PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2019, **21**, 18958

Assessing the performance of the MM/PBSA and MM/GBSA methods. 10. Impacts of enhanced sampling and variable dielectric model on protein–protein Interactions†

Ercheng Wang,[a] Gaoqi Weng,[a] Huiyong Sun,[a] Hongyan Du,[a] Feng Zhu,[a] Fu Chen,[a] Zhe Wang[a] and Tingjun Hou*[ab]

Enhanced sampling has been extensively used to capture the conformational transitions in protein folding, but it attracts much less attention in the studies of protein–protein recognition. In this study, we evaluated the impact of enhanced sampling methods and solute dielectric constants on the overall accuracy of the molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) and molecular mechanics/generalized Born surface area (MM/GBSA) approaches for the protein–protein binding free energy calculations. Here, two widely used enhanced sampling methods, including aMD and GaMD, and conventional molecular dynamics (cMD) simulations with two AMBER force fields (ff03 and ff14SB) were used to sample the conformations for 21 protein–protein complexes. The MM/PBSA and MM/GBSA calculation results illustrate that the standard MM/GBSA based on the cMD simulations yields the best Pearson correlation (r_p = −0.523) between the predicted binding affinities and the experimental data, which is much higher than that given by MM/PBSA (r_p = −0.212). Two enhanced sampling methods (aMD and GaMD) are indeed more efficient for conformational sampling, but they did not improve the binding affinity predictions for protein–protein systems, suggesting that the aMD or GaMD sampling (at least in short timescale simulations) may not be a good choice for the MM/PBSA and MM/GBSA predictions of protein–protein complexes. The solute dielectric constant of 1.0 is recommended to MM/GBSA, but a higher solute dielectric constant is recommended to MM/PBSA, especially for the systems with higher polarity on the protein–protein binding interfaces. Then, a preliminary assessment of the MM/GBSA calculations based on a variable dielectric generalized Born (VDGB) model was conducted. The results highlight the potential power of VDGB in the free energy predictions for protein–protein systems, but more thorough studies should be done in the future.

Received 22nd July 2019,
Accepted 15th August 2019

DOI: 10.1039/c9cp04096j

rsc.li/pccp

Introduction

Protein–protein interactions (PPIs) play important roles in most crucial biological processes in living cells,[1–3] and many PPIs have even been regarded as potential drug targets.[4–6] A complete picture of how two proteins interact can be provided by the three-dimensional structure of the protein–protein complex. There are three main experimental techniques to determine the structures of protein–protein complexes at molecular resolution, including X-ray crystallography, solution nuclear magnetic resonance (NMR), and cryogenic electron microscopy (cryo-EM).[7] But solving the high-resolution structures for all PPIs is extremely difficult or even impossible.[8] Moreover, a huge number of weak and transient PPIs cannot be captured by most experimental techniques. Alternatively, numerous computational approaches have been developed and employed to investigate how two proteins interact with each other.[9,10]

Molecular dynamics (MD) simulations have been widely used to explore the structural and dynamical properties of bio-molecular systems,[11] and the MM/PBSA and MM/GBSA (or MM/PB(GB)SA) methodologies based on MD simulations[12–16] have been regarded as powerful technologies to estimate the binding affinity between the interaction partners in a bio-molecular complex since they can often achieve a good balance between speed and accuracy compared with the alchemical methods (FEP/TI) with low speed and molecular docking scoring functions with low accuracy.[17–19] For instance, Maffucci *et al.*

[a] *Hangzhou Institute of Innovative Medicine, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China. E-mail: tingjunhou@zju.edu.cn*
[b] *State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang 310058, China*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9cp04096j

developed a modified MM/PBSA approach (called Nwat-MM/PBSA),[17] which explicitly incorporates the water effects into protein–protein binding, to determine the binding free energies for a dataset of 20 protein–protein complexes, and the results illustrate that Nwat-MM/PBSA yields a higher correlation between the predictions and the experimental data (r^2 = 0.77) than the traditional method (r^2 = 0.45). In the previous study, we have evaluated the performance of MM/PB(GB)SA on the predictions of the binding affinities for 46 protein–protein complexes,[19] and the calculation results show that MM/GBSA has a better capability to distinguish the correct binding structures from the decoys than ZDOCK.[20] Therefore, MM/PB(GB)SA that considers both computational efficiency and prediction accuracy simultaneously may have their own superiority over other methods to predict the binding affinities and identify the correct binding structures for protein–protein systems.

In the standard MM/PB(GB)SA calculations, the energy terms were computed based on a number of snapshots extracted from the MD trajectories, and therefore their prediction accuracy is partially determined by the sampling ability of MD to reach the different microstates in both the bound and unbound states for a given system. Conventional MD (cMD) simulations have limited sampling capability, which may fail to generate exhaustive and relevant conformational ensemble for accurately predicting receptor–ligand binding affinities. Nowadays several enhanced sampling methods have been developed to improve the MD sampling efficiency and explore the relevant regions of the multi-dimensional conformational landscape by simulating identical molecular systems at different temperatures, such as replica-exchange MD[21] (REMD), and modifying the potential energy surface, such as metadynamics,[22] umbrella sampling[23] (US), accelerated MD[24,25] (aMD) and Gaussian accelerated MD[26–28] (GaMD), to escape from free-energy minima. In these enhanced sampling approaches, REMD, metadynamics and US require the predefined reaction coordinates (or collective variables) to represent the process of interest and a potential or force bias is applied along the reaction coordinates to facilitate the biomolecular conformational transitions across high energy barriers.[29] Apparently, the reliability of REMD, metadynamics and US is determined by the correct choice of the reaction coordinates. However, two enhanced sampling methods, aMD and GaMD, avoid such requirements of the predefined reaction coordinates. In the aMD method, a boost potential (the boost potential following a nearly Gaussian distribution in GaMD) is added into the potential energy surface, thereby decreasing the energy barriers effectively and accelerating the transitions between the low-energy microstates.[28] Moreover, the aMD and GaMD methods are suitable for the automation of the simulation procedure with less hardware demand.[30,31] Therefore, in this study, aMD and GaMD were used to sample the conformation space for the MM/PB(GB)SA calculations of protein–protein complexes, and the impact of the aMD and GaMD sampling on the MM/PB(GB)SA calculations was examined.

The MM/PB(GB)SA methods compute a linear combination of energy terms for molecular mechanics (electrostatic and van der Waals), and polar and nonpolar solvation energies.[14] The polar contribution of solvation energy is calculated by solving the Poisson–Boltzmann (PB) equation[32,33] or further simplified by using the generalized Born (GB) model.[34,35] The solvent (exterior) dielectric constant (typically 80 for water at 300 K), a common user-tunable parameter for MM/PB(GB)SA, represents the nature of the solvent used in MD simulations. Compared with the solvent dielectric constant, another parameter, the solute dielectric constant, seems much more important in calculating the polar solvation energy.[36] The solute (interior) dielectric constant is generally set to 1 by default,[37] but extensive studies illustrate that a solute dielectric of 1 is not a good choice for many systems[38–40] and the best dielectric constant strongly depends on the characteristics of the investigated system.[41–44] Moreover, since receptor–ligand complexes are not uniformly distributed continuous dielectric medium, PB or GB calculations based on a single solute dielectric constant are controversial, which may lead to large prediction errors.[38,41,43,45,46] Recently, a more rigorous dielectric approach by applying variable dielectric constants for different types of residues,[38,46] namely the variable dielectric model, has been developed, and it is considered to be a more rational way to describe the electrostatics of protein–ligand interactions. For example, a recent study shows that, for a set of 20 randomly selected protein–protein complexes, MM/GBSA based on a dielectric constant of 2.7 for charged residues and 1.1 for non-charged residues could yield the best predictions for protein–protein binding affinities,[47] indicating that the strategy of using a higher dielectric constant for charged residues, to some extent, may reduce the high prediction error of MM/GBSA for highly polar or charged molecules.[48]

In this study, cMD and two enhanced sampling methods (aMD and GaMD) with two AMBER force fields (ff03 and ff14SB) were used to sample the conformations for 21 protein–protein complexes and the MM/PB(GB)SA predictions based on different conformational trajectories and different solute dielectric constants were compared. Finally, a preliminary assessment of the MM/GBSA calculations based on a variable dielectric Generalized Born (VDGB) model for protein–protein complexes was conducted, and the results illustrate that VDGB may be a potential strategy to improve the prediction accuracy of MM/GBSA for protein–protein systems but a more thorough investigation should be done in the future.

