PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2022, **24**, 14525

Comparative assessment of QM-based and MM-based models for prediction of protein–ligand binding affinity trends†

Sarah Maier, [a] Bishnu Thapa, [ab] Jon Erickson[b] and Krishnan Raghavachari *[a]

Methods which accurately predict protein–ligand binding strengths are critical for drug discovery. In the last two decades, advances in chemical modelling have enabled steadily accelerating progress in the discovery and optimization of structure-based drug design. Most computational methods currently used in this context are based on molecular mechanics force fields that often have deficiencies in describing the quantum mechanical (QM) aspects of molecular binding. In this study, we show the competitiveness of our QM-based Molecules-in-Molecules (MIM) fragmentation method for characterizing binding energy trends for seven different datasets of protein–ligand complexes. By using molecular fragmentation, the MIM method allows for accelerated QM calculations. We demonstrate that for classes of structurally similar ligands bound to a common receptor, MIM provides excellent correlation to experiment, surpassing the more popular Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) and Molecular Mechanics Generalized Born Surface Area (MM/GBSA) methods. The MIM method offers a relatively simple, well-defined protocol by which binding trends can be ascertained at the QM level and is suggested as a promising option for lead optimization in structure-based drug design.

Received 27th January 2022,
Accepted 21st May 2022

DOI: 10.1039/d2cp00464j

rsc.li/pccp

1. Introduction

The study of structure-function relationships in proteins has accelerated over the last decade due to newly emergent techniques in computational chemistry.[1,2] In particular, these techniques have become increasingly important for the study of protein–ligand binding. The binding of a ligand to a target protein is a thermodynamic process, whose direction is guided by the free energy of the system. Free energy steers all biochemical processes, and therefore its calculation is fundamental. However, obtaining experimentally significant enthalpies and free energies is often difficult, and for systems involving complex biological systems in their native aqueous environment, accurate free energy calculations remain a grand challenge for chemists.

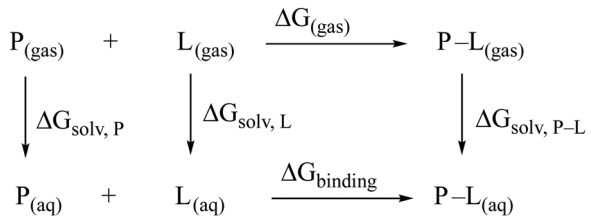
For computational investigations of protein–ligand interactions, there is a dearth of methods which calculate binding free energies while maintaining both accuracy and efficiency. Molecular docking, whose primary goal is to predict and rank ligand binding poses using computationally inexpensive scoring functions, is one of the most widely used computational approaches for studying protein–ligand binding.[3,4] While this technique is computationally economical, its accuracy and predictive ability are often inadequate. By contrast, alchemical free energy (AFE) methods, such as free energy perturbation (FEP) and thermodynamic integration (TI) are mathematically rigorous, though they are computationally demanding.[5–7] These methods involve extensive sampling of intermediate states, generated *via* small changes of the energy function, and are often slow to converge. While FEP-type methods are reliable and are used successfully in many applications, their accuracy still may suffer from the fundamental limitations of the underlying force fields.

The two sets of methods described above occupy opposite ends of the spectrum when considering both accuracy and efficiency. Representing a compromise between AFE methods and molecular docking are the so-called end-point free energy methods, which have become extremely popular in structure-based drug design.[8–17] As the name suggests, end-point methods sample only the bound and unbound states, unlike AFE methods which also sample intermediate states. By far, some of the most popular end-point free energy methods are the "Molecular Mechanics Poisson–Boltzmann Surface Area" (MM/PBSA) method and the "Molecular Mechanics Generalized Born Surface Area" (MM/GBSA) method developed by Kollman

[a] *Department of Chemistry, Indiana University, Bloomington, IN 47405, USA. E-mail: kraghava@indiana.edu*
[b] *Lilly Research Laboratories, Eli Lilly & Co., Indianapolis, Indiana 47285, USA*

† Electronic supplementary information (ESI) available: Description of molecular dynamics simulations as well as tables and figures. See DOI: https://doi.org/10.1039/d2cp00464j

et al.[18] In the MM/P(G)BSA approach, the free energy of binding is calculated using the following thermodynamic cycle:



Thus, the free energy of a ligand (L) binding to a target protein (P) to form a complex (P–L) is given as

$$\Delta G_{\text{binding}} = G_{\text{P-L}} - (G_{\text{P}} + G_{\text{L}}) \quad (1)$$

where the free energy of binding, $\Delta G_{\text{binding}}$, can be approximated as

$$\Delta G_{\text{binding}} \approx \Delta E_{\text{gas}} + \Delta G_{\text{solv}} - T\Delta S \quad (2)$$

where ΔE_{gas} is the gas phase binding energy typically calculated with MM potentials, ΔG_{solv} denotes the contribution from the energy of solvation, and $-T\Delta S$ is the change in configurational entropy upon ligand binding.[16,19] Terms in eqn (2) can be further decomposed:

$$\Delta E_{\text{gas}} \approx \Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{elec}} + \Delta E_{\text{vdW}} \quad (3)$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{SA}} \quad (4)$$

$$\Delta G_{\text{SA}} = \gamma \Delta \text{SASA} + b \quad (5)$$

where ΔE_{int} includes all contributions to the internal energy (bond, angle and dihedral terms), ΔE_{elec} represents the electrostatic contribution, and ΔE_{vdW} is the van der Waals contribution. The solvation term, ΔG_{solv}, is composed of an electrostatic (polar) contribution, $\Delta G_{\text{PB/GB}}$, and a nonpolar contribution, ΔG_{SA}. $\Delta G_{\text{PB/GB}}$ is calculated using either the PB or GB model, and ΔG_{SA} is approximated using the solvent-accessible surface area (SASA). For the MM/P(G)BSA methods, all terms in eqn (2) are typically taken as an average over many snapshots from a molecular dynamics (MD) trajectory, although a few studies have investigated the use of only a single minimized structure.[11,20] The change in configurational entropy, $-T\Delta S$, may be determined *via* normal-mode analysis on a set of MD snapshots; however, due to its high computational cost and larger uncertainty, entropy terms are often neglected when only relative free energies of structurally similar ligands bound to a common target receptor are desired. This choice assumes that entropic contributions are similar across a dataset, and therefore the neglect of such terms will have only a minimal effect on overall correlation.[8]

The accuracy of MM/P(G)BSA results, and more specifically of ΔE_{gas}, is directly related to the quality of the energy function used in the calculation.[10,21,22] Since ΔE_{gas} is most often calculated using MM potentials, the use of MM/P(G)BSA is limited to cases for which such a potential exists.[21,23–25] Moreover, for challenging and experimentally relevant cases, MM potentials often have significant deficiencies.[26–28] Conversely, QM potentials give a rigorous description of those electronic effects which play a critical role in intermolecular interactions (*e.g.*, charge-transfer, π–π interactions and many-body effects). As such, a high-level QM treatment of protein–ligand binding is highly desirable.

Several research groups have begun to study proteins using QM potentials.[10,29,30] However, because of their computational complexity, these calculations have typically been limited to minimal model systems, for example, including only a small number of residues around the ligand. In recent years, QM treatments of larger portions of protein–ligand complexes have begun to emerge, using either semiempirical methods or high-level methods with fragmentation techniques.[31,32] Among the many fragmentation-based methods, the fragment molecular orbital (FMO) technique, molecular fragmentation with conjugate caps (MFCC), and related approaches have been used in studying protein–ligand interactions.[33–37]

We have developed our own fragmentation method, the Molecules-in-Molecules (MIM) multilayer fragmentation method, to study protein–ligand interactions.[38,39] Our MIM fragmentation method is designed to produce high-level QM results efficiently for large systems. Similar to the ONIOM methodology developed by Morokuma and co-workers,[40] the MIM protocol is a hybrid method which employs layers of structural fragmentation to achieve highly accurate chemical calculations at a fraction of the typical cost. Over the course of several recent studies, we have successfully applied our method to over 100 protein–ligand complexes.[39,41–43] According to the MIM protocol, the full system is partitioned into smaller overlapping subsystems, and independent QM calculations are performed on each one. The energy contributions from all subsystems are then combined in such a way that the total energy of the full system is recovered. The MIM method is capable of providing high quality QM results at a substantially lower computational cost compared to traditional QM approaches. Moreover, the MIM method prescribes a well-defined protocol, which makes it particularly appealing for drug discovery purposes. The protocol has proven to be widely useful and has been effective in achieving high correlation (Spearman rank correlation (ρ) = 0.83–0.94; R^2 = 0.74–0.93) between calculated interaction energies and experimentally derived binding affinities for similar ligands binding to a target receptor.[39]

While we have demonstrated the effectiveness of our method in several studies, we have not yet made a direct comparison of the MIM method to the more popular end-point methods. MM/P(G)BSA methods achieve a reasonable balance between accuracy and computational cost; therefore, they serve as an appropriate standard to judge our own method for calculating and ranking binding interactions. In the present study, we make such a comparison, considering nearly 100 protein–ligand complexes, and demonstrate MIM's ability to achieve high correlation between theory and experiment.

