Enhancing MM/P(G)BSA Methods: Integration of Formulaic Entropy for Improved Binding Free Energy Calculations and Virtual Screening Efficiency

Lina Dong[1], Pengfei Li,[2] Binju Wang[1,*]

[1]State Key Laboratory of Physical Chemistry of Solid Surfaces and Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, iChEM, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 360015, P. R. China
[2] Department of Chemistry and Biochemistry, Loyola University Chicago, Chicago 60660, Illinois, United States.
*Correspondence to: wangbinju2018@xmu.edu.cn

KEYWORDS

Binding free energy, MM/PBSA, MM/GBSA, entropy, virtual screening

ABSTRACT. Balancing computational efficiency and precision, MM/P(G)BSA methods have been widely employed in the estimation of binding free energies within biological systems. However, the entropy contribution to the binding free energy is often neglected in MM/P(G)BSA calculations, due to the computational cost of conventional methods such as normal mode analysis (NMA). In this work, we develop an enhanced MM/P(G)BSA method by incorporating the entropy effect using a formulaic entropy. Extensive benchmarking reveals that the integration of formulaic entropy systematically elevates the performance of both MM/PBSA and MM/GBSA without incurring additional computational expenses. Notably, MM/PBSA_S, augmented with formulaic entropy, surpasses all other MM/P(G)BSA methods across a spectrum of datasets. Furthermore, the incorporation of MM/PBSA_S into a workflow can yield significantly improved results for the virtual screening, marked by a considerable enhancement in the enrichment factor. Our investigation furnishes a valuable and practical MM/P(G)BSA method, optimizing binding free energy calculations for a variety of biological systems.

1.INTRODUCTION

The accurate calculation and prediction of protein-ligand binding free energy constitute a pivotal phase in virtual screening for drug discovery.[1-3] Alchemical methods, encompassing thermodynamic integration (TI) and free energy perturbation (FEP), are renowned for their reliability in binding free energy calculations,[4-6] but their requirement for substantial computational resources renders them less feasible for high-throughput virtual screening scenarios. Scoring functions (SFs), which can be divided into physical-based, empirical-based, knowledge-based, and machine learning-based methods, excel in swiftly estimating the binding affinity of protein-ligand complexes.[7-12] However, the majority of these functions inadequately address the solvent effect intrinsic to binding free energy, leading to less than optimal performance. In stark contrast to alchemical methods and SFs, force field-based MM/P(G)BSA can achieve a superior equilibrium between computational expense and accuracy, thereby demonstrating wide applications in drug design and molecular modelling studies.[13-17]

Regarding MM/P(G)BSA, the binding free energy is a composite of gas-phase electrostatic and van der Waals interactions, solvent-mediated electrostatic and van der Waals differences at the terminal states, alongside the entropy component.[17-19] Among these critical energy constituents, the entropy contribution has persistently posed a computational challenge.[20, 21] A benchmark study

on 855 structures from PDBbind has shown that the ignorance of entropy effects in MM/GBSA can lead to either overestimation or underestimation of binding affinity.[22] Normal-mode analysis (NMA) is extensively employed for the calculations of the configurational entropy, yet its computational expense renders it unsuitable for large-scale screening applications. A benchmark study elucidated that, in comparison to isothermal titration calorimetry (ITC) experimental outcomes, NMA yields more accurate results than the quasi-harmonic approximation method.[23] However, a noteworthy observation was that the computed entropies derived from NMA did not exhibit a correlation with the entropy values obtained from ITC experiments, highlighting a discrepancy between theoretical calculations and experimental measurements.[23] In addition to NMA, the Interaction Entropy (IE) method is another widely adopted technique for entropy calculations. IE offers a theoretically robust framework grounded in molecular dynamics (MD) simulations, enabling the reliable computation of the entropic contribution to free energy.[24-26]

Beyond entropy computation, a multitude of strategies have been devised to enhance the precision of MM/P(G)BSA calculations.[27-32] Notably, Hou and colleagues have proposed the incorporation of a variable dielectric constant as a means to refine the performance of MM/P(G)BSA.[33, 34] In a recent work, Ding and co-worker introduced an exponential damping factor to the electrostatic interaction (ΔE_{ele}), which can significantly improve the performance of the MM/PBSA for the systems with charged ligands.[35] Meanwhile, several automatic workflows have been developed to facilitate the MM/GB(PB)SA calculations, which can cover the system setup, molecular docking, structure minimization and/or MD simulations and the free energy calculations with MM/P(G)BSA process.[36, 37]

In this work, we attempted to further improve the performance of MM/P(G)BSA by introducing the formulaic entropy into it. Our dual objectives encompass enhancing accuracy while concurrently managing computational expenses, thereby rendering the methodology amenable to high-throughput virtual screening applications. We propose an optimized MM/P(G)BSA method that approximates the entropy effect in binding free energy using formulaic entropy to achieve accuracy improvement. Molecular dynamics simulation is not a necessary step for formulaic entropy, thus balancing computational cost and enhancing its applicability for high-throughput screening tasks. In addition, we conducted benchmark tests and discussions on the parameter settings of MM/P(G)BSA, and applied them to downstream virtual screening tasks to provide suggestions and guidance for research related to the application of related methods.

2. THEORY

2.1. MMP(G)BSA.

In end-point MM/PBSA and MM/GBSA calculations, the free-energy change for binding of a ligand (L) to a protein (P) to form a complex (PL) can be expressed as:

$$\Delta G_{bind} = G_{PL} - G_P - G_L \qquad (1)$$

where the binding free energy (ΔG_{bind}) corresponds to the free energy difference between the bound state (G_{PL}) and two free states (G_P and G_L). The binding free energy can be further expressed as:

$$\begin{aligned}\Delta G_{bind} &= \Delta H - T\Delta S \\ &\approx \Delta E_{MM} + \Delta G_{solv,p} + \Delta G_{solv,np} - T\Delta S \\ &= \Delta E_{ele} + \Delta E_{vdw} + \Delta G_{PB/GB} + \Delta G_{Solv,np} - T\Delta S \qquad (2)\end{aligned}$$

where the gas-phase interaction energy (ΔE_{MM}) between protein and ligand can be decomposed into the electrostatic term (ΔE_{ele}) and the van der Waals term (ΔE_{vdw}), while ($\Delta G_{solv,p}$) and ($\Delta G_{solv,np}$) are the energy changes brought by the solution as polar solvation interactions and non-polar solvation interactions, and the entropy change (T Δ S). $\Delta G_{PB/GB}$ is the polar (electrostatic) contribution to the solvation free energy calculated by PB or GB model.[38, 39]

One-Term (SA-Only) Model for $\Delta G_{solv,np}$. The nonpolar solvation term is expressed by an empirical function of SASA in this model

$$\Delta G_{solv,np} = \gamma * \Delta SASA + b \qquad (3)$$

where γ and b are the surface tension constant and correction terms, respectively.