Materials and methods

Benchmark dataset

The benchmark dataset used to assess the binding free energy predictions of MM/PB(GB)SA was collected from our previous study.[19] The original dataset consists of 46 protein–protein complexes with experimentally determined binding constants. Here, to reduce the computational demand, the 21 protein–protein complexes with only two chains in each system (one chain in each protein) were chosen from the original dataset and their experimental binding constants cover an extremely broad range of 11 orders of magnitude (Table 1). The corresponding coordinates

Table 1 Some basic information and the experimental pK_d values for the 21 protein–protein complexes

PDB ID	R:L[a]	pK_d	PDB ID	R:L	pK_d	PDB ID	R:L	pK_d
1BVN	P:T	11.05	1DFJ	E:I	13.23	1E96	A:B	5.22
1GRN	A:B	6.41	1H1V	A:G	7.64	1HE8	B:A	5.60
1J2J	A:B	5.96	1KXP	A:D	9.00	1M10	A:B	8.24
1MAH	A:F	10.6	1ML0	A:D	8.51	1OPH	A:B	8.30
1R0R	A:C	9.47	1T6B	X:Y	9.40	1WQ1	R:G	4.77
2B42	A:B	8.97	2BTF	A:P	5.70	2C0L	A:B	6.96
2HLE	A:B	7.40	2OOB	A:B	4.22	7CEI	A:B	14.3

[a] R and L represent the names of the receptor and ligand chains, respectively.

of the studied complexes were downloaded from the Protein Data Bank (PDB entries: 1BVN,[49] 1DFJ,[50] 1E96,[51] 1GRN,[52] 1H1V,[53] 1HE8,[54] 1J2J,[55] 1KXP,[56] 1M10,[57] 1MAH,[58] 1ML0,[59] 1OPH,[60] 1R0R,[61] 1T6B,[62] 1WQ1,[63] 2B42,[64] 2BTF,[65] 2C0L,[66] 2HLE,[67] 2OOB,[68] and 7CEI[69]).

Structure preparation and minimization

For each complex in the dataset, the non-peptide molecules containing heteroatoms were removed and the missing residues were added by using MODELLER v9.15[70] (https://salilab.org/modeller/). Before the MM/PB(GB)SA calculations, each protein–protein complex was prepared using the standard procedure with the *tleap* program in Amber18.[71] Two different Amber force fields, including ff03[72] and ff14SB,[73] were assigned to each protein.

Each protein–protein system was then immersed into a TIP3P water box with the water molecules extended 10.0 Å from any solute atom in each direction. Counterions (Na^+ or Cl^-) were added to neutralize the unbalanced charge in each complex. A cutoff distance of 8.0 Å was used for the van der Waals and short-range electrostatic interactions. The particle mesh Ewald (PME) algorithm was used to compute the long-range electrostatic interactions under periodic boundary conditions.[74,75] Each complex was minimized by three steps: first, all the backbone heavy atoms of the protein were restrained with an elastic constant of 50 kcal mol^{-1} Å^{-2} and the system was optimized by 2000 cycles of steepest descent and 3000 cycles of conjugate gradient minimizations; next, the elastic constant was weakened to 10 kcal mol^{-1} Å^{-2} and the system was optimized by 2000 cycles of steepest descent and 3000 cycles of conjugate gradient minimizations; finally, the whole system was minimized without any restraint (2000 cycles of steepest descent and 3000 cycles of conjugate gradient minimizations). All the minimizations were performed by the *pmemd.MPI* program in Amber18.[76]

Molecular dynamics simulations

Each solvated complex was heated from 0 to 300 K during a period of 50 ps, and then 50 ps of density equilibration with a weak restraint of 2.0 kcal mol^{-1} Å^{-2} on the complex and 500 ps equilibration in the *NTP* (T = 300 K and P = 1 atm) ensemble were performed. Finally, 5 ns *NTP* MD simulations were conducted. In this study, cMD, aMD and GaMD simulations were used to sample the conformations for each system. The temperature was controlled by Langevin dynamics. All of the covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm.[77] The time step was set to 2 fs and the coordinates were saved every 10 000 steps. The last 250 snapshots for each system were used in the MM/PB(GB)SA calculations. The *pmemd.cuda* program in Amber18 was used for all the simulations.[76]

aMD and GaMD

The aMD algorithm modifies the energy landscape by adding a boost potential $V(r)$ to the original potential energy surface when $V(r)$ is below a predefined energy level E, as defined in eqn (1). The boost potential, V, decreases the energy barriers and thus accelerates the transitions between the different low-energy states. Therefore, aMD has improved capability to sample distinct bio-molecular conformations and rare barrier-crossing events that can hardly be sampled by cMD simulations. In the simplest form, the boost potential is given by

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E \\ \dfrac{(E - V(r))^2}{\alpha + (E - V(r))}, & V(r) < E \end{cases} \quad (1)$$

where α is the acceleration factor. As the acceleration factor α decreases, the energy barriers will be decreased. In Amber, the implementation of aMD includes the possibility of independently boosting only the whole potential at once (iamd = 1) or the torsional terms of the potential (iamd = 2). It also allows the possibility to boost the whole potential with an extra boost to the torsions (iamd = 3). The parameters required for the aMD simulations are determined based on the corresponding equilibration phase of the cMD simulations, from which the average values of the potential and torsion energy can be estimated. In our work, the three boosting methods were conducted with the boost potential according to eqn (3) in Miao's work.[25] The four coefficients (a_1, a_2; b_1, b_2) used to calculate the boost potentials in the aMD simulations are (4.0, 4.0; 0.16, 0.16) and the average dihedral and total potential energies were calculated from the foregoing 500 ps equilibration simulations. After the aMD simulation, the effects of the bias can be removed with a straightforward way by exponential reweighting each conformation to recover the canonical ensemble.[78] However, it often results in the large energetic noise during the reweighting of aMD, especially for large proteins.[24,79,80]

To alleviate the problem, GaMD was presented to reduce the energetic noise for simultaneous unstrained enhanced sampling and free energy calculation of biomolecules,[27] even for large proteins. GaMD makes use of harmonic functions to construct the boost potential that is adaptively added to the biomolecular potential energy surface. Considering a system with N atoms at position r, when the system potential $V(r)$ is lower than a threshold energy, E, a boost potential is added as

$$\Delta V(r) = \frac{1}{2}(E - V(r))^2, \quad V(r) < E \quad (2)$$

where k is the harmonic force constant.

A minimal set of simulation parameters are dynamically adjusted to control the magnitude and distribution width of the

boost potential. As such, the resulting boost potential follows a Gaussian distribution and allows for the accurate reweighting of the simulations using cumulant expansion to the second order. Similar to aMD, GaMD offers several options to add potentials for the system, such as only adding the total potential boost ΔV_P (igamd = 1), only adding the dihedral potential boost ΔV_D (igamd = 2), or adding the dual potential boost (both ΔV_P and ΔV_D) (igamd = 3). The dual boost simulation generally provides higher acceleration than the other two types of simulations for enhanced sampling.[81] The simulation parameters comprise of the threshold energy values and the effective harmonic force constants for the total (k_{0P}) and dihedral (k_{0D}) potential boosts, respectively. In this work, the three modes were employed to add boost potential with the threshold energy set to the lower bound while the upper limit of the standard deviation of the total potential boost was set to 6.0 kcal mol^{-1} if igamd = 1 or 3 and the upper limit of the standard deviation of the dihedral potential boost was set to 6.0 kcal mol^{-1} if igamd = 2 or 3.

