Theoretical understanding of the thermodynamics and interactions in transcriptional regulator TtgR-ligand binding

Zhaoxi Sun[1,2*], Xiaohui Wang[1,3], and John Z.H.Zhang[1,4,5*]

[1]*State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

[2]*Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich, Jülich 52425, Germany*

[3]*Institute of Computational Science, Università della Svizzera italiana (USI), Via Giuseppe Buffi 13, CH-6900, Lugano, Ticino, Switzerland*

[4]*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

[5]*Department of Chemistry, New York University, NY, NY 10003, USA*

*To whom correspondence should be addressed: proszx@163.com, john.zhang@nyu.edu

Abstract

The transcriptional regulator TtgR belongs to the TetR family of transcriptional repressors. It depresses the transcription of the TtgABC operon and itself and thus regulates the extrusion of noxious chemicals with efflux pumps in bacterial cells. As the ligand binding domain of TtgR is rather flexible, it can bind with a number of structurally diverse ligands, such as antibiotics, flavonoids and aromatic solvents. In the current work, we perform equilibrium and nonequilibrium alchemical free energy simulation to predict the binding affinities of a series of ligands targeting the TtgR protein and the agreement between the theoretical prediction and the experimental result is observed. End-point methods of MM/PBSA and MM/GBSA are also employed for comparison. We further study the interaction maps and identify important interactions in the protein-ligand binding cases. The current work sheds light on atomic and thermodynamic understanding on the TtgR-ligand interactions.

Introduction

With the development of computational forces, computer simulations are becoming cheaper and more efficient. Before relatively costly experimental studies, it is better to predict the possibility of success through computational investigations. In material science and biophysics, the power of computer simulation is widely acknowledged.[1-9] A series of computational models and algorithms have been proposed, and a number of software have been developed. For instance, VMD,[10] CHARMM-GUI[11-12] and built-in tools of MD software such as the leap module in AMBER have been developed for model constructions. Gaussian,[13] Q-Chem,[14] GAMESS,[15] NWchem,[16] ORCA,[17] and Terachem[18] have been developed for quantum mechanical (QM) calculations. Software such as AMBER,[19] CHARMM,[20] GROMACS,[21] LAMMPS,[22] NAMD,[23] and Tinker[24] are used for molecular dynamics (MD) simulations. The simulated systems become larger and larger. The descriptions of the system, the Hamiltonians, become more and more accurate. The resulted agreement between computational predictions and experimental outcomes is thus enhanced.

However, even with the most advanced computational power and the most efficient algorithms, MD accessible timescale still cannot fully satisfy the needs of researchers. We always want to build a box as large as possible to eliminate the finite-size effect,[25-29] use Hamiltonians as accurate as possible,[30-36] and simulate our systems as long as possible to get reliable and converged estimates.[37-41] Normally, if there is no breaking or formation of chemical bonds, the polarization effect is not significant and the phenomena of interest happen at about several ns to ms, all-atom force fields are preferred to describe the motion of molecules. The statistical mechanical insights obtained from MD simulations rely on the ergodic assumption, where the time-averaged quantities are used to estimate ensemble averages. To obtain well-converged results from brute force simulations, there should be no rare events in the system. However, there are often slow motions that hinder the convergence of the simulation. To overcome the free energy barriers causing rare events, a series of smart sampling techniques are proposed. Examples of enhanced sampling techniques include umbrella sampling,[42-46] nonequilibrium steered MD,[47-50] and various replica exchange methods in the temperature, Hamiltonian and pH spaces.[51-56] As the systems of interest in modern research are often complex, an accurate description of the process often requires defining several collective degrees of freedoms. These important slow degrees of freedoms are often represented as collective variable (CV), reaction coordinate or order parameter. Even with proper definition of the important CVs, simulation in the high-dimensional CV space is complex and the computational cost is very high. If the only quantity of interest is the free energy difference between different states, such as the difference between the binding affinities of different ligands with the same protein, the simulation can be simplified with the so-called alchemical method. Alchemical free energy simulations construct alternative transformation pathways connecting the thermodynamic states and are widely applied in drug discovery,[37, 57-64] pKa shift predictions,[41, 65-67] solvation free energy calculations,[68-71] protein-protein