2. Methods

2.1 Data sets and structure preparation

MIM-calculated binding affinities of 98 ligands, bound to five different target receptors, are compared to affinities obtained *via* MM/P(G)BSA analysis. A group of congeneric ligands bound

to a single target receptor comprises each of the seven datasets. Ligand structures were obtained from published cocrystallized structures or generated from a similar cocrystallized ligand. Assuming similar ligands bind similarly to a given receptor, each generated ligand was placed into the binding pocket of the published crystal structure. While modifying the ligand, the flexible alignment module, as implemented in the Molecular Operating Environment (MOE) program (version 2019.01), was used to conserve any key binding features observed in the cocrystallized structure. Missing hydrogen atoms in the crystal structure were added to the complexes at pH 7.0 with the Protonate 3D[44] tool, as implemented in MOE. Each protein–ligand complex was minimized in MOE with the AMBER10:EHT force field, using a generalized Born/volume integral implicit solvation model with an internal dielectric constant of 2 for the binding pocket and an external dielectric constant of 80. The minimization was performed using restraints *via* a harmonic potential centered on each atom, where the strength of the force constant was specified by the standard deviation (0.5 Å in this work) from the original coordinates at 300 K as implemented in MOE. Assuming the PDB structure is a good representation of the dominant binding pose, we minimize in order to relax the added hydrogens while keeping the overall structure close to the starting PDB structure. Protonation states of histidine residues present within 5 Å of the ligand were further determined by analyzing the possibility of hydrogen-bond formation with nearby residues. Our previous study considering 22 thrombin inhibitors showed that calculations which include residues within 5.0 Å of the ligand capture approximately 95 percent of the total interaction energy when compared to results calculated using the full complex.[43] In this same study, it was found that the coefficient of determination comparing theory to experiment is converged at about a 4 Å radius. Therefore, all residues and water molecules within 5.0 Å of the ligands were included in the MIM calculations. Any dangling bonds were saturated with hydrogen atoms using MOE. For MD, the full system was included in all simulations. Relevant information concerning protein–ligand structures considered in this work, along with corresponding PDB IDs are featured in Table 1 (set I,[45] set II,[31,46] set III,[47,48] set IV,[31,49] set V,[31,50] set VI,[51] and set VII[52,53]).

2.2 MIM calculations

The full MIM protocol has been detailed in previous works.[39,41,43] Herein, only the necessary protocol for our protein–ligand binding analysis is given. Similar to the ONIOM methodology, the MIM protocol is a hybrid method which employs layers of structural fragmentation to achieve highly accurate chemical calculations at a fraction of the typical cost. According to the MIM protocol, each subsequent fragmentation layer is comprised of increasingly smaller fragments (primary subsystems), with each subsequent layer treated at a higher (more accurate) level of theory. For example, the three-layer MIM protocol for calculating energies can be understood through eqn (6).

$$E^{\mathrm{MIM3}} = E^{\mathrm{R}}_{\mathrm{low}} + \left(E^{r'}_{\mathrm{medium}} - E^{r'}_{\mathrm{low}}\right) + \left(E^{r}_{\mathrm{high}} - E^{r}_{\mathrm{medium}}\right) \quad (6)$$

Eqn (6) involves three levels of theory (low, medium, and high) as well as three primary subsystem size parameters ($r < r' \ll R$). Resembling a telescoping series, each subsequent layer corrects errors present in the previous layer. According to the MIM protocol, proteins are first fragmented into small fragments (monomers) by cleaving single bonds between heavy atoms, as depicted in Fig. 2. Because of their partial double-bond character, peptide bonds are left uncut.[54]

Primary subsystems of each fragmentation layer are then created by grouping proximate monomers into larger, overlapping subsystems. Primary subsystems are created by expanding around an initiating monomer fragment, based on either a number- or distance-based criteria as described in previous works and as illustrated in Fig. 2. Under the number-based scheme, primary subsystems are generated by combining n covalently bonded monomers, with the initiating fragment as the central monomer. The prescription for the distance-based scheme involves the creation of primary subsystems *via* the grouping of all monomers within a certain cutoff distance, d, of the initiating monomer. In previous work, we have employed distance-based schemes in the middle layer in order to capture important long-range interactions at a satisfactory level of theory and number-based schemes in the high layer to capture bonded interactions with high accuracy. Additionally, we have found that a number-based high layer may be supplemented using a distance-based dimers scheme, under which interactions between pairs of primary subsystems whose distance falls within a certain cutoff ($\bar{d}$) are included in the high-layer (see Fig. 2). The distance-based dimer scheme ensures that certain short-range, nonbonded interactions (*e.g.* hydrogen bonding,

Table 1 Descriptors for each dataset

Set	Count	Ligand class	Receptor	PDB ID	Resolution (Å)
I	10	Biotin-based	Avidin	1AVD	2.70
I	14	Benzathiazole (BZT)-based	ITK	4MF0	2.67
III	13	—	CDK2	2VTA,2VTH, 2VTI, 2VTJ, 2VTL, 2VTM, 2VTN, 2VTO, 2VTQ, 2VTR, 2VU3, 2VTT, 2VTS	2.00, 1.90, 2.00, 2.20, 2.00, 2.25, 2.20, 2.19, 1.90, 1.89, 1.85, 1.68, 1.90
IV	11	Indazole (IND)-based	ITK	4PP9	2.58
V	18	Sulfonylpyridine (SAP)-based	ITK	4QD6	2.45
VI	16	D-Phe-Pro-based	Thrombin	2ZFF	1.47
VII	16	4-Aminopyridine benzamide-based	TYK2	4GIH	2.00

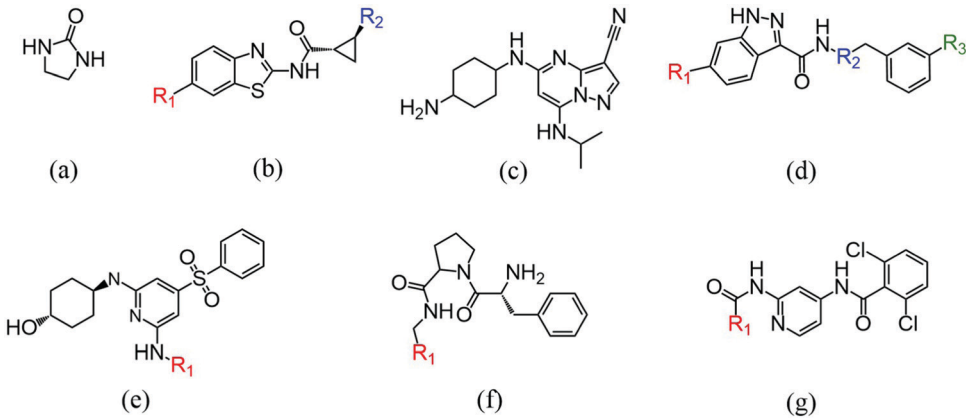


Fig. 1 Representative ligand structure for each dataset (a) biotin-analogue avidin ihibitors, (b) benzothiazle-based ITK inhibitors, (c) CDK2 inhibitors, (d) indazole-based ITK inhibitors, (e) sulfonylpyridine-based ITK inhibitors, (f) Thrombin inhibitors, (g) 4-aminopyridine benzamide-based TYK2 inhibitors.

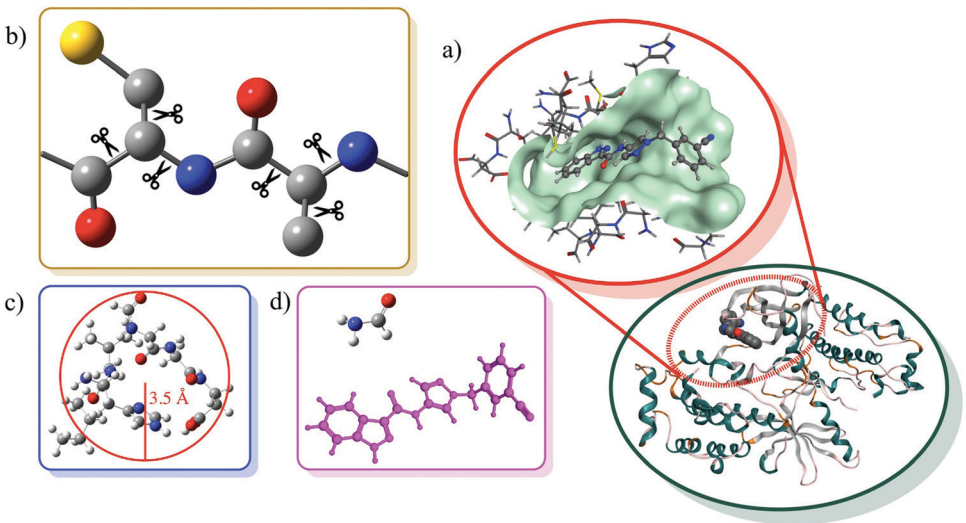


Fig. 2 Illustrations of the fragmentation scheme used in MIM3 calculations. (a) Depiction of 5.0 Å radius around ligand considered in MIM calculations. (b) Formation of monomers. (c) Example of distance-based middle layer subsystem. (d) Example of distance-based dimer pair used in high layer.