Two-Term (Cavity-Dispersion) Model for $\Delta G_{solv,np}$. The nonpolar solvation term comes from repulsive and attractive components between solute and explicit solvents

$$\Delta G_{solv,np} = \Delta G_{solv,rep} + \Delta G_{solv,att} \qquad (4)$$

where $\Delta G_{solv,rep}$ is the solvation free-energy contribution from solute−solvent repulsive interactions and the formation of solute cavity, while $\Delta G_{solv,att}$ is the solvent−solute attractive nonpolar interaction, which also includes the solvent−solvent reorganization component.

$$\Delta G_{solv,rep} \approx \Delta G_{cav} \qquad (5)$$

$$\Delta G_{solv,att} \approx \Delta G_{DP} \qquad (6)$$

$$\Delta G_{solv,np} = \Delta G_{cav} + \Delta G_{DP} \qquad (7)$$

ΔG_{cav} is the free energy of cavity formation corresponding to repulsive terms, and ΔG_{DP} is the dispersion free energy corresponding to attractive nonpolar interactions between solute and solvent. Therefore, nonpolar energy is further divided into cavity and dispersion contributions in the two-term model (Eq.7). In this work, the dispersion term is computed with a surface-based integration method,[40] which is closely related to the PCM solvent for quantum chemical programs.[41] Under this framework, the cavity term is still computed as a term linearly proportional to the molecular solvent accessible-surface area (SASA).

2.2. Formula entropy

Zhan and et al. proposed a method to estimate entropy.[42] In the method, the entropy is contributed by two terms: solvation free entropy (ΔS_{solv}) and conformational free entropy (ΔS_{conf}):

$$\Delta S = \Delta S_{solv} + \Delta S_{conf} \qquad (8)$$

The solvation entropy is gained by the tendency of water molecules to minimize their contacts with hydrophobic groups in protein.[43] It has been demonstrated that the solvation entropy is temperature-dependent and can be calculated with heat capacity.[44, 45]

$$\Delta S_{solv} = \Delta C_{p,ap}\, ln\frac{T}{T^*_{S,ap}} + \Delta C_{p,pol}\, ln\frac{T}{T^*_{S,pol}} \qquad (9)$$

where $\Delta C_{p,ap}$ and $\Delta C_{p,pol}$ are the apolar and polar heat capacity. $T^*_{S,ap}$ and $T^*_{S,pol}$ are the temperatures when the apolar and polar hydration entropies are zero, and their values are 385.15[46] and 335.15 K,[47] respectively. The value of temperature T used in Eq.9 was 298.15 K. For the calculation of apolar and polar heat capacities, the apolar and polar heat capacity changes for the protein–ligand binding can be expressed as a linear relationship of apolar and polar solvent-accessible surface area differences $\Delta SASA_{ap}$ and $\Delta SASA_{pol}$.

$$\Delta C_{p,ap} = a_c(T)\Delta SASA_{ap} \qquad (10)$$

$$\Delta C_{p,pol} = b_c(T)\Delta SASA_{pol} \qquad (11)$$

where $a_c(T)$ and $b_c(T)$ are temperature-dependent coefficients. In a low temperature (T < 353 K) situation, the heat capacities are temperature-independent, and the values of $a_c(T)$ and $b_c(T)$ are 0.45 and -0.26, respectively.[45]

As a first approximation, the conformational entropy change of the ligand binding, ΔS_{conf}, was written as a linear function of the number of rotatable bonds (N_{rb}) and total number of atoms (N_{atoms})[44]:

$$\Delta S_{conf} = K_1 N_{rb} + K_2 N_{atoms} \qquad (12)$$

The coefficient K_1 was suggested to be -1.76 cal/(K/mol), which is close to the conformational entropy value observed for C-C bonds in long chain paraffins.[48, 49] The coefficient K_2 was suggested to be 0.414 cal/(K/mol), which essentially accounts for the conformational entropy restrictions in the free inhibitor. [31]

3. METHODS

3.1. System Preparation.

The crystal structures from PDBbind-v2016[50] refined set and CASF-2016[51] were used in this study. All absent hydrogen atoms were added to the proteins using the LEAP module in AMBER20[52]. We assigned the protonation states of titratable residues (His, Glu, Asp) on the basis of pKa values from the PROPKA software.[53] The Amber ff14SB[54] FF was employed for the protein residues, while the general AMBER FF (GAFF)[55] and AM1-BCC[56] charges were used for ligands.

3.2. Structure Optimization.

The resulting protein–ligand complexes were solvated in a rectangular box of TIP3P[57] waters extending at least 10 Å from the protein surface. Counterions, Na^+ or Cl^-, were added to neutralize the total charge of each system. Afterwards, the whole system was fully minimized using combined steepest descent and conjugate gradient method. All optimizations were performed using the AMBER20 program.[58] The fully optimized structures were used in the subsequent MM/P(G)BSA calculations.

3.3. MM/P(G)BSA Calculations.

MM/P(G)BSA calculations were performed using the structures of AMBER minimum energy optimizations via MMPBSA.py[59] from AMBER. For the default MM/PBSA calculation, we use the AMBER built-in model and parameters are istrng=0.1, exdi=80, indi=4.0, radiopt=1 and inp=2. In the following chapters, the modified internal dielectric constant is indi here. For the default MM/GBSA calculation, we use the AMBER built-in model and parameters are igb=2 and saltcon=0.1. It is worth noting that MM/PBSA calculations default to the Two-Term Model for $\Delta G_{solv,np}$, whereas MM/GBSA calculations default to the One-Term (SA-Only) Model for $\Delta G_{solv,np}$ in AMBER.

