicit solvent (*top*) and implicit solvent (*bottom*) molecular dynamics simulations with Langevin dynamics and a friction coefficient of 50/ps. The explicit solvent simulation was carried out over 200 ns, the implicit simulation over 500 ns. In both simulations the CHARMM22 force field with the recent CMAP correction ref.10 was employed

The reduced computational cost compared to explicit solvent methods is the foremost selling point of implicit solvent methods. While some early GB implementations are certainly very fast, only requiring a few times the speed of simulations in vacuum, recently proposed GB methods trade speed for improved accuracy. First implementations of the GBMV method [141, 146] required as much as 20 times the speed of vacuum simulation [132]. Although the computational cost can be reduced to about half with careful optimizations

(unpublished results), the question arises whether such expensive GB methods are still competitive compared to the time required for standard explicit solvent simulations. In order to test this point we have obtained timing data for 200 steps of molecular dynamics simulation with both implicit solvent (optimized GBMV) and explicit solvent (optimal rectangular box with 10 Å minimum solvent margin, periodic boundaries, particle mesh Ewald summation, heuristic non-bonded list update) for a large set of proteins with different shapes and sizes. The results are shown in Fig. 6. We find that even the expensive GBMV method is more cost effective in most cases. For small systems it is clear that the implicit solvent is much faster because of the large difference in system size with and without explicit solvent. However, as the size of the biomolecules increases relatively fewer explicit solvent atoms are needed to generate a solvent box with a given margin. As a result explicit solvent simulations for large but compact proteins can become less expensive than the GBMV implicit solvent method used in this test. Figure 6 shows that the cost of GBMV simulations exhibits near-linear scaling as a function of the number of solute residues (or atoms) up to 1,500 residues. Actually, the best fit curve shown suggests $O(N \times \log(N))$ scaling with a small coefficient. Such favorable scaling is a result from using a cutoff in both the GB equation and when calculating Born radii according to Eq. (4). While cutoff calculations will not reproduce electrostatic solvation energies as given from Poisson theory, relative solvation energies (and consequently the forces needed for molecular dynamics simulations) converge with large enough cutoff distances (16–20 Å) (data not shown). Furthermore, non-bonded list updates are needed much less frequently without explicit solvent (about every 50–100 steps) and add relatively little extra cost.

The advantage of implicit solvent is greatest for extended systems, and molecules that have large internal or interstitial water cavities all of which require a much larger number of atoms with explicit solvent. Therefore, implicit solvent is most suited for simulations of protein folding, large-scale dynamic processes, studies of oligomeric assembly and protein-protein association, as well as any studies of very large biomolecular complexes. However, a direct cost-per-time-step comparison does not take into account that implicit solvent methods offer additional advantages over explicit solvent. Since solvent friction is available as a control parameter (and could be turned off entirely), it is possible to generate substantially accelerated dynamics so that relevant thermodynamic states of the biomolecule are sampled much faster than with explicit solvent. Furthermore, implicit solvent as a mean-field theory provides equilibrium solvation free energy estimates without the need for sampling solvent that are particularly useful as components of physically motivated scoring functions.

4 Implicit modeling of membranes

So far we have focused on modeling biomolecules in aqueous solvent, approximated with a single continuum dielectric in

Table 3 Barrier crossing rates in alanine dipeptide

	Explicit solvent	Implicit Solvent		
		$f = 50$/ps	$f = 10$/ps	$f = 5$/ps
$\alpha \rightarrow \beta$	2.53	1.13	3.75	4.84
$\beta \rightarrow \alpha$	2.67	1.69	5.13	6.46
$\alpha_L \rightarrow \beta$	2.73	0.82	1.43	1.43
$\beta \rightarrow \alpha_L$	0.32	0.02	0.06	0.10