π–π stacking) are included at the high-level of theory. The low layer includes the full molecule and thus captures any long-range effects not included in the high and middle layers. After subsystems are generated, all remaining cleaved bonds are then saturated with hydrogens. Finally, contributions from overlapping regions are subtracted out *via* the Inclusion-Exclusion principle:

$$\left|a_1 \ldots \cup_n^a\right| = \sum_i |a_i| - \sum_{i<j} |a_i \cap_j^a| + \sum_{i<j<k} |a_i \cap_j^a \cap_k^a| \ldots + (-1)^{n-1} |a_i \cap \cdots \cap_n^a| \quad (7)$$

In multilayer MIM, less computationally expensive QM methods are used in the low layer to describe long-range interactions and polarization effects not present in individual subsystems, while more expensive but highly accurate methods are used to locally correct any errors present at the low-level of theory. In the present study, we use three-layer MIM (MIM3) using a distance-based (d = 3.5 Å) scheme at the middle level and a number-based (n = 5) scheme supplemented with a distance-based dimer scheme ($\bar{d}$ = 3.5 Å) at the high level. This particular protocol (denoted as **N5D**) has been benchmarked on similar complexes in previous studies.[39,41,43] Ligand molecules are left uncut.

2.3 Protein–ligand binding energy calculation with MIM3 protocol

The starting point is the gas phase interaction energy between a ligand and a protein, given as

$$\Delta E_{\text{interaction}}^{\text{gas}} = E_{\text{PL}} - (E_P + E_{\text{L}}) \quad (8)$$

where E_{PL}, E_{P}, and E_{L} are dispersion-corrected gas-phase electronic energies of the complex, protein, and ligand,

respectively. The protocol outlined in the previous section was used to calculate energies for the bound and unbound structures. Both bound and unbound structures were obtained from the same energy-minimized structure. All QM calculations were performed in the gas phase, and due to overestimation of electrostatic interactions, charged ligands and residues (Lys, Arg, His, Asp, and Glu) were neutralized to better match charge stabilization observed in solution. The effectiveness of this treatment has been validated in several previous studies.[31,39,41–43] While QM calculations are performed in gas phase, the penalty for ligand desolvation is included in the MIM calculation. Thus, the total MIM3 binding energy reported herein includes $\Delta E^{\text{gas}}_{\text{interaction}}$ as well as the penalty term from ligand desolvation (the penalty of abstracting ligand from solvent upon binding). Therefore, the total MIM3 binding energy is given as

$$\Delta E^{r}_{\text{MIM3}} = \Delta E^{\text{gas}}_{\text{interaction}} + \Delta E^{\text{L}}_{\text{(desolv)}} \quad (9)$$

where $\Delta E^{\text{L}}_{\text{(desolv)}}$ is the desolvation penalty of ligand scaled based on the buried solvent accessible surface area (SASA). The method used for determining the desolvation penalty is detailed in our previous works.[41]

For the high level of theory, the dispersion-corrected B97-D3BJ functional (B97 functional[55,56] with Grimme's D3 dispersion correction[57] and Becke—Johnson damping[58]) with 6-311++G(d,p) basis set was used. For the middle layer, B97-D3BJ was again used, however this time with the smaller 6-31+G(d) basis set.[59–62] Finally, PM6-D3 was used for the low layer.[63] The ligand desolvation penalty was determined using the SMD implicit polarizable continuum solvation model at the high level of theory.[64] All DFT calculations were performed using the Gaussian 16 program suite[65] and our MIM external program was used to generate and sum the fragmented systems.

2.4 Molecular dynamics simulations

The electrostatic potential (ESP) was obtained *via* geometry optimization at the SMD/B3LYP/6-31+G(d,p) level of theory followed by a Merz–Kollman ESP-fitted charge calculation at the HF/6-31G* level of theory.[66,67] The RESP method was used for charge fitting. Ligand atoms were described by the generalized Amber force field (GAFF).[68,69] All protein residues were described with the ff14SB force field.[70] Each complex was solvated in a truncated octahedral periodic box of TIP3P water molecules extending 18 Å from the solute. The crystal structure water molecules were removed before solvating with the explicit water molecules. Counterions of Na^+ or Cl^- were added to neutralize the system. Specific details concerning minimization, heating, equilibration, and production are given in the ESI.†

One prominent issue with MM/P(G)BSA methods is that there may exist several energy states that are rarely sampled from the MD trajectory, and therefore, one may obtain energies with high standard errors.[71–73] Several recent publications have indicated that the results of MM/P(G)BSA calculations are highly dependent upon the length of MD simulation. Multiple authors have explored issues with convergence and reproducibility of MM/P(G)BSA calculations and have suggested that averaging over several independent MD trajectories, initiated with random starting velocities, provides greater statistical significance than a single long trajectory.[73–76] Therefore, we ran 40 × 1 ns independent trajectories using random starting velocities (see "ig", "ntx", and "irest" flags in AMBER 18). The last recorded geometry of the 10 ns equilibration was used as the initial structure for each 1 ns trajectory run. Coordinates were recorded every 10 ps.

2.5 MM/PBSA calculations

Studies describing the effects of various force fields, charge models, radii sets, *etc.*, for MM/P(G)BSA calculations are abundant. In this study, we elected to follow a computational protocol similar to that most often encountered in the literature. The MM/PBSA calculations were performed using MMPBSA.py in AMBERTools18.[77] The solute and solvent dielectric constants were set to 1 and 80, respectively, and the salt concentration was set to 0.1 M. It should be noted that MM/PBSA is known to be particularly sensitive to the choice of solute dielectric, especially for charged binding pockets.[15] 1000 snapshots were taken evenly from the 40 × 1 ns production runs and used for MM/PBSA calculations. This sampling interval is believed to ensure that subsequent MD snapshots are sufficiently uncorrelated.[73] The default MM/PBSA surface tension and non-polar free energy correction were used, and the mbondi2 radii set was used. Entropic terms calculated *via* normal mode analysis were not included. Because we consider similar ligands bound to a common target receptor, entropic contributions are expected to be similar across each data set.

2.6 MM/GBSA calculations

The MM/GBSA calculations were also performed using MMPBSA.py in AMBERTools18. For this study, the polar contribution to solvation was obtained by solving the generalized Born solvation model of Onufriev *et al.* with α, β, and γ to 0.8, 0.0, and 2.909125, respectively (GB^{OBCI}/"igb = 2" flag in AMBER18).[78] The solute and solvent dielectric constants were set to 1 and 80, respectively, and the salt concentration was set to 0.1 M. As for the case of MM/PBSA calculations, 1000 snapshots were evenly extracted from the 40 × 1 ns production runs and used for MM/GBSA calculations. The default MM/GBSA surface tension and non-polar free energy correction terms were used, and the mbondi2 radii set was used. As in the case of MM/PBSA calculations, entropic terms were not included.

3. Results and discussion

In the lead optimization stage of drug discovery, the goal is to develop drug molecules with improved selectivity and potency relative to the lead compound. Therefore, the ability to rank a series of structurally similar molecules in terms of their capacity to bind to target receptors becomes an invaluable tool. In

this section, we make a direct comparison of the QM-based MIM model to the more popular MM/P(G)BSA methods and determine their relative performance when it comes to correctly ranking binding free energies. To make this comparison, we have elected to study a series of structurally and chemically diverse ligands, bound to one of five different target receptors. A representative structure for each dataset is given in Fig. 1. For a complete structural description of each ligand, see Fig. S1–S7 (ESI†).

