J Mol Model (2013) 19:3647–3657
DOI 10.1007/s00894-013-1879-8

ORIGINAL PAPER

A coupled two-dimensional main chain torsional potential for protein dynamics: generation and implementation

Yongxiu Li · Ya Gao · Xuqiang Zhang · Xingyu Wang · Lirong Mou · LiLi Duan · Xiao He · Ye Mei · John Z. H. Zhang

Received: 29 March 2013 / Accepted: 1 May 2013 / Published online: 14 June 2013

Abstract Main chain torsions of alanine dipeptide are parameterized into coupled 2-dimensional Fourier expansions based on quantum mechanical (QM) calculations at M06 2X/aug-cc-pvtz//HF/6-31G** level. Solvation effect is considered by employing polarizable continuum model. Utilization of the M06 2X functional leads to precise potential energy surface that is comparable to or even better than MP2 level, but with much less computational demand. Parameterization of the 2D expansions is against the full main chain torsion space instead of just a few low energy conformations. This procedure is similar to that for the development of AMBER03 force field, except unique weighting factor was assigned to all the grid points. To avoid inconsistency between quantum mechanical calculations and molecular modeling, the model peptide is further optimized at molecular mechanics level with main chain dihedral angles fixed before the calculation of the conformational energy on molecular mechanical level at each grid point, during which generalized Born model is employed. Difference in solvation models at quantum mechanics and molecular mechanics levels makes this parameterization procedure less straightforward. All force field parameters other than main chain torsions are taken from existing AMBER force field. With this new main chain torsion terms, we have studied the main chain dihedral distributions of ALA dipeptide and pentapeptide in aqueous solution. The results demonstrate that 2D main chain torsion is effective in delineating the energy variation associated with rotations along main chain dihedrals. This work is an implication for the necessity of more accurate description of main chain torsions in the future development of ab initio force field and it also raises a challenge to the development of quantum mechanical methods, especially the quantum mechanical solvation models.

Keywords Alanine oligopeptide · Density functional theory · Free energy landscape · Force field · J coupling · Solvation effect

Yongxiu Li and Ya Gao contributed equally to this work

Y. Li · Y. Gao · X. He (✉) · Y. Mei (✉) · J. Z. H. Zhang (✉)
Center for Laser and Computational Biophysics, State Key Laboratory of Precision Spectroscopy and Department of Physics and Institute of Theoretical and Computational Science, East China Normal University, Shanghai 200062, China
e-mail: xhe@phy.ecnu.edu.cn
e-mail: ymei@phy.ecnu.edu.cn
e-mail: john.zhang@nyu.edu

X. Zhang · X. Wang · J. Z. H. Zhang
Department of Chemistry, New York University, New York, NY 10003, USA

L. Mou
Institutes for Advanced Interdisciplinary Research, East China Normal University, Shanghai 200062, China

L. Duan
College of Physics and Electronics, Shandong Normal University, Jinan 250014, China

Introduction

High level quantum mechanical (QM) methods have been proved to be quite successful in the study of structures and energies of small molecules. However, QM calculations of biological molecules such as protein, DNA, RNA are hindered by the extremely large CPU time and storage requirements. Parameterization of the QM data into classical energy components for model systems and extending to macromolecules afterward are more practical for contemporary in silico study of biological systems. Because the energy and its derivative with respect to the coordinates are generally evaluated millions of times or even more in one simulation, the functional forms of the energy are relatively simple to keep the computational expense as low as possible. Currently, force fields like AMBER [1], CHARMM [2], OPLS [3], GROMOS [4] etc. have been widely used in the computational biology

community. In these force fields, the total energy from high level ab initio calculations is decomposed into various bonded (bond, angle, torsion, etc.) and nonbonded (Coulomb, van der Waals) terms, as proposed first by Lifson [5]. Further calibration of force field parameters in accordance with experimental measurement may also be applied [6–9]. For recent reviews of current trend and philosophy under the force field development please refer to refs [10–12].

Utilizing inappropriate parameters, especially imprecise main chain torsion parameters leading to inaccurate secondary structure propensity, is a likely origin of the failure in some protein folding simulations [13]. During the parameterization of AMBER force field, the main chain torsion terms serve to account for the "high order" terms not present explicitly. It is informally called the garbage terms. However this is unfair, considering its profound impact on the quality of a force field. The deficiency in the parameters of main chain torsions may accumulate along the peptide chain, and leads to large deviations of the structure. For instance, obsolete AMBER94 and its refined version AMBER99 have been proved to bias the helix structure [14, 15], while AMBER96 overcorrected this problem and it overpopulates the β strand conformation [16–19]. AMBER03 force field also biases helix [20], and AMBER99SB is believed to disfavor helical conformation [7]. Thus, refinement of the main chain torsion potential has been the focus of recent developments in protein force fields [6–8, 15, 21–30]. In the widely used force fields, the main chain torsional terms are treated individually for ϕ ($C-N-C_\alpha-C$) and ψ ($N-C_\alpha-C-N$), shown in Fig. 1, as Fourier series. Due to the limited number of parameters in the simple functional form, parameterization mainly focuses on a small portion of space around the energy basins. This leads to less accuracy for the inter-basin barrier height, which is deleterious to the study of large scale conformational change like protein folding and loop dynamics, as well as the calculation of conformational entropy [23, 31, 32]. Including ϕ/ψ crossing terms has been attempted by Mackerell et al., but it led to overly populated π-helix, although the accuracy has been significantly improved. And finally it was replaced by a grid-based energy correction map (CMAP) [23, 24, 33], but still too helical according to a recent study [13]. Since these two main chain torsions are not separable in the potential energy map, the torsion energy function should be expanded in an orthogonal basis set of both ϕ and ψ. With the advancing of new computer technology, the computational expense is less and less of a bottleneck than it was. Therefore, it is worth trading efficiency a little bit for accuracy. This idea has been applied by Okamoto et al. in their ad hoc force field refinement work [26, 34]. In this work, the new force field is parameterized by quantum mechanical calculations with respect to the existing AMBER force fields. No further optimization according to the experimental observations is conducted. By implementation of this new torsional term in molecular dynamics package AMBER11 [35], the conformation distributions of alanine oligopeptides are studied.