MM/PBSA and MM/GBSA calculations

In the MM/PB(GB)SA methods, the free energy of a ligand (L) binding to the receptor (R) to produce the complex (RL) can be decomposed into different energy terms and expressed as:

$$\Delta G_{bind} = G_{RL} - G_R - G_L \quad (3)$$

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (4)$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW} \quad (5)$$

$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA} \quad (6)$$

$$\Delta G_{SA} = \gamma \cdot SASA + b \quad (7)$$

where ΔE_{MM}, ΔG_{sol} and $-T\Delta S$ are the changes of the gas phase molecular mechanism (MM) energy, the solvation free energy, and the conformational entropy upon binding, respectively. ΔE_{MM} includes the internal energy ΔE_{int} (bond, angle, and dihedral energies), the electrostatic energy ΔE_{ele} and the van der Waals energy ΔE_{vdW}. ΔG_{sol} is the sum of the polar ($\Delta G_{PB/GB}$) and non-polar (ΔG_{SA}) contributions. $\Delta G_{PB/GB}$ is calculated using either the PB or GB model, while ΔG_{SA} is estimated by solvent accessible surface area (SASA). The conformational entropy change $-T\Delta S$ is usually computed by normal-mode analysis,[14] but it is usually neglected considering its high computational cost and low prediction accuracy.[82,83] Here, the MM/PB(GB)SA calculations for each system were performed based on the last 250 snapshots extracted from the 5 ns MD trajectory with the *MMPBSA.py*[84] module in Amber18.

In the MM/PBSA calculations, the polar solvation energy was calculated by the PB model developed by Tan and Luo,[85] in which the atomic radii were optimized with respect to the reaction field energies computed in the TIP3P explicit solvents. The grid resolution was set to 0.5 Å, and the partial charges of the proteins used in the PB calculations were taken directly from the force field parameters. The surface tension (γ) and correction term (β) for ΔG_{SA} were set to 0.00542 kcal mol^{-1} Å^{-2} and 0.92, respectively.

In the MM/GBSA calculations, the polar solvation energy was calculated by the modified GB model (GB^{OBC1}) developed by Onufriev *et al.*[86] According to our previous studies, the predictions of protein–ligand binding free energies are quite sensitive to the interior dielectric constant, and therefore, four solute dielectric constants, 1, 2, 4 or 6, were used for the MM/PB(GB)SA calculations. The non-polar component of desolvation was estimated by using the LCPO algorithm,[87] where γ and b were set to 0.0072 and 0.0, respectively.

The variable dielectric MM/GBSA model

In the framework of the conventional GB model, the protein–ligand electrostatic interactions and the solvation free energies are described as:

$$E_{ele} = \sum_{i,j} \frac{q_i q_j}{\varepsilon_{in} r_{ij}} \quad (8)$$

$$E_{GB} = -\left(\frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{sol}}\right) \sum_{i,j} q_i q_j \left(r_{ij}^2 + \alpha_{ij}^2 \exp\left(\frac{r_{ij}^2}{4\alpha_{ij}^2}\right)\right)^{-\frac{1}{2}} \quad (9)$$

where q_i and q_j are the partial charges of atom i in the protein and atom j in the ligand, respectively, r_{ij} is their distance, ε_{in} is the solute dielectric constant, ε_{sol} is the solvent dielectric constant, and α_{ij} is the geometric average of the Born radii α_i and α_j. If ε_{in} adopts different values depending on the residue type k interacting with the ligand, the eqn (8) and (9) can be rewritten as shown in eqn (10) and (11).

$$E_{ele} = \sum_k \frac{1}{\varepsilon_{in(k)}} \sum_{i \in k,j} \frac{q_i q_j}{r_{ij}} \quad (10)$$

$$E_{GB} = -\sum_k \left(\frac{1}{\varepsilon_{in(k)}} - \frac{1}{\varepsilon_{sol}}\right) \sum_{i \in k,j} q_i q_j \left(r_{ij}^2 + \alpha_{ij}^2 \exp\left(\frac{r_{ij}^2}{4\alpha_{ij}^2}\right)\right)^{-\frac{1}{2}} \quad (11)$$

where $\varepsilon_{in(k)}$ is the dielectric constant assigned for the interactions between atom i belonging to the residue type k and atom j in the ligand. In our study, the protein–protein binding affinities were also predicted by the MM/GBSA method based on the VDGB model as a comparison.

Results and discussions

Performance of MM/PB(GB)SA in predicting binding affinities

At first, based on the 5 ns cMD trajectories, the binding free energies of the 21 protein–protein complexes were predicted by using the standard MM/PB(GB)SA methods. The Pearson correlation coefficients between the predicted binding free energies and the experimental data (pK_d) are listed in Table 2. For the MM/GBSA calculations, when ε_{in} was set to 1, cMD based on the ff03 force field yields the best prediction ($r_p = -0.523$). For the MM/PBSA calculations, when ε_{in} was set to 6, cMD based on the ff03 force field yields the best prediction ($r_p = -0.615$). It is interesting to observe that, with the increase of the interior dielectric constant, the MM/PBSA predictions

Table 2 Pearson correlation coefficients between the predicted binding free energies and the experimental data for MM/PB(GB)SA based on the 5 ns MD trajectories[a]

		MM/GBSA (r_p)				MM/PBSA (r_p)			
Force field	MD	ε_{in} = 1	ε_{in} = 2	ε_{in} = 4	ε_{in} = 6	ε_{in} = 1	ε_{in} = 2	ε_{in} = 4	ε_{in} = 6
ff03	cMD	**−0.523**	−0.501	−0.468	−0.456	−0.193	−0.421	−0.575	**−0.615**
	aMD1	−0.357	−0.386	−0.382	−0.380	−0.087	−0.332	−0.531	−0.586
	aMD2	−0.385	−0.396	−0.39	−0.387	−0.212	−0.413	−0.553	−0.590
	aMD3	−0.374	−0.398	−0.395	−0.393	−0.125	−0.334	−0.488	−0.536
	GaMD1	−0.451	−0.421	−0.387	−0.375	−0.206	−0.395	−0.538	−0.579
	GaMD2	−0.427	−0.452	−0.441	−0.435	−0.095	−0.314	−0.494	−0.550
	GaMD3	−0.436	−0.429	−0.413	−0.406	−0.147	−0.378	−0.538	−0.580
ff14SB	cMD	−0.405	−0.372	−0.337	−0.325	−0.248	−0.361	−0.452	−0.479
	aMD1	−0.292	−0.319	−0.316	−0.314	−0.008	−0.155	−0.303	−0.364
	aMD2	−0.142	−0.208	−0.231	−0.238	0.151	−0.032	−0.203	−0.272
	aMD3	−0.343	−0.334	−0.324	−0.320	−0.097	−0.288	−0.411	−0.446
	GaMD1	−0.353	−0.332	−0.306	−0.297	−0.162	−0.278	−0.378	−0.413
	GaMD2	−0.390	−0.380	−0.357	−0.349	−0.117	−0.276	−0.412	−0.458
	GaMD3	−0.377	−0.374	−0.356	−0.348	−0.043	−0.188	−0.319	−0.370

[a] aMD1/GaMD1: only the total potential boost (ΔV_P) added; aMD2/GaMD2: only dihedral potential boost (ΔV_D) added; aMD3/GaMD3: the dual potential boost (both ΔV_P and ΔV_D) added.

would become better. However, for MM/GBSA, the predictions are not quite sensitive to the change of the interior dielectric constant. Apparently, relatively low interior dielectric constant (ε_{in} = 1) is preferred to the MM/GBSA calculations, while relatively high interior dielectric constant (ε_{in} = 6) is preferred to the MM/PBSA calculations.

The Pearson correlation coefficients between the predicted binding free energies and the experimental data given by MM/PB(GB)SA with ε_{in} = 1 based on two different force fields, five different MD simulation lengths and seven different sampling processes are shown in Fig. 1. Generally, the force field plays a critical role in the conformational sampling. The results indicate that the MM/GBSA predictions for the ff03 force field based on all the combinations of different MD simulation lengths and sampling methods are obviously better than those for the ff14SB force field (Fig. 1A) and the MM/PBSA predictions for the ff03 force field based on most combinations of different MD simulations lengths and sampling methods are obviously better than those for the ff14SB force field (Fig. 1B). It is worth noting that the predictions of MM/PBSA based on the aMD3 and GaMD2 simulations for the two force fields are quite similar. Although the ff14SB force field is recommended by Amber14–18 for protein simulations, it appears that, according to our calculation results, ff14SB is not the best choice. Therefore, ff03 is more attractive to predict the binding affinities for protein–protein systems.