binding[8, 72-75] and protein-DNA interactions.[76] The free energy differences between different systems or states are obtained by alchemically switching the Hamiltonian from one state to another, and integrating the ensemble averages of the partial derivative of the alchemical Hamiltonian or reweighting via the Zwanzig equation or its derivatives.[8, 41, 60, 68, 75, 77-81] The integration methods are termed as thermodynamic integration (TI)[79-80] and the reweighting techniques are referred to as free energy perturbation (FEP)[82] and acceptance ratio methods.[83-85] Among all the equilibrium reweighting methods in the alchemical transformation, the acceptance ratio method of Bennett Acceptance Ratio (BAR) and its multistate variant named multi-state BAR (MBAR) have the best efficiency.[41] The nonequilibrium extensions of FEP and BAR are Jarzynski's Identity (JI)[86-87] and Crooks' Equation (CE).[88] The nonequilibrium technique uses a different simulation protocol. The system initiates from equilibrated configurations. Its Hamiltonian is switched gradually from one state to another. The nonequilibrium work accumulated during the nonequilibrium pulling is used as input for reweighting. The nonequilibrium alchemical methods have been applied in a number of cases.[41, 60, 68] Although the alchemical method is theoretically rigorous, it is often computationally demanding in large-scale applications. As a result, the less-detailed end-point methods named MM/PBSA and MM/GBSA[89] are also popular in drug discovery. They can be very efficient but are less accurate.

The bacteria resist naturally occurring deleterious compounds and antimicrobial agents of semi-synthetic antibiotics and biocides by expelling them with efflux pumps. The active major cell protective behavior transports deleterious chemicals from the cytoplasmic membrane to the external environment and thus reduces the concentration of the noxious compounds in the cell.[90-92] The multidrug resistance efflux pump in Gram-negative bacteria mostly belongs to one of the five bacterial efflux transporter families, the resistance-nodulation-cell division (RND) superfamily.[93-95] The expression of the multidrug resistance efflux pump is controlled by transcriptional regulators, such as TtgR of the non-pathogenic bacterium Pseudomonas putida DOT-T1E.[96-97] They directly bind with the toxic chemicals and activate the efflux pump genes.

The DOT-T1E strain has high resistance to noxious organic solvents.[98] It has three essential RND efflux pumps, including TtgABC, TtgDEF, and TtgGHI.[99] The transcriptional regulator of TtgR depresses not only the transcription of TtgABC operon but also the expression of itself. As an important member of the TetR family of transcriptional repressors,[100-103] the TtgR operator has two functional domains: a highly conserved N-terminal helix-turn-helix DNA binding domain, and a less conserved C-terminal domain for dimerization and ligand binding.[104-107] There is an angle of about 80° between the two domains. As the ligand binding domain has less sequence conservation, a number of structurally diverse ligands can be recognized by this site. Specifically, TtgR is able to bind with antibiotics, flavonoids and aromatic solvents.[108-110] For instance, the antimicrobial properties of the flavonoids of quercetin, naringenin and phloretin lead to their abilities of binding to TtgR.[111] The binding between the effector and TtgR often leads to the dissociation and the

transcriptional activation of TtgABC and TtgR.[96] The size, shape and the charge distributions of the ligands can differ significantly. The only common feature of the ligands is the existence of aromatic rings.

All-atom understanding of the interactions between the transcriptional regulator TtgR and inhibitors is indispensable. Although experimental structures provide snapshots of the interaction networks in the complex, dynamics and short-lived interactions are missing. Therefore, in the current work, free energy simulations are performed to calculate the binding affinities between TtgR and its inhibitors. The comprehensive equilibrium and nonequilibrium alchemical methods are used and agreements between the predictions and the experimental values are obtained. We further test the performance of the end-point methods of MM/PBSA and MM/GBSA in the current case. After checking the accuracy of the predicted thermodynamics, we investigate the conformational ensemble and identify important interactions stabilizing the protein-ligand complex. We expect the current work to shed light on the understanding the TtgR-inhibitor interactions and the further development of possible drugs targeting this protein.

