

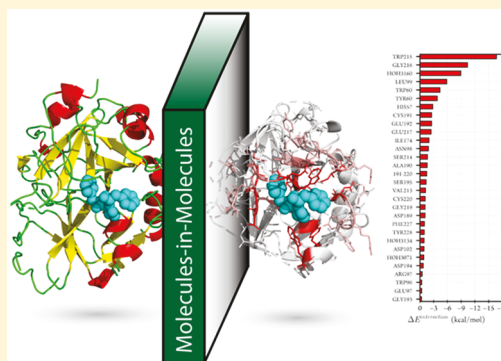
Energy Decomposition Analysis of Protein–Ligand Interactions Using Molecules-in-Molecules Fragmentation-Based Method

Bishnu Thapa[✉] and Krishnan Raghavachari^{*✉}

Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

S Supporting Information

ABSTRACT: Accurate prediction of protein–ligand binding affinities and their quantitative decomposition into residue-specific contributions represent challenging problems in drug discovery. While quantum mechanical (QM) methods can provide an accurate description of such interactions, the associated computational cost is normally prohibitive for broad-based applications. Recently, we have shown that QM-based protein–ligand interaction energies in the gas phase can be determined accurately using our multilayer molecules-in-molecules (MIM) fragmentation-based method at a significantly lower computational cost. In this paper, we present a new approach for calculating protein–ligand interactions using our three-layer model (MIM3) that allows us to decompose the total binding affinity into quantitative contributions from individual residues (or backbone and side chain), crystal water molecules, solvation energy, and entropy. In our approach, the desolvation energy and entropy changes during protein–ligand binding are modeled using simple and inexpensive empirical models while intermolecular interactions are computed using an accurate QM method. The performance of our approach has been assessed on a congeneric series of 22 thrombin inhibitors, all with experimentally known binding affinities, using a binding pocket cutout of 120 residues with more than 1550 atoms. Comparison of our MIM3-calculated binding affinities calculated at the B97-D3BJ/6-311++G(2d,2p) level with experiment shows a good correlation with an R^2 range of 0.81–0.88 and a Spearman rank correlation coefficient (ρ) range of 0.84–0.89 while providing a quantitative description of residue-specific interactions. We show that such residue-specific interaction energies can be employed to identify and rationalize both obvious (e.g., hydrogen bonds, $\pi\cdots\pi$) and nonobvious (e.g., $\text{CH}\cdots\pi$) interactions that play a critical role in protein–ligand binding. We suggest that such quantitative information can be used to identify the key residues that determine the comparative binding affinities of different ligands in order to improve and optimize the effectiveness of computational drug design.



1. INTRODUCTION

Molecular recognition is an important aspect of drug discovery that occurs through collective contributions from several key factors, viz., the binding interaction between a ligand and a receptor, solvation/desolvation processes, dynamic vibrational and conformational changes, etc.^{1,2} An in-depth analysis of these components leading to the formation of a stable protein–ligand complex is therefore central to understanding the drug binding process at the molecular level.³ Partitioning of the observed protein–ligand binding energy into the individual contributions from various parts of the receptor molecule, solely from experimental analysis, is essentially impossible. In this regard, computational methods can play an important role in guiding the drug discovery process robustly and cost effectively^{4,5} by using structural as well as energetic data on protein–ligand binding.⁶ Thus, a full description and quantification of the magnitude and significance of the energetic contributions of the different components is crucial for optimizing protein–ligand binding, and for effective drug design.

A range of computational tools with different levels of complexity and sophistication have been developed to facilitate the drug discovery and development process.^{5,7–9} Most computational methods estimate binding affinities based on some energy function derived from statistical analysis (knowledge-based scoring function or empirical scoring function) or physics-based methods (e.g., molecular mechanics (MM) force field).^{10,11} More rigorous approaches estimate the binding affinities as averages of protein–ligand interaction energies over a sample of configurations from molecular dynamics (MD) or Monte Carlo (MC) simulations.^{12,13} These methods are capable of providing intra- and intermolecular interaction energies along with conformational sampling and solvation. However, even with the high-resolution protein structures in hand, it may be difficult to achieve high accuracy using MM-based methods alone because the different electronic effects such as electrostatics, polarization, charge transfer, and many-body effects that play a pivotal role in protein–ligand binding

Received: May 23, 2019

Published: July 29, 2019

are often quite difficult to accurately model via the MM parameters.^{14,15}

Quantum mechanical (QM) methods, on the other hand, offer a realistic and reliable electronic level description of the noncovalent interactions compared to the classical methods. There has been a growing interest in using QM methods in protein–ligand binding studies.^{9,16,17} In principle, QM methods include all contributions to the protein–ligand interaction energies and therefore are ideal for such studies. However, the complete QM treatment of all components of the ligand binding process with explicit treatment of the protein and solvent conformations is still intractable. Therefore, in practice, only the ligand and a small portion of protein near ligand are usually treated with accurate QM methods (such as in QM/MM or ONIOM approaches).

The computational cost associated with conventional QM methods can be significantly reduced by using the linear scaling fragmentation-based methods that have evolved rapidly in the past few years.^{17–19} These methods have made QM calculations of large protein–ligand complexes containing more than 1000 atoms computationally tractable. In addition to the calculation of total interaction energies, some of the fragmentation-based methods also offer a further decomposition of interaction energy into residue-specific interactions to facilitate a more profound understanding of protein–ligand interactions.^{20–27} So far, the fragmentation-based QM calculations of protein–ligand binding energies are performed using a few representative structures obtained from the crystal structures, docking, MD simulations, or structure minimization.⁹ Though such approximations based on a single structure (or a few structures) may seem remarkably simple, several studies have shown that they provide a reasonably good quantitative description of the fundamental nonbonded interactions responsible for protein–ligand binding.⁹ However, it is important to note that there exist many systems where a complete conformational sampling may be necessary to obtain a good correlation with the experiment.

To help understand the protein–ligand interactions using QM methods at a lower computational cost, we have developed a multilayered, fragmentation-based method, molecules-in-molecules (MIM).^{28–30} In MIM, the full molecule is divided into smaller overlapping subsystems, and independent QM calculations are performed on them. The energy contributions from all subsystems are combined to obtain the total energy of the full system. In our recent study,³¹ we showed that our MIM method is capable of providing protein–ligand interaction energies at a substantially lower computational cost compared to the traditional quantum mechanical approach. Using a wide range of protein–ligand complexes, and the single-structure approximation, we calculated interaction energies for seven different data sets, each consisting of 7–18 structurally similar protein–ligand complexes. In particular, we derived a simple and broadly applicable computational protocol to obtain a good correlation (Spearman rank correlation (ρ) = 0.83–0.94; R^2 = 0.74–0.93) between the calculated interaction energies and their experimentally derived binding affinities. We showed that this protocol is appropriate to investigate a set of structurally similar ligands bound to the same receptor.

In this study, we present a new approach to studying the protein–ligand interactions using our multilayered MIM method. In this protocol, the total protein–ligand interaction energy calculated using MIM is partitioned into the

contributions from individual amino acid residues (or separately from backbones and side chains), crystal water molecules, ligand desolvation energy, and the entropic changes. The desolvation energy and entropy changes during the protein–ligand binding are modeled using simple and inexpensive empirical models, while intermolecular interactions are computed using an accurate QM method. These data are essential to identify the residues that play an essential role in the ligand binding process so that refinements could be made to improve the effectiveness of new drug design. To demonstrate the application of our protocol, we investigate a set of 22 structurally similar D-phenyl-proline-based (D-Phe-Pro) thrombin inhibitors.³² These inhibitors are selected mainly for two reasons: First, high-quality experimental binding affinities are available for these compounds so that the calculated trends can be directly compared. Second, the ligand modifications are relatively minor; thereby the modified ligands are structurally similar to the native, cocrystallized ligand. This helps to simplify the analysis and to validate the MIM-calculated binding affinities, and allows a fairly rigorous comparison.