Compared with MM/PBSA based on ε_{in} = 1 ($-0.268 \leq r_p \leq 0.151$), MM/GBSA based on ε_{in} = 1 shows a much better predictive power for protein–protein complexes ($-0.527 \leq r_p \leq -0.142$). In most cases, for either MM/GBSA or MM/PBSA, the predictive power decreases with the extension of the simulation time, especially for aMD2 with the ff14SB force field. However, the predictive power of MM/GBSA slightly increases with the extension of the simulation time for the

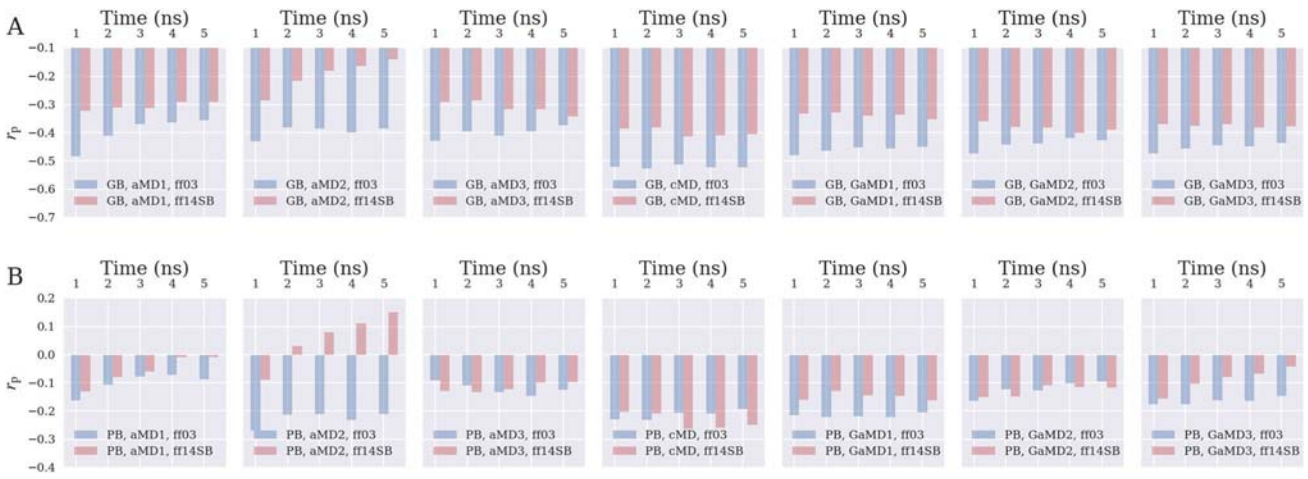

Fig. 1 Pearson correlation coefficients between the predicted binding free energies and the experimental data based on two different force fields, five different MD simulation lengths and seven different sampling processes for (A) MM/GBSA and (B) MM/PBSA.

Table 3 Correlation coefficients between the predicted results based on the 1 ns MD simulations and the experimental data for the 21 protein–protein complexes

Method	Force field	aMD1	aMD2	aMD3	cMD	GaMD1	GaMD2	GaMD3
MM/GBSA	ff03	−0.483	−0.430	−0.428	**−0.522**	−0.480	−0.474	−0.475
	ff14SB	−0.322	−0.286	−0.291	**−0.386**	−0.333	−0.359	−0.369
MM/PBSA	ff03	−0.163	−0.268	−0.092	**−0.230**	−0.215	−0.164	−0.176
	ff14SB	−0.130	−0.089	−0.128	**−0.203**	−0.160	−0.150	−0.157

aMD3, cMD and all GaMD simulations with the ff14SB force field. Therefore, it seems that 1 ns MD sampling is enough for the MM/PB(GB)SA calculations, and much longer MD simulations (>5 ns) are not quite necessary. Similar conclusions have also been made in our previous studies.[88] It is well-known that sufficient conformational sampling with a reliable force field provides the fundamental basis for the accurate prediction of binding free energies. However, because most force fields are inherently not precise enough, long-time conformational sampling may amplify the force field errors. Therefore, longer MD simulations do not necessarily lead to more accurate predictions of binding free energies when the single trajectory protocol is used in MM/PB(GB)SA.

Can enhanced sampling improve the MM/PB(GB)SA predictions?

Enhanced sampling methods, such as aMD and GaMD, are designed to sample more extensive bio-molecular conformational space and detect rare barrier-crossing events that are not easily accessible by cMD simulations.[27] It has been shown in the previous study that aMD exhibits much more efficient conformational sampling capability than cMD for the IL-1R1 ectodomain with the same simulation time.[89] GaMD could also successfully capture the complete binding process of benzene to the deeply buried ligand-binding cavity in the Leu99Ala T4-lysozyme within ~100 ns in one of the five independent 800 ns simulations.[27] Therefore, it is quite interesting to explore if the enhanced sampling capability of aMD and GaMD can improve the prediction accuracy of MM/PB(GB)SA based on the conformations sampled by the cMD simulations.

Here, we compared the performances of the MM/PB(GB)SA calculations based on the cMD, aMD and GaMD trajectories. As shown in Fig. 1, MM/GBSA based on the 2 ns cMD simulations with the ff03 force field yields the best prediction (r_p = −0.527). When the ff03 force field was used, the simulation length of cMD slightly affects the performance of MM/GBSA ($-0.527 \leq r_p \leq -0.514$ for 1–5 ns simulations). Unfortunately, the MM/PBSA calculation based on the trajectories given by the enhanced sampling of aMD1 with ff14SB yields the worst predictions, especially for the 4–5 ns simulations (the correlation coefficients are −0.009 and −0.008, respectively). As shown in Table 3, compared with the 1 ns aMD and GaMD simulations, the 1 ns cMD simulations with the ff03 or ff14SB force field yield the best performance for both of the MM/GBSA (r_p = −0.522 and −0.386 for ff03 and ff14SB, respectively) and MM/PBSA (r_p = −0.230 and −0.203 for ff03 and ff14SB, respectively) calculations. For the MM/GBSA calculations, the aMD simulations with the ff03 force field give the worst prediction (r_p = −0.092). Our results suggest that the enhanced sampling methods, such as aMD and GaMD, cannot enhance the prediction accuracy of MM/PB(GB)SA.

In order to figure out the reason to interpret the low predictive accuracy of the MM/PB(GB)SA calculations based on the enhanced sampling trajectories, the average root-mean-square deviations (RMSDs) of the heavy atoms for the snapshots sampled by the 1 ns MD simulations were analyzed and shown in Fig. 2. The average RMSD represents the overall conformational change for a protein–protein system. For examples, as shown in Fig. 2A, when the ff03 force field was used, the conformations of 1E96 sampled by aMD1 and aMD3 underwent more significant change than those sampled by cMD (the average RMSD = 2.408 Å, 2.614 Å and 1.956 Å for aMD1, aMD3 and cMD, respectively) and the conformations of 1T6B sampled by aMD2 and aMD3 also underwent more significant change (the average RMSD = 3.456 Å, 3.195 Å and 2.566 Å for aMD2, aMD3 and cMD, respectively). As shown in Fig. 2B, when the ff14SB force field was used, the conformations of 1OPH sampled by aMD3 have the most significant change (the average RMSD = 3.803 Å).

Interestingly, the cMD simulations with the two force fields give the lowest conformation changes for most of the 21 protein–protein complexes. The reason may be attributed to the fact that the structural adjustment from the enhanced sampling simulations potentially disturb the best binding structures of the complexes in the crystal structures. Therefore, in our study, the aMD and GaMD simulations indeed enhance the conformational sampling, but they may give worse binding free energy predictions because of the "over-sampling" of local structures.

