PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2018, **20**, 14450

Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches†

Huiyong Sun,[ab] Lili Duan,[c] Fu Chen,[a] Hui Liu,[a] Zhe Wang,[a] Peichen Pan,[a] Feng Zhu,[a] John Z. H. Zhang[cd] and Tingjun Hou*[ab]

Entropy effects play an important role in drug–target interactions, but the entropic contribution to ligand-binding affinity is often neglected by end-point binding free energy calculation methods, such as MM/GBSA and MM/PBSA, due to the expensive computational cost of normal mode analysis (NMA). Here, we systematically investigated entropy effects on the prediction power of MM/GBSA and MM/PBSA using >1500 protein–ligand systems and six representative AMBER force fields. Two computationally efficient methods, including NMA based on truncated structures and the interaction entropy approach, were used to estimate the entropic contributions to ligand–target binding free energies. In terms of the overall accuracy, we found that, for the minimized structures, in most cases the inclusion of the conformational entropies predicted by truncated NMA ($\text{enthalpy}_{\text{nmode\_min\_9Å}}$) compromises the overall accuracy of MM/GBSA and MM/PBSA compared with the enthalpies calculated based on the minimized structures ($\text{enthalpy}_{\text{min}}$). However, for the MD trajectories, the binding free energies can be improved by the inclusion of the conformation entropies predicted by either truncated-NMA for a relatively high dielectric constant (ε_{in} = 4) or the interaction entropy method for ε_{in} = 1–4. In terms of reproducing the absolute binding free energies, the binding free energies estimated by including the truncated-NMA entropies based on the MD trajectories ($\Delta G_{\text{nmode\_md\_9Å}}$) give the lowest average absolute deviations against the experimental data among all the tested strategies for both MM/GBSA and MM/PBSA. Although the inclusion of the truncated NMA based on the MD trajectories ($\Delta G_{\text{nmode\_md\_9Å}}$) for a relatively high dielectric constant gave the overall best result and the lowest average absolute deviations against the experimental data (for the ff03 force field), it needs too much computational time. Alternatively, considering that the interaction entropy method does not incur any additional computational cost and can give comparable (at high dielectric constant, ε_{in} = 4) or even better (at low dielectric constant, ε_{in} = 1–2) results than the truncated-NMA entropy ($\Delta G_{\text{nmode\_md\_9Å}}$), the interaction entropy approach is recommended to estimate the entropic component for MM/GBSA and MM/PBSA based on MD trajectories, especially for a diverse dataset. Furthermore, we compared the predictions of MM/GBSA with six different AMBER force fields. The results show that the ff03 force field (ff03 for proteins and *gaff* with AM1-BCC charges for ligands) performs the best, but the predictions given by the tested force fields are comparable, implying that the MM/GBSA predictions are not very sensitive to force fields.

Received 12th November 2017,
Accepted 1st May 2018

DOI: 10.1039/c7cp07623a

rsc.li/pccp

[a] *College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China. E-mail: tingjunhou@zju.edu.cn; Tel: +86-571-88208412*
[b] *State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang 310058, China*
[c] *School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*
[d] *NYU-ECNU Center for Computational Chemistry, NYU Shanghai, Shanghai 200062, China*

† Electronic supplementary information (ESI) available: Fig. S1. Structure of thrombin-like proteins. The protein and co-crystallized ligand are shown in a yellow cartoon model and green stick model, respectively. A PDB code of 1C4U is used for the illustration. Table S1. Information on the used dataset. Table S2. Overall accuracy of MM/GBSA results based on various calculation protocols. Table S3. Overall accuracy of MM/PBSA results based on various calculation protocols. Table S4. Prediction accuracy of MM/GBSA results of kinases based on various calculation protocols. Table S5. Prediction accuracy of MM/GBSA results of HIV proteases based on various calculation protocols. Table S6. Prediction accuracy of MM/GBSA results of thrombin-like proteins based on various calculation protocols. See DOI: 10.1039/c7cp07623a

Introduction

Both chemical thermodynamics and kinetics play essential roles in drug–target interactions, in which the thermodynamics determines how tight a drug binds to its target, while the kinetic process, usually characterized by drug–target residence time,[1,2] determines how long a drug binds with its target. Increasing attention has been focused on the determination of drug residence times in recent years,[3–5] and numerous studies have shown that the drug residence time of a drug is highly correlated with its binding affinity,[6–8] suggesting that, besides the high kinetic barrier of the ligand–receptor interaction which keeps the ligand bound (enhancing the k_{off} value), tight binding (deep energetic basin or high binding affinity of the ligand–receptor interaction) is also needed to enhance the drug–target residence time (or drug efficacy).[2] Therefore, it is essential to accurately predict the binding affinity of a drug to its target. In past decades, numerous binding affinity estimation methods, from the theoretically simplified but computationally cheaper scoring functions for molecular docking to the theoretically rigorous but time-consuming alchemical approaches, have been proposed and used in solving various problems.[9–12] Roughly, the binding free energy calculation methods can be divided into two categories, that is, pathway-based methods and end-point methods.[13] The pathway-based methods are usually conducted directly with molecular dynamics (MD) simulations (that is, all the parameters are derived from the original simulations), where the system can be driven from one thermodynamic state to another by physical (such as umbrella sampling[14] and metadynamics[15] that usually use the drug–target distance or protein conformational change as the sampling pathways[16–21]) or non-physical pathways (such as alchemical methods of free energy perturbation (FEP)[22–25] and thermodynamic integration (TI)[26] that use a series of non-physical intermediate states to connect one state and another[27–29]). Different from the pathway-based methods, the end-point methods, including most scoring functions of molecular docking, the semi-empirical method of linear interaction energy (LIE),[30,31] and implicit solvent based methods such as Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) and Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA),[32,33] only sample the conformations for the initial and final states (free and bound states) and then calculate the free energy difference between the two states with different post-processing strategies. The pathway-based methods are generally more theoretically rigorous than the end-point methods, whereas, they are more time-consuming as well, which hinders the large-scale practical applications of the pathway-based methods. Therefore, the end-point methods represented by the MM/GBSA and MM/PBSA approaches, which employ a more physically meaningful framework than docking scoring functions,[32] are more promising to be used in the era of '*big data*', such as for virtual screening.

In the theoretical framework of MM/GBSA and MM/PBSA, the free energy can be split into several components, including the gas-phase potential energy, the polar and non-polar solvation free energies, and the entropy upon ligand–receptor interactions, which can be calculated independently based on the conformations extracted from the produced MD trajectories. The potential energy can be computed by molecular mechanics (MM) based on different force fields. The polar solvation free energy can be calculated by using either a Generalized-Born (GB) model or the Poisson–Boltzmann (PB) equation. The entropy is usually estimated by normal mode analysis (NMA). In the past decade, MM/GBSA and MM/PBSA may be the most popular methods for large-scale binding free energy calculations due to their acceptable accuracy, relatively low computational cost, and widely applicable scopes,[34] such as for small-ligand-protein systems,[35–42] protein–protein systems,[43–45] and protein–RNA/DNA systems,[46–48] which represent almost the whole interaction-omics of life science. In practice, many applications using MM/GBSA and MM/PBSA simply neglect the entropy change of protein–ligand binding due to the extremely expensive computational cost and relatively low prediction accuracy of NMA.[49]

Nevertheless, numerous attempts for accurately predicting entropy have been carried out by methods ranging from using post-processing approaches[50–54] to the simulation-synchronized methodologies.[55–57] For instance, Genheden *et al.* have shown that it may be a feasible method to use the truncated structures for the NMA entropy calculation to save computational cost.[50] And the same group also showed that most entropy estimation methods besides NMA can hardly converge even with a very long simulation time (such as 1 ms),[53,58] which hinders their applications in practice. To speed up the estimation of entropy with ligand–receptor interactions, Duan and co-workers recently proposed a new entropy estimating method called the interaction entropy calculation, where the entropy can be derived directly from the partition function of the ligand–receptor interactions and represents the fluctuations of the energetic components upon ligand–receptor interactions during MD simulations.[52,59] Moreover, by using the Clausius–van't Hoff method, Sharp calculated the molecular entropy with temperature integration and found that this method can give results consistent with the Boltzmann-quasiharmonic[60] method.[55] Choi *et al.* estimated the hydration entropy of a set of organic molecules with the free energy perturbation (FEP) method,[56] where the electrostatic and hydrophobic parts of the hydration entropy were estimated by the temperature-based finite difference method and the scaled-particle theory, respectively, and a high correlation ($r^2 > 0.8$) between the predicted hydration entropies and the experimental data was observed. Gyimesi and colleagues provided a method to calculate the conformational entropy based on the structure ensemble regarded as a continuous space and treated with the Gaussian mixture function,[57] and they found that this method is effective in discriminating structures with little conformational entropy difference. However, most reported entropy calculation methods need hundreds of nanoseconds simulation time to precisely estimate the entropies and are only tested on a set of small molecules with at most several hundreds of atoms, and thus may be not feasible for the entropy calculations for more complicated systems such as drug–target systems.