In this work, the binding energies have been calculated using a single, energy-minimized structure of the protein–ligand complex obtained from the X-ray crystal structure or docking with a new protocol. As shown in our previous study, this is a reasonable approximation to estimate the relative binding affinity trends (or ranking) of structurally similar ligands, thereby avoiding the high computational cost associated with the protein conformational sampling.³¹ The correlation between the three-layered MIM method (MIM3) calculated interaction energies and the experimental binding affinities has been analyzed, and the role of different energy components in determining protein–ligand binding is discussed. The results show that our new protocol provides a comprehensive quantitative description of the interaction energy profiles of the residues present in the ligand binding pocket. Furthermore, a comparative difference energy analysis has been performed to show that the residues responsible for an experimentally observed binding affinity change between pairs of ligands can be identified, and their contributions quantified.

2. METHOD

2.1. Structure Preparation. In this study, we perform a comparative analysis of the protein–ligand binding energies for a set of 22 congeneric D-phenyl-proline-based (D-Ph-Pro) inhibitors bound to serine protease thrombin, a key enzyme involved in blood coagulation and platelet aggregation.³² The binding pocket with a cocrystallized ligand is shown in Figure 1 along with the complete set of ligands considered in this study. The inhibitors differ only in the substituents on the phenyl moiety portion of the ligand binding in the S1 pocket of the enzyme. Despite the high structural similarities among the ligands, the observed significant differences in the associated binding affinities (range -4.1 to -11.5 kcal/mol) make it an interesting system for a quantum mechanical investigation.

A well-resolved X-ray crystal structure with a cocrystallized ligand (PDB ID 2ZFF; resolution 1.47 Å) was used to prepare the set of protein–ligand complexes. The cocrystallized ligand (**1m**) was modified to obtain the rest of the ligands in the set using a systematic protocol. To ensure that all important interactions are preserved during the ligand modification, the modified ligands were aligned with the cocrystallized ligand using the flexible alignment module as implemented in the

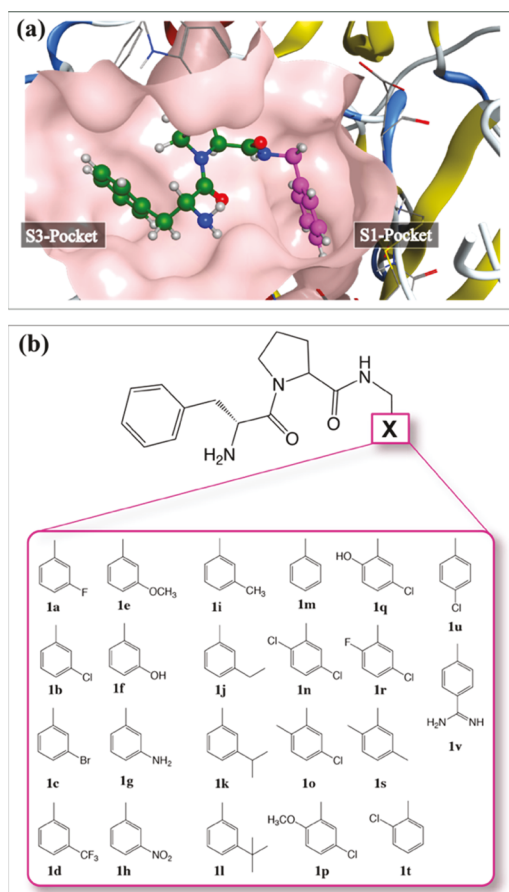


Figure 1. (a) Active site of the thrombin receptor with a cocrystallized ligand. The receptor cavity is shown as a light-pink surface, and the carbon atoms in the ligand are shown in green and pink colors. The phenyl moiety of the ligand present in S1 pocket is shown in pink color. (b) Structures of 22 inhibitors used in this study.

Molecular Operating Environment (MOE) program (version 2018.01).³³ Missing hydrogen atoms in the crystal structure were added with the Protonate3D³⁴ tool as implemented in MOE at a neutral pH 7. The incidental crystal water, HOH3098, present in the S1 pocket, near the meta-position of the phenyl moiety of the cocrystallized ligand, was removed before performing the ligand modifications. This was necessary because the modification of the ligand by any atom/group (other than the hydrogen atom of the cocrystallized ligand) results in an unfavorable overlap with HOH3098, indicating that the water molecule would get pushed out from the S1 pocket upon substitution at the meta-position of the phenyl ring. Each protein–ligand complex was minimized in MOE with the AMBER10:EHT force field,³⁵ 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 under a 0.5 Å restraint for every atom with respect to the starting structure. All the charged residues such as lysine, asparagine, glutamate, and aspartate were neutralized by adding/removing protons to better match the stabilization trends observed in the solution. Since the interaction energy calculations are performed in the gas phase, neutralizing the charged residues is necessary due to the substantial overestimation of the electrostatic interactions in the gas phase. We have shown in our previous study (and also here) that this is

an appropriate approximation for calculating protein–ligand interaction energy calculation in the gas phase correlating well with the corresponding experimental binding affinities. The information about the neutralized residues is listed in the [Supporting Information](#). For the QM calculations, all residues along with the crystal water molecules within 11.0 Å of the ligand were selected. The selected regions were then extended to include the full residues at the boundary, which makes the effective radius to be more than 13.0 Å from the ligand. The final QM region includes a total 1560–1575 atoms with 92 amino acid residues, 28 crystal water molecules, and the ligand being considered.

2.2. MIM Method. In this work, we present a new approach to study the protein–ligand interactions using our recently developed multilayered, molecules-in-molecules (MIM) fragmentation-based method. Multilayered MIM is capable of providing protein–ligand interaction energies at a substantially lower computational cost compared to traditional quantum mechanical approaches. We have used the three-layered MIM method (MIM3) with high, middle, and low layers to estimate the total interaction energy. To obtain the initial nonoverlapping “monomer” fragments, the receptor molecule is divided by cutting the C–C α and C α –C β bonds, as shown in [Figure 2](#). In this scheme, each amino acid residue is divided into two monomers, containing the backbone atoms (HC(=O)–NH–CH $_3$) and the side chain, respectively. Depending upon the size, the ligand can also be divided into

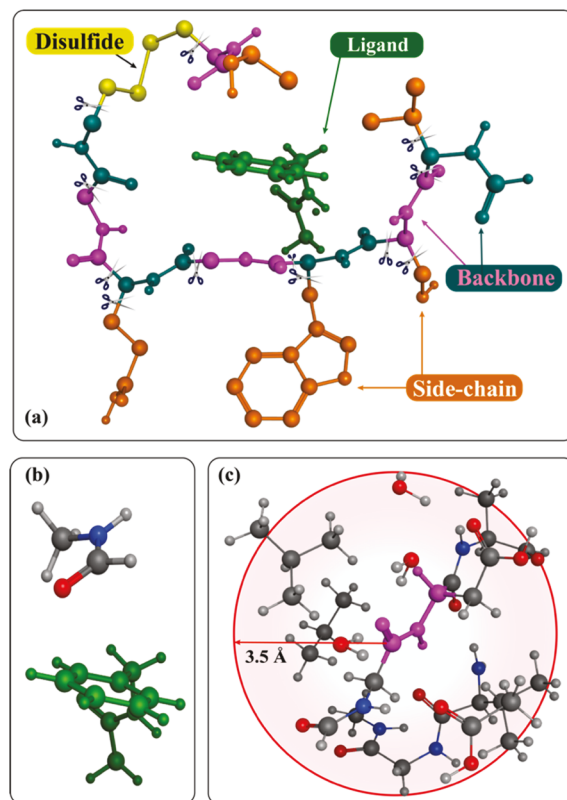
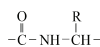


Figure 2. (a) Illustration of the fragmentation scheme used in the MIM3 calculation. Scissors across the bonds denote the bonds being broken during MIM fragmentation. Side-chain atoms of the amino acid residues are shown in orange, backbone atoms are shown in alternating pink and green colors, and the fragment with a disulfide bond is shown in yellow. (b) and (c) are examples of high and middle layer fragments, respectively.

multiple nonoverlapping monomers. This fragmentation scheme is slightly different from our previous approaches where every single, non-hydrogen, bond was cut. However, the new scheme is convenient to obtain residue-specific interaction energies (*vide infra*), one of the main objectives of this study. Since the peptide bonds between the amino acid residues have partial double bond character and therefore are not broken during the MIM fragmentation, the definition of amino acid residue is slightly different, i.e.



compared to the conventional structure, i.e.