3.4. Performance evaluation

The Pearson correlation coefficient (Rp) is a measure of the linear dependence of the predicted binding free energy values on the experimental values according to Eq. 8. The Spearman correlation coefficient (Rs) measures the strength of an association between the predicted and experimental binding free energy values according to a monotonic function (Eq. 9). RMSE is the root mean square error between the predicted binding free energy and the experimental value (Eq. 10). These three quantities are calculated as follows:

$$R_p = \frac{\sum_{i=1}^{N}(pre_i - pre_{ave})(exp_i - exp_{ave})}{\sqrt{\sum_{i=1}^{N}(pre_i - pre_{ave})^2 \sum_{i=1}^{N}(exp_i - exp_{ave})^2}} \quad (8)$$

$$R_s = 1 - \frac{6\sum_{i=1}^{N}(rpre_i - rexp_i)}{N(N^2-1)} \quad (9)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(pre_i - exp_i)^2} \quad (10)$$

where pre_i is the binding free energy from the given scoring function on the ith complex in the test set; exp_i is the experimental binding constant (in logarithm units, $logK_a$) of this complex; pre_{ave} and exp_{ave} are the corresponding averages; $rpre_i$ is the rank of the binding free energy of the ith complex; $rexp_i$ is the rank of the experimental binding free energy of this complex and N is the total number of samples.

Enrichment factor (EF) is a measure that accounts for the number of true binders among the top x% ranked molecules and is calculated as follows:

$$EF_{x\%} = \frac{actives\ at\ x\%}{ligands\ at\ x\%} \times \frac{total\ ligands}{total\ actives}$$

The screening power of methods are assessed based on their enrichment performance that relates to enrichment factor.

4. RESULTS

4.1. MM/P(G)BSA with and without dispersion free energy

As detailed in the Theory section, MM/P(G)BSA calculations can be executed using either one-term (SA-Only) model or two-term (Cavity-Dispersion) model for $\Delta G_{solv,np}$. When the one-term model is applied, the calculations are denoted as MM/PBSA or MM/GBSA. For the two-term approach, we append _DP to the original nomenclature (resulting in MM/PBSA_DP and MM/GBSA_DP). As illustrated in Figure 1, the results without DP correction were better in Rp, while the results with DP correction were better in RMSE. It was found that the one-term model

has better correlation for linear aliphatic molecules,[60] while the two-term model has better correlation for branched and cyclic organic molecules.[61] Unlike many previous studies, we have employed a substantial and structurally diverse dataset in our test. In addition, the data set is quite diverse structurally. Through large-scale tests, we have shown that the one-term can get better correlation for the calculation of the binding free energy of various systems (higher Rp), while the two-term model for the dispersion energy calculation is more accurate for the prediction of the absolute binding free energy (lower RMSE).

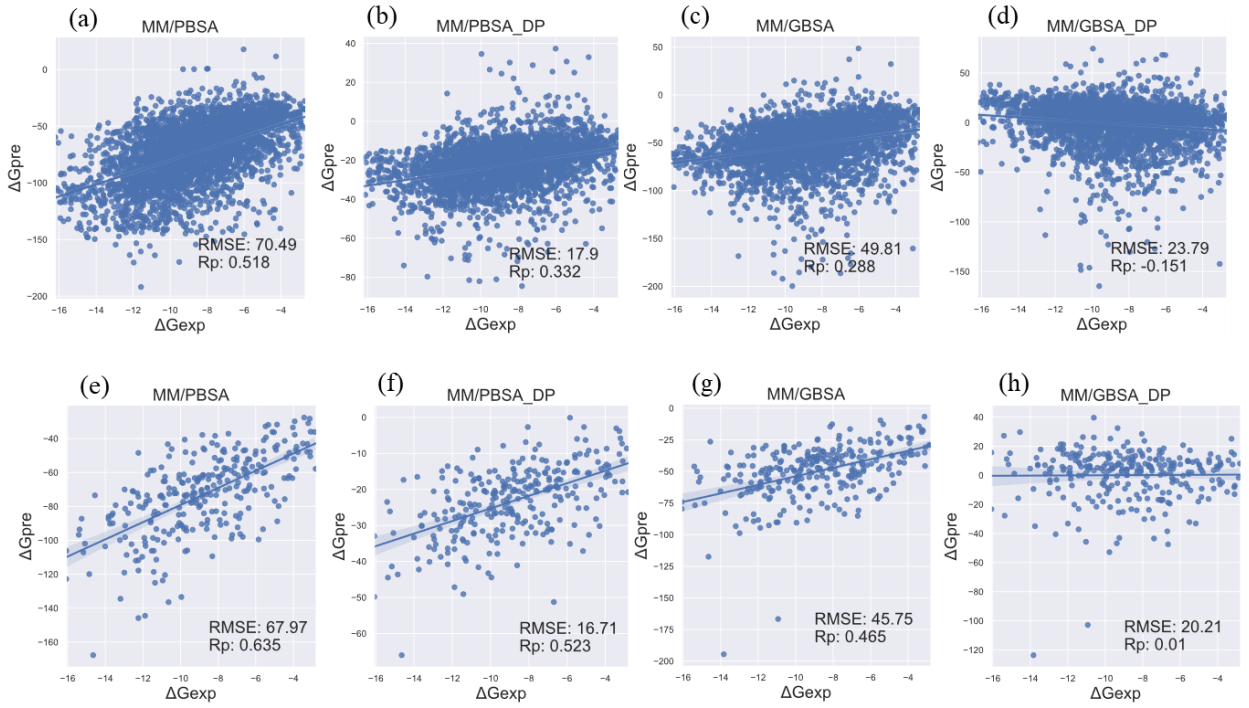

Figure 1. Results of MM/PBSA and MM/GBSA in one-term (SA-Only) model and two-term (Cavity-Dispersion) model for $\Delta G_{solv,np}$. For one-term model, the calculations are denoted as MM/PBSA or MM/GBSA. For the two-term model, _DP is added at the end of their original name (MM/PBSA_DP and MM/GBSA_DP). (a), (b), (c) and (d) are the results of refined set (3186 complexes) and (e), (f), (g) and (h) are the results of CASF-2016 (285 complexes).

we also evaluated the influence of varying internal dielectric constants on MM/PBSA calculations, employing both the one-term and two-term models for $\Delta G_{solv,np}$. Across different dielectric constants, we observed that the one-term model yields better correlation than the two-term model (Figure S1). Strikingly, for the one-term model, there appears to be no consistent relationship between performance and dielectric constant. On the contrary, for the two-term model, higher dielectric constants correlate with improved performance. In addition, these above

conclusions are not dependent on whether the ligands were charged or not (Figure S2 and S3).