As far as we know, until now, no comprehensive analyses have been reported to recommend persuasive guidelines for

entropy calculations with MM/GBSA and MM/PBSA. Therefore, in this study, by using various calculating protocols, we systematically investigated the entropy effects on MM/GBSA and MM/PBSA based on a dataset with more than 1500 protein–ligand systems. Two entropy calculating methods, including NMA based on truncated structures (referred to as truncated NMA) and the interaction entropy method (referred to as IE), were used in our analyses because these two methods are computationally efficient and do not need too many parameters to be set up. The whole study is organized as follows: (1) to test whether the conformational entropies calculated by truncated NMA are stable and accurate, we compared the NMA entropies (NMEs) based on the 9 Å-truncated structures and the whole structures of 99 kinase systems; (2) we evaluated the impact of the entropy changes for protein–ligand binding predicted by truncated NMA and IE on the predictions of MM/GBSA and MM/PBSA using >1500 crystal structures under three calculating conditions (different force fields, dielectric constants, and simulation protocols); (3) to examine the system dependence of the calculation methods, we analyzed the entropy effects on the three largest protein families in our dataset (kinases, HIV proteases, and thrombin-like proteins).

Materials and methods

Dataset preparation

1508 systems derived from the refined PDBbind database[61–63] were used for the evaluation, and they are all metal-free protein–ligand complexes with the sequence length of the proteins less than 1000 (the PDB codes and the corresponding experimental data are listed in Table S1 in the ESI†). The *antechamber* and *tleap* modules in the *amber14* simulation package[64] were used to prepare the initial systems.[65] Due to the relatively high computational speed and good performance of binding free energy calculations,[66] the AM1-BCC charges (AM1 with bond charge corrections)[67] for each small molecule were calculated and used in the following calculations. The general Amber force field (*gaff*, version 1.7)[68] was used to parameterize the small molecule ligands. Six different Amber force fields were used for the proteins, including ff02[69] (with *ipol* = 0, namely, the non-polarized version, which performs the best for the protein–protein systems for MM/GBSA calculations),[43] ff03[70] (which performs the best for the protein–ligand systems for MM/GBSA calculations),[66] ff99,[71] ff99SB,[72] ff99SBildn,[73] and ff14SB (the latest force field in *amber14*).[74] All the residues were protonated according to the default parameters in *antechamber* (for example the HIS residue was parameterized to HIE and CYS involved in a disulfide bond was replaced by CYX). Counterions of Na^+ and Cl^- were added to neutralize the redundant charges of the protein–ligand complexes. Each protein–ligand complex was immersed in a cuboid TIP3P water box[75] extended by 10 Å out of the boundary of the solute.

Molecular mechanics (MM) minimizations and molecular dynamics (MD) simulations

As shown in our previous study, compared with long-time MD trajectories, the minimized structures and short-time MD trajectories may give a better prediction accuracy for both MM/PBSA and MM/GBSA.[49,66] Thus, here, molecular mechanics (MM) minimization and 1 ns MD simulation were performed for each system (6 force fields × 1508 systems). For both MM minimization and MD simulation, the cutoff of the nonbonded interactions (van der Waals interaction and short-range electrostatic interaction) was set to 10 Å and the long-range electrostatic interactions were handled by the PME (particle mesh Ewald) algorithm.[76] All the MM minimizations and MD simulations were performed with the *pmemd* and *pmemd.cuda* modules in *amber14*, respectively.

In the stage of MM minimization, the following four-step procedure was employed: (1) all the hydrogen atoms were relaxed for 1000 steps with all the heavy atoms restrained at 5 kcal mol^{-1} $Å^2$ (500 cycles of steepest descent and 500 cycles of conjugate gradient minimization); (2) the heavy atoms in the protein and ligand were restrained at 5 kcal mol^{-1} $Å^2$ and the other atoms were free to move (500 cycles of steepest descent and 500 cycles of conjugate gradient minimization); (3) only the heavy atoms in the backbone of the protein were restrained (5 kcal mol^{-1} $Å^2$) with the other atoms set free (500 cycles of steepest descent and 500 cycles of conjugate gradient minimization); and (4) the whole system was released and minimized for 5000 steps (1000 cycles of steepest descent and 4000 cycles of conjugate gradient minimization).

In the stage of MD simulation, all the covalent bonds between the heavy atoms and the hydrogen atoms were constrained with the SHAKE algorithm.[77] Another three-step MD simulation was performed for each system: (1) each system was at first heated from the minimized structure in an *NVT* ensemble, where the temperature was increased from 0 to 300 K within 250 ps with the backbone atoms of the protein restrained at 2 kcal mol^{-1} $Å^2$; next, the system was equilibrated for another 250 ps in an *NPT* ensemble (P = 1 atm and T = 300 K) with the heavy atoms in the protein backbone restrained as well (2 kcal mol^{-1} $Å^2$); finally, 1 ns MD simulation was carried out for each system in the *NPT* ensemble without any restraints. The time step was set to 2 fs, and the time interval for data collection was set to 5 ps (200 snapshots were collected for each system in each force field).

End-point binding free energy calculations

The end-point binding free energy including MM/GBSA and MM/PBSA in conjunction with two entropy estimating methods (truncated NME and IE) was calculated for each system. The formulas for calculating the end-state binding free energies and their decomposed energetic components are shown in eqn (1)–(4), where the total binding free energy (ΔG_{bind}) represents the free energy difference between the bound-state complex (G_{com}) and the free-state individuals of the receptor and the ligand (G_{rec} + G_{lig}) (eqn (1)). According to the second law of thermodynamics, ΔG_{bind} can be decomposed into the enthalpy part (ΔH) and the entropy part ($-T\Delta S$) (eqn (1)). Here, the enthalpies were calculated by MM/GBSA and MM/PBSA, and the entropies were estimated with NMA and IE. The calculation details can be found below.

$$\Delta G_{bind} = G_{com} - (G_{rec} + G_{lig}) = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (1)$$

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{ele} + \Delta E_{vdW} \quad (2)$$

$$\Delta G_{sol} = \Delta G_{pol} + \Delta G_{np} \quad (3)$$

$$\Delta G_{np} = \gamma \Delta A + b \quad (4)$$

Enthalpy calculations with MM/GBSA and MM/PBSA

As shown in eqn (1), the enthalpy part can be further expressed as the summation of the molecular mechanical energy (ΔE_{MM}) and the solvation free energy (ΔG_{sol}), where ΔE_{MM} is composed of the intra-molecular energy ($\Delta E_{internal}$, including the bond, angle, and dihedral energies of the system), the electrostatic energy (ΔE_{ele}), and the van der Waals interactions (ΔE_{vdW}) (eqn (2)). In this study, $\Delta E_{internal}$ can be exactly canceled because the single MD trajectory protocol is used in the MM/GBSA and MM/PBSA calculations. The solvation free energy (ΔG_{sol}) is also composed of two parts, namely the polar part (ΔG_{pol}) and non-polar (ΔG_{np}) part (eqn (3)), where ΔG_{pol} is usually computed by the Generalized Born (GB) model or by solving the Poisson–Boltzmann (PB) equation, while ΔG_{np} is estimated by the solvent accessible surface area (SASA)-based approach. Here, the modified GB model developed by Onufriev (GB^{OBC1} with the default radii)[78] and the radii optimized by Tan (PB^{pbsa})[79] were used for the MM/GBSA and MM/PBSA calculations, respectively, as these models performed well in our previous study.[49,66] A series of interior dielectric constants (ε_{in} = 1–4 with the interval 0.5) were used for the MM/GBSA calculations, while three different interior dielectric constants (1, 2, and 4) were used for the MM/PBSA calculations. It should be noted that, in this study, we did not try to use all the calculation protocols for the MM/PBSA calculations because MM/PBSA performs worse than MM/GBSA in all cases in our previous studies.[49,66,80–82] The exterior dielectric constant was set to 80 and the ionic strength was set to 0. As for the non-polar part of the solvation effect (ΔG_{np}), ΔG_{np} was calculated by eqn (4) with the LCPO algorithm based on SASA (ΔA),[83] where γ and b were set to 0.0072 and 0, respectively. All the MM/GBSA (including 6 force fields) and MM/PBSA (only for the ff03 force field) calculations were carried out based on the minimized structures and the 1 ns MD trajectories (200 frames).