Often used in structure-based drug design, MM/P(G)BSA methods are regarded as well suited for ranking structurally similar ligands according to their ability to bind to a target receptor, rather than for obtaining absolute binding energies.[9,13,17,79] Indeed, systematic inaccuracies of MM potentials and the common neglect of entropic contributions renders MM/P(G)BSA methods perhaps more appropriate for identifying correlation between theory and experiment. Similarly, the MIM method has been proven effective at producing binding energy trends that are well correlated with experiment. Therefore, we have elected to compare the MM/P(G)BSA and MIM methods based on their ability to calculate binding strength trends of structurally similar ligands bound to a common target receptor. We employ two different metrics to assess the performance of the methods tested: the Pearson correlation coefficient, r_p, and the Spearman rank-order correlation coefficient, r_s.

The correlation plots comparing the experimental ΔG_{bind} or pK_i and the calculated interaction energies for MM/GBSA with sampling over 40 × 1 ns trajectory runs are shown in Fig. 3. The corresponding plots for MM/PBSA are shown in Fig. 4. The MM/GBSA results show an overall strong correlation, with five of the seven test sets achieving r_s values above 0.6. The Pearson correlation coefficient for set I, calculated with MM/GBSA, is particularly impressive at r_p = 0.95. However, it is perhaps expected that the dataset with the widest range of experimental affinities achieves the highest correlation. As experimental affinities become more closely spaced, the task of ranking ligands *via* computational means becomes more difficult, as in the case of set V, which contains the most datapoints as well as the narrowest range of experimental affinities. The precision and rank-ordering capability of a particular method is determined by the inherent errors and uncertainties present within the method. If differences between affinities is smaller than the precision of the method, then erroneous performance may result. This idea is discussed further below.

In contrast to MM/GBSA, MM/PBSA results (Fig. 4) are somewhat mixed, with the degree of correlation varying

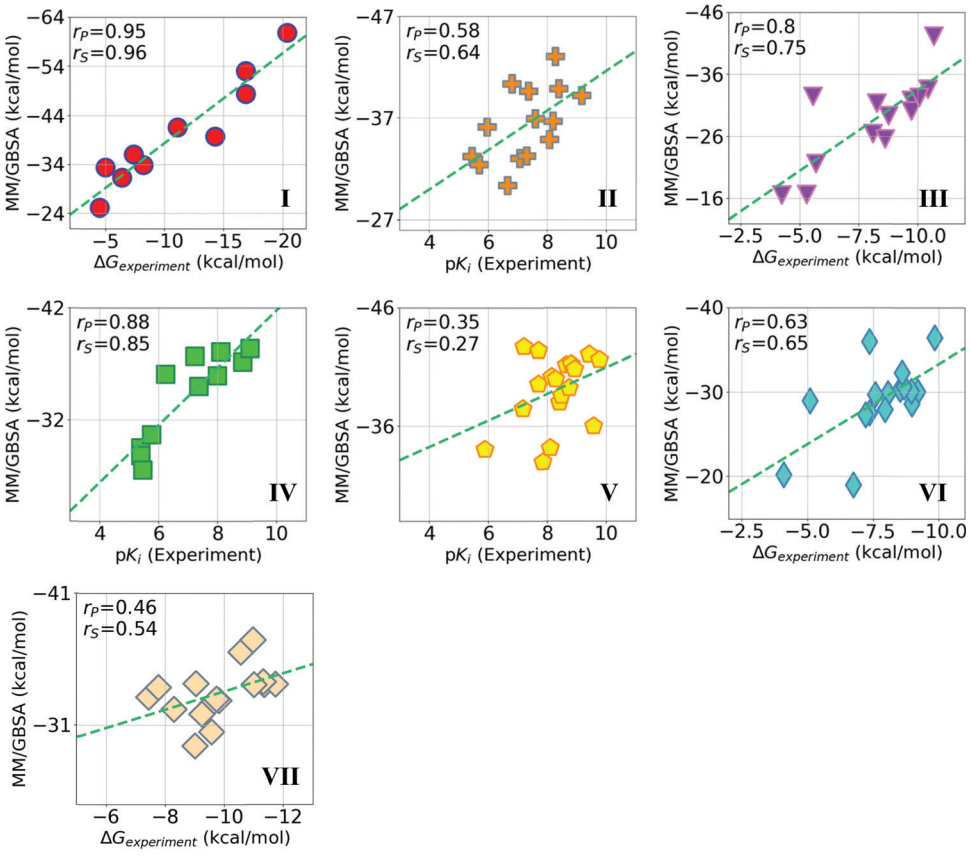


Fig. 3 Correlation (r_p = Pearson, r_S = Spearman) between experimentally measured binding affinities and **MM/GBSA** (40x1 ns trajectory run with ε = 1) calculated binding energies for various protein-ligand complexes (I) biotin-analogue avidin inhibitors, (II) benzothiazole-based ITK inhibitors, (III) CDK2 inhibitors, (IV) indazole-based ITK inhibitors, (V) sulfonylpyridine-based ITK inhibitors, (VI) Thrombin inhibitors, (VII) 4-aminopyridine benzamide-based TYK2 inhibitors.

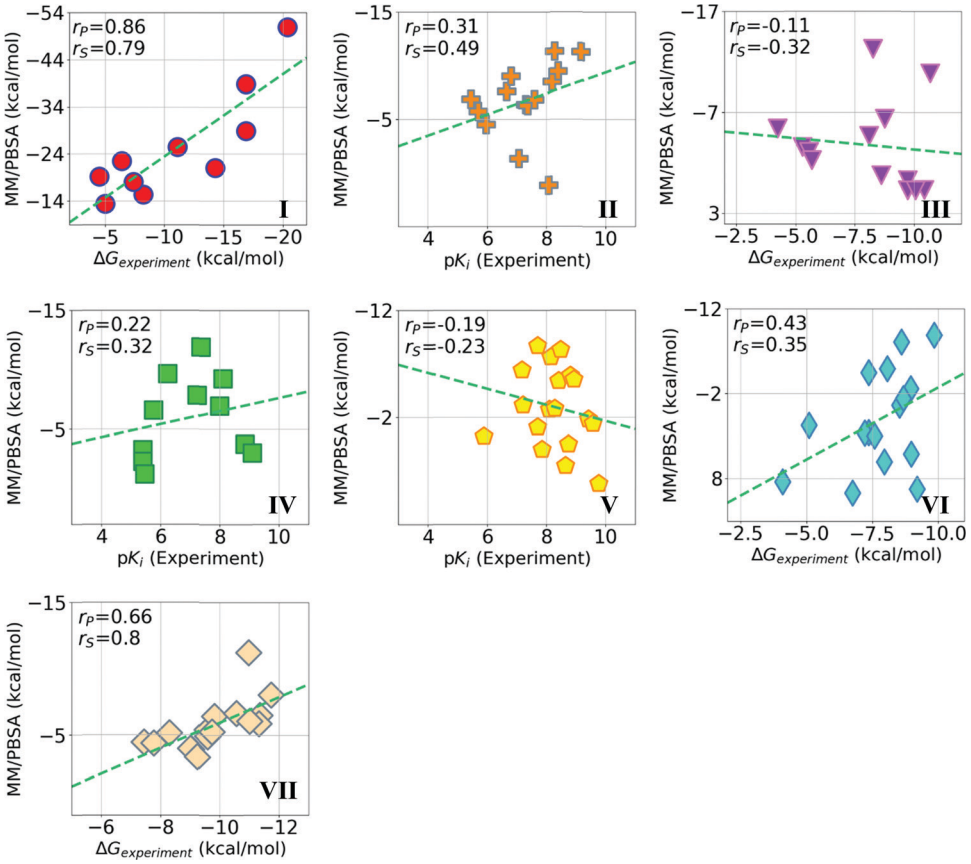


Fig. 4 Correlation (r_p = Pearson, r_s = Spearman) between experimentally measured binding affinities and **MM/PBSA** (40x1 ns trajectory run with ε = 1) calculated binding energies for various protein-ligand complexes (I) biotin-analogue avidin inhibitors, (II) benzothiazole-based ITK inhibitors, (III) CDK2 inhibitors, (IV) indazole-based ITK inhibitors, (V) sulfonylpyridine-based ITK inhibitors, (VI) Thrombin inhibitors, (VII) 4-aminopyridine benzamide-based TYK2 inhibitors.

between datasets. For MM/PBSA, the biotin-analogue avidin inhibitors, set I, and the 4-aminopyridine benzamide-based TYK2 inhibitors, set VII, achieve good correlation, with Pearson coefficients of 0.86 and 0.66, respectively. Here again, the dataset with the widest range of experimental affinities (set I) achieves the highest correlation. Although set I and VII show good correlation using MM/PBSA, the remainder of the datasets display rather poor correlation, with r_p values of 0.31, −0.11, 0.22, −0.19, and 0.43 for set II, III, IV, V, and VI, respectively.