Method and Computational Details

System preparation. The structure of 2UXH, 2UXI, 2UXP and 2UXU are used to obtain the coordinate of the protein atoms.[111] The missing residues in protein are added with the Modloop web interface.[112] The AMBER14SB force field is used to describe the protein. The structures of ligands are obtained from the above-mentioned pdb files and the PubChem Compound Database. The 2D chemical structures of the ligands considered in this work along with an illustration of the protein-ligand binding case are given in Figure 1. In the parameterization of the force field parameters for ligands, we perform geometry optimization on the ligand atoms at AM1-BCC[113] level and then derive the atomic charges. The other parameters such as bonded terms and vdW radius are obtained from the GAFF force field.[114] The systems are solvated with TIP3P[115-116] water molecules following the 12 Å criteria, namely the minimum distance between the surface of the protein-ligand complex and the edge of the periodic box being larger than 12 Å. The truncated octahedron cell is replicated in whole space by periodic boundary conditions. Non-polarizable spherical counter ions of Na^+ parameterized for TIP3P water model by Joung and Cheatham[117-118] are added for neutralization.

Free energy simulation. In order to compute the relative binding affinities between different inhibitors, we construct the thermodynamic cycle in Figure 2a. The closure of the thermodynamics cycle gives $\Delta G_{bound,PA->PB} - \Delta G_{unbound,PA->PB} = \Delta G_{binding,PB} - \Delta G_{binding,PA}$. The right-hand side of the equation is the difference between binding affinities of protein-B and protein-A complexes, which is our aim in the free energy simulation. The double free energy difference in the left-hand side of the equation is the relative binding affinities between the ligand A and the ligand B, which is the quantity we extract from simulations. Two types

of simulations are performed to obtain the two free energy differences in the left-hand side of the equation. The first term $\Delta G_{unbound, PA \to PB}$ is computed from the transformation between ligands in the absence of protein. In solution, ligand A is gradually mutated to ligand B along the alchemical pathway. Therefore, this calculation can also be understood as the calculation of the relative solvation free energy of the two ligands. The second term $\Delta G_{bound, PA \to PB}$ is obtained by similar mutation procedures in the presence of the protein.

As the differences in structures of the ligands are relatively large, even with the help of the staging regime, the convergence behavior of transformations between different ligands is expected to be poor. To make the alchemical free energy simulation easier to converge, instead of selecting a reference ligand and then performing the transformation from it to the other ligands, we construct a transformation map shown in Figure 2b, according to chemical intuition. In this transformation map, neighboring ligands share high similarities and thus the alchemical mutation introduces smaller perturbations to the system.

The similar atom mapping technique is often used to improve the convergence behavior of alchemical free energy simulation. The method maps similar atoms in different inhibitors and sets the non-bonded parameters (i.e. atomic charges and vdW radius) in different ligands to be the same. To effectively and realistically represent the variation of force field parameters of ligands upon the alchemical mutation, in our transformation, we do not use this scheme. The whole ligand is included in the alchemical region and all non-bonded parameters are allowed to vary. As a result, no two atomic charges share the same value in different ligands. This setting introduces much larger fluctuations compared with maximum similarity mixing regimes[41, 60] and thus has worse convergence behavior, but the true variation in Hamiltonians in mutating the ligand is represented more realistically.

Equilibrium and nonequilibrium free energy simulation are performed along the alchemical pathway. As the difference between the full A state and the full B state is large, to increase the phase space overlap between neighboring states in equilibrium sampling and to reduce the dissipation in nonequilibrium pulling, we employ the staging regime or the stratification scheme. As the creation and annihilation of atoms at the end states in the vdW transformation results in vdW singularity,[119-128] which hinders the convergence of the simulation, we employed the nonlinear separation-shifted softcore-potential[119, 122-124, 129] to avoid the problem. Further, in order to simplify the procedure of the transformation, the softcore scheme is applied to both the vdW transformation and the charge transformation and the two types of mutation are performed altogether. Also due to the use of the soft-core potential, the linear mixing rule can be used to combine the Hamiltonians of two ligands to determine the Hamiltonians of the intermediate states. Another technical note is that the net charges of all ligands considered in the current work are zeros and thus the alchemical mutation does not introduce new charges to the system. Therefore, there is no need to add corrections for the non-equal charges

of the two ligands for simulations with periodic boundary condition.