4.2. The entropy calculation with different forms of formulaic entropy

As mentioned in Section 2.2, when the formulaic entropy is included, the methods are denoted as MM/PBSA_S or MM/GBSA_S. In a previous study, the conformational entropy was approximated using only the number of rotatable bonds, not the total number of atoms.[42] In light of this, we also conducted tests with the conformational entropy term that excluded the number of atoms in Eq.12, denoting these methods with the suffix _S-A (MM/PBSA_S-A and MM/GBSA_S-A).

The comparison of formulaic entropy calculation for MM/PBSA is summarized in Figure 2. It can be seen that no matter which method is used to calculate the formulaic entropy, the correlation (Rp) between the calculated values of binding free energy and the experimental ones can be improved a little bit, which is true for the refined set (a) (b) (c) with larger data volume and the benchmark set (d) (e) (f) with smaller data volume. However, the estimation of formulaic entropy elevates errors of absolute binding free energy (RMSE). In general, MM/PBSA_S shows a higher Rp and a lower RMSE than MM/PBSA_S-A, resulting in better performance. Therefore, the MM/PBSA_S method is recommended to estimate the entropy due to its improvement of the correlation between the calculated binding free energy and experimental values for MM/PBSA.

As for MM/GBSA (Figure 3), we can see that no matter which method is used to calculate the formulaic entropy, the correlation (Rp) can be improved, which is true for the refined set (a) (b) (c) with larger data volume and the benchmark set (d) (e) (f) with smaller data volume. However, the estimation of formulaic entropy increases the error of absolute binding free energy (RMSE). Overall, MM/GBSA_S-A shows both higher Rp and RMSE than those of MM/GBSA_S. Therefore, we recommend the MM/GBSA_S-A method for entropy estimation for MM/GBSA.

Overall, incorporating formulaic entropy significantly enhances the correlation performance of both MM/PBSA and MM/GBSA. On the refined set, Rp increases from 0.532 to 0. 541 for MM/PBSA, while it increases from 0.305 to 0.43 for MM/GBSA. In CASF-2016 test set, for MM/PBSA, Rp increases from 0.635 to 0.644 for MM/PBSA, and from 0.465 to 0.568 for MM/GBSA. Obviously, the addition of formulaic entropy yields a more pronounced improvement in the performance of MM/GBSA compared to MM/PBSA.

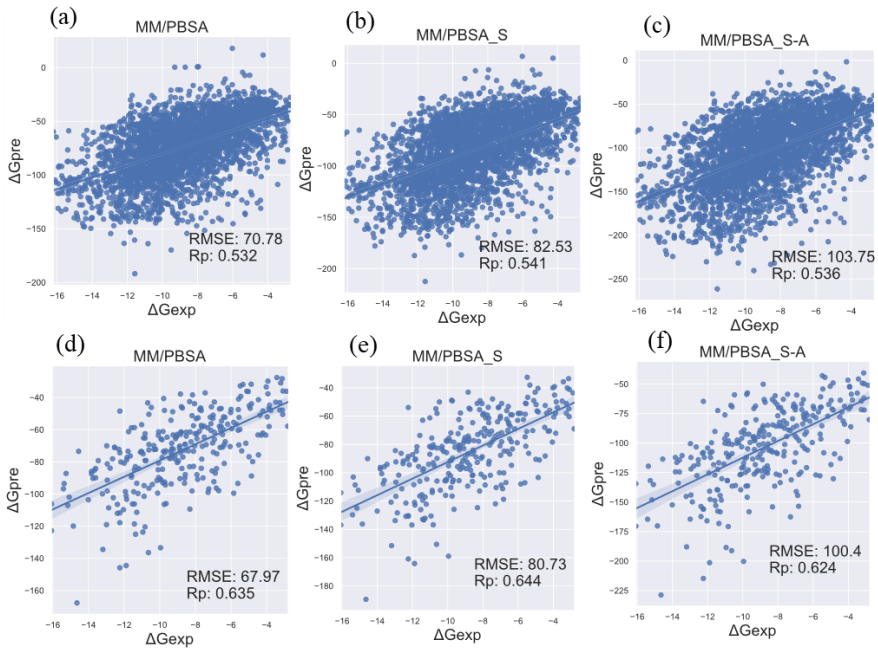

Figure 2. The entropy calculation results of MM/PBSA in different ways. (a), (b) and (c) are the results of refined set (2948 complexes) and (d), (e) and (f) are the results of CASF-2016 (285 complexes). (b) and (e) use the number of rotatable bonds and atoms of a ligand to approximate the conformational entropy. (c) and (f) the number of rotatable bonds of the ligand to approximate the conformational entropy.

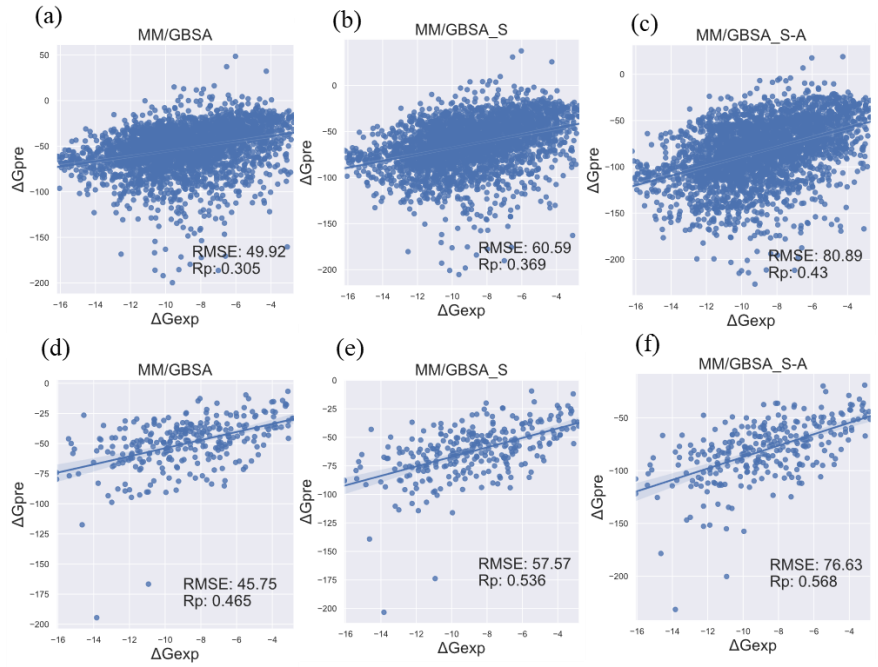

Figure 3. The entropy calculation results of MM/GBSA in different ways. (a), (b) and (c) are the results of refined set (2948 complexes) and (d), (e) and (f) are the results of CASF-2016 (285 complexes). (b) and (e) use the number of rotatable bonds and atoms of a ligand to approximate the conformational entropy. (c) and (f) the number of rotatable bonds of the ligand to approximate the conformational entropy.