It is perhaps counterintuitive that GBSA gives better overall correlation than PBSA, given that the former is a more approximate method. A few studies have also found cases for which GBSA gives higher correlation with experiment than PBSA.[9,80,81] In this context, it should be noted that the results of MM/P(G)BSA calculations are dependent on parameter and protocol choice. Several studies have explored the effect of varying parameters.[13,17,81] These studies often indicate that optimum parameter and protocol choice may be system dependent. Since we elected to follow a standardized protocol for all seven test sets in this study,[15] it is possible that parameter optimization may improve results in some cases. A particularly relevant study in 2013 showed that for an internal dielectric constant of 1 (ε = 1), the GBSA method outperforms PBSA when it comes to rank-ordering.[82] However, setting ε = 10 resulted in significantly better rank-ordering for PBSA. In fact, even doubling ε led to a significant improvement in performance.

Given these results, we elected to run MM/PBSA calculations for ε = 4. Results are shown in Fig. 5. We saw a similar overall improvement in correlation for PBSA when compared to results for ε = 1. The improvement was particularly noticeable for complexes in set III that feature a charged binding pocket. After the increase in ε, PBSA results are on par with those obtained using GBSA. Despite the improvement in PBSA results, the 2013 study found that the improvement with higher ε was attributed to a dulled effect of the electrostatic contribution due to screening. Moreover, the authors found no correlation between the electrostatic components of the experimental free energy and the calculated free energy. We explored this possibility for set III and had similar findings. Results are shown in Fig. S16 (ESI†). The electrostatic part of the MM/PBSA free energy for ε = 4 was anticorrelated to the experimental free energy, while the van der Waals term was well correlated.

An acknowledged deficiency of MM/P(G)BSA methods is the large variation in per snapshot energies, which can lead to high standard deviations—typically 5–40 kcal mol^{-1} for MM/GBSA if a single trajectory is used for all three reactants.[73] If one hopes

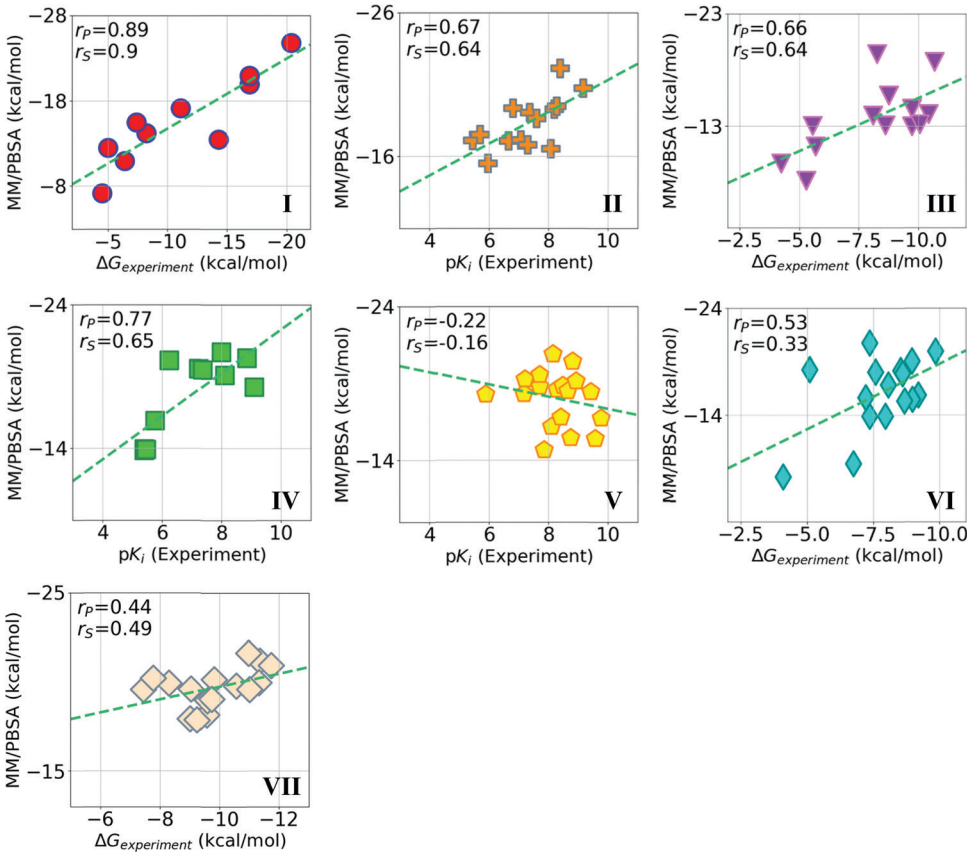


Fig. 5 Correlation (r_p = Pearson, r_S = Spearman) between experimentally measured binding affinities and **MM/PBSA** (40x1 ns trajectory run with ε = 4) calculated binding energies for various protein-ligand complexes (I) biotin-analogue avidin inhibitors, (II) benzothiazole-based ITK inhibitors, (III) CDK2 inhibitors, (IV) indazole-based ITK inhibitors, (V) sulfonylpyridine-based ITK inhibitors, (VI) thrombin inhibitors, (VII) 4-aminopyridine benzamide-based TYK2 inhibitors.

to achieve statistically significant analyses, converged energies are needed. Fig. 6 shows per snapshot binding energy distributions calculated using MM/PBSA over 40 × 1 ns trajectory runs for the 4-aminopyridine benzamide-based TYK2 inhibitor dataset (set VII), along with corresponding Gaussian fits. Overall, the 1000 energies sampled for each protein-ligand complex form a near Gaussian distribution with a standard error (S. E., standard deviation divided by $\sqrt{1000}$) below or near ~0.25 kcal mol^{-1}, indicating the energies are sufficiently converged. Per snapshot MM/PBSA energy distributions over 40 × 1 ns trajectory runs for the other six datasets can be found in Fig. S10–S15 (ESI†). Expected near Gaussian distributions are seen in all datasets with few exceptions.

Next, we make a direct comparison of MIM to the more popular MM/P(G)BSA methods and determine the relative performance of the three models when it comes to correctly ranking binding free energies. Therefore, we have determined the binding affinities of each of the seven datasets using our MIM3 method. Fig. 7 displays the correlation between MIM3 calculated interaction energies and experimental affinities for each of the seven tests sets. As mentioned earlier, the reported MIM3 values include the gas-phase interaction energy along with the contribution from ligand desolvation. The MIM3 method yields well correlated results for all seven datasets, with Pearson coefficients ranging from 0.81 for the 4-aminopyridine benzamide-based TYK2 inhibitors to 0.97 for the indazole-based ITK inhibitors (Spearman rank-order coefficients of 0.84 and 0.92, respectively). Using MIM, the indazole-based ITK inhibitors, set IV, achieve the best r_p value of 0.97. The MM/P(G)BSA methods deliver weaker overall correlation to experiment when compared to MIM3. The difference in performance between these methods is perhaps most obvious in the case of the sulfonylpyridine-based ITK inhibitors, as its Pearson correlation improves from 0.35, and ~−0.2 for MM/GBSA and MM/PBSA, respectively, to 0.86 for MIM3. While the absolute interaction energies calculated by MIM3 are much larger than their experimental counterparts, the overestimation of binding strength is clearly systematic within a group of similar ligands, as seen in the excellent correlation achieved by MIM3.

Fig. 8(a) shows a linear plot between experimental binding energies and MIM3 calculated energies for all seven datasets combined together instead of individually as in Fig. 7. Since different datasets have different receptor binding sites, error cancellation is expected to be poor across datasets. Indeed, neglected receptor-specific effects, such as contributions to the energy from configurational entropy, result in different linear