Firstly, we provide details about equilibrium free energy simulations. In equilibrium alchemical transformation, the intermediate states are equally spaced from $\lambda_{min} = 0.00$ to $\lambda_{max} = 1.00$ with increments of $\Delta\lambda_{eq} = 0.05$, leading to 21 states in total. The phase space overlap between neighboring states is checked with the quantitative estimator of phase space overlap named overlap matrix.[130] The matrix relates non-linearly with the covariance matrix of MBAR. The matrix O in the current case is a 21 x 21 matrix and its element O_{ij} gives the average possibility of finding a sample from state i at state j.[130] According to previous experience, reliable free energy estimates can be obtained when the main diagonal and its neighbors are appreciably larger than 0.03.[130] In each alchemical intermediate state, 9000 cycles energy minimization, 600 ps NVT heating from 0 K to 303 K and 2 ns NPT equilibration are performed in order to get rid of the bias due to the initial configuration. After equilibration, 2 ns production run with a sampling interval of 2 ps is performed to accumulate the times-eries data. Isotropic position scaling and Berendsen barostat is implemented to regulate the pressure. As there are 21 windows and 2 types of simulation (solvated ligand and protein-ligand systems), for each mutation of ligand the total sampling time is 84 ns. To get a theoretically rigorous estimate of statistical uncertainty, we calculate the autocorrelation time τ of the partial derivative of the alchemical Hamiltonian $\left.\dfrac{\partial H}{\partial \lambda}\right|_{\lambda=\lambda_i}$ [130-131] and the statistical inefficiency $\phi = 1 + 2\tau$. The whole dataset is then subsampled by ϕ to extract the independent data points.[41, 60, 68] Three post-processing methods including TI with trapezoid rule for numerical integration,[79-80, 132] the bidirectional reweighting regime of BAR,[133] and multistate reweighting estimator of MBAR are used to extract the free energy estimates from alchemical free energy simulations.[84-85] According to our previous experience, these three methods are among the most reliable and efficient free energy estimators based on equilibrium dynamics.[41, 71, 134-135]

Secondly, we provide details about the nonequilibrium transformation. The initial configurations for nonequilibrium transformations are obtained during the above equilibrium alchemical transformation. As the window spacing used in equilibrium sampling is already dense enough for reliable reweighting with equilibrium perturbation-based estimators, to illustrative the feature of nonequilibrium pulling, we use a larger increment of $\Delta\lambda_{non-eq} = 0.1$, with which the instantaneous perturbation is relatively large. We calculate the statistical inefficiency and extract 100 independent configurations in each alchemical intermediate. Bidirectional pulling simulations are initiated from these uncorrelated configurations and nonequilibrium works are accumulated. The alchemical order parameter λ is changed by 0.001 every 200 fs. According to our previous experience on nonequilibrium stratification in the alchemical space, this transformation speed is

already very slow and able to obtain converged results with reasonable computational costs even in hard-to-converge cases.[41, 60, 68] The resulting overall computational cost in the nonequilibrium alchemical transformation is similar to that in the equilibrium one. According to our previous work, the overall statistical uncertainty is non-linearly dependent on the overall simulation time.[48, 68] Thus, the overall statistical uncertainties in equilibrium and nonequilibrium transformation should be similar.