4.3. Results of subsets clustered by different properties of protein pockets

As noted in the preceding sections, MM/PBSA_S, with the inclusion of formulaic entropy, outperforms both the default MM/PBSA_DP (which is default in AMBER with an adjusted dielectric constant) and MM/PBSA_S-A (which includes formulaic entropy but excludes dispersion). Conversely, MM/GBSA_S-A, which incorporates formulaic entropy without dispersion, surpasses the performance of the default MM/GBSA and MM/GBSA_S (which includes formulaic entropy). In Figure 4, we present a comparison of the performance of these four representative methods on the CASF-2016 subsets, categorized by three properties of protein pockets 1):the excluded volume inside the binding pocket upon ligand binding, 2): the percentage of the solvent-accessible surface area of the ligand molecule that was buried upon binding, 3): the hydrophobic scale of the binding pocket). Nine subsets were obtained by clustering based on these three properties.[51] It is evident that MM/PBSA_S and MM/GBSA_S-A exhibit markedly improved scoring and ranking power compared to their default counterparts (MM/PBSA_DP and MM/GBSA). Specifically, for scoring power (as measured by Rp), MM/PBSA_S emerges as the superior method among the four (Figure 4a). In terms of ranking power (Rs), MM/PBSA_S again leads the pack, delivering the best performance. For a more detailed analysis, please refer to Tables S1, S2, S3, and S4.

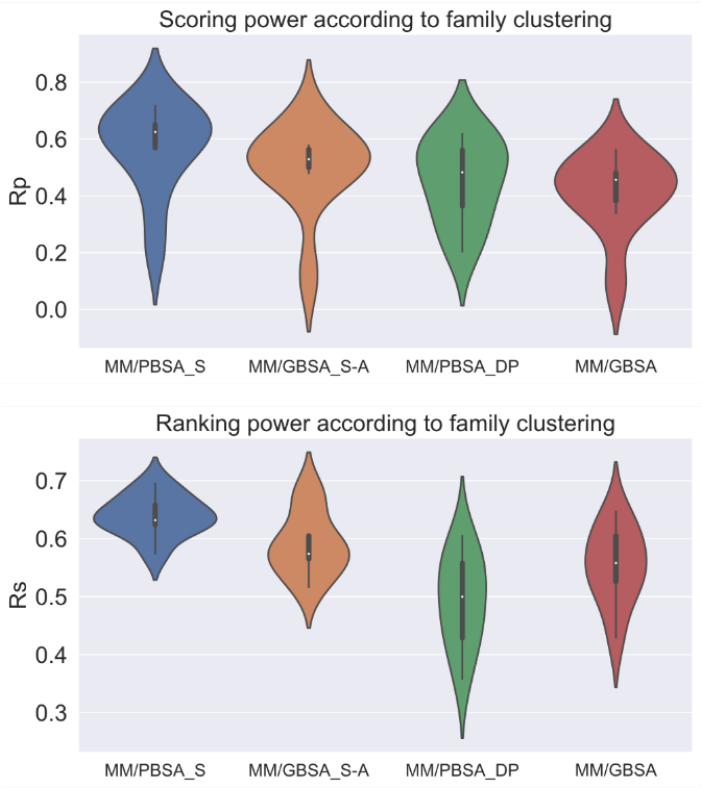

Figure 4. The results of adjusted and default methods on CASF-2016 according to different subsets clustered by three properties of protein pockets. MM/PBSA_S and MM/GBSA_S-A are the best adjusted MM/PBSA and MM/GBSA, respectively. MM/PBSA_DP and MM/GBSA are the default MM/PBSA and MM/GBSA in AMBER, respectively. (a) Scoring power based on Rp. (b) Ranking power based on Rs.

4.4. Workflow for drug screening with the improved MM/P(G)BSA

MM/P(G)BSA methods are widely utilized for binding free energy calculations in drug discovery screening. To further evaluate the performance of our improved MM/P(G)BSA methodologies, we conducted binding free energy calculations for actives and decoys from the DUD-E[62] target AKT2 following the workflow depicted in Figure 5a. Initially, molecular docking was performed using AutoDock Vina, and subsequently, the protein-ligand complexes were constructed using the docked top-ranking structure with AMBER. This process yielded a total of 4898 complexes, including 72 actives and 4826 decoys. The actives referred to here are molecules that have been experimentally detected to have activity, as mentioned in the DUD-E dataset.[62] Subsequently, geometry optimization of the protein-ligand complexes was performed under a molecular force field, as detailed in Section 3.2. Finally, various MM/P(G)BSA calculations were executed. Through the binding free energy ranking, we can distinguish between actives and decoys

(Figure 5b) and then evaluate the virtual screening power of each method using enrichment factor (EF) (Figure 5c). The results of EF are shown in Figure 5c. We have evaluated three EFs, namely EF1%, EF5% and EF10% (Figure 5c)., more details can be found in Table S5. Among four compared methods, we can see that MM/PBSA_S achieves the highest enrichment factor, indicating that it has the best screening power. Unlike MM/PBSA, we found the adjusted MM/GBSA_S-A did not yield the improvement in enrichment factor than the default MM/GBSA for this test set. In summary, our large-scale tests have demonstrated that the entropy-corrected MM/PBSA_S method exhibits significantly better performance than other MM/P(G)BSA methods, making it highly recommended for binding free energy calculations and virtual screening tasks.