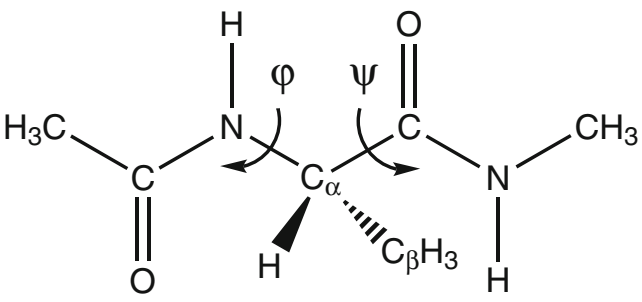

Fig. 1 Main chain torsions in alanine dipeptide

Methods

Quantum mechanical and molecular mechanical calculations

The alanine dipeptide (AD), shown in Fig. 1, was chosen as the model system for parameterization. The structures of AD were optimized at HF/6-31G** level on 24×24 grid points, with both ϕ and ψ fixed at a series of designated values in the range from -180° to 180°. Møller-Plesset (MP2) [36] level and CCSD(T) [37] with large basis set and polarizable continuum model (PCM) are too demanding for such study. While with appropriate parameterization, density functional theory has been proved to be both efficient and accurate in electronic structure calculation. Therefore energy calculations were carried out at M06 2X/aug-cc-pvtz level [38, 39]. Although the reliability of M06 2X functional has been examined in several studies, we also investigated its performance by comparing to MP2. A smaller basis set (aug-cc-pvdz) was employed when comparing M06 2X and MP2. Solvation effect was modeled by the integral equation formalism variant of the PCM (IEFPCM) [40, 41] in both the optimization and single point calculations. The dielectric constant of the solvent was set to 80 and default radii of solute atoms were utilized [41]. Employing solvation effect in the structure optimization is very important, without which the conformations with any of the intramolecular main chain hydrogen bonds, especially the five-membered ring (C5), will be overpopulated. The free energy of AD can be written as

$$E_{QM} = E_{int} + G_{PCM} = < \rho(f)|H|\rho(f) > + < \rho(f)\left|V\middle/2\right|\rho(f) >, \quad (1)$$

in which H is the Hamiltonian of AD in gas phase, V is the solute-solvent interaction, $\rho(f)$ is the eigenfunction of Hamiltonian $H+V$, and $<\rho(f)|V/2|\rho(f)>$ is the solvation free

energy. All the quantum mechanical calculations were carried out by Gaussian 09 [42].

On molecular mechanical (MM) level, the free energy of AD is decomposed into various bonded and nonbonded terms as

$$E_{MM} = E_{bond} + E_{angle} + E_{dihedral} + E_{vdw} + E_{ele} + G_{GB}, \quad (2)$$

where E_{bond}, E_{angle}, $E_{dihedral}$, E_{vdw} and E_{ele} are respectively the bond stretching, angle bending, torsion, van der Waals and Coulomb interaction energies. G_{GB} is the solvation free energy by employing the Hawkins, Cramer, Truhlar pairwise generalized Born model (GB1) [43]. Nonpolar contribution was included in both PCM and GB free energies. We can rewrite Eq. 2 as

$$E_{MM} = E_{mct} + E_{oth} + G_{GB}, \quad (3)$$

in which E_{mct} and E_{oth} are the main chain torsion related energy terms and other internal contributions respectively. By equalizing the MM and QM energies, the main chain torsion term can be calculated as

$$E_{mct} = E_{int} - E_{oth} + \alpha(G_{PCM} - G_{GB}), \quad (4)$$

where α can be either 0 or 1 to exclude or include the difference of solvation free energies. Including solvation effect in parameter fitting has also been conducted by Cao et al. [44]. Because the PCM and GB descriptions of solvation effect are based on different formalisms, the difference in solvation energies might be nonnegligible. To save computational expense, QM and MM calculations were carried out only on 24×24 grid points with 15° interval. We further refined the resolution of this potential energy surface by 2-dimensional spline interpolation with periodic boundary conditions [45].

Another discrepancy between QM and MM calculations was that the QM optimized structures with certain main chain torsions were not necessarily the lowest energy conformations under MM potential. For some energetically unfavorable conformations (e.g., $\{\phi,\psi\}=\{-180.0°, 0.0°\}$), the QM optimized structures had the four atoms around amide bond (H-N-C-O in the atom nomenclature of Protein Data Bank) deviated from plane by about 20°. These conformations exerted a strong stress on this torsion under MM potential, and they are scarcely observed in molecular dynamics (MD) simulations. In other words, the quantum mechanically (molecular mechanically) optimized structures were not fully relaxed under the Hamiltonian of molecular mechanics (quantum mechanics). And this discrepancy may have profound impact on the fitted 2D torsional potential.

2D torsion terms

In a general practice of additive force field, the potential energy is decomposed into bond, angle, torsion, electrostatic and van der Waals terms as in Eq. 2. The torsion energy in AMBER force field is expressed as a series of Fourier expansions as

$$E_{torsion} = \sum_{dihedral} (V_n/2)(1 + cos[n\theta - \delta]), \quad (5)$$

in which V_n, n and δ are the depth of the potential energy well, periodicity and the phase angle respectively. Usually, these expansions are truncated to 3 or 4 terms, which leads to insufficient parameters that can give an accurate description all over the main chain torsion space. It is practical to focus more on the low energy conformations. Toward this end, during the fitting of the parameters for torsion energy in AMBER03 force field [1], unequal weights were used for different (ϕ,ψ), the lower the potential energy the larger the weight. Therefore, this expansion does not give an equally accurate description of potential energies everywhere in the (ϕ, ψ) map, which may cast uncertainty in the study of conformation change.

A more precise way is to discretize the main chain torsion energy map with a double Fourier series, and the expansion coefficients can be shown in matrix form as

$$C(m,n) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} d\phi \int_{-\pi}^{\pi} d\psi E(\phi,\psi) e^{im\phi} e^{in\psi}, \quad (6)$$

in which $E(\phi,\psi)$ are the potential energy map ascribed to main chain torsions. The integral is calculated on a grid with (ϕ,ψ) in the range of $[-\pi,\pi)$ in radian using the trapezoidal rule. All grid points are equally weighted, which is different from the fitting of AMBER03 force field parameters [1]. And for any (ϕ,ψ) occurred during the MD simulation, the potential energy and the force over atom k can be calculated respectively as

$$E(\phi,\psi) = \sum_{m=-N_\phi}^{N_\phi} \sum_{n=-N_\psi}^{N_\psi} C(m,n) e^{-im\phi} e^{-in\psi}, \quad (7)$$

and

$$F_k(\phi,\psi) = -\frac{\partial E(\phi,\psi)}{\partial \mathbf{R}_k} = -\frac{\partial E(\phi,\psi)}{\partial \phi}\frac{\partial \phi}{\partial \mathbf{R}_k} - \frac{\partial E(\phi,\psi)}{\partial \psi}\frac{\partial \psi}{\partial \mathbf{R}_k}. \quad (8)$$

The original code for the calculation of the first derivative of E with respect to the atomic coordinates in sander module of AMBER11 utilizes the chain rule through torsion angle cosine, in which the calculation of $\partial\phi/\partial cos\phi$ is required. It might encounter mathematical difficulty when the four atoms determining ϕ or ψ are in the same plane. Also, it requires that the phase angles δ in Eq. 5 must be either 0 or π in radian. Therefore, in AMBER force field extra torsions related to ϕ and ψ must be defined to mimic the shift of the minima (see below). To avoid the possible singularity in calculating the force as in Eq. 8, we used the approach

developed by Blondel and Karplus [46], which is based on derivatives of the torsion angle itself.