Conformational entropy calculations based on NMA

NMA was used for the conformational entropy calculations. To save computational resource, truncated structures were employed to calculate the NMA conformational entropies (truncated NMEs), where a cutoff of 9 Å extended out of the ligand boundary was set to truncate the proteins, and the whole residue is incorporated into the corresponding truncated structure if any heavy atom in this residue is located within the truncation sphere. All the discontinuous residues were treated as charged terminals ($-NH3^+$ or $-COO^-$), and the whole truncated system was re-treated by the *tleap* module in *amber14* to generate the initial coordinates and parameters. The maximum optimizing steps and convergence condition were set to 10 000 and 1×10^{-4}, respectively. The truncated NMEs were calculated for all the minimized structures (6 force fields) and the MD trajectories of the ff03 force field (including 1508 systems, with the snapshot interval of 10, namely, 20 frames for each system). In addition, this method has also been carried out for kinase systems of the other 5 force fields (including 99 systems for each force field). Moreover, to give a comparison, we also calculated the NMEs of the full-length MD structures of kinases based on the ff03 force field (referred to as full-length NME, 20 frames for each system). All the parameters of the full-length NME calculations were similar to those of the truncated ones.

Interaction entropy calculations

The newly developed entropy calculation method (called interaction entropy, IE) developed by Duan *et al.*[52] was also used for calculation of the binding free energy. In the spirit of IE, it can be derived from the partition function of the gas-phase ligand–receptor interactions (the derivation process can be found in ref. 52) and represents the fluctuations of ligand–receptor interactions during MD simulations. Eqn (5) shows the components for calculating IE, where $\langle\rangle$ represents the ensemble average of the ligand–receptor interactions, and ΔE_{ele} and ΔE_{vdW} are the electrostatic and van der Waals interactions between the ligand and receptor, respectively. β, K, and T are the inverse of the temperature ($1/kT$), the Boltzmann constant, and the simulation temperature (300 K), respectively. Note that the calculation of IE incorporates electrostatic interactions (ΔE_{ele}) which vary according to the choice of interior dielectric constants. All the IEs were computed based on the 1 ns MD trajectories.

$$-T\Delta S = KT \ln \left\langle e^{\beta(\Delta E_{ele} + \Delta E_{vdW})} \right\rangle \quad (5)$$

Results and discussion

Validation of truncated NME calculations

Genheden *et al.* have shown that the truncation of a whole protein–ligand complex into a sphere will not markedly affect the predicted conformational entropies.[50] However, they only compared the relative conformational entropies for the 8–16 Å truncated structures using 18 ligand–receptor complexes. Thus, we do not know the absolute difference of the conformational entropies for a full-length structure and a truncated one. In order to answer this question, we calculated the NMEs based on the full-length and the 9 Å-truncated structures (using Genheden's procedure[50] without a fixed buffer region) for 99 kinases contained in our dataset due to their feasible sequence length for NMA (230–360 amino acids). As shown in Fig. 1A, a high Pearson correlation (r_p = 0.86) is observed between the NMEs predicted by these two calculation strategies, where the slope of the regression line is close to 1 (blue solid line in Fig. 1A) and 81 systems (81.8%) with 95%-confidence error-bars cover the baseline of the absolute entropy difference (red dotted line in Fig. 1B), suggesting that the truncation strategy can be used in large-scale NME calculations. Thus, the following NMEs were calculated based on the 9 Å-truncated structures for all the minimized structures with 6 force fields (1508 systems × 1 frame × 6 force fields)

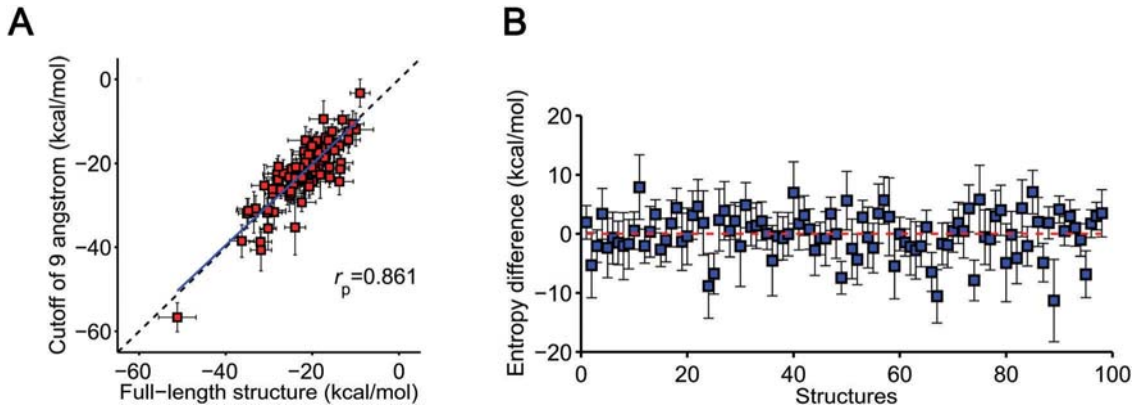

Fig. 1 Correlation between the full-length NMEs and the 9 Å-truncated NMEs (panel A) and the entropy difference of the two calculation strategies (entropy_difference = $NME_{9Å} - NME_{full\text{-}length}$, panel B), where all the errors were estimated from 95% confidence of the standard errors of the mean (s.e.m). The combined s.e.m is shown in panel B.

and MD structures with the ff03 force field (1508 systems × 20 frames × 1 force field).

Entropy effects on the performance of MM/GBSA and MM/PBSA

Fig. 2 summarizes the results of the current issues of wide concern on the performance of MM/GBSA, including entropy effects, force field effects, and dielectric constant effects. Panels A–C in Fig. 2 show the r_p-based heat maps (Pearson correlation coefficient) of the entropy effects, force field effects, and dielectric constant effects (only the dielectric constants of 1, 2, and 4 are shown) on MM/GBSA. To clearly explore the dependence of the dielectric constants on the performance of MM/GBSA for each simulation protocol, linear illustrations are shown in panels D–G in Fig. 2 as well. Detailed discussion is given below:

The entropy effects on MM/GBSA

Here, we added two entropy terms (IE and truncated NME) for the MM/GBSA calculations (IE was calculated based on the 1 ns MD trajectories and NME was estimated based on the 9 Å-truncated minimized structures for 6 force fields and 1 ns MD trajectories for the ff03 force field) and the corresponding binding free energies are termed $\Delta G_{interaction\_entropy}$, $\Delta G_{nmode\_min\_9Å}$, and $\Delta G_{nmode\_md\_9Å}$, respectively. To give a comparison, the enthalpies based on the minimized structure and the 1 ns MD trajectory were also calculated for each system