As shown in Fig. 2, the conformational fluctuations of several complexes, especially 1HE8, are significantly larger than those of the others. By utilizing the *InterProSurf* server,[90] the residues on the protein–protein binding interface of 1HE8 were identified and then the RMSDs and the native contacts of the interface residues were calculated. The RMSD can reflect the global conformational changes of the interface residues, while the native contacts represent the changes of the interface residues in detail. The distributions of the RMSDs *versus* the native contacts of the interface residues shown in Fig. 3 illustrate that cMD with ff03 yields the most stable conformations, while aMD3 with ff03 exhibits the largest conformational changes. The number of the native contacts between the interface residues in the two protein monomers decrease dramatically from ~200 to ~20, indicating that the interface residues deviate far away from the crystal structure in the end-phase of the aMD3 simulations, which can also be proven by the decomposition of the total MM/GBSA binding free energies. As shown in Table 4,

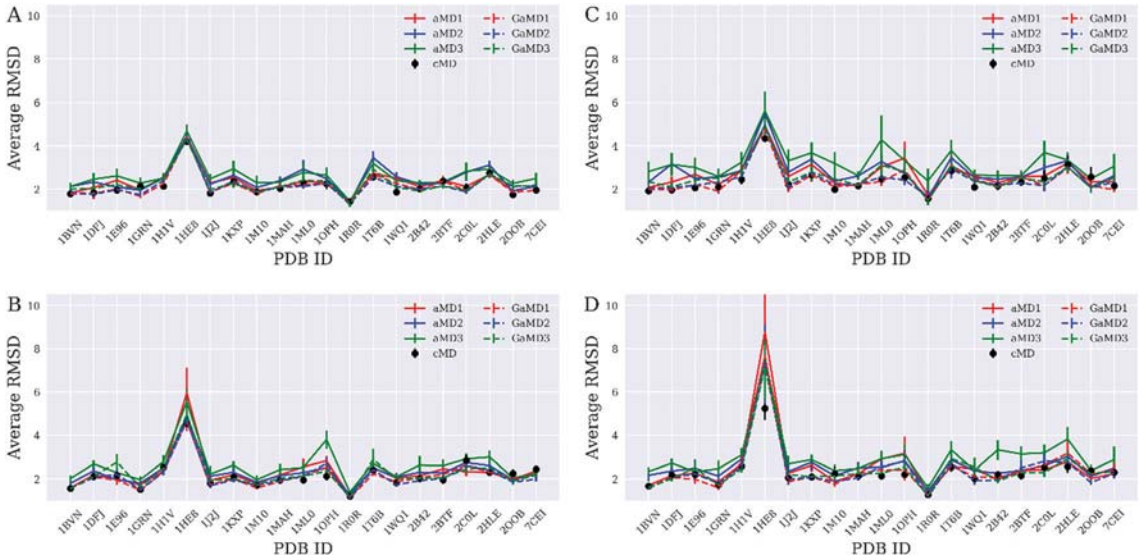

Fig. 2 The average RMSDs (Å) of the heavy atoms for all the snapshots relative to the initial structures sampled by (A) the 1 ns MD simulations with ff03, (B) the 1 ns MD simulations with ff14SB, (C) the 5 ns MD simulations with ff03, or (D) the 5 ns MD simulations with ff14SB. The error bars are displayed in the plots.

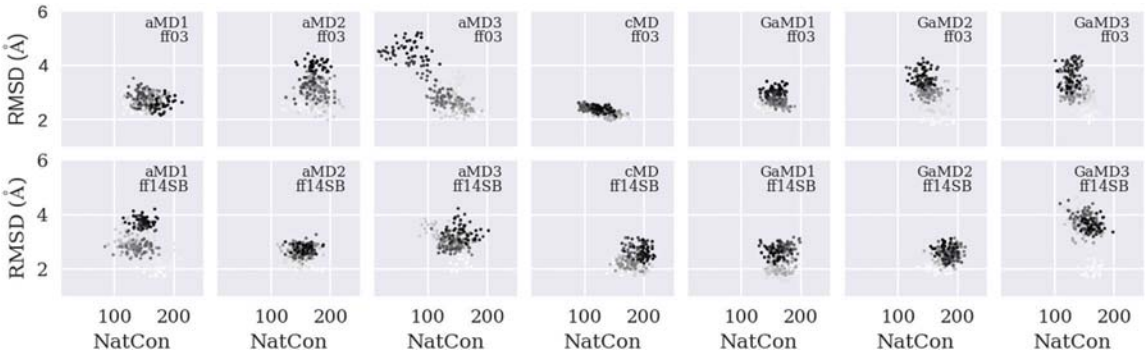

Fig. 3 2-D scatter plot of the RMSDs *versus* the native contacts (NatCon) of the interface residues in 1HE8 for all kinds of 5 ns MD simulations. The grayscale depth represents the simulation length. The darker is the color, the longer is the simulation.

the electrostatic and polar solvation energies (ΔE_{ele} = −143.46 kcal mol^{-1} and ΔG_{GB} = 182.20 kcal mol^{-1}) play critical roles in the protein–protein interaction for 1HE8 when the cMD simulation was performed with the ff03 force field. However, for the aMD and GaMD simulations, the contributions of the two interactions are reduced obviously, especially for aMD2, aMD3 and GaMD2. The electrostatic interactions between the two protein monomers of 1HE8 for the aMD3 simulations are even slightly repulsive (ΔE_{ele} = 7.40 kcal mol^{-1}). When the ff14SB force field was used, the GaMD1 and GaMD2 simulations yield much more weakened electrostatic and polar solvation energies. However, as shown in Fig. 3, the conformational changes of the interface residues in 1HE8 are lower. Hence, it is possible that some residues out of the binding interface would undergo large conformational changes in the GaMD1 and GaMD2 simulations.

Another possible reason is that, for large protein–protein systems, 5 ns aMD or GaMD simulations may be not long enough to sample the transitions between the microstates successfully. In addition, our simulation results suggest that aMD may not be a good choice for the short timescale conformational sampling for protein–protein systems.

In MM/PB(GB)SA calculations, the entropic changes are usually computed by normal mode analysis (NMA),[14] but they are neglected in the above analysis due to high computational cost and low prediction accuracy of NMA. Recently, Duan *et al.* developed a new entropy estimation method called the interaction entropy (IE) approach,[91] which can compute the entropic contributions by correlating with the fluctuations of the ligand–receptor interaction energy (both electrostatic and van der Waals interactions) around its average from MD simulations without any extra computational cost.[91–94] Unfortunately, as shown in Table S1 (ESI†), the predicted protein–protein binding free energies with the addition of the entropic term become even worse. For example, as shown in Table 2, when ε_{in} = 1 was used, cMD with ff03 and ff14SB yield good predictions (r_p = −0.523 and r_p = −0.405, respectively), but when the interaction entropies were taken into consideration, the predictions become much worse (r_p = 0.248 for ff03 and r_p = 0.078 for ff14SB). It seems that the IE

Table 4 Energy decomposition of the MM/GBSA calculations based on the 5 ns simulations for 1HE8

Force field	MD	ΔE_{vdW}	ΔE_{ele}	ΔG_{GB}	$\Delta\Delta G_{surf}$	E_{gas}	ΔG_{solv}	ΔG_{bind}
ff03	cMD	−67.69	−143.46	182.20	−10.71	−211.14	171.48	−39.66
	aMD1	−58.46	−83.19	122.18	−8.55	−141.65	113.62	−28.02
	aMD2	−61.57	−10.91	48.54	−8.63	−72.48	39.91	−32.57
	aMD3	−56.28	7.40	32.02	−8.11	−48.88	23.91	−24.97
	GaMD1	−74.21	−113.10	163.35	−10.80	−187.31	152.56	−34.76
	GaMD2	−52.46	−6.00	38.51	−7.28	−58.46	31.24	−27.22
	GaMD3	−51.15	−66.18	97.43	−7.64	−117.34	89.79	−27.55
ff14SB	cMD	−62.74	−115.79	149.28	−9.80	−178.53	139.48	−39.05
	aMD1	−59.88	−109.23	147.46	−8.75	−169.11	138.71	−30.40
	aMD2	−62.93	−101.62	141.16	−8.93	−164.55	132.23	−32.32
	aMD3	−55.26	−119.36	147.70	−8.21	−174.62	139.49	−35.13
	GaMD1	−56.78	−48.30	85.56	−8.50	−105.08	77.06	−28.01
	GaMD2	−59.67	−48.15	80.72	−9.04	−107.82	71.67	−36.15
	GaMD3	−55.92	−108.55	141.26	−8.38	−164.47	132.88	−31.59

method is not sufficient to estimate the protein–protein entropic changes, especially for lower interior dielectric constants.