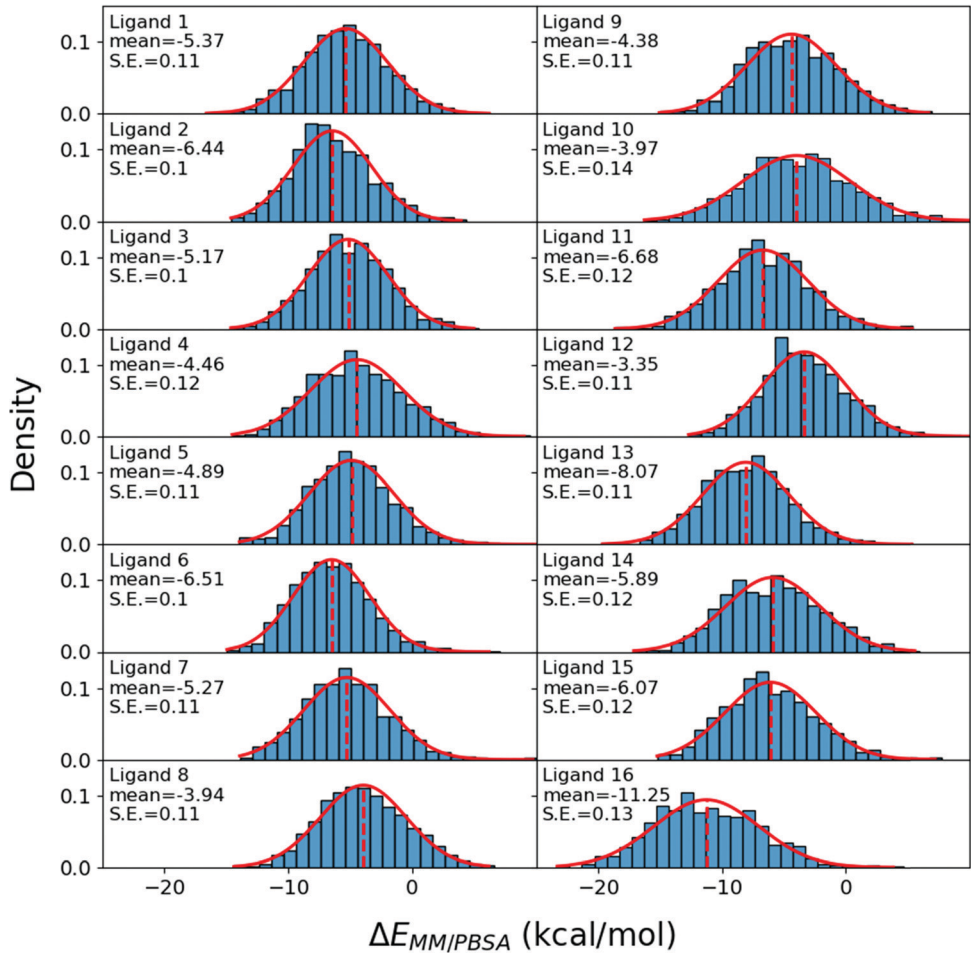


Fig. 6 Per snapshot distribution of MM/PBSA (ε = 1) energies in kcal mol^{-1} over the 40 × 1 ns trajectory runs for 4-aminopyridine benzamide-based TYK2 inhibitors.

slopes for each dataset, leading to a poor overall fit in Fig. 8(a). However, the complexes of set II, IV, and V, despite having differing ligand scaffolds share a common receptor. Linear trends for these three datasets are well matched, and so a predictive model can be made, as seen in Fig. 8(b). Therefore, MIM stands as a good candidate for lead optimization, where the more relevant goal is to optimize an initial lead compound for a particular receptor.

The resolution of a particular method is related to the variability of the data. Methods which have low standard error are considered more precise. Uncertainty estimates for MM/P(G)BSA methods, which sample over multiple frames, are given by the standard error, shown in Fig. 6 and Fig. S10–S15 (ESI†). In contrast, MIM starts from a single PDB-based geometry and employs a well-defined protocol for structure preparation and minimization. Therefore, aspects such as the quality of the crystal structure and the flexibility around the binding cavity will affect the resolution of the method. However, for a given dataset involving several ligands in a specific binding site, a single energy-minimized receptor structure is used in modelling. Our protocol guarantees that the overall binding pose is close to the one found in the PDB crystal structure. Thus, errors specific to a particular structure are expected to be mostly systematic across a dataset. However, such error cancellation is unlikely for comparisons across different receptors, as evident in the differing linear trends in Fig. 8(a).

It is important to determine whether the improved performance of MIM is indeed due to the accuracy of QM methods rather than the strategy of using single frames. Therefore, we also performed MM/P(G)BSA calculations on single snapshots for each complex in dataset II (benzothiazole-based inhibitors) and dataset V (sulfonylpyridine-based ITK inhibitors). Dataset V was specifically chosen as correlation between theory and experiment for this set showed drastic improvement upon the use of MIM3. For MM/P(G)BSA calculations, the minimized structures used in the MD simulations were used as the single frames. PBSA calculations were run with ε = 4. While there is a slight improvement in the correlations, the r_p values (Pearson rank-order coefficients) for MM/GBSA (0.31 and 0.61 for sets II and V) and MM/PBSA (0.63 and 0.31) are substantially worse than MIM3 (0.87 and 0.86). Results for these calculations are shown in Fig. S17 and S18 (ESI†). Additionally, we performed low-level PM6-D3 calculations on the 5 Å cutout used in the

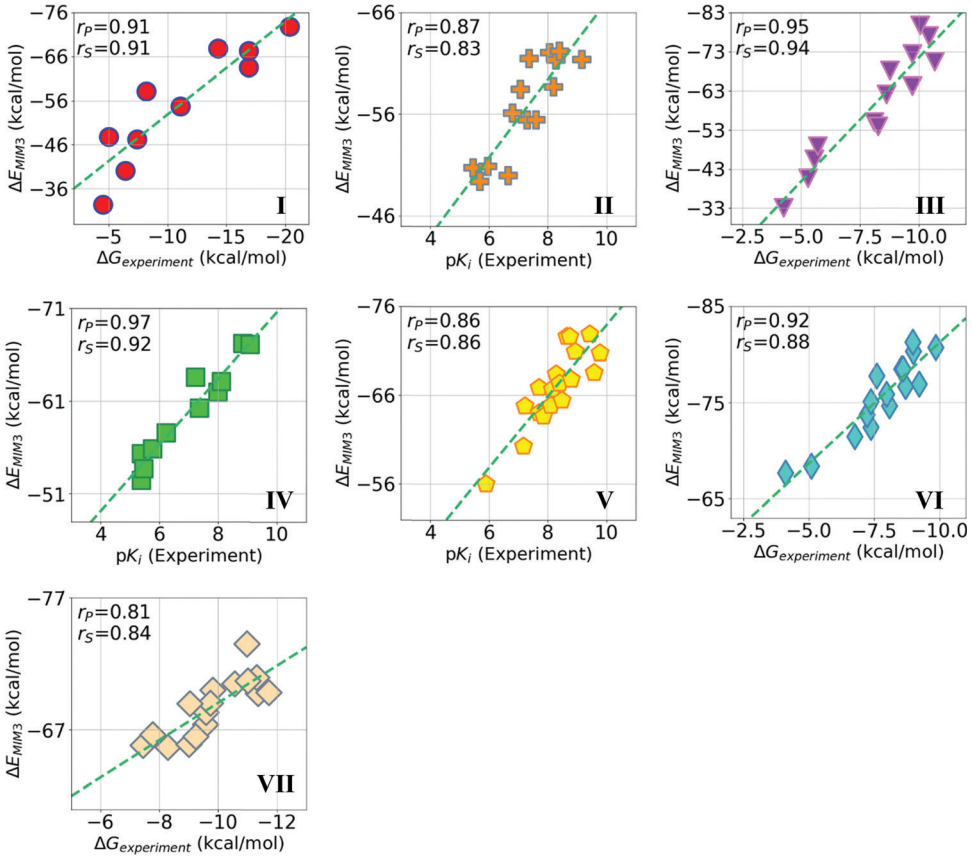


Fig. 7 Correlation (r_p = Pearson, r_S = Spearman) between experimentally measured binding affinities and **MIM3** calculated binding energies for various protein-ligand complexes (I) biotin-analogue avidin inhibitors, (II) benzothiazole-based ITK inhibitors, (III) CDK2 inhibitors, (IV) indazole-based ITK inhibitors, (V) sulfonylpyridine-based ITK inhibitors, (VI) Thrombin inhibitors, (VII) 4-aminopyridine benzamide-based TYK2 inhibitors.

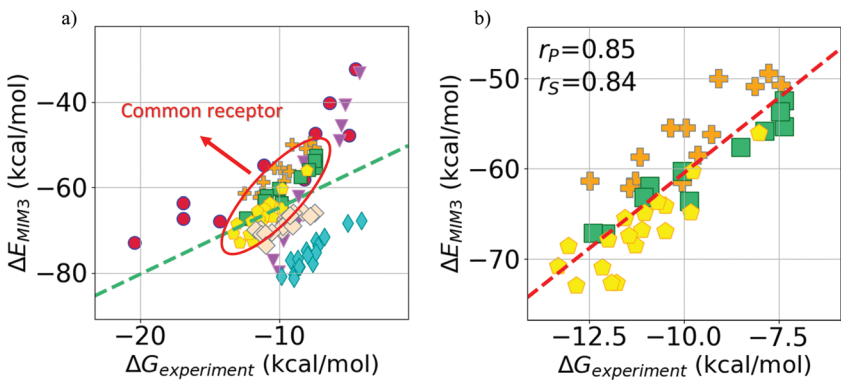


Fig. 8 (a) Linear plots between experimentally measured binding affinities and MIM3 calculated binding energies for all seven datasets (b) linear plots showing correlation for datasets involving the ITK receptor, *i.e.* (II) benzothiazole-based ITK inhibitors, indazole-based ITK inhibitors, (V) sulfonylpyridine-based ITK inhibitors.