Kinetic rates are compared between explicit solvent and implicit solvent simulations of alanine dipeptide. Transitions between the α, β, and α_L basins are considered. Explicit solvent molecular dynamics simulations were carried out with periodic boundaries and particle-mesh Ewald over 200 ns. The implicit solvent simulations employed the GBMV generalized Born model with Langevin dynamics and were carried out over 0.5 μs simulation time each. The friction constant f was varied as indicated in the table. All rates are in ns^{-1}

the implicit solvent model. While the formalism introduced above also allows low-dielectric environments, an extension to heterogeneous environments is less straightforward. While Poisson theory is equally well suited to describe a set of charges in any non-uniform spatial distribution of the dielectric function, the GB approximation in its original form is limited to a single external dielectric. This poses a particular problem for modeling biological membrane environments, which can be approximated as a layered dielectric system from the low-dielectric lipid tail region to the high-dielectric lipid head groups and water (see Fig. 7b). As a first approximation one can maintain a two-dielectric model by extending the solute cavity ($\epsilon = 1$) into the hydrophobic tail region (see Fig. 7a) [176]. Such a model captures the basic idea of a hydrophobic slab in a high-dielectric environment and has been used successfully in a number of test applications [176–178]. Other, similar models for the implicit modeling of membranes have also been proposed [97,179].

A more realistic model of biological membranes would allow weak polarization within the hydrophobic part of the membrane ($\epsilon = 1$–2) but also include a region of intermediate dielectric response ($\epsilon = 4$–8) near the location of the ester groups [180]. In order to implement such a model within the GB formalism we have modified Eq. (2) in the following way [181]:

$$\Delta G^{\text{elec}} = -\frac{1}{2} \sum \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_{ij}} \right)$$
$$\times \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i(\varepsilon_i)\alpha_j(\varepsilon_j) \exp(-r_{ij}^2 / F\alpha_i(\alpha_i)\alpha_j(\varepsilon_j))}}$$
$$(7)$$

where ϵ_{ij} is calculated as the arithmetic mean of effective local dielectric constants ϵ_i and ϵ_j at two interacting charge sites i and j. In the case of a membrane bilayer, the local dielectric constant only depends on z, the direction perpendicular to the membrane layer. The effective value of the dielectric constant can be obtained from the Born equation when the electrostatic solvation energy profile of a probe ion is calculated from Poisson theory for the given set of dielectric layers. The resulting dielectric profile would depend on the choice of dielectric constants and widths of the dielectric layers, but it only needs to be calculated once for a given type of membrane. Interpolated with a spline function, the fixed dielectric profile is then used throughout the
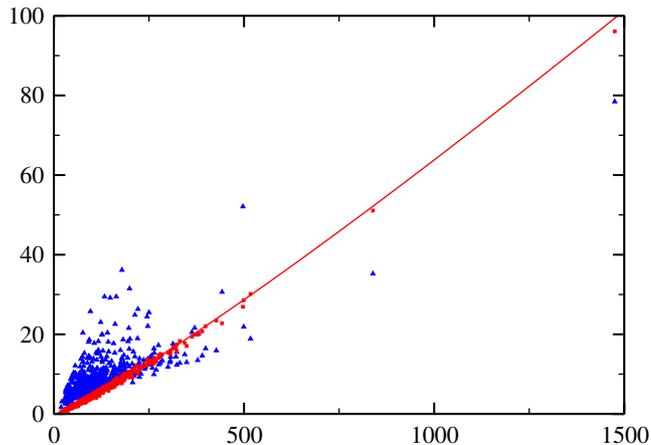

Fig. 6 Computational cost from 200 steps of molecular dynamics with explicit (*blue*) and implicit solvent (*red*) for 611 test proteins used and described previously [132]. In addition timing data for the 1,475 residue MutS–DNA complex [194, ?] was included as well. The implicit solvent simulations were carried out with a recently optimized GBMV implementation. An electrostatic cutoff of 16 Å was used. The non-bonded list was updated heuristically. The integration grid was adjusted to four angular grid points and an increased density in the radial direction. For the explicit solvent simulations the proteins were solvated in a rectangular box that would provide at least a 10 Å water layer to the edge of the box. Particle-mesh Ewald was used to calculate electrostatic interactions. The time in s required for ten steps of molecular dynamics is shown as a function of the number of residues. All timings were performed on a 2.8 GHz Intel Xeon CPU

simulation to assign ϵ_i as a function of z. With such formalism, it is possible to accurately reproduce electrostatic solvation free energies for macromolecular systems in heterogeneous dielectric environments as compared with solutions from Poisson theory [181].