In the high layer of MIM3, the primary subsystems are formed by pairing the ligand with each of the nonoverlapping residue monomers as well as the crystal water molecules. In the middle layer, subsystems are obtained using a distance-based fragmentation with a cutoff radius of 3.5 Å. All the monomers within the specified cutoff distance are combined to form a subsystem. The full molecule is used in the low layer. From the overlaps between the primary subsystems, derivative subsystems are obtained, and the energy is summed according to the *inclusion–exclusion principle*. Each dangling bond in the subsystems is capped with a hydrogen atom. The three levels of theory used in MIM3 are B97-D3BJ/6-311++G(2d,2p) as the “high” level, B97-D3BJ/6-31+G(d) as the “medium” level, and PM6-D3³⁶ as the “low” level. The interaction energies calculated in each layer are summed up using the following equation to obtain the total interaction energy at the target B97-D3BJ/6-311++G(2d,2p) level of theory [i.e., the B97^{37,38} functional with Grimme’s D3 dispersion correction³⁹ and Becke–Johnson damping⁴⁰ and 6-311++G(2d,2p)^{41–44} basis set]:

$$\Delta E_{\text{MIM3}}^{\text{interaction}} = \Delta E_{\text{high}}^{\text{interaction},r} - (\Delta E_{\text{med}}^{\text{interaction},r} - \Delta E_{\text{med}}^{\text{interaction},r'}) - (\Delta E_{\text{low}}^{\text{interaction},r'} - \Delta E_{\text{low}}^{\text{interaction},R}) \quad (1)$$

where r , r' , and R with ($r < r' \ll R$) are symbolic representations of the relative subsystem sizes; $\Delta E_{\text{high}}^{\text{interaction},r}$ and $\Delta E_{\text{med}}^{\text{interaction},r}$ are the sum of the ligand–residue pairwise interaction energies calculated at the high and medium levels of theory, respectively, using the high layer fragmentation scheme; $\Delta E_{\text{med}}^{\text{interaction},r'}$ and $\Delta E_{\text{low}}^{\text{interaction},r'}$ are the total interaction energies calculated at the medium and low levels of theory, respectively, using the distance-based middle layer fragmentation scheme; and $\Delta E_{\text{low}}^{\text{interaction},R}$ is the interaction energy of the full molecule calculated at the low level of theory.

The total protein–ligand interaction energy as computed above can now be broken down into residue-specific interactions at the high level of theory. Since high layer calculations are already available for ligand–residue (monomer) pairwise interactions, their sum yields the total residue-specific interactions at the high level of theory.

$$\Delta E_{\text{high}}^{\text{interaction},r} = \sum_k^N \Delta E_{\text{high}}^k \quad (2)$$

Here N is the number of primary subsystems (i.e., ligand–residue pairs) formed in the high layer, which increases linearly with the system size, and ΔE_{high}^k is the interaction energy of the k th primary subsystem. However, since the pairwise interaction energies calculated in the high layer will be nonadditive (i.e.,

will not add up to the total MIM3 interaction energy) due to the missing many-body effects, we use a single uniform scaling factor to correct the per-residue interactions. This factor is obtained from the ratio of total interaction energy calculated using MIM3 and the sum of the per-residue interaction energies calculated at the high layer.

$$\text{scaling factor } (\alpha) = \frac{\Delta E_{\text{MIM3}}^{\text{interaction}}}{\Delta E_{\text{high}}^{\text{interaction},r}} \quad (3)$$

The corrected high-layer interaction energy and the scaled per-residue interaction energies are

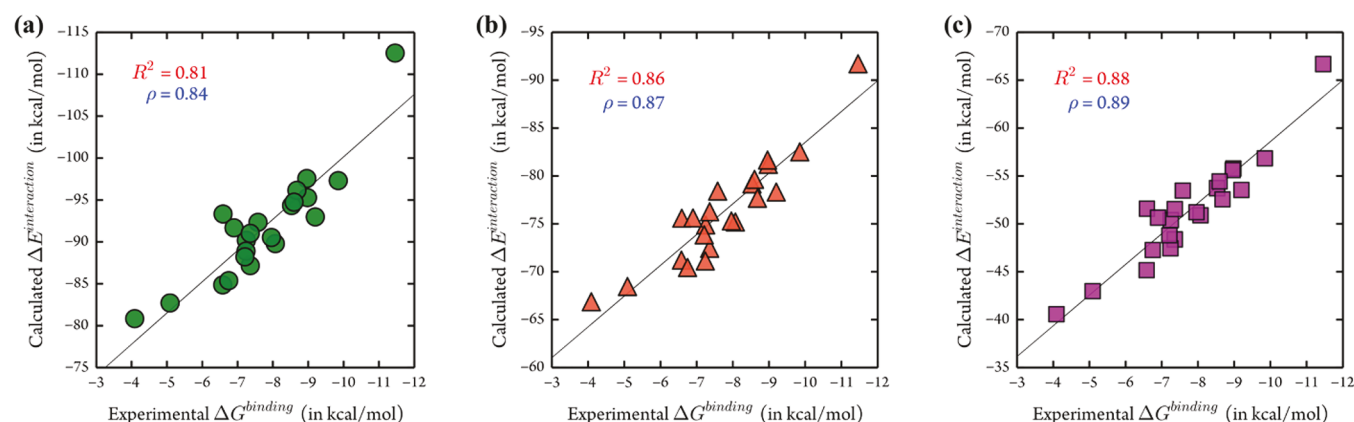
$$(\Delta E_{\text{high}}^{\text{interaction},r})_{\text{corrected}} = \alpha \Delta E_{\text{high}}^{\text{interaction},r} = \sum_k^N \alpha \Delta E_{\text{high}}^k \quad (4)$$

Since there is a possibility of having spatial variability in the polarization due to the nearby groups, it is possible that the use of a single uniform scaling factor per individual fragment derived from eq 3 may not appear to be the ideal solution. More importantly, this could be a potential problem in the case of charged residues where the polarization effects are significantly larger. However, since our study only deals with the neutral residues, such an issue is expected to be minimal and therefore should not affect the interpretation comparative study of residue-specific interactions. To assess the validity of our choice of using the uniform scaling factor per ligand, we performed test calculations by systematically increasing the fragment size from one residue (backbone + side chain) to two, four, and five residues per subsystem and computed the difference in interaction energy as a function of the size of the subsystems (see the [Supporting Information](#) for details). As expected, the difference is calculated to be relatively small ($< |\pm 1.3|$ kcal/mol). To make sure that this does not result any bias in the residue-specific interaction energy analysis, we also performed the subsystem-specific difference energy analysis on selected ligand pairs using the interaction energy calculated for larger vs smaller subsystems (see [Results and Discussion](#) for the details about the MIM difference energy analysis). Analysis showed that the fragment-specific difference in interaction energy remains reasonably small (largest error $< |\pm 0.30|$ kcal/mol). More importantly, the subsystems (or the residues) that contribute significantly to the calculated binding energy change for a pair of ligands are correctly identified even when smaller subsystems were used to calculate residue-specific interactions. This demonstrates that the use of a single scaling factor is correctly justified within our current protocol to study the relative strength of residue-specific interactions of structurally similar ligands. Furthermore, since the residue-specific pairwise interaction energies are already available from the MIM3 calculation, no additional calculations are required to perform.