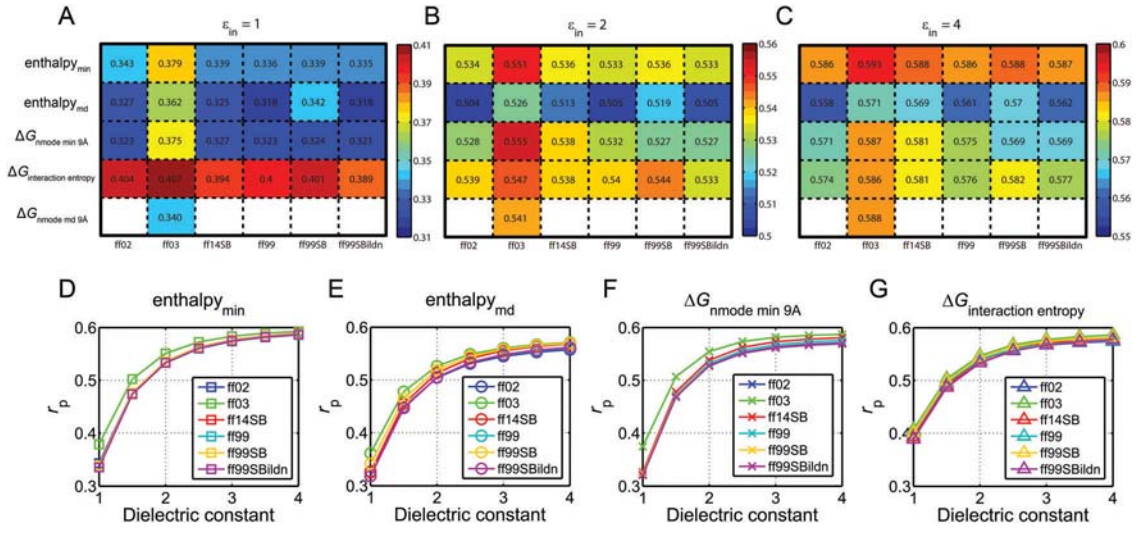

Fig. 2 Overall accuracies of MM/GBSA results based on various calculation protocols. Comparisons of the force fields and entropies on the performance of MM/GBSA are shown in the style of heat map in panels A–C (corresponding to dielectric constants of 1, 2, and 4, respectively). To capture the full impacts of the various simulation protocols on the performance of MM/GBSA, a comparison is also shown in line form in panels D–G (which correspond to the enthalpies of the minimized structures ($enthalpy_{min}$), enthalpies of the MD trajectories ($enthalpy_{md}$), binding free energies based on 9 Å-truncated NMEs of the minimized structures ($\Delta G_{nmode\_min\_9Å}$), and binding free energies based on IEs of the MD trajectories ($\Delta G_{interaction\_entropy}$), respectively). Note that the all-dataset truncated NMEs were only performed for the ff03 force field, thus no line form was shown of $\Delta G_{nmode\_md\_9Å}$. All the data reported here are Pearson correlation coefficients between the predicted binding affinities (enthalpy or binding free energy) and the experimental data. The detailed data and the corresponding standard deviations are illustrated in Table S2 (ESI†).

(termed as enthalpy$_{min}$ and enthalpy$_{md}$, respectively). As shown in each column in Fig. 2A, at the low interior dielectric constant (ε_{in} = 1), $\Delta G_{interaction\_entropy}$ performs the best among the five calculation strategies for each force field (enthalpy$_{min}$, enthalpy$_{md}$, $\Delta G_{interaction\_entropy}$, $\Delta G_{nmode\_min\_9Å}$, and $\Delta G_{nmode\_md\_9Å}$), where r_p reaches up to nearly or higher than 0.4 and is higher than the corresponding r_p resulting from the minimized structures (enthalpy$_{min}$). With the increase of the solute dielectric constant, the overall accuracy of MM/GBSA increases remarkably for each calculation strategy (Fig. 2D–G). However, the increase of the solute dielectric constant attenuates the advantage of the IE effect on the accuracy of MM/GBSA because the calculation of IE is associated with the electrostatic effects that can be attenuated by the increase of the solute dielectric constant (eqn (5)). Nevertheless, for a relatively high interior dielectric constant (ε_{in} = 4), the accuracy of the IE-based binding free energies ($\Delta G_{interaction\_entropy}$) is still higher than the corresponding MD-based enthalpies (enthalpy$_{md}$) for each force field (Fig. 2C), suggesting that MM/GBSA can benefit from the addition of IEs in the MD-based binding free energy calculations. Compared with IEs, the addition of the truncated NMEs does not have obvious impact on the overall accuracy of MM/GBSA for the minimized structures ($\Delta G_{nmode_min_9Å}$) in most cases, but shows benefit for the MD trajectories ($\Delta G_{nmode_md_9Å}$), where the r_ps improves markedly with the increase of dielectric constant (Fig. 2C). Thus, when estimating binding affinities for a diverse dataset, we do not recommend to include NME for MM/GBSA calculations for the minimized structures, whereas, we recommend to include IE for any dielectric constant calculations and NME for high dielectric constant (such as ε_{in} = 4) calculations for the MD trajectories.

The force field effects on MM/GBSA

Our previous work has shown that the ff03 force field may be the best choice for MM/GB(PB)SA.[66] However, this conclusion is made based on the analyses of a small dataset with only 5 drug targets in complex with 46 small molecules, and therefore the result may be biased by the limited dataset. To further validate this issue, 6 force fields were used for the MM/GBSA calculations based on the 1508 complexes. As shown in Fig. 2D–G, the ff03 force field (green lines) always performs the best across all the tested force fields for each calculation strategy at any dielectric constant, though little discrimination is shown for each force field at a relatively high dielectric constant (0.586–0.593 for enthalpy$_{min}$, 0.558–0.571 for enthalpy$_{md}$, 0.574–0.586 for $\Delta G_{interaction\_entropy}$, and 0.569–0.587 for $\Delta G_{nmode\_min\_9Å}$ at ε_{in} = 4). The results indicate that the ff03 force field may be the best choice for the MM/GBSA calculations based on short-time MD simulations (or the minimized structures), but the other force fields are also reasonable as they can give comparable results compared with the ff03 force field as well (especially at a relatively high dielectric constant).

The entropy effects on MM/PBSA

In addition to the calculation of the binding free energies with MM/PBSA without considering the entropic contribution (the case of enthalpy), MM/PBSA performs worse than MM/GBSA in many aspects such as ranking the binding free energies for a set of diverse drug–target complexes[80] or derivatives targeting the same target.[49,66] To investigate the entropy effects on the performance of MM/PBSA, we calculated the binding free energies with the above illustrated five strategies (enthalpy$_{min}$, enthalpy$_{md}$, $\Delta G_{interaction\_entropy}$, $\Delta G_{nmode\_min\_9Å}$, and $\Delta G_{nmode\_md\_9Å}$). To save computational cost, the MM/PBSA calculations were only carried out for the ff03 force field because this force field performs the best among the 6 tested force fields in the MM/GBSA calculations. As shown in Fig. 3, the r_p of the enthalpies calculated based on the minimized structures (enthalpy$_{min}$, red bar) is lower than that based on the 1 ns MD trajectories (enthalpy$_{md}$, blue bar) at each dielectric constant (ε_{in} = 1, 2, and 4) and is consistent with our previous study.[80] As for the impact of the truncated NMEs on MM/PBSA for the minimized structures, the r_ps of $\Delta G_{nmode\_min\_9Å}$ (orange bars in Fig. 3) are all worse than the enthalpy based on the minimized structures (enthalpy$_{min}$, red bars in Fig. 3) at the tested dielectric constants, implying that, similar to MM/GBSA, inclusion of the truncated NMEs cannot improve the MM/PBSA results, either, for a diverse dataset with the minimized structures. For the MD trajectories, although the inclusion of NME makes the prediction accuracy worse at a low dielectric constant (such as for ε_{in} = 1, the cyan bar is much lower than the blue bar, Fig. 3), it is encouraging to see that the predictions could gain benefit from the addition of NME at a relatively high dielectric constant (ε_{in} = 4, the cyan bar is a bit higher than the blue bar). Besides, consistent with the MM/GBSA results,