In order to arrive at a complete implicit model of a membrane bilayer, the non-polar component of the solvation free energy plays a crucial role. If only the electrostatic energy is considered, even a hydrophobic molecular system with small partial charges would be relatively more favorable in the high-dielectric environment where the charge–charge interactions are screened more strongly. The non-polar component to the solvation free energy takes into account that the cost of cavity formation nearly vanishes in a non-polar environment compared to the substantial cost of reorienting a polar solvent in order to accommodate the cavity of a biomolecule. This difference in non-polar interactions offsets
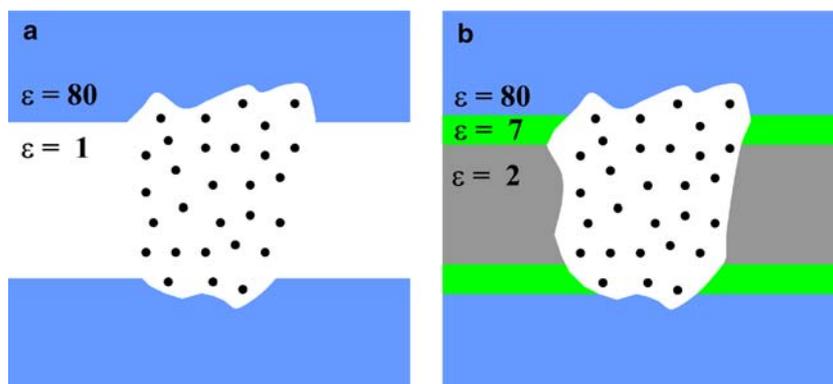
Fig. 7 Schematic illustration of a biomolecule in different implicit membrane environments

the reduced screening of charge–charge interactions in the lipid interior for hydrophobic molecular entitities. We have used the energetic profile of molecular oxygen insertion into explicit lipid membranes [182,183] as the starting point for defining the non-polar component as a linear function of the solvent-accessible surface area with the prefactor γ a function of z according to the profile. The profile was then scaled and shifted so that $\gamma = 0$ at the center of the membrane while reaching $\gamma = 15\,\text{cal/mol/Å}^2$ in aqueous solvent, the same value that we have used in the implicit aqueous solvent simulations shown above.

A realistic implicit model of membrane bilayers is then formed by combining the electrostatic and non-polar contributions. Such a model gives reasonable free energy profiles for the insertion of small amino acid analogs into lipid bilayers in agreement with experimental data and explicit lipid simulations [181], but more exciting is the possibility to carry out simulations of integral membrane proteins without the need for either explicit water or lipids. Figure 8 shows the root mean square deviation from experiment in a simulation of the bacteriorhodopsin trimer with the newly developed implicit membrane model. The results are very similar to explicit lipid/water simulations of the same system [184,185] suggesting that an implicit model can be used successfully for the simulation of such complex environments.

5 Summary and outlook

In this paper we have described new methods for the realistic modeling of biomolecules in implicit solvent. In particular, recent GB methods appear to be very effective in capturing the essential effects of the environment through a continuum dielectric treatment. While implicit solvent simulations of proteins and peptides in aqueous solvent have been described before, we are demonstrating here that implicit solvent treatments can also be extended successfully to more challenging systems such as nucleic acids and integral membrane proteins. These developments are exciting as they ultimately pave the road for the simulation of much more complex molecular systems in full atomic detail that are otherwise
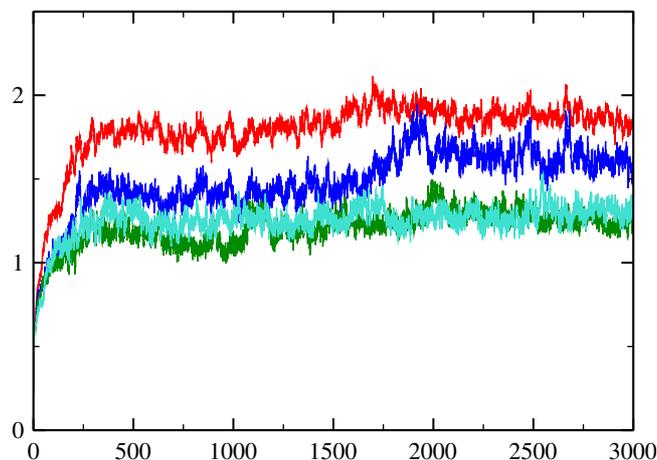

Fig. 8 $C\alpha$ root mean square deviation in Å as a function of time in ps for molecular dynamics simulation of the bacteriorhodopsin trimer including three retinal copies with respect to the experimental structure. The CHARMM22 force field [192] was used for the protein, retinal parameters were obtained from Jan Saam and Klaus Schulten. Implicit solvent simulations were carried out with a new implicit membrane model based on the GBMV method [181]. The RMSD value for the entire system is shown in *red*, individual RMSD values for the monomers are shown in *blue*, *green*, and *cyan*

considered inaccessible to conventional explicit solvent simulations.