In addition to the gas-phase interaction energies, we have also included the contributions from desolvation energy and entropy ($-T\Delta S$) in the interaction energy calculations. The desolvation energy for protein–ligand binding is approximated using the solvation energy of the ligand and a scaling factor obtained from the solvent accessible surface area (SASA) calculations, as described in our previous work.³¹ The solvation energies of ligands in aqueous solution are calculated using the B97-D3/6-311++G(2d,2p) method and the SMD⁴⁵ implicit solvation model. The entropic contribution ($-T\Delta S$) to the total interaction energy is calculated from an empirical model based on the atomic solvent accessible surface (SAS) area and

Table 1. Experimental Binding Energy ($\Delta G^{\text{binding}}$) and Calculated Interaction Energies of 22 Thrombin Inhibitors at B97-D3BJ/6-311++G(2d,2p) Level of Theory^a

ligand	$\Delta G^{\text{binding}}(\text{expt})$	$\Delta E_{\text{gas}}^{\text{interaction}}$	$\Delta E_{\text{ligand}}^{\text{desolv}}$	$-T\Delta S$	$\Delta E_{\text{gas}}^{\text{interaction}} + \Delta E_{\text{ligand}}^{\text{desolv}}$	$\Delta E_{\text{gas}}^{\text{interaction}} + \Delta E_{\text{ligand}}^{\text{desolv}} - T\Delta S$
1a	−7.36	−87.16	14.72	24.07	−72.44	−48.37
1b	−9.20	−92.97	14.66	24.79	−78.31	−53.52
1c	−8.53	−94.33	15.17	25.44	−79.16	−53.72
1d	−6.58	−84.86	13.66	26.04	−71.21	−45.17
1e	−7.25	−90.18	15.30	24.50	−74.88	−50.38
1f	−7.24	−88.89	17.76	23.69	−71.12	−47.43
1g	−6.59	−93.32	17.74	24.00	−75.58	−51.58
1h	−6.90	−91.67	16.06	24.97	−75.61	−50.64
1i	−8.07	−89.78	14.59	24.30	−75.19	−50.89
1j	−7.21	−88.21	14.36	25.05	−73.85	−48.80
1k	−5.09	−82.71	14.25	25.48	−68.46	−42.98
1l	−4.09	−80.84	13.97	26.30	−66.87	−40.57
1m	−6.75	−85.37	14.93	23.17	−70.44	−47.27
1n	−9.85	−97.30	14.81	25.68	−82.50	−56.82
1o	−8.98	−95.27	14.08	25.43	−81.20	−55.77
1p	−8.96	−97.55	15.90	26.06	−81.65	−55.59
1q	−8.68	−96.14	18.48	25.11	−77.66	−52.55
1r	−8.60	−94.73	15.09	25.21	−79.63	−54.42
1s	−7.58	−92.33	13.92	24.95	−78.41	−53.46
1t	−7.96	−90.52	15.23	24.09	−75.29	−51.20
1u	−7.36	−91.01	14.75	24.72	−76.26	−51.54
1v	−11.46	−112.51	20.81	25.03	−91.69	−66.66

^aAll energy values are given in kcal/mol.**Figure 3.** Correlation between the experimental binding energy ($\Delta G^{\text{binding}}$) and (a) gas-phase interaction energy ($\Delta E_{\text{gas}}^{\text{interaction}}$), (b) sum of the gas phase interaction energy and the ligand desolvation energy ($\Delta E_{\text{gas}}^{\text{interaction}} + \Delta E_{\text{ligand}}^{\text{desolv}}$), and (c) sum of the gas phase interaction energy, the ligand desolvation energy, and the entropy of binding ($\Delta E_{\text{gas}}^{\text{interaction}} + \Delta E_{\text{ligand}}^{\text{desolv}} - T\Delta S$). Black diagonal lines are the lines of best fit. ρ is the Spearman rank correlation.

buried solvent accessible surface (BSAS) area, developed by Wang and co-workers (see the [Supporting Information](#) for more details).⁴⁶ Their method is shown to accurately reproduce the quantum mechanical TS for small molecules as well as the $-T\Delta S$ of protein–ligand binding compared to the MM normal-mode analysis (NMA). All the electronic structure calculations were performed using the Gaussian 16 program suite⁴⁷ and our in-house MIM external package.

3. RESULTS AND DISCUSSION

3.1. Correlation with Experiments. We have calculated the interaction energies of 22 thrombin inhibitors using the MIM3 method at the B97-D3BJ/6-311++G(2d,2p) level of theory and compared the results with the corresponding experimental binding affinities. [Table 1](#) shows the various

components of protein–ligand binding energies calculated in this study.

[Figure 3](#) visualizes the correlation between the calculated interaction energies and the experimental binding affinities. As can be seen in [Table 1](#), the total binding free energies are the sum of large individual contributions, some with opposite signs, which can be analyzed separately. The calculated gas phase interaction energies range from -80.8 to -112.5 kcal/mol, which are about an order of magnitude larger than the experimental binding affinities. Despite this overestimation, a very good correlation of $R^2 = 0.81$ is obtained between the experimental binding affinities and the gas phase interaction energies ([Figure 3](#)), indicating that the gas phase interaction energies reflect the *trends* in the observed binding affinities reasonably well. This observation is somewhat expected since

the modifications made in the ligand are relatively small and possess similar rigidity compared to the cocrystallized ligand.

The addition of the ligand desolvation energy lowers the total interaction energy by 13–21 kcal/mol. The variation in the desolvation energies among the alkyl- and halogen-substituted ligands is relatively small (a spread of 1.6 kcal/mol) compared to their experimental binding free energies. On the other hand, the ligands substituted with $-\text{OH}$, $-\text{OMe}$, $-\text{NO}_2$, or $-\text{NH}_2$ group show a more substantial desolvation energy penalty upon binding (larger by 5–8 kcal/mol) compared to the unsubstituted ligand. The inclusion of the ligand desolvation energy ($\Delta E_{\text{ligand}}^{\text{desolv}}$) in the interaction energy calculation improves the correlation coefficient (R^2) from 0.81 to 0.86. This indicates that the protein–ligand (P–L) binding, to some extent, correlates with the differential ligand desolvation penalty of binding; however, it is not the only factor to dictate the overall protein–ligand binding. The entropy contributions lie between +23.2 and +26.3 kcal/mol and do not fluctuate as strongly. Not surprisingly, a good correlation ($R^2 = 0.88$) is maintained when the $-T\Delta S$ of binding is added to the total interaction energy. However, those contributions are still important for lowering the absolute binding energy toward the experiment, though there is still a significant gap between the two. It is important to note that, even with the ligand desolvation energy and entropy of binding included, the calculated values are still significantly larger compared to the experimental binding energies possibly due to the missing large contribution from the protein desolvation energy, and other effects such as ligand and protein reorganization upon binding. A good correlation obtained between the calculated values and the experiment suggests that those effects are somewhat systematic for the structurally similar ligands used in this study. However, a proper treatment of those contributions is required to obtain the accurate absolute energy of protein–ligand binding.

The correlation of the calculated total interaction energies with the experiment can also be analyzed as a function of the cutoff radius. Figure 4 shows the average cumulative

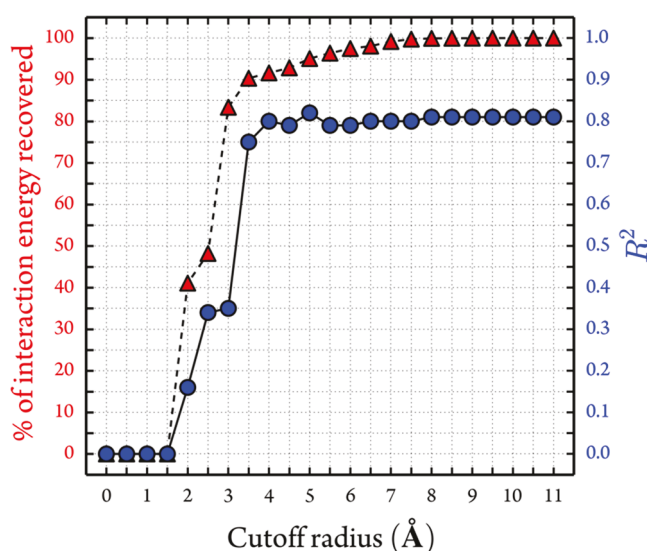


Figure 4. Percentage of average interaction energy recovered compared to the full molecule (red, triangles) and its correlation with the experimental binding affinity (blue, circles) as a function of the protein cutoff radius.

interaction energy versus increasing shortest cutoff radius (or contact distance) between the residues and the ligand. The shortest contact radius indicates that only the residues within that distance contribute to the calculated total interaction energy. The initial convergence of the interaction energy is relatively rapid; about 95% of the total interaction energy is already captured from the contributions from residues within 5.0 Å of the ligand. This observation is similar to a previous study of HIV-II protease complex, investigated with the functional group symmetry-adapted perturbation theory (F-SAPT).²⁷ Interestingly, the correlation coefficient (R^2) also converges significantly fast. At the closest contact radius of 4.0 Å, R^2 reaches close to the optimum value of 0.81 corresponding to the full molecule and remains similar afterward. This observation indicates that including residues within 5–6 Å of the ligand is sufficient to capture the most significant portion of protein–ligand interactions to estimate the relative binding trends.