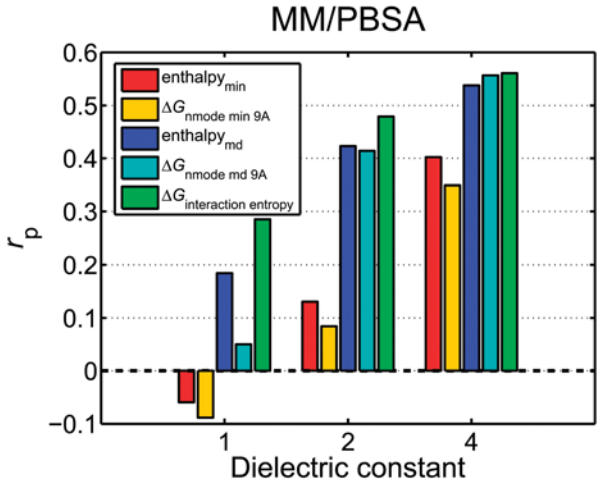

Fig. 3 Pearson correlation coefficients of the MM/PBSA results based on various calculation protocols (enthalpies of the minimized structures (enthalpy$_{min}$, red bar), binding free energies based on 9 Å-truncated NMEs of the minimized structures ($\Delta G_{nmode_min_9Å}$, orange bar), enthalpies of the MD trajectories (enthalpy$_{md}$, blue bar), binding free energies based on 9 Å-truncated NMEs of the MD trajectories ($\Delta G_{nmode_md_9Å}$, cyan bar), and binding free energies based on IEs of the MD trajectories ($\Delta G_{interaction_entropy}$, green bar)). The ff03 force field was used for the comparison (r_p = 0.561, 0.557, and 0.538 for $\Delta G_{interaction\_entropy}$, $\Delta G_{nmode\_md\_9Å}$, and enthalpy$_{md}$, respectively, at ε_{in} = 4). The detailed data and the corresponding standard deviations are illustrated in Table S3 (ESI†).

the addition of IEs can improve the MM/PBSA results as well at each dielectric constant (green bar in Fig. 3), indicating that IE is feasible for use in MM/PBSA binding free energy calculations based on MD trajectories. Nevertheless, although remarkable improvement has been achieved by adding IE to the MM/PBSA calculations, the resulted r_ps are still lower than the corresponding results provided by MM/GBSA (such as 0.559 *versus* 0.586 for MM/PBSA and MM/GBSA, respectively, at ε_{in} = 4). Thus, to a large extent, MM/GBSA may be a more suitable option for end-point binding free energy calculations (especially for a diverse dataset).

Capability to reproduce the absolute binding free energy for each strategy

Beside analyzing the Pearson correlation between the five strategies, we also compared the capability of MM/GBSA and MM/PBSA to reproduce the absolute binding affinity for the investigated strategies based on the ff03 force field. As illustrated in Table 1, consistent results are shown for MM/GBSA and MM/PBSA for all the calculating strategies, where the binding free energies estimated based on the MD trajectories ($\Delta G_{nmode\_md\_9Å}$) give the lowest absolute deviation on average compared with the other methods for both MM/GBSA and MM/PBSA. The reason why the average absolute deviations of the binding free energies calculated by the minimized structures ($\Delta G_{nmode\_min\_9Å}$) are larger than the corresponding $\Delta G_{nmode_md_9Å}$s may be attributed to the fact that the optimized single structure can always give the lowest global energy, which usually leads to the tightest binding between the ligand and receptor. Compared with the average absolute deviation of $\Delta G_{nmode_md_9Å}$s, the average absolute deviations of $\Delta G_{interaction\_entropy}$ are larger and increase markedly with the increase of the solute dielectric constant. This is because the IE algorithm decreases the electrostatic effect (eqn (5)) with the increase of the solute dielectric constant. Therefore, taken as a whole, in terms of reproducing the absolute binding affinities, the NME-MD based method ($\Delta G_{nmode\_md\_9Å}$) can give the best results for both MM/GBSA and MM/PBSA, though it still largely overestimates the binding affinities compared with the experimental data.

System dependence of the entropy effects on the performance of MM/GBSA

Numerous studies have shown that system dependence prevails for almost all ligand–target binding estimation methods, from the high-speed docking scoring approaches to the time-consuming alchemical approaches.[80,84,85] Thus, here, we also explored the system dependence of the entropy effects on the performance of MM/GBSA for several representative drug targets such as kinases (99 samples), HIV proteases (178 samples), and thrombin-like proteins (173 samples), which take up to ~30% of the systems in the dataset.

Table 1 Average absolute deviations between the experimental data and the MM/GB(PB)SA results based on the ff03 force field (kcal mol^{-1})

	MM/GBSA			MM/PBSA		
ε_{in}	1	2	4	1	2	4
$Enthalpy_{min}$	42.2	43.2	43.8	36.7	40.9	43.7
$Enthalpy_{md}$	35.1	38.1	39.7	31.9	37.4	40.3
$\Delta G_{nmode\_min\_9Å}$	20.8	21.0	21.7	22.6	20.1	21.6
$\Delta G_{nmode\_md\_9Å}$	14.7	16.0	17.4	17.6	16.1	18.0
$\Delta G_{interaction\_entropy}$	22.4	28.9	33.4	21.6	28.3	34.0

The entropy effects on kinases

Fig. 4 summarizes the entropy effects on the performance of MM/GBSA for the kinase systems, where the NMEs based on the full-length structures were only computed for the ff03 force field (purple line in Fig. 4A) because it performed the best among the 6 tested force fields. As shown in Fig. 4, to a large extent, the tendency of the entropy effects varies from each other in different force fields and calculation strategies. However, consistent results are observed at a low dielectric constant (ε_{in} = 1), where the MM/GBSA calculations with the addition of entropies (including NMEs and IEs; orange, cyan, purple, and green lines) perform worse than the corresponding results based on the minimized structures (red line) and the MD trajectories (blue line) for all the tested force fields. With the increase of the dielectric constant, different entropic effects exhibit for different force fields. In addition to the ff03 force field (6 calculation strategies were compared, Fig. 4A), although the binding free energies with the full-length NMEs ($\Delta G_{nmode\_md\_full\text{-}length}$, purple line in Fig. 4A) are better than those calculated with the 9 Å-truncated NMEs ($\Delta G_{nmode\_md\_9Å}$, cyan line in Fig. 4A), the full-length NMEs still perform worse than the MD-based enthalpies ($enthalpy_{md}$, blue line in Fig. 4A). And similar results are also observed for the three ff99 based force fields (ff99, ff99SB, and ff99SBildn force fields), where the cyan lines are all lower than the corresponding blue lines (panels D–E in Fig. 4). In addition, other consistent results are observed for the binding free energies calculated with the truncated NMEs based on the minimized structures ($\Delta G_{nmode\_min\_9Å}$, orange lines in Fig. 4), where all the orange lines are lower than the corresponding red lines. Thus, all the above data implies that, in most cases, the addition of NMEs cannot improve the prediction accuracy of MM/GBSA for the kinases systems. Nevertheless, $\Delta G_{nmode\_md\_9Å}$ performs well for the ff02 and ff14SB force fields at a relatively high dielectric constant (ε_{in} = 4). Moreover, although the performance of $\Delta G_{interaction\_entropy}$ (green lines in Fig. 4) is not satisfactory, either, even at a relatively high dielectric constant (ε_{in} = 4), the ff03 (Fig. 4A) and ff99SBildn (Fig. 4F) force fields perform the best among the tested strategies and the r_ps with $\Delta G_{interaction\_entropy}$ based on the ff14SB, ff99, and ff99SB force fields (green lines) are also higher than (or comparable to) the corresponding enthalpies calculated based on the MD trajectories ($enthalpy_{md}$, blue lines). In summary, considering the calculation of IEs will not incur any additional computational resource, it seems feasible to add IEs to the MD-based MM/GBSA calculations with the above recommended force fields (except the ff02 force field) for the kinase systems.