The effect of dielectric constant on MM/PBSA and MM/GBSA

In MM/PB(GB)SA calculations, the solute dielectric constant (ε_{in}) is set to 1. But for macromolecules, especially for highly charged binding interface, larger ε_{in} values are usually used for the consideration of the high electronic polarization effect. We then compared the binding free energies predicted by MM/PB(GB)SA based on four different solute dielectric constants (1, 2, 4, and 6) for the protein–protein systems. The Pearson correlation coefficients between the calculated binding free energies and the experimental data are listed in Table 1. It is obvious that the MM/PBSA predictions are quite sensitive to the solute dielectric constant and the r_p increases significantly with the increase of ε_{in}, and ε_{in} = 6 yields the highest correlations for the two force fields. However, for MM/GBSA, the increased solute dielectric constant results in inconsistent

Fig. 4 Pearson correlations coefficients between the experimental binding free energies and the predicted values (ε_{in} = 1) given by (A) MM/GBSA for the protein–protein complexes with lower polarity, (B) MM/GBSA for the protein–protein complexes with higher polarity, (C) MM/PBSA for the protein–protein complexes with lower polarity, or (D) MM/PBSA for the protein–protein complexes with higher polarity.

binding free energy predictions. For aMDs, the correlations slightly increase (such as $r_p = -0.357$ to -0.380 for aMD1 with ff03), whereas for cMDs and GaMDs, the correlations decrease (such as $r_p = -0.523$ to -0.456 for cMD with ff03). Interestingly, although the aMD simulations (aMD1, aMD2 and aMD3) yield relatively worse conformational sampling, they give improved performance for a higher solute dielectric constant. Moreover, the two different enhanced sampling approaches result in opposed predictions when the solute dielectric constant increases.

We then calculated the properties of the protein–protein binding interfaces by using the *COCOMAPS* online service.[95] The polarity of a complex is defined as the ratio of the polar interface area to the entire interface area. If the value of the ratio is lower than 0.6, the protein–protein interface is considered as lower polarity (fewer polar residues in the interface). That is to say, 0.6 is the criterion to distinguish lower and higher polarity. As shown in Fig. 4, the MM/GBSA calculations give much worse predictions to the low-polarized complexes ($|r_p| < 0.4$ in most cases, Fig. 3A) than the high-polarized protein–protein systems ($|r_p| > 0.6$), no matter which method was used for conformational sampling and even the ff14SB force field achieves slightly better results than ff03 (Fig. 4B). However, the performance of MM/PBSA is opposite, and it shows powerful predictive capability to the low-polarized systems (Fig. 4C). Moreover, the performance of MM/PBSA is not quite sensitive to the conformations sampled with the ff14SB force field.

For the high-polarized systems, the MM/GBSA calculations perform worse and the predictions are not improved with the increase of the solute dielectric constant (Table 1). We believe that the predictions for the systems with the lower interface polarity lead to the poor predictions for MM/GBSA (Fig. 5A). Theoretically, using a higher dielectric constant (*i.e.* $\varepsilon_{in} = 6$) for the non-polar residues is not quite reasonable. For the high-polarized systems, MM/PBSA with $\varepsilon_{in} = 1$ yields the worst predictions for the high-polarized systems (Fig. 4D), especially with the ff03 force field. However, the MM/PBSA predictions are obviously improved when a higher interior dielectric constant is used (Table 1). Therefore, it is suggested that higher dielectric constants are suitable for the high-polarized protein–protein complexes for the MM/PBSA calculations (Fig. 5B). In general, when the standard MM/PBSA or MM/GBSA method with $\varepsilon_{in} = 1$ is employed for binding free energy predictions, MM/GBSA is better for the high-polarized protein–protein complexes and MM/PBSA is better for the low-polarized protein–protein systems.

Can variable dielectric model improve the predictions of MM/GBSA?

A single dielectric constant is usually used for the whole solute in both PB and GB models, but it is well known that the use of a

Fig. 5 Pearson correlations coefficients between the experimental binding free energies and the predicted values ($\varepsilon_{in} = 6$) given by (A) MM/GBSA for the protein–protein complexes with lower polarity, (B) MM/GBSA for the protein–protein complexes with higher polarity, (C) MM/PBSA for the protein–protein complexes with lower polarity, or (D) MM/PBSA for the protein–protein complexes with higher polarity.

Table 5 Pearson correlation coefficients between the experimental binding free energies and the predictions given by the MM/GBSA calculations based on the standard GB (ε_{in} = 1) and VDGB models[a]

	ff03					ff14SB				
Mode	1 ns	2 ns	3 ns	4 ns	5 ns	1 ns	2 ns	3 ns	4 ns	5 ns
MM/GBSA	−0.522	−0.527	−0.514	−0.523	−0.523	−0.386	−0.382	−0.413	−0.409	−0.405
VDGB_112[b]	−0.438	−0.440	−0.439	−0.451	−0.443	−0.218	−0.222	−0.249	−0.240	−0.245
VDGB_122	−0.539	−0.538	−0.535	−0.540	−0.536	−0.395	−0.396	−0.416	−0.410	−0.410
VDGB_114	−0.347	−0.349	−0.357	−0.367	−0.359	−0.059	−0.067	−0.087	−0.076	−0.091
VDGB_124	−0.500	−0.496	−0.497	−0.500	−0.494	−0.321	−0.324	−0.338	−0.330	−0.335
VDGB_144	−0.509	−0.507	−0.509	−0.511	−0.510	−0.370	−0.371	−0.384	−0.377	−0.381
VDGB_224	−0.448	−0.447	−0.449	−0.453	−0.449	−0.273	−0.279	−0.294	−0.286	−0.293
VDGB_244	−0.484	−0.483	−0.484	−0.485	−0.483	−0.338	−0.342	−0.356	−0.348	−0.352
VDGB_116	−0.308	−0.311	−0.322	−0.331	−0.324	0.004	−0.005	−0.022	−0.011	−0.030
VDGB_126	−0.482	−0.479	−0.481	−0.482	−0.476	−0.289	−0.293	−0.305	−0.296	−0.303
VDGB_146	−0.494	−0.491	−0.495	−0.496	−0.495	−0.345	−0.347	−0.358	−0.350	−0.355
VDGB_166	−0.496	−0.494	−0.498	−0.499	−0.499	−0.359	−0.36	−0.371	−0.363	−0.369
VDGB_226	−0.423	−0.423	−0.428	−0.431	−0.427	−0.239	−0.247	−0.26	−0.251	−0.260
VDGB_246	−0.470	−0.468	−0.471	−0.471	−0.469	−0.315	−0.32	−0.331	−0.323	−0.329
VDGB_266	−0.474	−0.472	−0.475	−0.476	−0.474	−0.330	−0.334	−0.345	−0.338	−0.343
VDGB_446	−0.450	−0.449	−0.452	−0.453	−0.450	−0.296	−0.302	−0.314	−0.306	−0.312
VDGB_466	−0.461	−0.460	−0.462	−0.463	−0.461	−0.315	−0.32	−0.332	−0.324	−0.330

[a] The cMD trajectories were used in the MM/GBSA calculations. [b] The different dielectric constant set is represented by VDGB_*ABC*, in which *A*, *B* and *C* are the dielectric constants for non-polar, polar and charged residues, respectively.