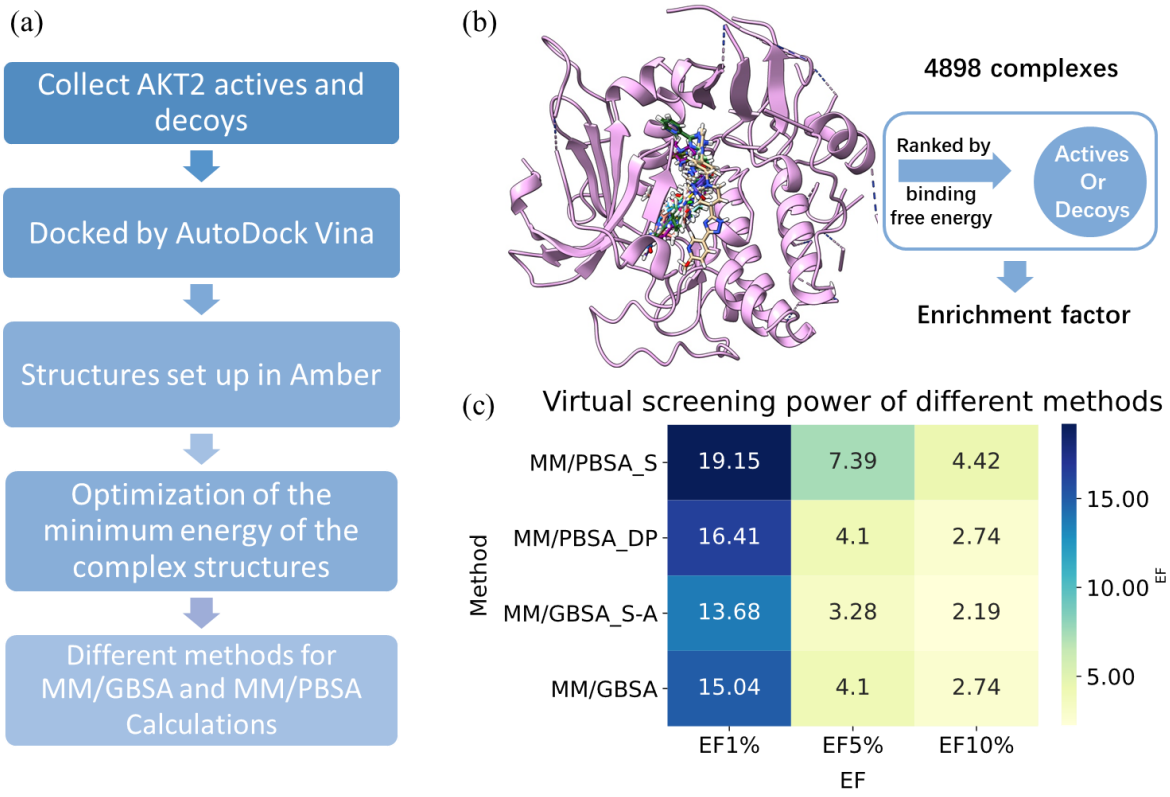

Figure 5. The test flow and results of the virtual screening power of MM/P(G)BSA before and after improvement. (a) Testing flow of virtual screening power (b) Schematic diagram of obtaining enrichment factors by calculating the binding free energy of the complex and ranking it (c) The results of three enrichment factors (EF1%, EF5% and EF10%) obtained by the improved method before and after.

CONCLUSION

In this study, we have optimized MM/P(G)BSA methods through the integration of formulaic entropy to account for the entropy effect. Unlike the computationally demanding entropy

calculations involving Normal Mode Analysis (NMA), the formulaic entropy approach offers a swift and efficient alternative. Upon extensive benchmark tests on refined set and CASF-2016 of PDBbind, we made discussions on the parameter settings of MM/P(G)BSA. Our analysis indicated that MM/PBSA_S with inclusion of formulaic entropy and MM/GBSA_S-A that includes formulaic entropy but excludes the dispersion shows better performance than their defaults methods. Our findings underscore that MM/PBSA_S, when incorporating formulaic entropy, surpasses all other MM/P(G)BSA methods in terms of calculated Rp and Rs values. Furthermore, a benchmark evaluation of various MM/P(G)BSA methods on the DUD-E target AKT2, utilizing a workflow, demonstrates that MM/PBSA_S, enhanced with formulaic entropy, achieves the highest enrichment factor for virtual screening purposes. Conclusively, our formulaic entropy-corrected MM/PBSA_S is highly recommended for the prediction of binding free energy in diverse biological systems, offering a robust and efficient computational tool for researchers in this field.

ASSOCIATED CONTENT

Supporting Information.

The Supporting Information is available free of charge at http://pubs.acs.org.
(PDF)

Data and Software Availability

All data used in this manuscript are sourced from publicly available databases. Specifically, datasets from PDBbind2016 (http://www.pdbbind.org.cn/) and DUD-E (https://dude.docking.org/datasets) were employed. Formulaic entropy calculations were based on delta solvent-accessible surface area (Δ SASA) values obtained from the Δ vinaXGB (https://github.com/jennienig/deltaVinaXGB) method. The source codes utilized in this paper are freely available online under an open-source license.
(https://github.com/LinaDongXMU/MM_S).

AUTHOR INFORMATION

Corresponding Author

Binju Wang − State Key Laboratory of Physical Chemistry of Solid Surfaces and Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, College of Chemistry and Chemical

Engineering, Xiamen University, Xiamen 361005, P. R. China; orcid.org/0000-0002-3353-9411; Email: wangbinju2018@xmu.edu.cn

Author

Lina Dong − State Key Laboratory of Physical Chemistry of Solid Surfaces and Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, iChEM, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 360015, P. R. China

Pengfei Li − Department of Chemistry and Biochemistry, Loyola University Chicago, Chicago 60660, Illinois, United States

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (22122305, 22121001 and 21933009).