Current work is now aimed at more extensive comparisons with explicit solvent and ultimately experimental data in order to show more clearly the strengths and limits of continuum dielectric models. In parallel, we are working on new hybrid methods that allow the incorporation of selected, fully mobile solvent molecules so that specific interactions with the environment can be considered when needed. A related direction is the incorporation of ionic effects. A scheme for the incorporation of bulk salt concentrations through an additional screening term has been proposed previously [186]. It is not clear, however, to what extent such a term is sufficient for including the additional screening due to salt in a realistic manner.

Another aspect that needs to be understood better is the modeling of kinetic rates with implicit solvent. While implicit

solvent offers the opportunity to accelerate conformational sampling through reduced friction, it is unclear how to interpret such dynamics. This issue becomes even more difficult in the context of heterogeneous environments where friction rates may vary spatially. While we think that all of these issues can be addressed in the near future, we hope to arrive at a robust method that offers the same level of realism as explicit solvent simulation but includes only as many explicit solvent molecules as needed, while treating the environment in a continuous fashion otherwise.

The primary motivation for using implicit solvent is of course the promise that a reduced number of particles translates into reduced computational cost and therefore allows the simulation of longer timescales and larger systems. We have tested whether this is in fact the case by comparing a recently optimized version of the GBMV method with conventional explicit solvent simulations. Somewhat surprisingly, we did not find as much of an advantage for the implicit solvent simulations as we would have hoped for, despite nearly linear scaling of the implicit solvent method as a function of the number of residues. The performance of GBMV, which is arguably the most accurate GB approximation to Poisson theory at this time [132, 172], can serve as the upper limit for what may be expected with other GB implementations. A number of alternative GB methods are available [86, 132], that offer better performance at a reduced level of agreement with Poisson theory but would be faster than explicit solvent even for very large systems. However, while further improvements in the efficiency of implicit solvent methods may be possible, the main advantage of implicit solvent methods lies in the fact that interactions with the solvent are represented as a mean-field response. Statistical sampling of a given biomolecule in implicit solvent thereby requires only conformational sampling of the molecule itself, while in explicit solvent simulations both the solute and solvent need to be sufficiently sampled in order to obtain meaningful averages. Therefore, implicit solvent simulations should be able to reach much longer timescales than with explicit solvent even if the cost per time step is similar [187]. An understanding of how to accelerate dynamics in implicit solvent in a controlled manner that allows the recovery of meaningful kinetic rates will be the key to establishing a highly efficient and realistic simulation methodology for the study of dynamic processes in large biomolecular complexes in atomic detail.