single dielectric constant to describe the heterogenetic dielectric environment of a solute is problematic. The variable dielectric GB model may be a feasible solution to this problem. In this study, the variable dielectric GB model reported by Ravindranathan *et al.*[38] was implemented by us within the AMBER package. In the framework of the variable dielectric GB model, a set of different dielectric constants were assigned to different types of residues. For example, the dielectric constant for each residue type reported by Ravindranathan *et al.*[38] is assigned by using the following scheme: a value of 1 is assigned for non-polar residues, 2 for polar residues, and 4 for charged (positive and negative) residues. In this study, the dielectric constants used in the MM/GBSA analysis range from 1 to 6 and a total of 16 sets of dielectric constants were tested. As shown in Table 5, the best prediction with the correlation coefficient of −0.54 is achieved by the VDGB_122 model based on the 4 ns (the simulation length has little effect on the predictions) cMD simulations with the ff03 force field. However, the predictions of VDGB_122 are only slightly better than the standard MM/GBSA calculations, and unfortunately, the other sets of dielectric constants even yield much worse predictions, especially for VDGB_114 and VDGB_116. When the ff14SB force field was used in the cMD simulations, VDGB_114 and VDGB_116 achieve extremely poor predictions ($|r_p| < 0.1$), suggesting that the values of the dielectric constants for the polar and charged residues should be similar. Although most of the tested sets of the solute dielectric constants yield poor predictions, the VDGB model exhibits its potential. Our results are consistent with the study reported by Zhang *et al.*[47] By combining the VDGB model and the interaction entropy method for efficient computation of entropic change, they proposed a similar strategy to achieve optimal predictions for protein–protein binding free energies. According to Zhang *et al.*'s work, the optimal dielectric constants for charged and non-charged residues are 2.7 and 1.1, respectively.[47] Certainly, in the future, a thorough research to optimize the VDGB model is quite necessary.

Conclusions

In this study, we examined the impact of enhanced sampling methods (aMD and GaMD) and the solute dielectric constants on the MM/PB(GB)SA binding free energy predictions for protein–protein complexes. Our conclusions are as follows: (1) The best prediction of the standard MM/GBSA (ε_{in} = 1) based on the cMD simulations yields the Pearson correlation coefficient of −0.523, which is much better than MM/PBSA (r_p = −0.212). (2) The enhanced sampling methods (aMD and GaMD) are indeed more efficient for conformational sampling, but they cannot improve the binding free energy predictions for protein–protein systems, indicating that aMD and GaMD are not the best choice (at least in short timescale MD simulations) to sample the conformations for the MM/PB(GB)SA calculations of protein–protein complexes. However, aMD or GaMD may be good choice for sufficient sampling in a certain length of simulations. (3) A higher interior dielectric constant is necessary to improve the rescoring accuracy for MM/PBSA, especially for the protein–protein systems with higher polarized binding interfaces, and an interior dielectric constant of 1.0 is recommended to MM/GBSA. (4) The VDGB model exhibits potential power in binding free energy predictions and more thorough research should be done in the near future.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This study was supported by National Key R&D Program of China (2016YFB0201700), National Science Foundation of China (81603031, 21575128, 81773632), and Zhejiang Provincial Natural Science Foundation of China (LZ19H300001).

References

1 T. Vreven, H. Hwang, B. G. Pierce and Z. Weng, *Protein Sci.*, 2012, **21**, 396–404.
2 J. Andreani and R. Guerois, *Arch. Biochem. Biophys.*, 2014, **554**, 65–75.
3 G. Ramakrishnan, N. R. Chandra and N. Srinivasan, *IUBMB Life*, 2014, **66**, 759–774.
4 A. Metz, E. Ciglia and H. Gohlke, *Curr. Pharm. Des.*, 2012, **18**, 4630–4647.
5 D. Gonzalez-Ruiz and H. Gohlke, *Curr. Med. Chem.*, 2006, **13**, 2607–2625.
6 B. Nisius, F. Sha and H. Gohlke, *J. Biotechnol.*, 2012, **159**, 123–134.
7 K. Takemura, H. Guo, S. Sakuraba, N. Matubayasi and A. Kitao, *J. Chem. Phys.*, 2012, **137**, 215105.
8 I. Anishchenko, P. J. Kundrotas, A. V. Tuzikov and I. A. Vakser, *Proteins*, 2014, **82**, 278–287.
9 P. L. Kastritis and A. M. Bonvin, *J. Proteome Res.*, 2010, **9**, 2216–2225.
10 I. A. Vakser and P. Kundrotas, *Curr. Pharm. Biotechnol.*, 2008, **9**, 57–66.
11 S. Z. Wan, B. Knapp, D. W. Wright, C. M. Deane and P. V. Coveney, *J. Chem. Theory Comput.*, 2015, **11**, 3346–3356.
12 J. Srinivasan, J. Miller, P. A. Kollman and D. A. Case, *J. Biomol. Struct. Dyn.*, 1998, **16**, 671–682.
13 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889–897.
14 J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 1998, **120**, 9401–9409.
15 J. M. Wang, T. J. Hou and X. J. Xu, *Curr. Comput.-Aided Drug Des.*, 2006, **2**, 287–306.
16 E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. H. Zhang and T. Hou, *Chem. Rev.*, 2019, DOI: 10.1021/acs.chemrev.1029b00055.
17 I. Maffucci and A. Contini, *J. Chem. Inf. Model.*, 2016, **56**, 1692–1704.
18 T. Feng, F. Chen, Y. Kang, H. Sun, H. Liu, D. Li, F. Zhu and T. Hou, *J. Cheminf.*, 2017, **9**, 66.
19 F. Chen, H. Liu, H. Sun, P. Pan, Y. Li, D. Li and T. Hou, *Phys. Chem. Chem. Phys.*, 2016, **18**, 22129–22139.
20 R. Chen, L. Li and Z. Weng, *Proteins*, 2003, **52**, 80–87.
21 W. Jiang and B. Roux, *J. Chem. Theory Comput.*, 2010, **6**, 2559–2565.
22 V. Limongelli, M. Bonomi and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 6358–6363.
23 F. Zeller and M. Zacharias, *J. Comput. Chem.*, 2014, **35**, 2256–2262.
24 K. Kappel, Y. Miao and J. A. McCammon, *Q. Rev. Biophys.*, 2015, **48**, 479–487.
25 Y. Miao, F. Feixas, C. Eun and J. A. McCammon, *J. Comput. Chem.*, 2015, **36**, 1536–1549.
26 Y. Miao and J. A. McCammon, *Annu. Rep. Comput. Chem.*, 2017, **13**, 231–278.
27 Y. Miao, V. A. Feher and J. A. McCammon, *J. Chem. Theory Comput.*, 2015, **11**, 3584–3595.
28 Y. T. Wang and Y. H. Chan, *Sci. Rep.*, 2017, **7**, 7828.
29 J. M. Johnston and M. Filizola, *Curr. Opin. Struct. Biol.*, 2011, **21**, 552–558.
30 R. C. Bernardi, M. C. R. Melo and K. Schulten, *Biochim. Biophys. Acta*, 2015, **1850**, 872–877.
31 S. Iida, H. Nakamura and J. Higo, *Biochem. J.*, 2016, **473**, 1651–1662.
32 K. A. Sharp and B. Honig, *J. Phys. Chem.*, 1990, **94**, 7684–7692.
33 A. H. Boschitsch and M. O. Fenley, *J. Chem. Theory Comput.*, 2011, **7**, 1524–1540.
34 N. Forouzesh, S. Izadi and A. V. Onufriev, *J. Chem. Inf. Model.*, 2017, **57**, 2505–2513.
35 W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127–6129.
36 C. N. Schutz and A. Warshel, *Proteins*, 2001, **44**, 400–417.
37 I. Massova and P. A. Kollman, *Perspect. Drug Discovery Des.*, 2000, **18**, 113–135.
38 K. Ravindranathan, J. Tirado-Rives, W. L. Jorgensen and C. R. Guimaraes, *J. Chem. Theory Comput.*, 2011, **7**, 3859–3865.
39 P. Mikulskis, S. Genheden, P. Rydberg, L. Sandberg, L. Olsen and U. Ryde, *J. Comput.-Aided Mol. Des.*, 2012, **26**, 527–541.
40 C. R. Guimaraes and A. M. Mathiowetz, *J. Chem. Inf. Model.*, 2010, **50**, 547–559.
41 T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.
42 T. Hou, J. Wang, Y. Li and W. Wang, *J. Comput. Chem.*, 2011, **32**, 866–877.
43 S. Genheden and U. Ryde, *Proteins*, 2012, **80**, 1326–1342.
44 T. Yang, J. C. Wu, C. Yan, Y. Wang, R. Luo, M. B. Gonzales, K. N. Dalby and P. Ren, *Proteins*, 2011, **79**, 1940–1951.
45 H. Sun, Y. Li, M. Shen, S. Tian, L. Xu, P. Pan, Y. Guan and T. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 22035–22045.
46 T. Venken, D. Krnavek, J. Munch, F. Kirchhoff, P. Henklein, M. De Maeyer and A. Voet, *Proteins*, 2011, **79**, 3221–3235.
47 X. Liu, L. Peng and J. Z. H. Zhang, *J. Chem. Inf. Model.*, 2019, **59**, 272–281.
48 J. Kongsted, P. Soderhjelm and U. Ryde, *J. Comput.-Aided Mol. Des.*, 2009, **23**, 395–409.
49 G. Wiegand, O. Epp and R. Huber, *J. Mol. Biol.*, 1995, **247**, 99–110.
50 B. Kobe and J. Deisenhofer, *Nature*, 1995, **374**, 183–186.
51 K. Lapouge, S. J. Smith, P. A. Walker, S. J. Gamblin, S. J. Smerdon and K. Rittinger, *Mol. Cell*, 2000, **6**, 899–907.
52 N. Nassar, G. R. Hoffman, D. Manor, J. C. Clardy and R. A. Cerione, *Nat. Struct. Biol.*, 1998, **5**, 1047–1052.