The entropy effects on HIV proteases

Compared with the kinase systems, very consistent results are observed for the HIV proteases for the four calculating

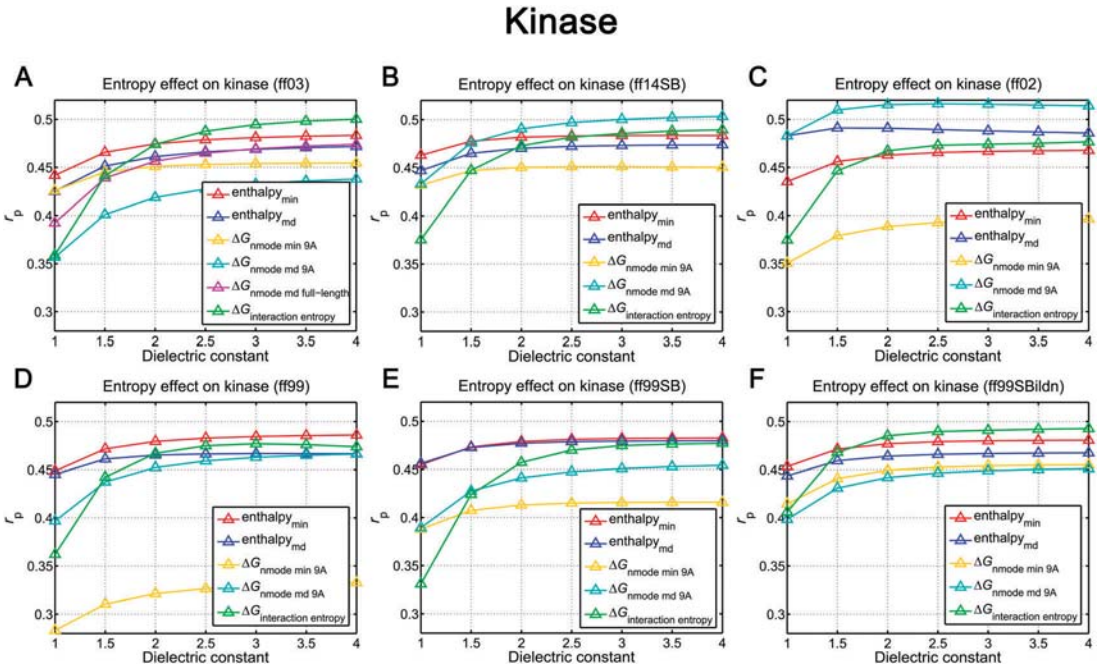

Fig. 4 Entropy effects on kinases (99 systems). Pearson correlation coefficients are used for the comparison, where binding free energies based on IEs ($\Delta G_{\text{interaction\_entropy}}$), 9 Å-truncated NMEs with minimized structures ($\Delta G_{\text{nmode\_min\_9Å}}$) and the MD trajectory ($\Delta G_{\text{nmode\_md\_9Å}}$), and enthalpies based on the minimized structure ($\text{enthalpy}_{\text{min}}$) and the MD trajectories ($\text{enthalpy}_{\text{md}}$) are shown with green, yellow, cyan, red, and blue lines, respectively. Full-length MD-based NMEs ($\Delta G_{\text{nmode\_md\_full-length}}$) were only used for the ff03 force field (purple line in panel A). The detailed data and the corresponding standard deviations are illustrated in Table S4 (ESI†).

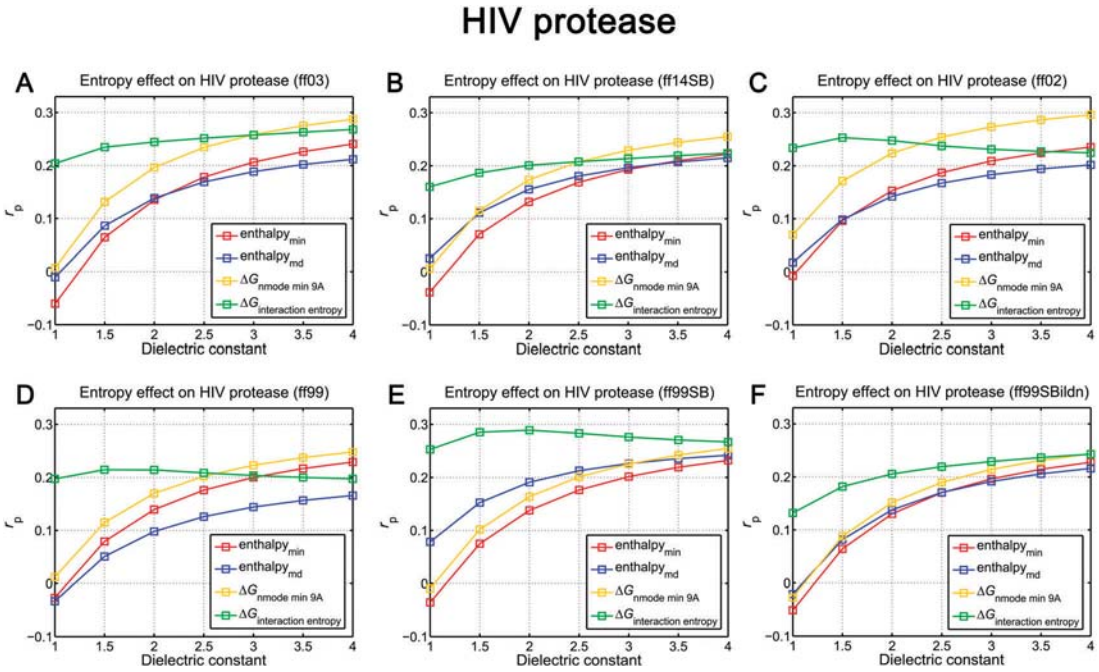

Fig. 5 Entropy effects on HIV proteases (178 systems). Pearson correlation coefficients are used for the comparison, where binding free energies based on IEs ($\Delta G_{\text{interaction\_entropy}}$), 9 Å-truncated NMEs with minimized structures ($\Delta G_{\text{nmode\_min\_9Å}}$), and enthalpies based on the minimized structures ($\text{enthalpy}_{\text{min}}$) and the MD trajectories ($\text{enthalpy}_{\text{md}}$) are shown by green, yellow, red, and blue lines, respectively. The detailed data and the corresponding standard deviations are illustrated in Table S5 (ESI†).

strategies with the 6 tested force fields (Fig. 5). For example, $\Delta G_{\text{interaction\_entropy}}$ (green lines in Fig. 5) always performs the best among the four calculation strategies at $\varepsilon_{\text{in}} = 1$ among all the tested force fields. Besides, the truncated NMEs also show a

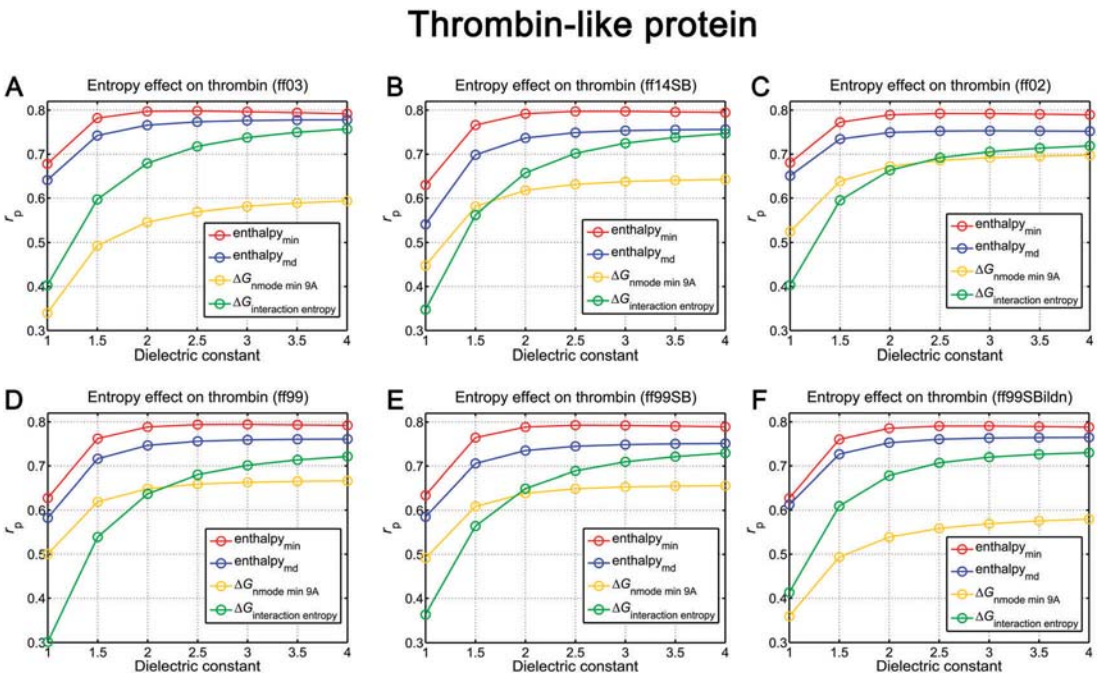

Fig. 6 Entropy effects on thrombin-like proteins (173 systems). Pearson correlation coefficients are used for the comparison, where binding free energies based on IEs ($\Delta G_{\text{interaction\_entropy}}$), 9 Å-truncated NMEs with minimized structures ($\Delta G_{\text{nmode\_min\_9Å}}$), and enthalpies based on the minimized structures ($\text{enthalpy}_{\text{min}}$) and the MD trajectories ($\text{enthalpy}_{\text{md}}$) are shown by green, yellow, red, and blue lines, respectively. The detailed data and the corresponding standard deviations are illustrated in Table S6 (ESI†).