References

1. York DM, Yang W, Lee H, Darden TA, Pedersen L (1995) J Am Chem Soc 117:5001–5002
2. Smith PE, Pettitt BM (1995) Comput Phys Commun 91:339–344
3. Luty BA, Davis ME, Tironi IG, Gunsteren WF (1994) Mol Simul 14:11
4. Cheatham TE, Miller JL, Fox T, Darden TA, Kollman PA (1995) J Am Chem Soc 117:4193–4194
5. Straub JE, Whitfield T, Andricioaei I (2002) Abstr Pap Am Chem Soc 223:U470–U470
6. Okamoto Y (2004) J Mol Graph Model 22:425–439
7. Hansmann UHE (1997) Chem Phys Lett 281:140–150
8. Hornak V, Simmerling C (2004) J Mol Graph Model 22:405–413
9. Micheletti C, Laio A, Parrinello M (2004) Phys Rev Lett 92:170601
10. MacKerell AD Jr, Feig M, Brooks CL III (2004) J Comput Chem 25:1400–1415
11. MacKerell AD Jr, Feig M, Brooks CL III (2004) J Am Chem Soc 126:698–699
12. Feig M, MacKerell AD Jr, Brooks CL III (2003) J Phys Chem B 107:2831–2836
13. Patel S, Brooks CL (2004) J Comput Chem 25:1–15
14. MacKerell AD Jr, Banavali NK, Foloppe N (2001) Biopolymers 56:257–265
15. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF (2004) J Comput Chem 25:1656–1676
16. Mackerell AD (2004) J Comput Chem 25:1584–1604
17. Patel S, Mackerell AD, Brooks CL (2004) J Comput Chem 25:1504–1514
18. Patel S, Brooks CL (2005) J Chem Phys 122:24508
19. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) J Comput Chem 25:1157–1174
20. Gao J, Truhlar DG (2002) Annu Rev Phys Chem 53:267–505
21. Li GH, Cui Q (2003) J Phys Chem B 107:14521–14528
22. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996) J Phys Chem 100:19357–19363
23. Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu K Y, Garcia AE (2003) Curr Opin Struct Biol 13:168–174
24. Pande VS, Rokhsar DS (1996) Proc Nat Acad Sci USA 96:9062–9067
25. Mayor U, Johnson CM, Daggett V, Fersht AR (2000) Proc Nat Acad Sci USA 97:13518–13522
26. Bozcko EM, Brooks CL III (1995) Science 269:393–396
27. Pitera JW, Swope WC (2003) Proc Nat Acad Sci USA 100:7587–7592
28. Simmerling C, Strockbine B, Roitberg AE (2002) J Am Chem Soc 124:11258–11259
29. Okur A, Roe DR, Cui GL, Hornak V, Simmerling C (2003) Abstr Pap Am Chem Soc 226:U450–U451
30. Zagrovic B, Sorin E, Pande V (2002) Biophys J 82:474a–474a
31. Brooks CL (2002) Acc Chem Res 35:447–454
32. Brooks CL (1998) Curr Opin Struct Biol 8:222–226
33. Andricioaei I, Goel A, Herschbach D, Karplus M (2004) Biophys J 87:1478–1497
34. Florian J, Goodman MF, Warshel A (2003) Biopolymers 68:286–299
35. Rittenhouse RC, Apostoluk WK, Miller JH, Straatsma TP (2003) Proteins Struct Funct Genet 53:667–682
36. Sanbonmatsu KY, Joseph S (2003) J Mol Biol 328:33–47
37. Trylska J, Konecny R, Tama F, Brooks CL, McCammon JA (2004) Biopolymers 74:423–431
38. Ramakrishnan V, Moore PB (2001) Curr Opin Struct Biol 11:144–154
39. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) Science 289:905–920
40. Westover KD, Bushnell DA, Kornberg RD (2004) Science 303:1014–1016
41. Bushnell DA, Kornberg RD (2003) Proc Nat Acad Sci USA 100:6969–6973
42. Gnatt AL, Cramer P, Fu JH, Bushnell DA, Kornberg RD (2001) Science 292:1876–1882
43. Cramer P, Bushnell DA, Kornberg RD (2001) Science 292:1863–1876
44. Zhirnov VV, Cavin RK III, Hutchby JA, Bourianoff GI (2003) Proc IEEE 91:1934–1939