53 H. Choe, L. D. Burtnick, M. Mejillano, H. L. Yin, R. C. Robinson and S. Choe, *J. Mol. Biol.*, 2002, **324**, 691–702.
54 M. E. Pacold, S. Suire, O. Perisic, S. Lara-Gonzalez, C. T. Davis, E. H. Walker, P. T. Hawkins, L. Stephens, J. F. Eccleston and R. L. Williams, *Cell*, 2000, **103**, 931–943.
55 T. Shiba, M. Kawasaki, H. Takatsu, T. Nogi, N. Matsugaki, N. Igarashi, M. Suzuki, R. Kato, K. Nakayama and S. Wakatsuki, *Nat. Struct. Biol.*, 2003, **10**, 386–393.
56 L. R. Otterbein, C. Cosio, P. Graceffa and R. Dominguez, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 8003–8008.
57 E. G. Huizinga, S. Tsuji, R. A. Romijn, M. E. Schiphorst, P. G. de Groot, J. J. Sixma and P. Gros, *Science*, 2002, **297**, 1176–1179.
58 Y. Bourne, P. Taylor and P. Marchot, *Cell*, 1995, **83**, 503–512.
59 J. M. Alexander, C. A. Nelson, V. van Berkel, E. K. Lau, J. M. Studts, T. J. Brett, S. H. Speck, T. M. Handel, H. W. Virgin and D. H. Fremont, *Cell*, 2002, **111**, 343–356.
60 A. Dementiev, M. Simonovic, K. Volz and P. G. Gettins, *J. Biol. Chem.*, 2003, **278**, 37881–37887.
61 J. R. Horn, S. Ramaswamy and K. P. Murphy, *J. Mol. Biol.*, 2003, **331**, 497–508.
62 E. Santelli, L. A. Bankston, S. H. Leppla and R. C. Liddington, *Nature*, 2004, **430**, 905–908.
63 K. Scheffzek, M. R. Ahmadian, W. Kabsch, L. Wiesmuller, A. Lautwein, F. Schmitz and A. Wittinghofer, *Science*, 1997, **277**, 333–338.
64 A. Pollet, S. Sansen, G. Raedschelders, K. Gebruers, A. Rabijns, J. A. Delcour and C. M. Courtin, *FEBS J.*, 2009, **276**, 4340–4351.
65 C. E. Schutt, J. C. Myslik, M. D. Rozycki, N. C. W. Goonesekere and U. Lindberg, *Nature*, 1993, **365**, 810–816.
66 W. A. Stanley, F. V. Filipp, P. Kursula, N. Schuller, R. Erdmann, W. Schliebs, M. Sattler and M. Wilmanns, *Mol. Cell*, 2006, **24**, 653–663.
67 J. E. Chrencik, A. Brooun, M. L. Kraus, M. I. Recht, A. R. Kolatkar, G. W. Han, J. M. Seifert, H. Widmer, M. Auer and P. Kuhn, *J. Biol. Chem.*, 2006, **281**, 28185–28192.
68 P. Peschard, G. Kozlov, T. Lin, I. A. Mirza, A. M. Berghuis, S. Lipkowitz, M. Park and K. Gehring, *Mol. Cell*, 2007, **27**, 474–485.
69 T. P. Ko, C. C. Liao, W. Y. Ku, K. F. Chak and H. S. Yuan, *Structure*, 1999, **7**, 91–102.
70 A. Sali and T. L. Blundell, *J. Mol. Biol.*, 1993, **234**, 779–815.
71 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, **25**, 247–260.
72 Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. M. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.
73 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
74 M. J. Harvey and G. De Fabritiis, *J. Chem. Theory Comput.*, 2009, **5**, 2371–2377.
75 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
76 R. Salomon-Ferrer, D. A. Case and R. C. Walker, *WIREs Comput. Mol. Sci.*, 2013, **3**, 198–210.
77 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
78 I. G. Rodriguez-Bussey, U. Doshi and D. Hamelberg, *Biopolymers*, 2016, **105**, 35–42.
79 T. Y. Shen and D. Hamelberg, *J. Chem. Phys.*, 2008, **129**, 034103.
80 Y. L. Miao, S. E. Nichols and J. A. McCammon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 6398–6406.
81 D. Hamelberg, C. A. F. de Oliveira and J. A. McCammon, *J. Chem. Phys.*, 2007, **127**, 155102.
82 T. J. Hou and R. Yu, *J. Med. Chem.*, 2007, **50**, 1177–1188.
83 H. Sun, L. Duan, F. Chen, H. Liu, Z. Wang, P. Pan, F. Zhu, J. Z. H. Zhang and T. Hou, *Phys. Chem. Chem. Phys.*, 2018, **20**, 14450–14460.
84 B. R. Miller, 3rd, T. D. McGee, Jr., J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.
85 C. H. Tan, L. J. Yang and R. Luo, *J. Phys. Chem. B*, 2006, **110**, 18680–18687.
86 A. Onufriev, D. Bashford and D. A. Case, *J. Phys. Chem. B*, 2000, **104**, 3712–3720.
87 J. Weiser, P. S. Shenkin and W. C. Still, *J. Comput. Chem.*, 1999, **20**, 217–230.
88 L. Xu, H. Sun, Y. Li, J. Wang and T. Hou, *J. Phys. Chem. B*, 2013, **117**, 8408–8421.
89 C. Y. Yang, *PLoS One*, 2015, **10**, e0118671.
90 S. S. Negi, C. H. Schein, N. Oezguen, T. D. Power and W. Braun, *Bioinformatics*, 2007, **23**, 3397–3399.
91 L. Duan, X. Liu and J. Z. Zhang, *J. Am. Chem. Soc.*, 2016, **138**, 5722–5728.
92 D. D. Huang, Y. F. Qi, J. N. Song and J. Z. H. Zhang, *J. Comput. Chem.*, 2019, **40**, 1045–1056.
93 Z. Sun, Y. N. Yan, M. Yang and J. Z. Zhang, *J. Chem. Phys.*, 2017, **146**, 124124.
94 H. Sun, L. Duan, F. Chen, H. Liu, Z. Wang, P. Pan, F. Zhu, J. Z. H. Zhang and T. Hou, *Phys. Chem. Chem. Phys.*, 2018, **20**, 14450–14460.
95 A. Vangone, R. Spinelli, V. Scarano, L. Cavallo and R. Oliva, *Bioinformatics*, 2011, **27**, 2915–2916.