positive effect on improving the performance of MM/GBSA for the HIV protease systems, where the orange lines ($\Delta G_{\text{nmode\_min\_9Å}}$) are always higher than the corresponding red lines ($\text{enthalpy}_{\text{min}}$). These results mean that the entropic contribution is necessary for the drugs binding to the HIV protease, and it has been validated by numerous studies that the ligand binding to the HIV protease is an entropy-enthalpy process.[86–88] Nevertheless, the r_ps of all the tested strategies are lower than 0.3, meaning that it is still urgent to develop system-specific calculation strategies for the worse-performing systems.

The entropy effects on thrombin-like proteins

Similar to the HIV protease systems, very consistent results are observed for the thrombin-like proteins for the four calculation strategies with the 6 force fields (Fig. 6). However, unfortunately, the entropic contributions exhibit negative effects on the performance of MM/GBSA in all cases (*i.e.* r_p of $\Delta G_{\text{nmode\_min\_9Å}} < \Delta G_{\text{interaction\_entropy}} < \text{enthalpy}_{\text{md}} < \text{enthalpy}_{\text{min}}$ at $\varepsilon_{\text{in}} = 4$ for all the tested force fields), meaning that the currently widely used entropy calculation methods may be not suitable for thrombin-like proteins. The deep reason may be attributed to the surface-located binding sites of the thrombin-like proteins (Fig. S1, ESI†), which need low conformational change when binding with a ligand, thus leading to lower conformational entropies that cannot be effectively captured by entropy calculation approaches. Thus, it depends, whether to incorporate entropy effects into the end-point binding free energy calculations, where entropic contributions may be necessary for the systems with deep or flexible binding pockets (such as kinases and HIV proteases) and less important for the systems with surface-located active sites (such as thrombin-like proteins).

Conclusion

In this study, we comprehensively investigated the entropy effects on the performance of the MM/GBSA and MM/PBSA approaches. Although the NMEs calculated with the truncated structures are quite consistent with those calculated with the full-length structures, NMA performs worse and cannot improve the accuracy of MM/GBSA and MM/PBSA in most cases for the minimized structures. Thus, one should be cautious of estimating ligand-binding free energies with NMA when using the minimized structures. Different from the results of NMEs on the minimized structures, for the MD trajectories, the inclusion of the truncated NMEs can improve the performance of both MM/GBSA and MM/PBSA results at a relatively high dielectric constant ($\varepsilon_{\text{in}} = 4$) compared with the MD-based enthalpies ($\text{enthalpy}_{\text{md}}$), but the calculation of NMEs based on MD trajectories needs much more computational resource. Compared with NMEs, the addition of IEs can reasonably improve the performance of MM/GBSA and MM/PBSA at a low dielectric constant ($\varepsilon_{\text{in}} = 1$) for any simulation strategy and can also totally defeat the enthalpy results based on the MD trajectories ($\text{enthalpy}_{\text{md}}$) for any dielectric constant. Considering the calculation of IEs does not incur additional computational cost, we recommend to calculate end-point binding free energies with IEs based on MD simulations. Nevertheless, system dependence

always exists, and for any case, it is better to do a prior test or at least use the publication-recommended methods for a specific system. The current conclusions of entropy calculations are mainly summarized from short-time MD simulations, and long-time MD simulations may be carried out to further validate the convergence problem of entropy calculations. Furthermore, in terms of reproducing the absolute binding free energy, the NME-MD based method ($\Delta G_{\text{nmode\_md\_9Å}}$) can give the lowest deviations against the experimental data for both MM/GBSA and MM/PBSA in all the tested strategies. Besides, although the ff03 force field performs the best in all the tested simulation protocols, no large difference is observed among the tested force fields, implying that all the force fields may be reasonable in the MM/GBSA binding free energy calculations (however, the ff03 force field may be the best choice).

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

This study was supported by the National Key R&D Program of China (2016YFA0501701), the National Science Foundation of China (81603031, 21575128, 81773632, 21433004, 11774207), the National Science Foundation for Post-doctoral Scientists of China (2016T90550, 2015M581953), the Fundamental Research Funds for the Central Universities (2017QNA7033, 2017QNA7034), and the Ministry of Science and Technology of China (2016YFA0501700).

References

1 R. A. Copeland, D. L. Pompliano and T. D. Meek, *Nat. Rev. Drug Discovery*, 2006, **5**, 730–739.
2 R. A. Copeland, *Nat. Rev. Drug Discovery*, 2016, **15**, 87.
3 H. Lu and P. J. Tonge, *Curr. Opin. Chem. Biol.*, 2010, **14**, 467–474.
4 P. J. Tummino and R. A. Copeland, *Biochemistry*, 2008, **47**, 5481–5492.
5 H. Sun, Y. Li, M. Shen, D. Li, Y. Kang and T. Hou, *J. Chem. Inf. Model.*, 2017, **57**, 1895–1906.
6 D. E. Hyre, I. Le Trong, E. A. Merritt, J. F. Eccleston, N. M. Green, R. E. Stenkamp and P. S. Stayton, *Protein Sci.*, 2006, **15**, 459–467.
7 B. Maschera, G. Darby, G. Palú, L. L. Wright, M. Tisdale, R. Myers, E. D. Blair and E. S. Furfine, *J. Biol. Chem.*, 1996, **271**, 33231–33235.
8 A. Basavapathruni, L. Jin, S. R. Daigle, C. R. Majer, C. A. Therkelsen, T. J. Wigle, K. W. Kuntz, R. Chesworth, R. M. Pollock and M. P. Scott, *Chem. Biol. Drug Des.*, 2012, **80**, 971–980.
9 H. Keränen, L. Pérez-Benito, M. Ciordia, F. Delgado, T. B. Steinbrecher, D. Oehlrich, H. W. van Vlijmen, A. A. Trabanco and G. Tresadern, *J. Chem. Theory Comput.*, 2017, **13**, 1439–1453.
10 H. Sun, P. Pan, S. Tian, L. Xu, X. Kong, Y. Li, D. Li and T. Hou, *Sci. Rep.*, 2016, **6**, 24817.
11 L. Xu, Y. Zhang, L. Zheng, C. Qiao, Y. Li, D. Li, X. Zhen and T. Hou, *J. Med. Chem.*, 2014, **57**, 3737–3745.
12 B. Kuhn, M. Tichy, L. Wang, S. Robinson, R. E. Martin, A. Kuglstatter, J. R. Benz, M. Giroud, T. Schirmeister and R. Abel, *J. Med. Chem.*, 2017, **60**, 2485–2497.
13 M. Aldeghi, M. J. Bodkin, S. Knapp and P. C. Biggin, *J. Chem. Inf. Model.*, 2017, **57**, 2203–2221.
14 J. Kästner, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 932–942.
15 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12562–12566.
16 L. Li, S. A. Martinis and Z. Luthey-Schulten, *J. Am. Chem. Soc.*, 2013, **135**, 6047–6055.
17 V. Limongelli, M. Bonomi and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 6358–6363.
18 H.-J. Woo and B. Roux, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6825–6830.
19 H. Sun, P. Chen, D. Li, Y. Li and T. Hou, *J. Chem. Theory Comput.*, 2016, **12**, 851–860.
20 H. Sun, Y. Li, S. Tian, J. Wang and T. Hou, *PLoS Comput. Biol.*, 2014, **10**, e1003729.
21 H. Sun, S. Tian, S. Zhou, Y. Li, D. Li, L. Xu, M. Shen, P. Pan and T. Hou, *Sci. Rep.*, 2015, **5**, 8457.
22 S. N. Rao, U. C. Singh, P. A. Bash and P. A. Kollman, *Nature*, 1987, **328**, 551–554.
23 P. A. Bash, M. J. Field and M. Karplus, *J. Am. Chem. Soc.*, 1987, **109**, 8092–8094.
24 P. Kollman, *Chem. Rev.*, 1993, **93**, 2395–2417.
25 W. L. Jorgensen and L. L. Thomas, *J. Chem. Theory Comput.*, 2008, **4**, 869–876.
26 M. Zacharias, T. Straatsma and J. McCammon, *J. Chem. Phys.*, 1994, **100**, 9025–9031.
27 J. C. Gumbart, B. t. Roux and C. Chipot, *J. Chem. Theory Comput.*, 2012, **9**, 794–802.
28 M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp and P. C. Biggin, *J. Am. Chem. Soc.*, 2017, **139**, 946–957.
29 M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp and P. C. Biggin, *Chem. Sci.*, 2016, **7**, 207–218.
30 J. Åqvist, C. Medina and J.-E. Samuelsson, *Protein Eng.*, 1994, **7**, 385–391.
31 S. Genheden and U. Ryde, *J. Chem. Theory Comput.*, 2011, **7**, 3768–3778.
32 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan and W. Wang, *Acc. Chem. Res.*, 2000, **33**, 889–897.
33 S. Genheden and U. Ryde, *Expert Opin. Drug Discovery*, 2015, **10**, 449–461.
34 W. Wang, O. Donini, C. M. Reyes and P. A. Kollman, *Annu. Rev. Biophys. Biomol. Struct.*, 2001, **30**, 211–243.
35 T. Hou and R. Yu, *J. Med. Chem.*, 2007, **50**, 1177–1188.
36 H. Sun, Y. Li, D. Li and T. Hou, *J. Chem. Inf. Model.*, 2013, **53**, 2376–2389.
37 H. Y. Sun and F. Q. Ji, *Biochem. Biophys. Res. Commun.*, 2012, **423**, 319–324.