45. Schlick T, Barth E, Mandziuk M (1997) Ann Rev Biophys Biomol Struct 26:179–220
46. Zhou RH, Harder E, Xu HF, Berne BJ (2001) J Chem Phys 115:2348–2358
47. Barth E, Schlick T (1998) J Chem Phys 109:1633–1642
48. Gront D, Kolinski A, Skolnick J (2001) J Chem Phys 115:1569–1574
49. Rao F, Caflisch A (2003) J Chem Phys 119:4035–4042
50. Rhee YM, Pande VS (2003) Biophys J 84:775–786
51. Sanbonmatsu KY, Garcia AE (2002) Proteins 46:225–234
52. Sugita Y, Okamoto Y (1999) Chem Phys Lett 314:141–151
53. Suenaga A (2003) J Mol Struct (Theochem) 634:235–241
54. Zhou R, Berne BJ, Germain R (2001) Proc Nat Acad Sci USA 98:14931–14936
55. Garcia AE, Sanbonmatsu KY (2002) Proc Nat Acad Sci USA 99:2782–2787
56. Garcia AE, Sanbonmatsu KY (2001) Proteins 42:345–354
57. Beutler TC, Gunsteren WF (1993) J Chem Phys 100:1492–1497
58. Fraternali F, Gunsteren WF (1994) Biopolymers 34:347–355
59. Schmidt RK, Teo B, Brady JW (1995) J Phys Chem 99:11339–11343
60. Torrie GM, Valleau JP (1974) Chem Phys Lett 28:578–581
61. Hansmann UHE, Okamoto Y (1999) Curr Opin Struct Biol 9:177–183
62. Nakajima N, Nakamura H, Kidera A (1997) J Phys Chem 101:817–824
63. Lin C-Y, Hu C-K, Hansmann UHE (2003) Proteins 52:436–445
64. Hansmann UHE (2002) Comput Phys Commun 147:604–607
65. Guo ZY, Thirumalai D, Honeycutt JD (1992) J Chem Phys 97:525–535
66. Camacho CJ, Thirumalai D (1993) Proc Nat Acad Sci USA 90:6369–6372
67. Dill KA (1985) Biochemistry 24:1501–1509
68. Dill KA, Fiebig KM, Chan HS (1993) Proc Nat Acad Sci USA 90:1942–1946
69. Dill KA, Chan HS, Sun SJ, Yue K (1995) Faseb J 9:A1240–A1240
70. Yue K, Fiebig KM, Thomas PD, Chan HS, Shakhnovich EI, Dill KA (1995) Proceedings of the National Academy of Sciences of the United States of America 92:325–329
71. Kolinski A, Skolnick J, Yaris R (1986) Proc Nat Acad Sci USA 83:7267–7271
72. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA (1997) J Comput Chem 18:849–873
73. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Oldziej S, Scheraga HA (1997) J Comput Chem 18:874–887
74. Kolinski A (2004) Acta Biochimica Pol 51:349–371
75. Kolinski A, Skolnick J (1994) Proteins 18:338–352
76. Skolnick J, Kolinski A, Ortiz AR (1997) J Mol Biol 265:217–241
77. Hinds DA, Levitt M (1992) Proc Nat Acad Sci USA 89:2536–2540
78. Malhotra A, Tan RK-Z, Harvey SC (1994) Biophyisics J 66:1777–1795
79. Langowski J, Olson WK, Yang Y (1996) Trends Biochem Sci 21:50
80. Fenley MO, Olson WK, Tobias I, Manning GS (1994) Biophys Chem 50:255–271
81. Feig M, Rotkiewicz P, Kolinski A, Skolnick J, Brooks CLI (2000) Proteins 41:86–97
82. Milik M, Kolinski A, Skolnick J (1997) J Comput Chem 18:80–85
83. Payne PW (1993) Protein Sci 2:315–324
84. Rey A, Skolnick J (1992) J Comput Chem 13:443–456
85. Feig M, Karanicolas J, Brooks CL III (2004) J Mol Graph Model 22:377–395
86. Feig M, Brooks CL III (2004) Curr Opin Struct Biol 14:217–224
87. Cramer CJ, Truhlar DG (1999) Chem Rev 99:2161–2200
88. Lazaridis T, Karplus M (2000) Curr Opin Struct Biol 10:139–145
89. Hao M-H, Scheraga HA (1999) Curr Opin Struct Biol 9:184–188
90. Hou T, Qiao X, Zhang W, Xu X (2002) J Phys Chem B 106:11295–11304
91. Wesson L, Eisenberg D (1992) Protein Sci 1:227–235
92. Zhou H, Zhou Y (2002) Proteins 49:483–492