REFERENCES

(1) King, E.; Aitchison, E.; Li, H.; Luo, R., Recent Developments in Free Energy Calculations for Drug Discovery. *Front. Mol. Biosci.* **2021**, 8, 712085.
(2) Ryde, U.; Söderhjelm, P., Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chem. Rev.* **2016**, 116, 5520-5566.
(3) Wang, J.; Miao, Y., Ligand Gaussian Accelerated Molecular Dynamics 2 (LiGaMD2): Improved Calculations of Ligand Binding Thermodynamics and Kinetics with Closed Protein Pocket. *J. Chem. Theory. Comput.* **2023**, 19, 733-745.
(4) Lee, T. S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D., Jr.; Pearlman, D. A.; Radak, B. K., et al., Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, 60, 5595-5623.
(5) Olsson, M. A.; Ryde, U., Comparison of QM/MM Methods To Obtain Ligand-Binding Free Energies. *J. Chem. Theory. Comput.* **2017**, 13, 2245-2253.
(6) Chen, W.; Cui, D.; Jerome, S. V.; Michino, M.; Lenselink, E. B.; Huggins, D. J.; Beautrait, A.; Vendome, J.; Abel, R., et al., Enhancing Hit Discovery in Virtual Screening through Absolute Protein-Ligand Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **2023**, 63, 3171-3185.
(7) Xiong, G.; Shen, C.; Yang, Z.; Jiang, D.; Liu, S.; Lu, A.; Chen, X.; Hou, T.; Cao, D., Featurization strategies for protein–ligand interactions and their applications in scoring function development. *Wires Comput. Mol. Sci.* **2022**, 12, e1567.

(8) Lu, J.; Hou, X.; Wang, C.; Zhang, Y., Incorporating Explicit Water Molecules and Ligand Conformation Stability in Machine-Learning Scoring Functions. *J. Chem. Inf. Model.* **2019**, 59, 4540-4549.
(9) Moon, S.; Zhung, W.; Kim, W. Y., Toward generalizable structure-based deep learning models for protein–ligand interaction prediction: Challenges and strategies. *Wires Comput. Mol. Sci.* **2024**, 14, e1705.
(10) Xia, S.; Chen, E.; Zhang, Y., Integrated Molecular Modeling and Machine Learning for Drug Design. *J. Chem. Theory. Comput.* **2023**, 19, 7478-7495.
(11) Yang, C.; Chen, E. A.; Zhang, Y., Protein-Ligand Docking in the Machine-Learning Era. *Molecules* **2022**, 27.
(12) Xue M, L. B., Cao S, Huang X., FeatureDock: Protein-Ligand Docking Guided by Physicochemical Feature-Based Local Environment Learning using Transformer. *ChemRxiv* **2024**, doi:10.26434/chemrxiv-2024-dh2rw.
(13) Chen, Y.; Zheng, Y.; Fong, P.; Mao, S.; Wang, Q., The application of the MM/GBSA method in the binding pose prediction of FGFR inhibitors. *Phys. Chem. Chem. Phys.* **2020**, 22, 9656-9663.
(14) Crean, R. M.; Pudney, C. R.; Cole, D. K.; van der Kamp, M. W., Reliable In Silico Ranking of Engineered Therapeutic TCR Binding Affinities with MMPB/GBSA. *J. Chem. Inf. Model.* **2022**, 62, 577-590.
(15) Xiang, S.; Wang, Z.; Tang, R.; Wang, L.; Wang, Q.; Yu, Y.; Deng, Q.; Hou, T.; Hao, H., et al., Exhaustively Exploring the Prevalent Interaction Pathways of Ligands Targeting the Ligand-Binding Pocket of Farnesoid X Receptor via Combined Enhanced Sampling. *J. Chem. Inf. Model.* **2023**, 63, 7529-7544.
(16) Wang, C.; Greene, D.; Xiao, L.; Qi, R.; Luo, R., Recent Developments and Applications of the MMPBSA Method. *Front. Mol. Biosci.* **2017**, 4, 87.
(17) Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, 10, 449-461.
(18) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T., End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019**, 119, 9478-9508.
(19) Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2011**, 51, 69-82.
(20) Summa, C. M.; Langford, D. P.; Dinshaw, S. H.; Webb, J.; Rick, S. W., Calculations of Absolute Free Energies, Enthalpies, and Entropies for Drug Binding. *J. Chem. Theory. Comput.* **2024**, 20, 2812-2819.
(21) Ekberg, V.; Ryde, U., On the Use of Interaction Entropy and Related Methods to Estimate Binding Entropies. *J. Chem. Theory. Comput.* **2021**, 17, 5379-5391.
(22) Greenidge, P. A.; Kramer, C.; Mozziconacci, J. C.; Wolf, R. M., MM/GBSA binding energy prediction on the PDBbind data set: successes, failures, and directions for further improvement. *J. Chem. Inf. Model.* **2013**, 53, 201-209.
(23) Wang, B.; Li, L.; Hurley, T. D.; Meroueh, S. O., Molecular recognition in a diverse set of protein-ligand interactions studied with molecular dynamics simulations and end-point free energy calculations. *J. Chem. Inf. Model.* **2013**, 53, 2659-2670.