It can be easily found from Fig. 1 that there are some dihedrals highly correlated to ϕ and ψ. We denote them as ϕ' for torsions ($C-N-C_\alpha-C_\beta$), ψ' for torsions ($C_\beta-C_\alpha-C-N$), ϕ'' for other torsions around $N-C_\alpha$ bond, and ψ'' for other torsions around $C_\alpha-C$ bond. These dihedral angles are very similar to ϕ and ψ, but with a phase shift of $1\pi/3$ or $2\pi/3$. We removed these terms from the original AMBER force fields, and their contributions were merged into 2D torsion terms. This new force field has been implemented into sander and pmemd modules in AMBER 11.

Molecular dynamics simulations

In order to accelerate the sampling in the phase space, replica exchange molecular dynamics [47] simulations were carried out for ALA dipeptide (ACE-ALA-NME). The initial structures were generated by LEaP module in AmberTools 1.5 from sequence command. Generalized Born models (denoted as GB1 [43], GB5 [48] and GB7 [49] in AMBER suite package) and TIP3P water model [50] were utilized to mimic the solvation effect. Therefore the dependence on the solvent models can be investigated. Integral time step was set to 1 fs. Temperature was regulated using Langevin dynamics with the collision frequency of 1 ps^{-1}. Swaps were attempted every 0.25 ps. MD simulations were extended to 240 ns for each replica in implicit solvent models and to 80 ns in TIP3P water box, which were long enough to show convergence. Snapshots were saved every 0.25 ps in implicit solvent models and every 5 ps in explicit solvent. Only REMD simulations in explicit water model were carried out for ALA pentapeptide (ACE-(ALA)4-NME). All the MD simulations were carried out by AMBER11 with some in-house modifications. Free energy was calculated by Weighted Histogram Analysis Method (WHAM) [51, 52] using density of state estimated by

$$\Omega_{mk} = \frac{H_{mk}}{\sum_{l=1}^{L} N_{kl}\Delta U exp[f_l - \beta_l U_m]}. \quad (9)$$

Results and discussion

Proof of method

In order to verify the validity of this representation of main chain torsion terms, we fitted C(m,n) in Eq. 6 to the total potential of 12 main chain torsion terms in AMBER99SB force field on 36×36 meshes. The results are depicted in Fig. 2. The potential was expanded up to $N_\phi=N_\psi=11$, and the root mean square deviation (RMSD) between the AMBER99SB energy and the fitted term was only 1.71×10^{-2} kcal mol^{-1}. Due to this high consistency, the calculated distributions of main chain torsions were quite close to each other and show only negligible differences. So, the 2D expansions of main chain torsional terms are applicable to the molecular dynamics simulations provided that the correct potential energy surface (PES) is available.

Total potential energy surface

Quantum mechanical calculation of the total potential energy surface was the most CPU intensive part in this work. High level ab initio methods such as MP2 and CCSD(T) with moderate to large basis set in condensed phase are generally too expensive to be utilized for the generation of total PES. More efficient methods such as density functional theory are more practical choices, so long as the accuracy is guaranteed by proper parameterization. Although the reliability of M06 2X functional had been studied in several works, we also investigated its performance by calculating the PES and comparing it with that from MP2 with the same basis set. The calculations were carried out on 24×24 meshes, and the structures on all the grid points were optimized with only $\{\phi,\psi\}$ fixed. PCM was utilized to mimic the solvation effect, and the basis set was reduced to aug-cc-pvdz for both M06 2X and MP2, due to the large expenses of MP2 calculation. Here we chose a relatively small basis set and we just assumed that their behaviors were consistent across double-ζ, triple-ζ and even complete basis sets. The calculated solvent energies are shown in Fig. 3. At a first glance, these two potential energy maps were quite close. To be quantitative, we calculated their RMSD by

$$R = \sqrt{\frac{\sum_{i,j}[E_{M062X}(i,j) - E_{MP2}(i,j)]^2}{\sum_{i,j}1}}, \quad (10)$$

which was only 0.32 kcal mol^{-1}. It indicated that M06 2X was accurate enough for the calculation of the total PES.

With M06 2X/aug-cc-pvtz PES, we fitted the main chain torsion terms with expansions up to $N_\phi=N_\psi=11$ (See Eq. 7). Due to the symmetry of matrix C, there were 265 complex parameters, which leads to RMSD of the fitted PES from the M06 2X one on the order of 0.01 kcal mol^{-1} and the largest deviation below 1 kcal mol^{-1}. This is remarkable, as compared to 1.9 kcal mol^{-1}, the unweighted RMSD of AMBER03 force field [1]. And we anticipate that this new potential gives a more accurate description of fully solvated peptides (without any buried residues) than AMBER force fields do.