As the alchemical method is computationally demanding for large-scale applications, often the more efficient but less accurate end-point free energy methods are employed. We thus perform end-point free energy calculations with MM/PBSA and MM/GBSA[89] to compare their accuracy with the alchemical method. The protein-ligand complex is prepared in the same way as the alchemical free energy simulation, namely AMBER14SB for protein, GAFF and AM1-BCC charge for ligand, Na^+ ions for neutralization and TIP3P water for solvation. For each protein-ligand complex, 5000 cycles of minimization and 2 ns NPT equilibration are performed. Then 100 ns production run with a sampling interval of 100 ps is performed to extract 1000 configurations. The 100 ns sampling time is the typical length of sampling in applying the end-point methods. The 100 ps sampling interval is very long in MD simulation and we expect the obtain uncorrelated samples with such a setting. The gas-phase enthalpy changes and solvation free energies are calculated with the 1000 independent configurations. The GB^{OBC} model[136-137] is used in GBSA calculations. All other settings remain their default values in AMBER.

In all simulations, the SHAKE[138] algorithm is employed to constrain bonds involving hydrogen atoms in protein, ligands and water molecules.[139] The time step used is 2 fs. Langevin dynamics[140] with the collision frequency of 2 ps^{-1} are implemented for temperature regulation. A cut off of 10 Å for non-bonded interactions in the real space is used and the long-range electrostatic interactions are treated with the PME method.[141] MD simulations[81] are performed with the AMBER[19] suite and all other calculations are obtained with homemade codes.

Result and discussion

Convergence behavior of the free energy simulation. Firstly, we check the convergence behavior of the alchemical free energy calculation. Here, the example we choose is the transformation from AGI to LU2. These two ligands are structurally similar. The alchemical mutation between them includes the –OH substitution. As we include the whole ligand into the alchemical region and do not use similar atom mapping, the alchemical transformation is performed actually by annihilating AGI and creating LU2 at the same time. The phase space overlap in this case is checked with the overlap matrices, which are given in Figure S1. We notice that the elements of the overlap matrices are large enough for reliable reweighting.

The ensemble averages and the time series of the partial derivative of the alchemical Hamiltonian, $\left. \frac{\partial U}{\partial \lambda} \right|_{\lambda_i}$, are shown in Figure 3. The charge distributions of the two ligands are also compared in Figure 3a, from which we know that the perturbation of the –OH substitution on the atomic charges is not very significant. In Figure 3a, the ensemble average of $\left. \frac{\partial U}{\partial \lambda} \right|_{\lambda_i}$ in the protein-ligand complex varies similarly with that in the ligand-only system, which indicates that both ligands perform similarly in protein-ligand complexes and solution. As a result, the overall free energy difference in the alchemical pathway is small. The time series of $\left. \frac{\partial U}{\partial \lambda} \right|_{\lambda_i}$ in Figure 3b are stable in all states and no multi-state behavior is observed. Therefore, the convergence of the alchemical free energy simulation is expected to be good.

The end-point free energy simulation is performed for the normal length of 100 ns. The standard errors of the mean of the MM/PBSA and MM/GBSA results are given in Table S1, from which we know that the statistical errors are very small, indicating good convergence behavior.

The quality of predictions. The quality of the computational predictions is assessed with the accuracy, the statistical uncertainty and ranking metrics. Firstly, we get an overview of the predictions. The ligand of AGI is used as the reference and we calculate the differences between binding affinities of it and all other ligands. The predictions from the alchemical and end-point free energy simulations are summarized in Table 1. More detailed results of end-point methods are given in Table S1.

Firstly, we discuss the alchemical predictions. The TI results suffer from large statistical uncertainties. Its bias is also large (c.a. 2 kcal/mol). This phenomenon is triggered by the non-linearity of $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda_i}$, which leads to integration bias inevitably. By contrast, the perturbation-based BAR and MBAR are successful in bias elimination and thus provide predictions in agreements with the experimental values, as they rely on statistically optimal multi-directional free energy perturbation rather than numerical integration. The free energy estimates obtained from these two methods are very similar, which is expected when the phase space overlap is not very large. The statistical errors of BAR and MBAR are also much smaller than those of TI. The predictions of nonequilibrium pulling simulations are in agreement with the equilibrium results (specifically BAR and MBAR predictions), which indicates the convergence of the nonequilibrium pulling. The statistical errors of the CE results are very similar to those of BAR and MBAR, which is expected as both simulations are of similar lengths. Compared with the alchemical results, the end-point results are less accurate but with significantly smaller statistical errors. Therefore, the elimination of bias is the major concern in end-point free energy simulations.