3.2. Residue-Specific Energy Decomposition Analysis. In addition to the calculation of the total interaction energy of binding, the MIM method can provide a better description of the noncovalent interactions in terms of per-residue interaction energies. In this approach, the MIM3 total interaction energy is partitioned into the contributions from individual amino acid residues allowing the quantification of their interaction energy contributions. This information is essential while designing a new drug since it helps to identify key residues essential for ligand binding and also allows for each interaction to be tuned and analyzed.

Figure 5 illustrates the residue-specific interaction energy partitioning for the ligand **1n**. All atoms of each amino acid residue are colored according to their contribution to the total interaction. The quantitative breakdown of the total interaction energy into per-residue contributions is also given. A similar partitioning of the interaction energies for other thrombin inhibitors is given in Figures S3–S24. Such residue-specific information makes it easy to identify the key residues that contribute significantly to the total interaction. The decomposition analysis shows that there are at least 15 significantly favorable (the magnitude of $\Delta E_{\text{per-residue}}$ greater than 2.0 kcal/mol) interactions. Some of the key residues that contribute significantly across all the considered ligands include Trp215 (−15.9 to −18.8 kcal/mol), Gly216 (−7.2 to −12.6 kcal/mol), HOH3160 (−8.3 to −9.4 kcal/mol), Leu99 (−5.6 to −6.0 kcal/mol), Trp60 (−4.0 to −4.6 kcal/mol), and Glu192 (−2.5 to −4.3 kcal/mol).

The residue-specific interaction energy contribution can be further decomposed into the contribution from the backbone ($\text{HC(=O)}-\text{NH}-\text{CH}_3$), side chain, disulfide ($\text{CH}_3-\text{S}-\text{S}-\text{CH}_3$), and crystal water molecules (Tables S3–S24). For example, the −15.9 to −18.8 kcal/mol interaction energy for residue Trp215 comes from two major interactions with the ligand: $\pi-\pi$ stacking and a hydrogen bond. The $\pi-\pi$ stacking interaction between the indole side chain and the phenyl moiety of ligand contributes in the range of −5.2 to −5.7 kcal/mol depending upon substituent on the ligand. A strong $\text{C}=\text{O}\cdots\text{H}-\text{N}$ hydrogen bonding between the peptide oxygen of the tryptophan backbone and the H–N group of the ligand contributes the remaining (−10.2 to −13.6 kcal/mol) of the ligand–Trp215 interaction energy. Similarly, a crystal water molecule (HOH3160) bridging between $\text{C}=\text{O}$ and NH_2 groups of the ligand forms two hydrogen bonds and contributes −8.3 to −9.4 kcal/mol to the total interaction

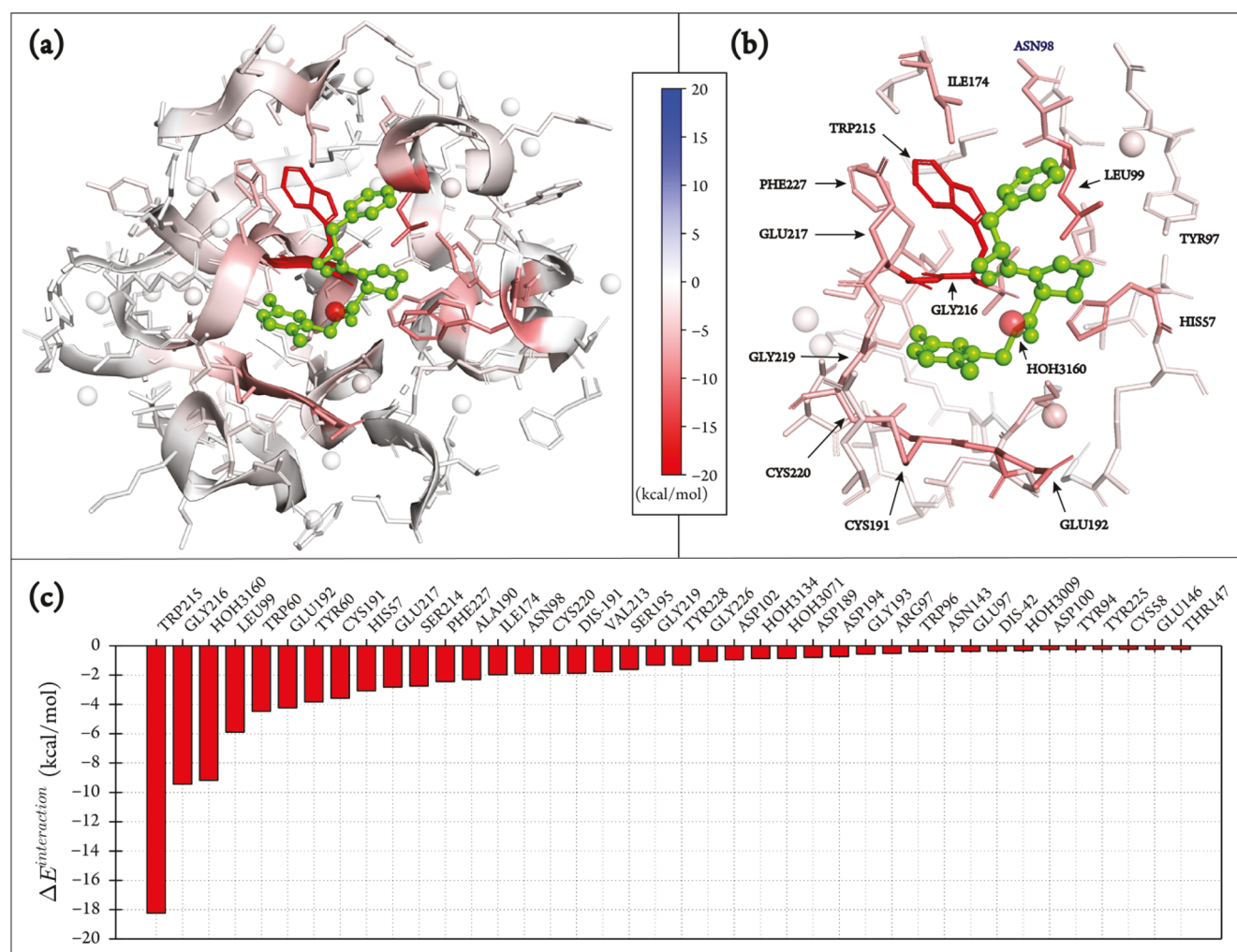


Figure 5. Illustration of residue-specific energy decomposition analysis. Ligand atoms are shown in green color. All atoms in each amino acid residue are colored based on the interaction energy contribution. (a) Ligand **1n** complexed with thrombin receptor, (b) ligand and selected amino acid residues, and (c) interaction energy contribution of some selected amino acid residues, more than ± 0.21 kcal/mol.