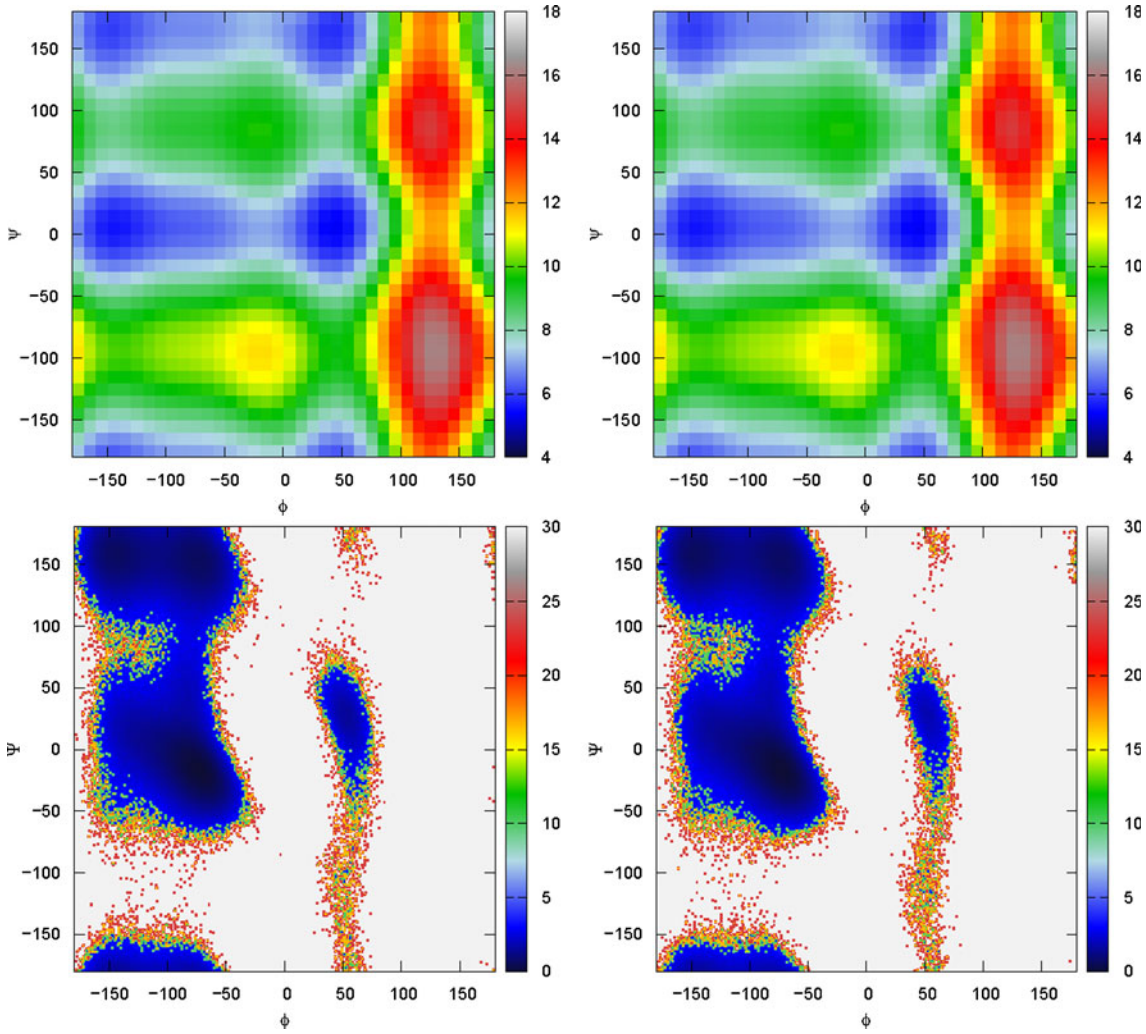

Fig. 2 Main chain torsion maps of alanine dipeptide in kcal mol^{-1} from AMBER99SB (*top left*) and from 2D expansions fitted to AMBER99SB (*top right*), and the free energy maps in kcal mol^{-1} generated in REMD simulations with AMBER99SB force field (*bottom left*) and AMBER99SB force field with main chain torsions replaced by 2D energy terms (*bottom right*)

Tuning of the main chain torsion potential energy map

As mentioned above, the QM optimized structures are not necessarily the minimums under MM Hamiltonian. This may cause inconsistency between the parameterization procedure and the follow-up simulations. In an ideal situation, the energies in full (3 N-6)-dimension space of the model system are consistent between QM and MM, but this is impossible due to the limited number of parameters in MM potential. There are various ways to (partially) surmount this difficulty, e.g., re-parameterizing some of the potential energy terms or further optimizing the structures at MM level with main chain dihedral angles fixed before the calculation of the main chain torsion PES. Here we chose the latter one. The calculated main chain torsion PES without and with MM optimization are depicted in Fig. 4, from which we can read that after optimization the main chain torsion energy around $(\phi,\psi)=(-180°,0°)$ was raised up by about 2 kcal mol^{-1}, and that around $(\phi,\psi)=(-75°,120°)$ was lowered by approximate 1 kcal mol^{-1}. These variations may cause significant shift in the distributions of main chain torsions at ambient condition.

The polarizable continuum model (PCM) has been widely used in the quantum mechanical calculation of small molecules. In computational biology community, solvation effect is often implemented implicitly by Poisson-Boltzmann (PB) or generalized Born (GB) model to save us from intensive sampling in the phase space of water molecules. In this work, we utilized GB model. However, PCM and GB models are formulated in different forms with different parameters. They might give different solvation free energies for the same solute structure. The solvation free energies from PCM and GB solvation models are shown in Fig. 5. It shows that regardless of whether the conformations were further optimized on MM level, there were large differences in the solvation free energies from PCM and GB models, which ranged from −3 to

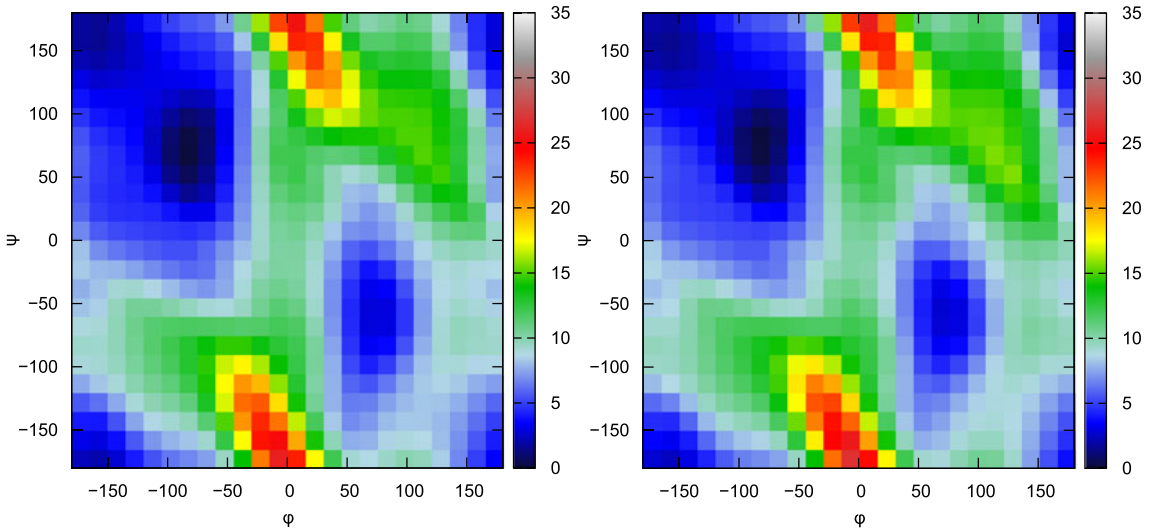

Fig. 3 Total potential energy surface map in kcal mol^{-1} from (*left*) M06 2X/aug-cc-pvdz and (*right*) MP2/aug-cc-pvdz calculations

2 kcal mol^{-1}. And important differences are on (ϕ,ψ)=(−180°,0°) and the right-handed helix domain.