After checking individual predictions, we focus on the error quantities of MSE, RMSE and MAE, which reflect the average errors of the predictions. Among alchemical methods, TI has the largest errors, while BAR, MBAR and CE share very similar error estimates. This phenomenon indicates that in the alchemical free energy calculation for protein-ligand binding, the statistically optimal perturbation-based estimators including the equilibrium methods of BAR and MBAR and the nonequilibrium CE are more efficient than that the integration-based TI. The values of the ranking metric PI are similar for all of the four estimators, indicating the correctness of the predicted rank. Namely, all of the alchemical methods can correctly predict the relative binding affinities. Therefore, if only the rank of binding affinities is of interest, TI is as effective as the other estimators. As for the end-point methods, the error quantities are much larger than the alchemical ones. The ranking metrics of MM/PBSA is smaller than 0, which indicates that the rank predicted is totally wrong. By contrast, the MM/GBSA has a PI of 0.6, which is as good as the alchemical methods. Therefore, the MM/GBSA method can be efficient in predicting the rank of binding affinities, although the absolute values of the end-point predictions are inaccurate.

The comparison between the computational predictions and the experimental results are shown in Figure 4a, from which we know that BAR, MBAR and CE predictions are the best among all methods, the TI results are a little worse than them and end-point methods provide the worst predictions. These observations are consistent with the previous discussions in the aspects of the error quantities and ranking information.

We further decompose the free energy difference in the alchemical transformations in Figure 2b with MBAR reweighting. The enthalpic and entropic contributions to the free energy differences are provided in Table 2. The double enthalpy difference of $\Delta\Delta H = \Delta H_{\text{complex}} - \Delta H_{\text{ligands}}$ and the double entropy difference $-T\Delta\Delta S$ give the enthalpic and entropic changes during the alchemical mutation, respectively. The statistical uncertainties of the enthalpic and entropic contributions are much larger than those of the free energy differences in Table 1, which is triggered by the correlation between the statistical uncertainties of ΔH and $-T\Delta S$. In each alchemical transformation, the ligand-only systems have smaller statistical uncertainties than the protein-ligand complexes. Despite the size of the statistical uncertainties, we can still obtain some insights from this decomposition. The changes in enthalpy and entropy are plotted in Figure 4b, from which we notice that the sum of the enthalpic change and the entropic change is almost zero in all alchemical transformations. This phenomenon indicates the existence of the entropy-enthalpy compensation,[142-144] which is widely observed in biological systems.[145-153] The existence of such compensation emphasizes the importance of using the free energy difference rather than its enthalpic or entropic component in predicting the binding thermodynamics in drug discovery.

Interaction patterns of the protein-ligand complexes. Aside from the thermodynamic information, we extract more detailed interaction patterns of the protein-ligand binding from simulations.

Firstly, we calculate the number of hydrogen bonds between the ligand and its surroundings in complex and solvated ligand systems. The time series, mean and standard error of the mean of the number of hydrogen bonds are shown in Figure S2 and summarized in Table S2. We notice that for ligands except QUE, the numbers of hydrogen bonds formed in the solvated ligand system are larger than those in complexes. There are obvious differences between the numbers of hydrogen bonds formed for different ligands.

Then, we investigate further details about protein-ligand interactions. As the crystal structures of protein-ligand complexes are only available for some of the ligands under investigation, we use equilibrated structures to study the interaction map for protein-ligand binding. In Figure 5a and Figure S3, the interaction maps for 9 ligands under consideration are shown. We can identify important residues contributing favorable interactions in protein-ligand binding.