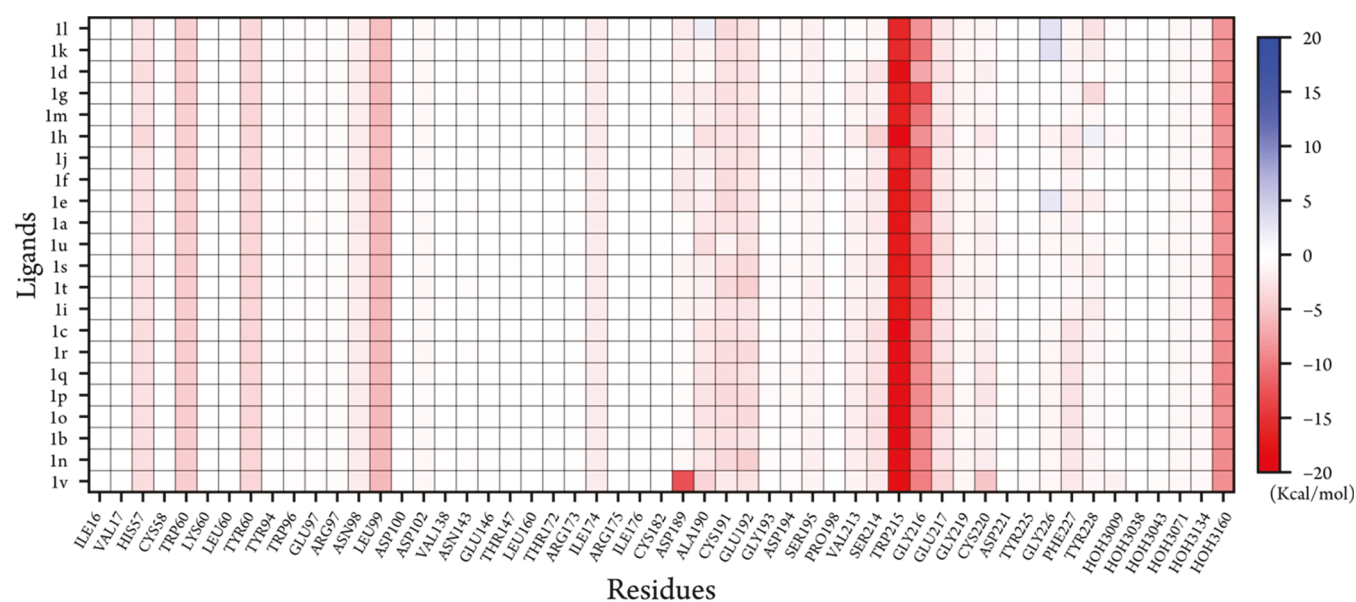


Figure 6. Interaction map for some of the key residues (x-axis). Ligands on the y-axis are ordered based on their experimental binding affinities.

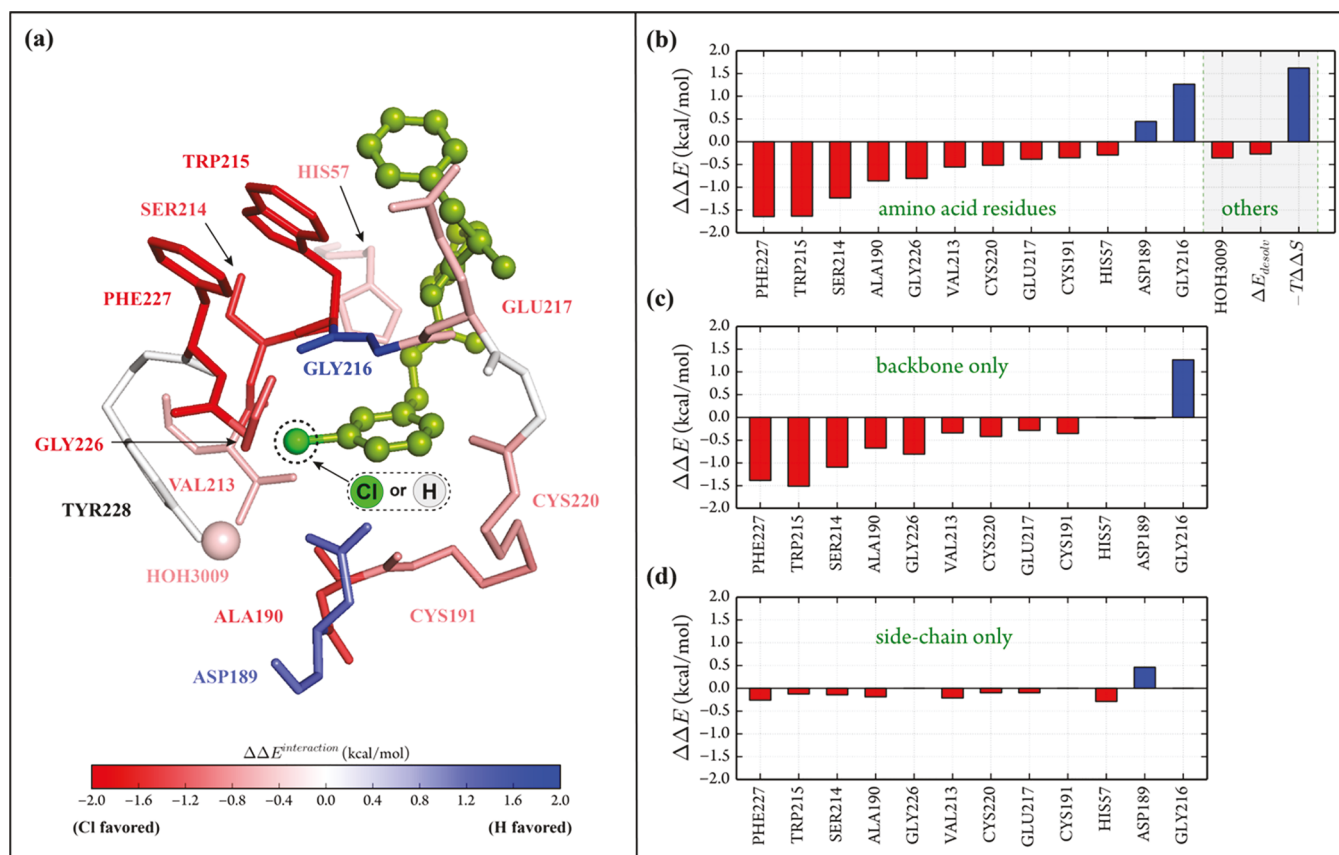


Figure 7. MIM difference energy analysis (MIM-DEA) results for ligands **1b** and **1m** at B97-D3/6-311++G(2d,2p) level of theory. (a) $\Delta\Delta E^{\text{interaction}}$ mapped to the thrombin S1 pocket complex geometry. All atoms in each amino acid residue are colored based on the interaction energy contribution. (b) Quantitative $\Delta\Delta E^{\text{interaction}}$ data for residues contributing more than $|\pm 0.2|$ kcal/mol. (c) $\Delta\Delta E^{\text{interaction}}$ contribution from residue backbone only. (d) $\Delta\Delta E^{\text{interaction}}$ contribution from residue side chain only.

energy. The contribution from other amino acid residues can also be analyzed in a similar manner.

In many structure–activity relationship (SAR) studies, the experimental binding affinities are analyzed based on the geometrical features that appear in the crystallographic structures. In favorable cases, the obvious changes in protein–ligand interactions such as the presence or absence of hydrogen bonding or π – π stacking interactions can be qualitatively identified from a visual inspection. However, a formal quantification of such changes and their contribution to the total interaction for different ligands are, in general, difficult to assess. The quantitative information provided by MIM can be useful to rationalize both the obvious and nonobvious changes in protein–ligand interactions. For example, the nonclassical interactions such as the CH– π interaction between the C–H of the proline moiety of the ligand and the imidazole side chain of His57 (~ -3 kcal/mol), the hydrophobic interaction between the side chain of Val213 and the phenyl moiety of the ligand (~ -1 kcal/mol), and the π – π interaction between the backbone of Glu192 and the phenyl group of the ligand (-2 to -4 kcal/mol), found in many of the investigated thrombin inhibitors, contribute significantly although these interactions are not easily apparent by visual inspection (see Tables S3–S24 for the quantitative details).

The residue-specific interactions can also be analyzed using an interaction map and can be correlated with the structural features. Figure 6 shows the heat map of the interaction energy

contributions from some of the important residues for all 22 ligands. Figure 6 highlights the residues with the significant as well as smaller contributions to the total interaction energy. Such an analysis becomes useful while identifying common critical residues that appear across a series of ligands. For example, columns that correspond to the residues Trp215, Gly216, His57, Trp60, Tyr60, and HOH3160 appear consistently in dark red, irrespective of the ligand, indicating that these residues are critical for the binding of these D-Phe-Pro-based thrombin inhibitors. On the other hand, the residue Asp189 which has a modest contribution for all but one ligand contributes significantly to the ligand **1v** binding due to the strong hydrogen bonding interaction that is unique to this ligand. Similarly, repulsive interactions resulting from the substitution of a hydrogen atom in the native ligand (ligand **1m**) by bulkier groups such as $-\text{OMe}$, $-\text{isopropyl}$, or $-\text{tert-butyl}$ can also be seen in the heat map (shown in blue squares).