38 H.-Y. Sun, T.-J. Hou and H.-Y. Zhang, *Drug Discovery Today*, 2014, **19**, 1836–1840.
39 H.-Y. Sun, F.-Q. Ji, L.-Y. Fu, Z.-Y. Wang and H.-Y. Zhang, *J. Chem. Inf. Model.*, 2013, **53**, 3343–3351.
40 W. Xue, X. Jin, L. Ning, M. Wang, H. Liu and X. Yao, *J. Chem. Inf. Model.*, 2012, **53**, 210–222.
41 W. W. Xue, D. B. Pan, Y. Yang, H. X. Liu and X. J. Yao, *Antiviral Res.*, 2012, **93**, 126–137.
42 Y.-L. Cui, Q.-C. Zheng, J.-L. Zhang, Q. Xue, Y. Wang and H.-X. Zhang, *J. Chem. Inf. Model.*, 2013, **53**, 3308–3317.
43 F. Chen, H. Liu, H. Sun, P. Pan, Y. Li, D. Li and T. Hou, *Phys. Chem. Chem. Phys.*, 2016, **18**, 22129–22139.
44 H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330**, 891–914.
45 T. J. Hou, N. Li, Y. Y. Li and W. Wang, *J. Proteome Res.*, 2012, **11**, 2982–2995.
46 S. Chang, D.-W. Zhang, L. Xu, H. Wan, T.-J. Hou and R. Kong, *RNA Biol.*, 2016, **13**, 1133–1143.
47 C. M. Reyes and P. A. Kollman, *J. Mol. Biol.*, 2000, **297**, 1145–1158.
48 A. V. Vargiu and A. Magistrato, *Inorg. Chem.*, 2012, **51**, 2046–2057.
49 T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.
50 S. Genheden, O. Kuhn, P. Mikulskis, D. Hoffmann and U. Ryde, *J. Chem. Inf. Model.*, 2012, **52**, 2079–2088.
51 I. Y. Ben-Shalom, S. Pfeiffer-Marek, K.-H. Baringhaus and H. Gohlke, *J. Chem. Inf. Model.*, 2017, **57**, 170–189.
52 L. Duan, X. Liu and J. Z. Zhang, *J. Am. Chem. Soc.*, 2016, **138**, 5722–5728.
53 S. Genheden, M. Akke and U. Ryde, *J. Chem. Theory Comput.*, 2013, **10**, 432–438.
54 S. Hikiri, T. Yoshidome and M. Ikeguchi, *J. Chem. Theory Comput.*, 2016, **12**, 5990–6000.
55 K. Sharp, *J. Chem. Theory Comput.*, 2013, **9**, 1164–1172.
56 H. Choi, H. Kang and H. Park, *J. Chem. Theory Comput.*, 2015, **11**, 4933–4942.
57 G. Gyimesi, P. Závodszky and A. Szilágyi, *J. Chem. Theory Comput.*, 2017, **13**, 29–41.
58 S. Genheden and U. Ryde, *Phys. Chem. Chem. Phys.*, 2012, **14**, 8662–8677.
59 Y. Yan, M. Yang, C. G. Ji and J. Z. H. Zhang, *J. Chem. Inf. Model.*, 2017, **57**, 1112–1122.
60 A. Di Nola, H. J. Berendsen and O. Edholm, *Macromolecules*, 1984, **17**, 2044–2050.
61 T. Cheng, X. Li, Y. Li, Z. Liu and R. Wang, *J. Chem. Inf. Model.*, 2009, **49**, 1079–1093.
62 R. Wang, X. Fang, Y. Lu and S. Wang, *J. Med. Chem.*, 2004, **47**, 2977–2980.
63 R. Wang, X. Fang, Y. Lu, C.-Y. Yang and S. Wang, *J. Med. Chem.*, 2005, **48**, 4111–4119.
64 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
65 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, 2006, **25**, 247–260.
66 L. Xu, H. Sun, Y. Li, J. Wang and T. Hou, *J. Phys. Chem. B*, 2013, **117**, 8408–8421.
67 A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623–1641.
68 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
69 P. Cieplak, J. Caldwell and P. Kollman, *J. Comput. Chem.*, 2001, **22**, 1048–1057.
70 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.
71 J. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074.
72 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712–725.
73 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 1950–1958.
74 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
75 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
76 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
77 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
78 A. Onufriev, D. Bashford and D. A. Case, *Proteins: Struct., Funct., Bioinf.*, 2004, **55**, 383–394.
79 C. Tan, L. Yang and R. Luo, *J. Phys. Chem. B*, 2006, **110**, 18680–18687.
80 H. Sun, Y. Li, S. Tian, L. Xu and T. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16719–16729.
81 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.
82 T. Hou, J. Wang, Y. Li and W. Wang, *J. Comput. Chem.*, 2011, **32**, 866–877.
83 J. Weiser, P. S. Shenkin and W. C. Still, *J. Comput. Chem.*, 1999, **20**, 217–230.
84 Z. Wang, H. Sun, X. Yao, D. Li, L. Xu, Y. Li, S. Tian and T. Hou, *Phys. Chem. Chem. Phys.*, 2016, **18**, 12964–12975.
85 P. Mikulskis, S. Genheden and U. Ryde, *J. Chem. Inf. Model.*, 2014, **54**, 2794–2806.
86 N. M. King, M. Prabu-Jeyabalan, R. M. Bandaranayake, M. N. Nalam, E. A. Nalivaika, A. E. L. Özen, T. R. Haliloğlu, N. E. K. Yılmaz and C. A. Schiffer, *ACS Chem. Biol.*, 2012, **7**, 1536–1546.
87 M. Ahmad, V. Helms, T. Lengauer and O. V. Kalinina, *J. Chem. Theory Comput.*, 2015, **11**, 1410–1418.
88 V. Lafont, A. A. Armstrong, H. Ohtaka, Y. Kiso, L. Mario Amzel and E. Freire, *Chem. Biol. Drug Des.*, 2007, **69**, 413–422.