For 1NP, the interaction map tells us that SER77 forms stable hydrogen bonds with the ligand. From simulations, we also observe short-lived protein-ligand hydrogen bonds involving GLU78 and MET89. The ligand 2NP is structurally similar to 1NP and is a –OH substitution of 1NP. As a result, the interaction map of 2NP is very similar to that of 1NP. Its extra –OH group forms weak hydrogen bonds with GLY140. For 27K, in the interaction map, the ALA74-27K hydrogen bonds seem to be stable. From the trajectories, we know that the GLU78 forms the most stable hydrogen bonds with the ligand, while the ALA74-27K hydrogen bonds are weaker. Short-lived hydrogen bonds involving GLY140 are also observed. For CLM, the interaction map shows two candidates of important residues, including ASP172 and HIS112. From simulations, we know that the ASP172-CLM interaction is strong, while the HIS112-CLM interaction is relatively weak. For G50, the interaction map gives ASN110 as the important residue for hydrogen bonding, while the trajectories give three important residues for hydrogen bonding, including ASN110, SER77 and CYS137. AGI is larger than all above ligands and its interactions are more complex. The interaction map gives three residues for hydrogen bonding, including ASP172, ASN110 and MET89. From trajectories, we identify one more residue of HIS112. For CUE, the interaction map gives ASN110, while the trajectories identify GLY140, LEU66 and HIS67 as important residues for protein-ligand binding. Short-lived CYS137-CUE hydrogen bonds are also observed. The interaction map of LU2 identifies ASN110, HIS67 and CYS137 as important residues, which is consistent with the observation from simulations. The last ligand of QUE is similar. The interaction map gives HIS67, ASN110, HIS114 and CYS137, which is consistent with the simulation results.

Therefore, considering the discussion about important residues stabilizing protein-ligand complexes, the residues are in the regions of LEU66-HIS67, SER77-GLU78, MET89, ASN110-HIS114, CYS137-GLY140, and ASP172.

As these interaction maps are extracted from a single structure, the dynamics of the system and the fluctuations of the interactions are not considered. We thus seek for ways to represent the fluctuation of the important stabilizing interactions during the simulations to get a better understanding. A better way to visualize the residue-specific interactions is the time series of contact numbers. The numbers of contacts between alpha-C atoms and the ligand atoms are plotted in Figure 5b and Figure S4. The red dots and green dots represent regions in close contacts with the ligand and thus have stronger interactions. From these plots, we know that the above-mentioned regions are in strong interactions with the ligand. Further, these interactions are formed in the entire simulation. These observations indicate that the above-mentioned interactions stabilize the protein-ligand interactions and the binding pose formed is very stable.

The time evolution of secondary structures are shown in Figure 6 and Figure S5, from which we can see how the secondary structures of protein change during the simulation. The secondary structures fluctuate and there is no obvious sign for conformational changes. The root-mean-squared fluctuation (RMSF) of C-α is calculated for different protein-ligand complexes, the results of which are shown in Figure 7. The RMSF plots for different complexes are similar, which indicates that the interactions between protein and ligands are similar. One exception is the protein-CLM complex. It seems that the protein-CLM complex has larger conformational fluctuations in several parts of the protein, such as the region of the 130th-140th and the 150th-170th residues. This phenomenon should be triggered by the unique chemical structure of CLM, which results in different strengths of protein-ligand interactions.

Conclusion

In this work, free energy calculations are employed to investigate the binding thermodynamics of several ligands targeting the transcriptional regulator TtgR. The alchemical method with equilibrium and nonequilibrium dynamics is coupled with the statistically optimal reweighting estimators to extract the relative binding affinities. The end-point methods including MM/PBSA and MM/GBSA are also employed for comparison. The results suggest that the accuracy of the alchemical method is very good and the predicted ranks of binding affinities agree with the experimental value, while the end-point methods fail to provide accurate predictions of the binding affinities. However, MM/GBSA does predict the correct rank of the binding affinities, which suggests that the method may be useful in finding the hits in drug discovery. The enthalpic and entropic contributions are extracted from the free energy differences with MBAR reweighting. The enthalpic change upon the alchemical mutation is almost completely compensated by the entropic change, indicating the existence of the entropy-enthalpy compensation. Therefore, we should rely on the binding free energies rather than the binding enthalpies of entropies to determine the rank of ligands. Detailed investigation including the interaction patterns from the interaction maps extracted from equilibrated structures, the

formations of intra-molecular hydrogen bonds, the number of contacts between protein backbone and ligand atoms, the fluctuation of secondary structures and the conformational fluctuations of protein backbone reveal the important residues stabilizing the ligands. These residues fall in the regions of LEU66-HIS67, SER77-GLU78, MET89, ASN110-HIS114, CYS137-GLY140, and ASP172. The current work sheds light on the atomic details of the interaction between the transcriptional regulator TtgR and its inhibitors.