(24) Duan, L.; Liu, X.; Zhang, J. Z., Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein-Ligand Binding Free Energy. *J. Am. Chem. Soc.* **2016**, 138, 5722-5728.
(25) Huang, K.; Luo, S.; Cong, Y.; Zhong, S.; Zhang, J. Z. H.; Duan, L., An accurate free energy estimator: based on MM/PBSA combined with interaction entropy for protein-ligand binding affinity. *Nanoscale* **2020**, 12, 10737-10750.
(26) Sun, H.; Duan, L.; Chen, F.; Liu, H.; Wang, Z.; Pan, P.; Zhu, F.; Zhang, J. Z. H.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. *Phys. Chem. Chem. Phys.* **2018**, 20, 14450-14460.
(27) Akkus, E.; Tayfuroglu, O.; Yildiz, M.; Kocak, A., Revisiting MMPBSA by Adoption of MC-Based Surface Area/Volume, ANI-ML Potentials, and Two-Valued Interior Dielectric Constant. *J. Phys. Chem. B* **2023**, 127, 4415-4429.
(28) Wei, H.; Zhao, Z.; Luo, R., Advancing MM/PBSA calculations with machine learning and cuda GPUs. *Biophys. J.* **2022**, 121, 527a-528a.
(29) King, E.; Qi, R.; Li, H.; Luo, R.; Aitchison, E., Estimating the Roles of Protonation and Electronic Polarization in Absolute Binding Affinity Simulations. *J. Chem. Theory. Comput.* **2021**, 17, 2541-2555.
(30) Greene, D.; Qi, R.; Nguyen, R.; Qiu, T.; Luo, R., Heterogeneous Dielectric Implicit Membrane Model for the Calculation of MMPBSA Binding Free Energies. *J. Chem. Inf. Model.* **2019**, 59, 3041-3056.
(31) Liu, X.; Zheng, L.; Qin, C.; Li, Y.; Zhang, J. Z. H.; Sun, Z., Screening Power of End-Point Free-Energy Calculations in Cucurbituril Host-Guest Systems. *J. Chem. Inf. Model.* **2023**, 63, 6938-6946.
(32) Jiang, D.; Du, H.; Zhao, H.; Deng, Y.; Wu, Z.; Wang, J.; Zeng, Y.; Zhang, H.; Wang, X., et al., Assessing the performance of MM/PBSA and MM/GBSA methods. 10. Prediction reliability of binding affinities and binding poses for RNA-ligand complexes. *Phys. Chem. Chem. Phys.* **2024**, 26, 10323-10335.
(33) Wang, E.; Liu, H.; Wang, J.; Weng, G.; Sun, H.; Wang, Z.; Kang, Y.; Hou, T., Development and Evaluation of MM/GBSA Based on a Variable Dielectric GB Model for Predicting Protein-Ligand Binding Affinities. *J. Chem. Inf. Model.* **2020**, 60, 5353-5365.
(34) Wang, E.; Fu, W.; Jiang, D.; Sun, H.; Wang, J.; Zhang, X.; Weng, G.; Liu, H.; Tao, P., et al., VAD-MM/GBSA: A Variable Atomic Dielectric MM/GBSA Model for Improved Accuracy in Protein-Ligand Binding Free Energy Calculations. *J. Chem. Inf. Model.* **2021**, 61, 2844-2856.
(35) Zhu, Y. X.; Sheng, Y. J.; Ma, Y. Q.; Ding, H. M., Assessing the Performance of Screening MM/PBSA in Protein-Ligand Interactions. *J. Phys. Chem. B* **2022**, 126, 1700-1708.
(36) Yang, M.; Bo, Z.; Xu, T.; Xu, B.; Wang, D.; Zheng, H., Uni-GBSA: an open-source and web-based automatic workflow to perform MM/GB(PB)SA calculations for virtual screening. *Brief. Bioinform.* **2023**, 24, bbad218.
(37) Barreto Gomes, D. E.; Galentino, K.; Sisquellas, M.; Monari, L.; Bouysset, C.; Cecchini, M., ChemFlow—From 2D Chemical Libraries to Protein-Ligand Binding Free Energies. *J. Chem. Inf. Model.* **2023**, 63, 407-411.
(38) Qi, R.; Luo, R., Robustness and Efficiency of Poisson-Boltzmann Modeling on Graphics Processing Units. *J. Chem. Inf. Model.* **2019**, 59, 409-420.
(39) Onufriev, A. V.; Case, D. A., Generalized Born Implicit Solvent Models for Biomolecules. *Annu Rev Biophys* **2019**, 48, 275-296.

(40) Tan, C.; Tan, Y. H.; Luo, R., Implicit nonpolar solvent models. *J. Phys. Chem. B* **2007**, 111, 12263-12274.
(41) Floris, F.; Tomasi, J., Evaluation of the dispersion contribution to the solvation energy. A simple computational model in the continuum approximation. *J. Comput. Chem.* **1989**, 10, 616-627.
(42) Pan, Y.; Gao, D.; Zhan, C. G., Modeling the catalysis of anti-cocaine catalytic antibody: competing reaction pathways and free energy barriers. *J. Am. Chem. Soc.* **2008**, 130, 5140-5149.
(43) Raha, K.; Merz, K. M., Jr., Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. *J. Med. Chem.* **2005**, 48, 4558-4575.
(44) Bardi, J. S.; Luque, I.; Freire, E., Structure-based thermodynamic analysis of HIV-1 protease inhibitors. *Biochemistry* **1997**, 36, 6588-6596.
(45) Gómez, J.; Freire, E., Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. *J. Mol. Biol.* **1995**, 252, 337-350.
(46) Baldwin, R. L., Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci.* **1986**, 83, 8069-8072.
(47) D'Aquino, J. A.; Gómez, J.; Hilser, V. J.; Lee, K. H.; Amzel, L. M.; Freire, E., The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **1996**, 25, 143-156.
(48) Schellman, J. A., The thermodynamics of urea solutions and the heat of formation of the peptide hydrogen bond. *C. R. Trav. Lab Carlsberg. Chim.* **1955**, 29, 223-229.
(49) Schellman, J. A., The stability of hydrogen-bonded peptide structures in aqueous solution. *C. R. Trav. Lab Carlsberg. Chim.* **1955**, 29, 230-259.
(50) Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S., The PDBbind database: methodologies and updates. *J. Med. Chem.* **2005**, 48, 4111-4119.
(51) Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R., Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf. Model.* **2019**, 59, 895-913.
(52) D.A. Case, H. M. A., K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, N. Forouzesh, M. Ghazimirsaeed, G. Giambaşu, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, J. Harris, Z. Huang, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, A. Kovalenko, T. Kurtzman, T.S. Lee, P. Li, Z. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D.M. York, C. Zhao, Q. Zhu, and P.A. Kollman, Amber 2020; University of California: San Francisco. **2020**.
(53) Søndergaard, C. R.; Olsson, M. H.; Rostkowski, M.; Jensen, J. H., Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory. Comput.* **2011**, 7, 2284-2295.
(54) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory. Comput.* **2015**, 11, 3696-3713.
(55) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, 25, 1157-1174.

(56) Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, 23, 1623-1641.
(57) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, 79, 926-935.
(58) Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. H., A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* **2001**, 22, 501-508.
(59) Miller, B. R., 3rd; McGee, T. D., Jr.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory. Comput.* **2012**, 8, 3314-3321.
(60) Sitkoff, D.; Sharp, K. A.; Honig, B., Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, 98, 1978-1988.
(61) Gallicchio, E.; Kubo, M. M.; Levy, R. M., Enthalpy−Entropy and Cavity Decomposition of Alkane Hydration Free Energies: Numerical Results and Implications for Theories of Hydrophobic Solvation. *J. Phys. Chem. B* **2000**, 104, 6271-6285.
(62) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K., Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **2012**, 55, 6582-6594.

TOC