MIM3 protocol, *i.e.*, using the same geometry, protonation states, system truncation and solvation models. Again, the r_p values for PM6-D3 (0.74 and 0.27) show that high-level methods such as MIM3 are needed for an improved correlation. Thus, MIM3 results outperform all methods, even in the single frame case. Overall, these results suggest that while improvements may be seen in individual cases by using single frames, consistent high performance in predicting rank-orders is only obtained by using a high-level QM method.

Overall, MIM3 offers the strongest correlation with experiment among the three methods. The strength of correlation achieved by the MIM3 method as compared to MM/P(G)BSA methods demonstrates the advantages of QM potentials when it comes to studying important interactions in the

protein–ligand complexes. Moreover, the MIM method calls for a well-defined protocol, which makes it particularly attractive for studying large datasets.

It is important to note that there are some significant differences in the protocols used in our QM-based MIM model and MM-based P(G)BSA models. A particularly noteworthy aspect is that excellent performance is obtained with MIM3 despite our use of only a single energy-minimized structure. We have found that by using a well-resolved crystal structure as a template for a series of congeneric ligands binding to a common receptor, high correlation with experiment has been obtained by MIM, even without sampling. This suggests that other factors such as conformational change or entropy change are significantly conserved in our comparisons between different ligands binding to the same receptor site. This is fortunate since the use of a QM-based method such as MIM will be computationally prohibitive to explore in conjunction with extensive sampling. In this context, a few MM-based studies have also explored the possibility of using a single energy-minimized structure for the calculation of free energies. A 2008 study showed that reasonable accuracy could be achieved for the calculation of free energies of reactions in proteins using QM/MM-MMPBSA, even when no sampling was performed.[11] Similarly, a 2009 study showed that using a single minimized protein–ligand structure in binding free energy calculations resulted in accuracies similar to those obtained after averaging over multiple MD snapshots.[20] Furthermore, the authors found that the use of a single structure proved effective and useful when applied to a virtual screening experiment. It will be interesting to explore the validity of such ideas in future work. Nevertheless, for a class of structurally similar ligands bound to a common target receptor, a common scenario in lead optimization, MIM is a promising option.

4. Computational efficiency of MIM

Reduction of computational cost is central to the MIM method. By fragmenting the complex and treating the full low-level calculation with semiempirical methods, the method achieves substantially improved scaling. Parallelization of fragment calculations lends added efficiency to MIM. The larger fragments in the middle layer of the MIM3 fragmentation scheme are typically the bottleneck in the calculations. As the size of the protein–ligand complex grows, the ratio between fragment size and full system size shrinks, revealing MIM's true computational advantage. As an illustration, for the first complex in dataset II at the B97D3/6-311++G(d,p) level, the computational cost for the costliest fragment is 1.5% of the full system high-level calculation. Nevertheless, MIM calculations are still computationally demanding so that only single-point energy evaluations on selected structures are feasible, as carried out in this study.

5. Conclusions

In the context of drug discovery, the ability to quickly rank binding strengths necessitates a method which is capable of capturing the important physics of a problem at a reasonable rate. By taking advantage of chemical fragmentation and QM potentials, the MIM method is able to provide a physically accurate description of electronic structure at a fraction of the typical cost. Moreover, when applied to the problem of structurally similar ligands bound to a common target receptor, the MIM method proves extremely effective. In this study we test our method against two popular endpoint methods for computing and ranking protein–ligand binding strengths, MM/PBSA and MM/GBSA. In total, we have calculated binding energies for nearly 100 chemically diverse ligands using the three methods mentioned.

Overall, the MM/GBSA method gave results more closely correlated with experiment than MM/PBSA. Most notably, however, the MIM method outperforms both MM/GBSA and MM/PBSA methods. MM/P(G)BSA methods involve extensive sampling over a MD trajectory using MM methods, while the MIM method uses QM potentials and considers only a single energy-minimized structure. Because we limit our studies to a set of structurally similar ligands bound to a common target receptor, where conformational and entropic changes are expected to be conserved, the use of a single energy-minimized structure proves sufficient and yields good correlation with experiment. To summarize, our results show the advantage of using a well-defined QM-based protocol where the computational cost is greatly reduced *via* molecular fragmentation.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We acknowledge support from the National Science Foundation grants CHE-1665427 and CHE-2102583 at Indiana University. This work was also supported by Eli Lilly and Company through the Lilly Research Award Program (LRAP) at Indiana University. Big Red III supercomputing facility at Indiana University was used for most of the calculations in this study.

References

1 N. Hansen and W. F. van Gunsteren, *J. Chem. Theory Comput.*, 2014, **10**, 2632–2647.
2 D. Mucs and R. A. Bryce, *Expert Opin. Drug Discovery*, 2013, **8**, 263–276.
3 C. J. Morris and D. Della Corte, *Mod. Phys. Lett. B*, 2021, **35**, 2130002.
4 L. Pinzi and G. Rastelli, *Int. J. Mol. Sci.*, 2019, **20**, 4331.
5 S. Boresch and H. L. Woodcock, *Mol. Phys.*, 2017, **115**, 1200–1213.
6 M. T. Wang, Y. Mei and U. Ryde, *J. Chem. Theory Comput.*, 2019, **15**, 2659–2671.
7 B. J. Williams-Noonan, E. Yuriev and D. K. Chalmers, *J. Med. Chem.*, 2018, **61**, 638–649.