3.3. MIM Difference Energy Analysis (MIM-DEA). To gather a more precise quantitative understanding of the observed change in the experimentally measured ligand dissociation constants across the various ligands, we performed the MIM difference energy analysis (MIM-DEA). In the MIM-DEA, the change in the total interaction energy for a pair of ligands ($\Delta\Delta E^{\text{interaction}}$) due to a structural modification is obtained by taking the difference in the energy contributions from individual residues ($\Delta\Delta E_{\text{gas}}^{\text{interaction}}$), from solvation ($\Delta\Delta E^{\text{desolv}}$), and from the entropy ($-T\Delta\Delta S$). The interaction energy difference between two ligands A and B is calculated as

$$\Delta\Delta E_{B \rightarrow A}^{\text{interaction}} = \Delta\Delta E_{B \rightarrow A, \text{gas}}^{\text{interaction}} + \Delta\Delta E_{B \rightarrow A}^{\text{desolv}} - T\Delta\Delta S_{B \rightarrow A} \quad (5)$$

where

$$\Delta\Delta E_{B \rightarrow A, \text{gas}}^{\text{interaction}} = \sum_k^N (\Delta E_B^k - \Delta E_A^k)_{\text{gas}} \quad (6)$$

$$\Delta\Delta E_{B \rightarrow A}^{\text{desolv}} = \Delta E_B^{\text{desolv}} - \Delta E_A^{\text{desolv}} \quad (7)$$

$$-T\Delta\Delta S_{B \rightarrow A} = -T(\Delta S_B - \Delta S_A) \quad (8)$$

In the equations above, N is the total number of ligand–residue pairs, ΔE^k is the interaction energy of the k th ligand–residue pair, ΔE^{desolv} is the desolvation energy, and ΔS is the entropy of binding for each of the ligands.

The difference interaction energy calculation allows identifying the residues that contribute significantly to the observed differences in the experimental binding affinities of the two ligands. To demonstrate this premise, we compared the interaction of a pair of thrombin inhibitors with a small difference in structure but a significant change in binding affinity: H-substituted versus Cl-substituted ligands (**1m** vs **1b**, respectively) with a 62-fold change in the experimental binding affinity.

The difference MIM interaction energy between the ligands **1b** (with –Cl) and **1m** (with –H) is presented in Figure 7. Figure 7 demonstrates the change in interaction energies ($\Delta\Delta E^{\text{interaction}}$) when the –H group is replaced with a strong electron withdrawing group, –Cl. MIM calculations show that this substitution is energetically favorable ($\Delta\Delta E^{\text{interaction}} = -7.60$ kcal/mol). The enhancement in the binding energy of the chloride-substituted ligand (**1b**) is primarily dominated by a strongly favorable contribution from electronic energy ($\Delta\Delta E^{\text{interaction}} = -7.60$ kcal/mol), slightly favored in terms of ligand desolvation energy ($\Delta\Delta E_{\text{ligand}}^{\text{desolv}} = -0.27$ kcal/mol), but disfavored in terms of the entropy ($-T\Delta\Delta S = +1.62$ kcal/mol), resulting in a net favorable substitution (calculated -6.25 kcal/mol; experimental -2.45 kcal/mol). Thirteen residues surrounding the substituted phenyl moiety in the S1-binding pocket are found to experience an appreciable change in the pairwise interaction energy (with $\Delta\Delta E^{\text{pairwise}} > |\pm 0.20|$ kcal/mol). The minor contributions (ranging from ± 0.02 to ± 0.20 kcal/mol) are found to extend up to 30 nearby amino acid residues; however, their collective contribution is relatively small ($< |\pm 0.20|$ kcal/mol). Among the significantly contributing 13 residues, chloride substitution results in an increase in favorable interaction for 11 residues, namely Phe227, Trp215, Ser214, Ala190, Gly226, Val213, Cys220, Glu217, HOH3009, Cys191, and His57. For two residues, Asp189 and Gly216, the interaction energy is found to be lowered by 0.44 and 1.26 kcal/mol upon the chloride substitution. Interestingly, most of the interaction energy change for the H- versus Cl-substitution is dominated by backbone–ligand interactions (Figure 7b) relative to the ligand–side-chain interaction (Figure 7c).

Overall, the comparative quantitative description of the protein–ligand interaction discussed above demonstrates how MIM can be useful to understand the change in the strength of protein–ligand interactions due to chemical substitution. Additionally, it shows that understanding noncovalent interactions based on visual inspection may not always be sufficient enough to identify the critical changes in the underlying interactions for structurally similar ligands.

4. CONCLUSIONS

We have presented a large-scale application of our multilayer MIM method to calculate protein–ligand interaction energies of significantly large systems with over 1550 atoms and more than 22 000 basis functions. A key goal of this study is to demonstrate that our fragmentation-based MIM method can be useful to understand the protein–ligand interactions which will decrease the need for a trial-and-error approach while adding to the toolbox of intelligent and informed drug design. Using a test set of 22 structurally similar but chemically diverse thrombin inhibitors, we demonstrated the capability of our approach to provide total interaction energies along with the partitioning of energy contribution from solvation, entropy, and individual residue fragments. As in our previous work, a good correlation has been obtained between the experimental binding affinities and calculated interaction energies (R^2 0.81–0.88; ρ 0.84–0.89), though the calculated interaction energies are higher by as much as 1 order of magnitude than the experiment. The solvation energy contribution is found to improve the correlation slightly (gas phase $R^2 = 0.81$ vs gas phase + solvation $R^2 = 0.86$). Our calculations also show that the entropic contribution does not vary by much for a set of structurally similar ligands (usually the case in a lead optimization process); however, its contribution is necessary to lower the gap between theoretical and experimental binding energies. We also show that, in order to obtain the accurate absolute binding energies that are comparable to the experiments, it is necessary to include the contributions from other components of protein–ligand binding including the proper treatment of protein desolvation. Analysis of the distance-dependent contribution to the MIM3 total interaction energy revealed that the initial convergence of the interaction energy is quite rapid. Residues within ~ 5.0 Å of closest contact contribute about 95% of the total interaction energy. Interestingly, the change in the correlation coefficient remains minimal after 4.5 Å, suggesting that the closest contact radius of 5.0–6.0 Å between ligand and residues is sufficient to obtain a good correlation. However, the contributions from residues farther away are necessary to calculate the absolute binding energy reliably.

The residue-specific energy decomposition analysis greatly enabled us to identify critical residues that contribute significantly to the total interaction energy. The quantitative information provided by MIM can be used to identify and rationalize both the obvious (e.g., hydrogen bonds, π – π interactions) and nonobvious (e.g., $\text{CH}\cdots\pi$ interactions) in the protein–ligand binding. Analysis of the difference interaction energy of a pair of ligands showed that residues responsible for the observed enhancement in experimental binding affinity can be quantitatively identified. Such information will be useful in designing new ligands since the chemical-intuition-based identification of noncovalent interactions may not always be sufficient to estimate the extent of binding affinity change.

Overall, the quantitative description of the protein–ligand interaction presented in this study demonstrates how MIM can be useful to understand the change in the strength of protein–ligand interactions upon chemical substitution. Although the interaction energies calculated in the gas phase are highly overestimated compared to the experimental binding affinities, these energies nevertheless provide valuable information about the relative binding strengths among different ligands. While the method presented in this study can be further improved by

introducing the conformational sampling and a better description of solvation energy contributions, we believe that our protocol is a promising tool to assist in structure-based drug design.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.9b00432.

Solvation energy and entropy approximation, entropy, assessment of scaling factor (α) for residue-specific interactions, figures and tables with residue-specific interaction energies (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kraghava@indiana.edu.

ORCID

Bishnu Thapa: 0000-0003-3521-1062

Krishnan Raghavachari: 0000-0003-3275-1426

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Eli Lilly and Company through the Lilly Research Award Program at Indiana University. The methods used in this project were developed based on the support from the National Science Foundation Grant CHE-1665427 at Indiana University. We thank Dr. Jon Erickson for stimulating discussions. The Big Red II supercomputing facility at Indiana University was used for most of the calculations in this study.