93. Ferrara P, Apostolakis J, Caflish A (2002) Proteins 46:24–33
94. Ooi T, Obatake M, Nemethy G, Scheraga HA (1987) Proc Nat Acad Sci USA 84
95. Mallik B, Masunov A, Lazaridis T (2002) J Comput Chem 23:1090–1099
96. Lazaridis T, Karplus M (1999) Proteins 35:133–152
97. Lazaridis T (2003) Proteins 52:176–192
98. Lazaridis T, Karplus M (1998) J Mol Biol 288:477–487
99. Forrest LR, Woolf TB (2003) Proteins 52:492–509
100. Feig M, Brooks CL III (2002) Proteins 49:232–245
101. Sitkoff D, Sharp KA, Honig B (1994) J Phys Chem 98:1978–1988
102. Hassan SA, Mehler EL (2002) Proteins 47:45–61
103. Sharp KA, Honig B (1990) Ann Rev Biophys Biophys Chem 19:301–332
104. Luque FJ, Alhambra C, Orozco M (1995) J Phys Chem 99:11344–11349
105. Hecht JL, Honig B, Shin YK, Hubbell WL (1995) J Phys Chem 99:7782–7786
106. Misra VK, Honig B (1996) Biochemistry 35:1115–1124
107. Srinivasan J, Miller J, Kollman PA, Case DA (1998) J Biomol Struct Dyn 16:671–682
108. Srinivasan J, Cheatham TE, Cieplak P, Kollman P A, Case DA (1998) J Am Chem Soc 120:9401–9409
109. Levy RM, Zhang LY, Gallicchio E, Felts AK (2003) J Am Chem Soc 125:9523–9530
110. Gallicchio E, Levy RM (2004) J Comput Chem 25:479–499
111. Hawkins GD, Cramer CJ, Truhlar DG (1996) J Phys Chem 100:19824–19839
112. Floris F, Tomasi J (1989) J Comput Chem 10:616–627
113. Felts AK, Gallicchio E, Wallqvist A, Levy RM (2002) Proteins 48:404–422
114. Lee MR, Duan Y, Kollman PA (2000) Proteins 39:309–316
115. Lee MR, Kollman PA (2001) Structure 9:905–916
116. Dominy BN, Brooks CL III (2001) J Comput Chem 23:147–160
117. Rapp CS, Friesner RA (1999) Proteins 35:173–183
118. Lee MC, Duan Y (2004) Proteins 55:620–634
119. Gohlke H, Kiel C, Case DA (2003) J Mol Biol 330:891–913
120. Gohlke H, Case DA (2003) J Comput Chem 25:238–250
121. Mardis KL, Luo R, Gilson MK (2001) J Mol Biol 309:507–517
122. Zhang LY, Gallicchio E, Friesner RA, Levy RM (2001) J Comput Chem 22:591–607
123. Zhou R, Friesner RA, Ghosh A, Rizzo RC, Jorgensen WL, Levy RM (2001) J Phys Chem B 105:10388–10397
124. Wang JM, Morin P, Wang W, Kollman PA (2001) J Am Chem Soc 123:5221–5230
125. Jang S, Shin S, Pak Y (2002) J Am Chem Soc 124:4976–4977
126. Krol M (2003) J Comput Chem 24:531–546
127. Zagrovic B, Sorin EJ, Pande V (2001) J Mol Biol 313:151–169
128. Luo R, David L, Gilson MK (2002) J Comput Chem 23:1244–1253
129. Zhou R (2003) Proteins 53:148–161
130. Nymeyer H, Garcia AE (2003) Proc Nat Acad Sci USA 100:13934–13939
131. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL III (2004) J Comput Chem 25:265–284
132. Bashford D, Case DA (2000) Annu Rev Phys Chem 51:129–152
133. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) J Am Chem Soc 112:6127–6129
134. Born M (1920) Zeitschrift für Physik 1:45–48
135. Constanciel R, Contreras R (1984) Theor Chim Acta 65:1–11
136. Onufriev A, Case DA, Bashford D (2002) J Comput Chem 23:1297–1304
137. Hawkins GD, Cramer CJ, Truhlar DG (1995) Chem Phys Lett 246:122–129
138. Qiu D, Shenkin PS, Hollinger FP, Still WC (1997) J Phys Chem A 101:3005–3014
139. Schaefer M, Karplus M (1996) J Phys Chem 100:1578–1599
140. Lee MS, Salsbury FR Jr, Brooks CL III (2002) J Chem Phys 116:10606–10614
141. Liotard DA, Hawkins GD, Lynch GC, Cramer CJ, Truhlar DG (1995) J Comput Chem 16:422–440