8 S. Genheden and U. Ryde, *Expert Opin. Drug Discovery*, 2015, **10**, 449–461.
9 T. Hou, J. Wang, Y. Li and W. Wang, *J. Comput. Chem.*, 2011, **32**, 866–877.
10 S. K. Mishra and J. Koca, *J. Phys. Chem. B*, 2018, **122**, 8113–8121.
11 G. Rastelli, A. Del Rio, G. Degliesposti and M. Sgobba, *J. Comput. Chem.*, 2010, **31**, 797–810.
12 P. C. Su, C. C. Tsai, S. Mehboob, K. E. Hevener and M. E. Johnson, *J. Comput. Chem.*, 2015, **36**, 1859–1873.
13 H. Y. Sun, Y. Y. Li, S. Tian, L. Xu and T. J. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16719–16729.
14 T. Tuccinardi, *Expert Opin. Drug Discovery*, 2021, 1–5, DOI: 10.1080/17460441.2021.1942836.
15 C. H. Wang, P. H. Nguyen, K. Pham, D. Huynh, T. B.-N. Le, H. L. Wang, P. Y. Ren and R. Luo, *J. Comput. Chem.*, 2016, **37**, 2436–2446.
16 E. C. Wang, H. Y. Sun, J. M. Wang, Z. Wang, H. Liu, J. Z.-H. Zhang and T. J. Hou, *Chem. Rev.*, 2019, **119**, 9478–9508.
17 L. Xu, H. Sun, Y. Li, J. Wang and T. Hou, *J. Phys. Chem. B*, 2013, **117**, 8408–8421.
18 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, Accounts Chem, *Res*, 2000, **33**, 889–897.
19 T. Tuccinardi, *Expert Opin. Drug Discovery*, 2021, **16**, 1233–1237.
20 M. Kaukonen, P. Soderhjelm, J. Heimdal and U. Ryde, *J. Phys. Chem. B*, 2008, **112**, 12537–12548.
21 L. Gundelach, T. Fox, C. S. Tautermann and C. K. Skylaris, *Phys. Chem. Chem. Phys.*, 2021, **23**, 9381–9393.
22 M. Retegan, A. Milet and H. Jamet, *J. Chem. Inf. Model.*, 2009, **49**, 963–971.
23 R. T. Bradshaw, J. Dziedzic, C. K. Skylaris and J. W. Essex, *J. Chem. Inf. Model.*, 2020, **60**, 3131–3144.
24 G. Konig and S. Riniker, *Interface Focus*, 2020, 10.
25 P. Pokorna, H. Kruse, M. Krepl and J. Sponer, *J. Chem. Theory Comput.*, 2018, **14**, 5419–5433.
26 V. S.-S. Inakollu, D. P. Geerke, C. N. Rowley and H. B. Yu, *Curr. Opin. Struct. Biol.*, 2020, **61**, 182–190.
27 A. L. Tran and E. B. Guidez, *ACS Omega*, 2020, **5**, 7446–7455.
28 Z. Y. Yan, X. Li and L. W. Chung, *J. Chem. Theory Comput.*, 2021, **17**, 3783–3796.
29 H. J. Kulik, J. Y. Zhang, J. P. Klinman and T. J. Martinez, *J. Phys. Chem. B*, 2016, **120**, 11381–11394.
30 H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198–1229.
31 A. Heifetz, G. Trani, M. Aldeghi, C. H. MacKinnon, P. A. McEwan, F. A. Brookfield, E. I. Chudyk, M. Bodkin, Z. H. Pei, J. D. Burch and D. F. Ortwine, *J. Med. Chem.*, 2016, **59**, 4352–4363.
32 J. F. Liu, X. W. Wang, J. Z.-H. Zhang and X. He, *RSC Adv.*, 2015, **5**, 107020–107030.
33 M. A. Collins and R. P.-A. Bettens, *Chem. Rev.*, 2015, **115**, 5607–5642.
34 D. G. Fedorov, *WIREs Comput. Mol. Sci.*, 2017, 7.
35 M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, *Chem. Rev.*, 2012, **112**, 632–672.
36 J. M. Herbert, *J. Chem. Phys.*, 2019, **151**, 170901.
37 X. S. Jin, J. Z.-H. Zhang and X. He, *J. Phys. Chem. A*, 2017, **121**, 2503–2514.
38 N. J. Mayhall and K. Raghavachari, *J. Chem. Theory Comput.*, 2011, **7**, 1336–1343.
39 B. Thapa, D. Beckett, K. V.-J. Jose and K. Raghavachari, *J. Chem. Theory Comput.*, 2018, **14**, 1383–1394.
40 L. W. Chung, W. M.-C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. F. Ke, F. Y. Liu, H. B. Li, L. N. Ding and K. Morokuma, *Chem. Rev.*, 2015, **115**, 5678–5796.
41 B. Thapa, D. Beckett, J. Erickson and K. Raghavachari, *J. Chem. Theory Comput.*, 2018, **14**, 5143–5155.
42 B. Thapa, J. Erickson and K. Raghavachari, *J. Chem. Inf. Model.*, 2020, **60**, 2924–2938.
43 B. Thapa and K. Raghavachari, *J. Chem. Inf. Model.*, 2019, **59**, 3474–3484.
44 P. Labute, *Proteins: Struct., Funct., Genet.*, 2009, **75**, 187–205.
45 N. M. Green, *Advances in Protein Chemistry*, Elsevier, 1975, vol. 29, pp. 85–133.
46 C. H. MacKinnon, K. Lau, J. D. Burch, Y. Chen, J. Dines, X. Ding, C. Eigenbrot, A. Heifetz, A. Jaochico, A. Johnson, J. Kraemer, S. Kruger, T. M. Krulle, M. Liimatta, J. Ly, R. Maghames, C. A.-G. N. Montalbetti, D. F. Ortwine, Y. Perez-Fuertes, S. Shia, D. B. Stein, G. Trani, D. G. Vaidya, X. L. Wang, S. M. Bromidge, L. C. Wu and Z. H. Pei, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 6331–6335.
47 M. Congreve, G. Chessari, D. Tisi and A. J. Woodhead, *J. Med. Chem.*, 2008, **51**, 3661–3680.
48 P. G. Wyatt, A. J. Woodhead, V. Berdini, J. A. Boulstridge, M. G. Carr, D. M. Cross, D. J. Davis, L. A. Devine, T. R. Early, R. E. Feltell, E. J. Lewis, R. L. McMenamin, E. F. Navarro, M. A. O'Brien, M. O'Reilly, M. Reule, G. Saxty, L. C.-A. Seavers, D. M. Smith, M. S. Squires, G. Trewartha, M. T. Walker and A. J.-A. Woolford, *J. Med. Chem.*, 2008, **51**, 4986–4999.
49 R. M. Pastor, J. D. Burch, S. Magnuson, D. F. Ortwine, Y. Chen, K. De La Torre, X. Ding, C. Eigenbrot, A. Johnson, M. Liimatta, Y. C. Liu, S. Shia, X. L. Wang, L. C. Wu and Z. H. Pei, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 2448–2452.
50 G. Trani, J. J. Barker, S. M. Bromidge, F. A. Brookfield, J. D. Burch, Y. Chen, C. Eigenbrot, A. Heifetz, M. H.-A. Ismaili, A. Johnson, T. M. Krulle, C. H. MacKinnon, R. Maghames, P. A. McEwan, C. A.-G. N. Montalbetti, D. F. Ortwine, Y. Perez-Fuertes, D. G. Vaidya, X. L. Wang, A. A. Zarrin and Z. H. Pei, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 5818–5823.
51 B. Baum, M. Mohamed, M. Zayed, C. Gerlach, A. Heine, D. Hangauer and G. Klebe, *J. Mol. Biol.*, 2009, **390**, 56–69.
52 J. Liang, V. Tsui, A. Van Abbema, L. Bao, K. Barrett, M. Beresini, L. Berezhkovskiy, W. S. Blair, C. Chang, J. Driscoll, C. Eigenbrot, N. Ghilardi, P. Gibbons, J. Halladay, A. Johnson, P. B. Kohli, Y. J. Lai, M. Liimatta,

P. Mantik, K. Menghrajani, J. Murray, A. Sambrone, Y. S. Xiao, S. Shia, Y. Shin, J. Smith, S. Sohn, M. Stanley, M. Ultsch, B. R. Zhang, L. C. Wu and S. Magnuson, *Eur. J. Med. Chem.*, 2013, **67**, 175–187.

53 J. Liang, A. van Abbema, M. Balazs, K. Barrett, L. Berezhkovsky, W. Blair, C. Chang, D. Delarosa, J. DeVoss, J. Driscoll, C. Eigenbrot, N. Ghilardi, P. Gibbons, J. Halladay, A. Johnson, P. B. Kohli, Y. J. Lai, Y. Z. Liu, J. Lyssikatos, P. Mantik, K. Menghrajani, J. Murray, I. Peng, A. Sambrone, S. Shia, Y. Shin, J. Smith, S. Sohn, V. Tsui, M. Ultsch, L. C. Wu, Y. S. Xiao, W. Q. Yang, J. Young, B. R. Zhang, B. Y. Zhu and S. Magnuson, *J. Med. Chem.*, 2013, **56**, 4521–4536.

54 A. Saha and K. Raghavachari, *J. Chem. Theory Comput.*, 2015, **11**, 2012–2023.

55 A. D. Becke, *J. Chem. Phys.*, 1997, **107**, 8554–8560.

56 H. L. Schmider and A. D. Becke, *J. Chem. Phys.*, 1998, **108**, 9624–9631.

57 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, 132.

58 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

59 T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. Schleyer, *J. Comput. Chem.*, 1983, **4**, 294–301.

60 R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728.

61 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.

62 W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. Defrees, J. A. Pople and J. S. Binkley, *J. Am. Chem. Soc.*, 1982, **104**, 5039–5048.

63 J. J.-P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173–1213.

64 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.

65 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, 2016.

66 B. H. Besler, K. M. Merz and P. A. Kollman, *J. Comput. Chem.*, 1990, **11**, 431–439.

67 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129–145.

68 A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623–1641.

69 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

70 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

71 M. Adler and P. Beroza, *J. Chem. Inf. Model.*, 2013, **53**, 2065–2072.

72 M. Aldeghi, M. J. Bodkin, S. Knapp and P. C. Biggin, *J. Chem. Inf. Model.*, 2017, **57**, 2203–2221.

73 S. R.-U. Genheden, *J. Comput. Chem.*, 2010, **31**, 837–846.

74 N. Cheron and E. I. Shakhnovich, *J. Comput. Chem.*, 2017, **38**, 1941–1951.

75 S. K. Sadiq, D. W. Wright, O. A. Kenway and P. V. Coveney, *J. Chem. Inf. Model.*, 2010, **50**, 890–905.

76 D. W. Wright, B. A. Hall, O. A. Kenway, S. Jha and P. V. Coveney, *J. Chem. Theory Comput.*, 2014, **10**, 1228–1241.

77 H. J.-C. Berendsen, J. P.-M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

78 A. Onufriev, D. Bashford and D. A. Case, *Proteins*, 2004, **55**, 383–394.

79 T. J. Hou, J. M. Wang, Y. Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.

80 F. Chen, H. Liu, H. Y. Sun, P. C. Pan, Y. Y. Li, D. Li and T. J. Hou, *Phys. Chem. Chem. Phys.*, 2016, **18**, 22129–22139.

81 H. Y. Sun, Y. Y. Li, M. Y. Shen, S. Tian, L. Xu, P. C. Pan, Y. Guan and T. J. Hou, *Phys. Chem. Chem. Phys.*, 2014, **16**, 22035–22045.

82 B. Wang, L. Li, T. D. Hurley and S. O. Meroueh, *J. Chem. Inf. Model.*, 2013, **53**, 2659–2670.