■ REFERENCES

- (1) Leach, A. R. *Molecular Modeling: Principles and Applications*; Addison Wesley Longman: 1996.
- (2) Williams, D. H.; Westwell, M. S. Aspects of Weak Interactions. *Chem. Soc. Rev.* **1998**, 27, 57–64.
- (3) Perozzo, R.; Folkers, G.; Scapozza, L. Thermodynamics of Protein–Ligand Interactions: History, Presence, and Future Aspects. *J. Recept. Signal Transduction Res.* **2004**, 24, 1–52.
- (4) Jorgensen, W. L. Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009**, 42, 724–733.
- (5) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, 66, 334–395.
- (6) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, 303, 1813–1818.
- (7) Gohlke, H.; Klebe, G. Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem., Int. Ed.* **2002**, 41, 2644–2676.
- (8) Ou-Yang, S.-s.; Lu, J.-y.; Kong, X.-q.; Liang, Z.-j.; Luo, C.; Jiang, H. Computational Drug Discovery. *Acta Pharmacol. Sin.* **2012**, 33, 1131.
- (9) Ryde, U.; Soderhjelm, P. Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chem. Rev.* **2016**, 116, 5520–5566.
- (10) Coupez, B.; Lewis, R. Docking and Scoring-Theoretically Easy, Practically Impossible? *Curr. Med. Chem.* **2006**, 13, 2995–3003.
- (11) Kontoyianni, M.; Madhav, P.; Suchanek, E.; Seibel, W. Theoretical and Practical Considerations in Virtual Screening: A Beaten Field? *Curr. Med. Chem.* **2008**, 15, 107–116.
- (12) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. *J. Am. Chem. Soc.* **1998**, 120, 9401–9409.
- (13) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, 10, 449–461.
- (14) Raha, K.; Merz, K. M. A Quantum Mechanics-Based Scoring Function: Study of Zinc Ion-Mediated Ligand Binding. *J. Am. Chem. Soc.* **2004**, 126, 1020–1021.
- (15) Merz, K. M., Jr Limits of Free Energy Computation for Protein–Ligand Interactions. *J. Chem. Theory Comput.* **2010**, 6, 1769–1776.
- (16) Mucs, D.; Bryce, R. A. The Application of Quantum Mechanics in Structure-Based Drug Design. *Expert Opin. Drug Discovery* **2013**, 8, 263–276.
- (17) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, 112, 632–672.
- (18) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, 115, 5643–5677.
- (19) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, 115, 5607–5642.
- (20) Raha, K.; van der Vaart, A. J.; Riley, K. E.; Peters, M. B.; Westerhoff, L. M.; Kim, H.; Merz, K. M. Pairwise Decomposition of Residue Interaction Energies Using Semiempirical Quantum Mechanical Methods in Studies of Protein–Ligand Interaction. *J. Am. Chem. Soc.* **2005**, 127, 6583–6594.
- (21) Peters, M. B.; Merz, K. M. Semiempirical Comparative Binding Energy Analysis (Se-Combine) of a Series of Trypsin Inhibitors. *J. Chem. Theory Comput.* **2006**, 2, 383–399.
- (22) Zhang, X.; Gibbs, A. C.; Reynolds, C. H.; Peters, M. B.; Westerhoff, L. M. Quantum Mechanical Pairwise Decomposition Analysis of Protein Kinase B Inhibitors: Validating a New Tool for Guiding Drug Design. *J. Chem. Inf. Model.* **2010**, 50, 651–661.
- (23) Mazanetz, M. P.; Ichihara, O.; Law, R. J.; Whittaker, M. Prediction of Cyclin-Dependent Kinase 2 Inhibitor Potency Using the Fragment Molecular Orbital Method. *J. Cheminf.* **2011**, 3, 2.
- (24) Heifetz, A.; Chudyk, E. I.; Gleave, L.; Aldeghi, M.; Cherezov, V.; Fedorov, D. G.; Biggin, P. C.; Bodkin, M. J. The Fragment Molecular Orbital Method Reveals New Insight into the Chemical Nature of GPCR–Ligand Interactions. *J. Chem. Inf. Model.* **2016**, 56, 159–172.
- (25) Heifetz, A.; Trani, G.; Aldeghi, M.; MacKinnon, C. H.; McEwan, P. A.; Brookfield, F. A.; Chudyk, E. I.; Bodkin, M.; Pei, Z.; Burch, J. D.; Ortwine, D. F. Fragment Molecular Orbital Method Applied to Lead Optimization of Novel Interleukin-2 Inducible T-cell Kinase (ITK) Inhibitors. *J. Med. Chem.* **2016**, 59, 4352–4363.
- (26) Parrish, R. M.; Sitkoff, D. F.; Cheney, D. L.; Sherrill, C. D. The Surprising Importance of Peptide Bond Contacts in Drug–Protein Interactions. *Chem. - Eur. J.* **2017**, 23, 7887–7890.
- (27) Parrish, R. M.; Thompson, K. C.; Martínez, T. J. Large-Scale Functional Group Symmetry-Adapted Perturbation Theory on Graphical Processing Units. *J. Chem. Theory Comput.* **2018**, 14, 1737–1753.
- (28) Mayhall, N. J.; Raghavachari, K. Molecules-in-Molecules: An Extrapolated Fragment-Based Approach for Accurate Calculations on Large Molecules and Materials. *J. Chem. Theory Comput.* **2011**, 7, 1336–1343.
- (29) Saha, A.; Raghavachari, K. Analysis of Different Fragmentation Strategies on a Variety of Large Peptides: Implementation of a Low Level of Theory in Fragment-Based Methods Can Be a Crucial Factor. *J. Chem. Theory Comput.* **2015**, 11, 2012–2023.
- (30) Thapa, B.; Beckett, D.; Jovan Jose, K.; Raghavachari, K. Assessment of Fragmentation Strategies for Large Proteins Using the Multilayer Molecules-in-Molecules Approach. *J. Chem. Theory Comput.* **2018**, 14, 1383–1394.

- (31) Thapa, B.; Beckett, D.; Erickson, J.; Raghavachari, K. Theoretical Study of Protein–Ligand Interactions Using the Molecules-in-Molecules Fragmentation-Based Method. *J. Chem. Theory Comput.* **2018**, *14*, 5143–5155.
- (32) Baum, B.; Mohamed, M.; Zayed, M.; Gerlach, C.; Heine, A.; Hangauer, D.; Klebe, G. More Than a Simple Lipophilic Contact: A Detailed Thermodynamic Analysis of Nonbasic Residues in the S1 Pocket of Thrombin. *J. Mol. Biol.* **2009**, *390*, 56–69.
- (33) *Molecular Operating Environment (MOE)*; Chemical Computing Group ULC: Montreal, QC, Canada, 2018.
- (34) Labute, P. Protonate3D: Assignment of Ionization States and Hydrogen Coordinates to Macromolecular Structures. *Proteins: Struct., Funct., Genet.* **2009**, *75*, 187–205.
- (35) The standard Amber 10 force field to describe the protein and extended Hückel theory (EHT) for the ligand to include the electronic effect.
- (36) Stewart, J. J. Optimization of Parameters for Semiempirical Methods V: Modification of Nddo Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.
- (37) Becke, A. D. Density-Functional Thermochemistry. V. Systematic Optimization of Exchange-Correlation Functionals. *J. Chem. Phys.* **1997**, *107*, 8554–8560.
- (38) Schmider, H. L.; Becke, A. D. Optimized Density Functionals from the Extended G2 Test Set. *J. Chem. Phys.* **1998**, *108*, 9624–9631.
- (39) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H–Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (40) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (41) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.
- (42) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (43) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theoret. Chim. Acta* **1973**, *28*, 213–222.
- (44) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XXIII. A Polarization-Type Basis Set for Second-Row Elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
- (45) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (46) Wang, J.; Hou, T. Develop and Test a Solvent Accessible Surface Area-Based Model in Conformational Entropy Calculations. *J. Chem. Inf. Model.* **2012**, *52*, 1199–1212.
- (47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.;
- Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, revision A.03; Gaussian, Inc.: Wallingford, CT, 2016.