Molecular dynamics simulation of ALA di- and penta-peptide

Assessing the reliability of this representation of main chain torsion energy, especially the balance between various structured and random conformations, is indispensable before it can be widely used in the computational biology community. Alanine dipeptide, which serves as a paradigm of protein backbone, has invoked extensive studies by various experimental means, such as two-dimensional (2D) infrared (IR) [53], vibrational spectroscopy [54], Raman [55, 56], circular dichroism [56], and NMR [56, 57] in aqueous solution. There are still discrepancies in the exact population of various conformations among these studies, due to different models and parameters for the interpretation of signals. Besides, post Hartree-Fock quantum mechanical calculations and semi empirical methods have also been carried out to study the relative energies of some highly populated conformations [24, 58] and the whole potential or free energy surface [32, 59–65]. However, these studies relied highly on the specific QM level and MM parameters. Comparison among experiment measurements, QM calculation and MM study also assess the quality of force fields [31, 32, 58, 62, 63, 66–68]. However, although consensus on the exact distribution of conformations has not been reached so far, there is still some generally accepted view that extended structures including PPII and β conformations

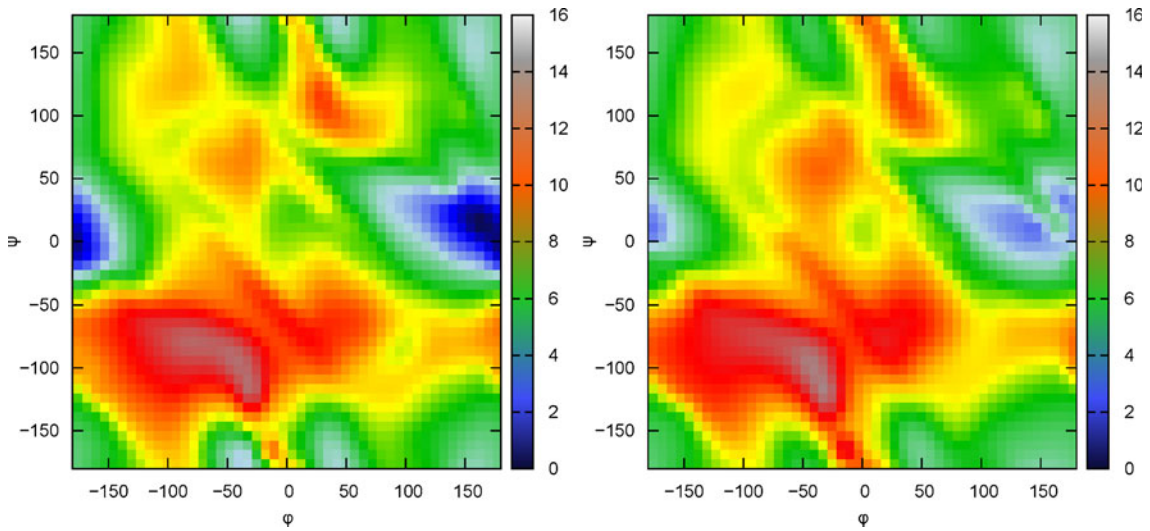

Fig. 4 Main chain torsion potential energy maps in kcal mol^{-1} generated by Eq. 4 with α=0. *Left*: The structures were not further optimized under MM Hamiltonian. *Right*: The structures were further optimized under AMBER03 force field

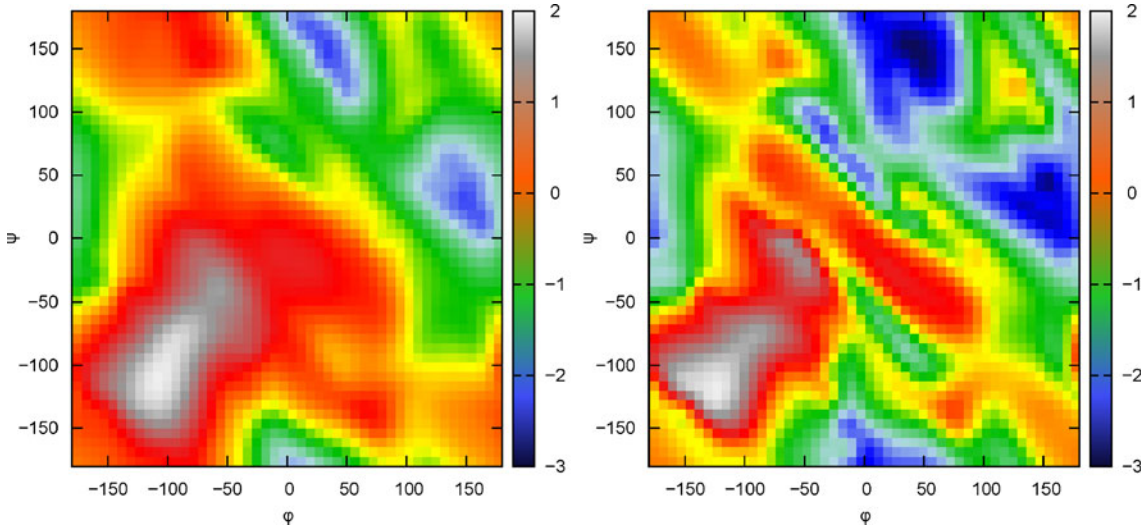

Fig. 5 Difference maps of generalized Born solvation model and polarized continuum model (in kcal mol^{-1}). *Left*: The structures were not further optimized under MM Hamiltonian. *Right*: The structures were further optimized under AMBER03 force field

are dominant, where α helical conformation occupies only a small portion.

With the original AMBER force fields and the main chain torsion energy surfaces calculated above, we studied the distribution of main chain torsions for ALA dipeptide using replica exchange molecular dynamics. The calculated free energy landscapes are shown in Fig. 6. It can be easily found that the distribution of main chain torsions under original AMBER99SB (Fig. 6a) and AMBER03 (Fig. 6b) are qualitatively consistent. The most sampled structures are β, PPII and right-handed α -helical conformations. Elaborative investigation of these two free energy landscapes finds that under AMBER03 the population of extended structures, i.e., β and PPII conformations, are less than that under AMBER99SB, and the left-handed helix under AMBER99SB is more populated than that under AMBER03 force field. The formal observation is consistent with existing knowledge that AMBER03 biases helical conformation, while AMBER99SB underestimates the helix propensity. To make a quantitative comparison, we split the main chain torsion space into "α_+" ($-160^\circ<\phi<-20^\circ$ and $-120^\circ<\psi<50^\circ$), "α_h" with a more stringent definition ($-100^\circ<\phi<-30^\circ$ and $-67^\circ<\psi<-7\phi$), "β" ($-180^\circ<\phi<-90^\circ$ and $50^\circ<\psi<240^\circ$; or $160^\circ<\phi<180^\circ$ and $110^\circ<\psi<180^\circ$), "PPII" ($-90^\circ<\phi<-20^\circ$ and $50^\circ<\psi<240^\circ$), and others, in the same way as that used by Best [20]. The result is shown in Table 1. It clearly indicates that the AMBER03 force field is more helical than AMBER99SB with α_h population about 13 % higher (26.5 % vs 13.4%). Experimental studies already showed that the most observed conformations were β and PPII, and helix occupied only a small portion (below 20 %). Therefore, AMBER99SB gives a more quantitative depiction of the conformation distribution than AMBER03 does, at least for AD model system.