142. Ghosh A, Rapp CS, Friesner RA (1998) J Phys Chem B 102:10983–10990
143. Feig M, Im W, Brooks CL III (2004) J Chem Phys 120
144. Grycuk T (2003) J Chem Phys 119:4817–4826
145. Lee MS, Feig M, Salsbury FR Jr, Brooks CL III (2003) J Comput Chem 24:1348–1356
146. Kirkwood JG (1934) J Chem Phys 2:351–361
147. Lee B, Richards FM (1971) J Mol Biol 55:379
148. Honig B, Nicholls A (1995) Science 268:1144–1149
149. Connolly ML (1983) J Appl Crystallogr 16:548–558
150. Im W, Lee MS, Brooks CL III (2003) J Comput Chem 24:1691–1702
151. Nina M, Beglov D, Roux B (1997) J Phys Chem 101:5239–5248
152. Banavali NK, Roux B (2002) J Phys Chem B 106:11026–11035
153. Zhang W, Hou T, Qiao X, Xu X (2003) J Phys Chem B 107:9071–9078
154. Zhu J, Shi Y, Liu H (2002) J Phys Chem B 106:4844–4853
155. Dominy BN, Brooks CL III (1999) J Phys Chem B 103:3765–3773
156. Calimet N, Schaefer M, Simonson T (2001) Proteins 45:144–158
157. Tsui V, Case DA (2000) J Am Chem Soc 122:2489–2498
158. Lu BZ, Chen WZ, Wang CX, Xu X-j (2002) Proteins 48:497–504
159. Sharp K (1991) J Comput Chem 12:454–468
160. Fogolari F, Brigo A, Molinari H (2003) Biophys J 85:159–166
161. Prabhu NV, Zhu PJ, Sharp KA (2004) J Comput Chem 25:2049–2064
162. Feig M, Pettitt BM (1998) Biophys J 75:134–149
163. Feig M, Pettitt BM (1997) J Phys Chem B 101:7361–7363
164. Bursulaya BD, Brooks CL III (2000) J Phys Chem B 104:12378–12383
165. Feig M, Pettitt BM (1999) J Mol Biol 286:1075–1095
166. Berman HM (1994) Curr Opin Struct Biol 4:345–350
167. Makarov V, Pettitt BM, Feig M (2002) Acc Chem Res 35:376–384
168. Pal SK, Zewail AH (2004) Chem Rev 104:2099–2123
169. Lee MS, Salsbury FR, Olson MA (2004) J Comput Chem 25:1967–1978
170. Lin J-H, Baker NA, McCammon JA (2002) Biophys J 83:1374–1379
171. Zhu J, Alexov E, Honig B (2005) J Phys Chem B 109:3008–3022
172. Adelman SA, Brooks CL (1982) J Phys Chem 86:1511–1524
173. Zagrovic B, Pande V (2003) J Comput Chem 24:1432–1436
174. Shen M-y, Freed KF (2002) Biophys J 82:1791–1808
175. Im W, Feig M, Brooks CL III (2003) Biophys J 85:2900–2918
176. Im W, Brooks CL (2004) J Mol Biol 337:513–519
177. Spassov VZ, Yan L, Szalma S (2002) J Phys Chem B 106:8726–8738
178. Chambers CC, Giese DJ, Hawkins GD, Vaes WH J, Cramer CJ, Truhlar DG (1999) In: Truhlar DG, Howe WJ, Hopfinger AJ, Blaney JM, Dammkoehler RA (eds) Rational drug design. Springer, Berlin Heidelberg New York, pp 51–72
179. Stern HA, Feller SE (2003) J Chem Phys 118:3401–3412
180. Tanizaki S, Feig M (2005) J Chem Phys 122:124706
181. Marrink SJ, Berendsen HJC (1996) J Phys Chem 100:16729–16738
182. Marrink SJ, Berendsen HJC (1994) J Phys Chem 98:4155–4168
183. Kandt C, Schlitter J, Gerwert K (2004) Biophys J 86:705–717
184. Baudry J, Tajkhorshid E, Molnar F, Phillips J, Schulten K (2001) J Phys Chem B 105:905–918
185. Srinivasan J, Trevathan MW, Beroza P, Case DA (1999) Theor Chim Acta 101:426–434
186. Cheng XL, Hornak V, Simmerling C (2004) J Phys Chem B 108:426–437
187. Im W, Beglov D, Roux B (1998) Comput Phys Commun 111:59–75
188. Roux B (1997) Biophys J 73:2980–2989
189. Dickerson RE (1998) Nucl Acids Res 26:1906–1926
190. MacKerell AD Jr, Banavali NK (2000) J Comput Chem 21:105–120
191. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack JD, Evanseck MJ, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau F TK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) J Phys Chem B 102:3586–3616
192. Foloppe N, MacKerell AD Jr (2000) J Comput Chem 21:86–104
193. Obmolova G, Ban C, Hsieh P, Yang, W (2000) Nature 407:703–710
194. Lamers MH, Perrakis A, Enzlin JH, Winterwerp H HK, de Wind N, Sixma TK (2000) Nature 407:711–717