Then we replaced the original uncorrelated main chain torsion energy term in AMBER force fields with 2D expansions. In Fig. 6c and Fig. 6d, we utilized AMBER99SB and AMBER03 force field parameters, respectively, for energy terms other than the main chain torsions. The main chain torsion energies were directly fitted from QM calculations without any optimization on MM level nor the contribution of solvation energy. Because the force fields we used in the parameterization procedure of 2D main chain torsions and the REMD simulations follow-up were consistent, the final result was only weakly dependent on the specific force field utilized, but highly relied on the quantum mechanical potential energy surface and parameterization procedure. We found an abnormally deep well in both of the free energy maps near $(\phi,\psi)=(-180°,0°)$. This erroneous distribution was due to the mismatch between QM and MM PES that the optimized structures under QM Hamiltonian were not necessarily the minimums under MM potential, as we have mentioned above. Therefore we optimized the structures of the model peptide under AMBER03 potential before fitting the 2D main chain torsion potential according to Eq. 4 with $\alpha=0$, and the corresponding free energy landscape with this new 2D potential is shown in Fig. 6e. It shows that the population near $(\phi,\psi)=(-180°,0°)$ was notably reduced, but it was still distinctly overpopulated than those form the original AMBER force fields. Further, we set α in Eq. 4 to 1 to account for the difference of solvation free energy, and the population near (-180°, 0°) flowed to the right-handed α helix region (See Fig. 6f). The population of left-handed helix was also reduced. Comparing this free energy landscape to that under AMBER99SB force field, the difference was mainly on the region with ϕ around -70° and ψ between 50° and 100°. This region is usually denoted as C7, named after the 7-membered ring adopted by the ALA

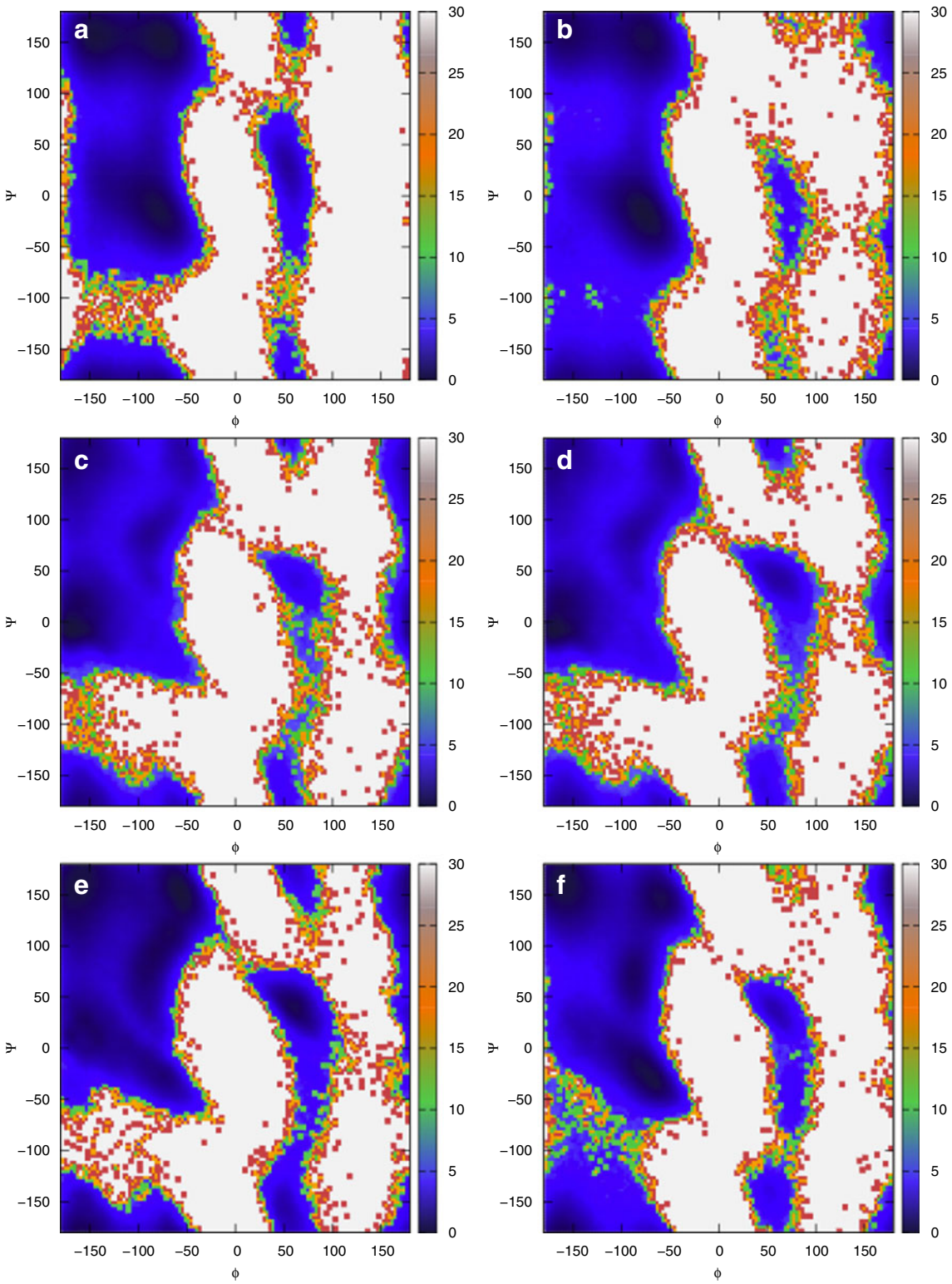

Fig. 6 Free energy landscapes of ALA dipeptide at 300 K from REMD simulations employing (**a**) AMBER99SB force field, (**b**) AMBER03 force field and (subfigures **c-f**) AMBER force field with 2D main chain torsions. **c** AMBER99SB force field was utilized for other terms. The structures were not further optimized and α=0 during parameterization. **d** AMBER03 force field was utilized for other terms. The structures were not further optimized and α=0 during parameterization. **e** AMBER03 force field was utilized for other terms. The structures were further optimized and α=0 during parameterization. **f** AMBER03 force field was utilized for other terms. The structures were further optimized and α=1 during parameterization

Table 1 Percent population of conformations at 300 K from replica exchange molecular dynamics simulations of ALA dipeptide employing GB1 solvation model

Force field	α_+	α_h	β	PPII
AMBER03	43.8	26.5	33.5	22.0
AMBER99SB	31.3	13.4	40.8	26.2
2D AMBER03	27.5	17.7	49.3	21.5
2D AMBER99SB	29.1	18.0	48.3	21.3

dipeptide with a hydrogen bond formed between the acetyl and N-methylamine groups. Statistical investigation of the structures deposited in Protein Data Bank also shows notable population in this region [7], indicating that proteins have some tendency in forming this 7-membered ring through a hydrogen bond. When fully exposed in aqueous solvent as AD in this work, this hydrogen bond might be disrupted by single water molecule that hydrogen-bonded to both ACE and NME groups. Unfortunately, implicit solvation models have difficulties in describing discrete solvent molecules. Therefore the reliability of this observation needs further investigation in explicit solvation. However, prohibitively expensive sampling of solvents' degrees of freedom (DoF) in QM calculations impedes the application of this idea. The percentage of PPII is about 5% lower than that from original AMBER99SB force field, while the β conformation is more frequently sampled under 2D potential, sacrificing part of the α_+ conformation. The population of α_h is between those from AMBER99SB and AMBER03 force fields. Since AMBER03 force field is thought to be helical biasing and AMBER99SB underestimates the helical propensity, this result is quite encouraging.

Although GB1 model was utilized during the parameterization of the main chain torsions, it is also of interest to see the dependence of the conformation distribution on solvation models. We carried out REMD simulations in various GB solvation models and TIP3P water box, and the results are shown in Table 2. The dependence on solvation models are nonnegligible. GB5 is the most helix-biasing model, and GB7 is the least one. All the GB models, especially GB7, bias the β conformation and disfavor the PPII conformation

Table 2 Dependence of the percent population of conformations at 300 K on the solvation models from replica exchange molecular dynamics simulations of ALA dipeptide employing 2D AMBER03 force field

Conformation	GB1	GB5	GB7	TIP3P
α_+	27.5	30.2	22.1	29.5
α_h	17.7	19.9	12.9	18.3
β	49.3	46.0	55.9	39.6
PPII	21.5	21.8	20.3	27.8

Table 3 $^3J(H_N,H_\alpha)$ Coupling (Hz) from the simulations in explicit water box in this work and the experimental measurement

Parameter	AMBER03	2D AMBER03	AMBER99SB	2D AMBER99SB	Exp [69].
Hu [70]	6.81	6.50	7.42	6.53	6.06
DFT2 [71]	7.20	6.51	7.90	6.52	

when compared with the TIP3P water model. $^3J(H_N,H_\alpha)$ coupling is very sensitive to the conformation distribution. In Table 3, the computed J coupling based on the trajectories in explicit water box and the experimental measurement by Avbelj et al. [69] are listed. Both of the parameters from Hu [70] and Case [71] were utilized. The results show that employing these coupled main chain torsion potentials improved the agreement between the simulations and the experiment by 0.3 to 1.4 Hz.

Employing the 2D main chain torsion potential fitted with conformations further optimized at MM level and α= 1 in Eq. 4, we studied ALA pentapeptide (ACE-(ALA)4-NME) in TIP3P water box. This is the smallest model that can form main chain hydrogen bonds with helical conformation. In Table 4, we list the populations of α_+, α_h, β, and PPII conformations for ALA pentapeptide under AMBER03 and AMBER99SB with coupled main chain torsions. Unlike alanine dipeptide, we find moderate differences in the conformation distributions of ALA pentapeptide between 2D AMBER99SB and 2D AMBER03 force fields, especially for the α and β conformations. This difference results from the terminal effect during the parameterization process, and it adds more difficulties to the force field development. The population of α_h and PPII regions increased by about 5 % and 2 % as compared to ALA dipeptide. While the population in β region decreased by about 5 %.

Conclusions

Although high level quantum mechanical methods have been widely used in the study of small-sized molecules, its application in macromolecules such as proteins is still quite limited. Parameterization of the quantum mechanical calculations into molecular mechanics relevant terms and applying them in the simulation of macromolecules is more practical nowadays. However, it is generally accepted that fitting molecular

Table 4 Percent population of conformations at 300 K from replica exchange molecular dynamics simulations in TIP3P water box of ALA pentapeptide

Force field	α_+	α_h	β	PPII
2D AMBER03	33.1	24.2	33.6	30.5
2D AMBER99SB	29.8	20.0	38.1	29.9

mechanical parameters merely from quantum mechanics is not a good choice, due to some discrepancies between QM and MM calculations. In this work, we implemented 2-dimensional main chain torsions into AMBER force field and studied the distribution of conformations of ALA oligopeptides. Inconsistencies between QM and MM made this parameterization procedure less straightforward.

Elaborative investigation of the potential energy maps, solvation free energy maps and free energy landscapes of ALA dipeptide indicated that possible deviation from "true" main chain torsions originates from the discrepancies of solvation models used in QM and MM calculations and of the relaxed structures under QM and MM Hamiltonians. A rigorous model utilizing explicit water models and lengthy sampling of water coordinates in QM and MM is still not feasible today and also in the foreseeable future. In this work, the solvation effect was considered by utilizing polarizable continuum model in QM calculation, while generalized Born model was invoked in MM modeling. Different parameters were employed in these two approaches. For the majority of proteins other than ALA oligopeptides, only a small portion of residues are fully solvated in water. More accurate QM calculations should be carried out in the environment mimicking the protein interior. For example, set the dielectric constant to between 2 and 4 as was in the AMBER03 force field parameterization, or to 1 with polarization effect turned on. For some energetically unfavorable conformations (e.g., $\{\phi,\psi\}=\{-180.0°,0.0°\}$), the QM optimized structures had the amide bond (H-N-C-O in the atom nomenclature of Protein Data Bank) deviated from plane by about 20°. These conformations exerted a strong stress on this torsion under MM potential energy surface, and they are rarely observed in MD simulations. In other words, the quantum mechanically optimized structures were not fully relaxed from the view of molecular mechanics. Here we simply optimized the model peptide further at MM level before the parameterization procedure, and made the result more reasonable. Another solution for this discrepancy is to soften the off-plane distortion term and reparameterize some of the van der Waals terms in MM.

With this new main chain torsions, we studied the distributions of the main chain dihedrals of ALA dipeptide and pentapeptide. The latter is the minimal system that can possess a main chain hydrogen bond in helical conformation. The helical propensities by 2D potential were between those from AMBER03 and AMBER99SB force field, which was quite encouraging. The computed $^3J(H_N,H_\alpha)$ coupling is also improved under this new force field. This work serves as an implication for the necessity of new functional forms for main chain torsions in order to trade efficiency for accuracy. It also poses a challenge to the development of quantum mechanical methods, especially the quantum mechanical solvation models. Further examinations of its performance in molecular dynamics such as protein folding, which are out of the scope of this work, are in progress.

Acknowledgments This work is supported by the National Natural Science Foundation of China (Grant No. 10974054, 20933002, 21173082 and 31200545), the Shanghai PuJiang Program (09PJ1404000), and the Shanghai Rising-Star Program (Grant No. 11QA1402000). We thank Supercomputer Center of East China Normal University for CPU time support.

References

1. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) J Comput Chem 24:1999–2012
2. MacKerell AD Jr et al. (1998) J Phys Chem B 102:3586–3616
3. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) J Phys Chem B 105:6474–6487
4. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF (2004) J Comput Chem 25:1656–1676
5. Bixon M, Lifson S (1967) Tetrahedron 23:769–784
6. Li D-W, Brüschweiler R (2010) Angew Chem Int Edit 49:6778–6780
7. Best RB, Hummer G (2009) J Phys Chem B 113:9004–9015
8. Nerenberg PS, Head-Gordon T (2011) J Chem Theor Comput 7:1220–1230
9. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Proteins 78:1950–1958
10. Jorgensen WL, Tirado-Rives J (2005) Proc Natl Acad Sci USA 102:6665–6670
11. Mackerell AD (2004) J Comput Chem 25:1584–1604
12. Ponder JW, Case DA (2003) Protein simulations. Academic, New York 66:27–85
13. Freddolino PL, Park S, Roux B, Schulten K (2009) Biophys J 96:3772–3780
14. Okur A, Strockbine B, Hornak V, Simmerling CJ (2003) J Comput Chem 24:21–31
15. Garcia AE, Sanbonmatsu KY (2002) Proc Natl Acad Sci USA 99:2782–2787
16. Kamiya N, Higo J, Nakamura H (2002) Protein Sci 11:2297–2307
17. Higo J, Ito N, Kuroda M, Ono S, Nakajima N, Nakamura H (2001) Protein Sci 10:1160–1171
18. Ono S, Nakajima N, Higo J, Nakamura HJ (2000) J Comput Chem 21:748–762
19. Wang L, Duan Y, Shortle R, Imperiali B, Kollman PA (1999) Protein Sci 8:1292–1304
20. Best RB, Buchete N-V, Hummer G (2008) Biophys J 95:L07–L09
21. Simmerling C, Strockbine B, Roitberg AE (2002) J Am Chem Soc 124:11258–11259
22. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Proteins 65:712–725
23. MacKerell AD, Feig M, Brooks CL (2004) J Am Chem Soc 126:698–699
24. Mackerell AD, Feig M, Brooks CL (2004) J Comput Chem 25:1400–1415
25. Wang Z-X, Zhang W, Wu C, Lei H, Cieplak P, Duan Y (2006) J Comput Chem 27:781–790
26. Sakae Y, Okamoto Y (2010) Mol Simulat 36:138–158
27. Sorin EJ, Pande VS (2005) Biophys J 88:2472–2493
28. Piana S, Lindorff-Larsen K, Shaw DE (2011) Biophys J 100:L47–L49
29. Kamiya N, Watanabe YS, Ono S, Higo J (2005) Chem Phys Lett 401:312–317

30. Iwaoka M, Tomoda S (2003) J Comput Chem 24:1192–1200
31. Zaman MH, Shen M-Y, Berry RS, Freed KF, Sosnick TR (2003) J Mol Biol 331:693–711
32. Liu ZW, Ensing B, Moore PB (2011) J Chem Theor Comput 7:402–419
33. Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD (2006) Biophys J 90:L36–L38
34. Sakae Y, Okamoto Y (2006) J Phys Soc Jpn 75:054802
35. Case DA et al. (2010) AMBER 11. University of California, San Francisco
36. Head-Gordon M, Pople JA, Frisch MJ (1988) Chem Phys Lett 153:503–506
37. Scuseria GE, Janssen CL, Schaefer HF III (1988) J Chem Phys 89:7382–7387
38. Zhao Y, Truhlar DG (2007) Theor Chem Accounts 120:215–241
39. Zhao Y, Truhlar DG (2008) Accounts Chem Res 41:157–167
40. Tomasi J, Mennucci B, Cancès E (1999) J Mol Struct (Theochem) 464:211–226
41. Tomasi J, Mennucci B, Cammi R (2005) Chem Rev 105:2999–3094
42. Frisch MJ et al. (2010) Gaussian 09, revision B.01. Gaussian, Inc, Wallingford
43. Tsui V, Case DA (2000) Biopolymers 56:275–291
44. Cao Z, Lin Z, Wang J, Liu H (2009) J Comput Chem 30:645–660
45. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1986) Numerical recipes. Cambridge University Press, Cambridge, UK
46. Blondel A, Karplus M (1996) J Comput Chem 17:1132–1141
47. Sugita Y, Okamoto Y (1999) Chem Phys Lett 314:141–151
48. Onufriev A, Bashford D, Case DA (2004) Proteins 55:383–394
49. Mongan J, Simmerling C, McCammon JA, Case DA, Onufriev A (2007) J Chem Theor Comput 3:156–169
50. Jorgensen WL, Candrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
51. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) J Comput Chem 13:1011–1021
52. Chodera JD, Swope WC, Pitera JW, Seok C, Dill KA (2007) J Chem Theor Comput 3:26–41
53. Kim YS, Wang JP, Hochstrasser RM (2005) J Phys Chem B 109:7511–7521
54. Grdadolnik J, Golic Grdadolnik S, Avbelj F (2008) J Phys Chem B 112:2712–2718
55. Mukhopadhyay P, Zuber G, Beratan DN (2008) Biophys J 95:5574–5586
56. Schweitzer-Stenner R, Measey T, Kakalis L, Jordan F, Pizzanelli S, Forte C, Griebenow K (2007) Biochemistry 46:1587–1596
57. Mehta MA, Fry EA, Eddy MT, Dedeo MT, Anagnost AE, Long JR (2004) J Phys Chem B 108:2777–2780
58. Beachy MD, Chasman D, Murphy RB, Halgren TA, Friesner RA (1997) J Am Chem Soc 119:5908–5920
59. Smith PE (1999) J Chem Phys 111:5568–5579
60. Vargas R, Garza J, Hay BP, Dixon DA (2002) J Phys Chem A 106:3213–3218
61. Hu H, Elstner M, Hermans J (2003) Proteins 50:451–463
62. Wang Z-X, Duan Y (2004) J Comput Chem 25:1699–1716
63. Kwac K, Lee K-K, Han JB, Oh K-I, Cho M (2008) J Chem Phys 128:105106
64. de Seabra GM, Walker RC, Roitberg AE (2009) J Phys Chem A 113:11938–11948
65. Gaigeot M-P (2009) J Phys Chem B 113:10059–10062
66. Mu YG, Kosov DS, Stock G (2003) J Phys Chem B 107:5064–5073
67. Graf J, Nguyen PH, Stock G, Schwalbe H (2007) J Am Chem Soc 129:1179–1189
68. Woutersen S, Pfister R, Hamm P, Mu YG, Kosov DS, Stock G (2002) J Chem Phys 117:6833–6840
69. Avbelj F, Grdadolnik SG, Grdadolnik J, Baldwin RL (2006) Proc Natl Acad Sci USA 103:1272–1277
70. Hu J-S, Bax A (1997) J Am Chem Soc 119:6360–6368
71. Case DA, Scheurer C, Brüschweiler R (2000) J Am Chem Soc 122:10390–10397