Acknowledgement

Conflicts of interest

There are no conflicts of interest to declare.

Supporting Information

The supporting information includes detailed results of MM/PBSA and MM/GBSA, overlap matrices of alchemical transformation from AGI to LU2, the time series, mean and standard errors of the number of hydrogen bonds formed between ligands and their surroundings, the interaction maps, the fluctuations of protein-ligand contact numbers, and the time series of secondary structures.

References

1. Piccini, G.; McCarty, J. J.; Valsson, O.; Parrinello, M., Variational flooding study of a SN2 reaction. *The journal of physical chemistry letters* **2017,** *8* (3), 580-583.

2. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008,** *100* (2), 020603.

3. Vanden-Eijnden, E.; Venturoli, M., Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J Chem Phys* **2009,** *130* (19), 194103.

4. Allen, R. J.; Frenkel, D.; ten Wolde, P. R., Forward flux sampling-type schemes for simulating rare events: efficiency analysis. *J. Chem. Phys.* **2006,** *124* (19), 194111/1-194111/17.

5. Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977,** *23* (2), 187-199.

6. Sun, Z.; Zhu, T.; Wang, X.; Mei, Y.; Zhang, J. Z., Optimization of convergence criteria for fragmentation methods. *Chem. Phys. Lett.* **2017,** *687*, 163-170.

7. Yan, Y.; Wang, W.; Sun, Z.; Zhang, J. Z.; Ji, C., Protein–Ligand Empirical Interaction Components for Virtual Screening. *J. Chem. Inf. Model.* **2017,** *57* (8), 1793-1806.

8. Qiu, L.; Yan, Y.; Sun, Z.; Song, J.; Zhang, J. Z. H., Interaction entropy for computational alanine scanning in protein–protein binding. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2017**, e1342-n/a.

9. Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S., Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101* (48), 16789-16794.

10. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graphics* **1996,** *14* (1), 33-38.

11. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008,** *29* (11), 1859-1865.

12. Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M., CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem.* **2014,** *35* (27), 1997-2004.

13. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G., GAUSSIAN09, Gaussian, Inc., Wallingford, CT, USA,(2009). *Google Scholar* **2016**.

14. Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P., Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* **2006,** *8* (27), 3172-3191.

15. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S., General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993,** *14* (11), 1347-1363.

16. Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L., NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **2010,** *181* (9), 1477-1489.

17. Neese, F., The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012,** *2* (1), 73-78.

18. Ufimtsev, I. S.; Martinez, T. J., Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics. *J. Chem. Theory Comput.* **2009,** *5* (10), 2619-2628.

19. Case, D. A.; Cheatham, T. E.; Tom, D.; Holger, G.; Luo, R.; Merz, K. M.; Alexey, O.; Carlos, S.; Bing, W.; Woods, R. J., The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005,** *26* (16), 1668-1688.

20. Brooks, B. R.; Brooks III, C. L.; Mackerell Jr, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009,** *30* (10), 1545-1614.

21. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1*, 19-25.

22. Plimpton, S., Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995,** *117* (1), 1-19.

23. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781-1802.

24. Ponder, J. W., TINKER: Software tools for molecular design.

25.   Villamaina, D.; Trizac, E., Thinking outside the box: fluctuations and finite size effects. *European Journal of Physics* **2014**, *35* (3), 035011.

26.   Heidari, M.; Cortes-Huerto, R.; Kremer, K.; Potestio, R., Concurrent coupling of realistic and ideal models of liquids and solids in Hamiltonian adaptive resolution simulations. *The European Physical Journal E* **2018**, *41* (5), 64.

27.   Román, F.; White, J.; Velasco, S., Fluctuations in an equilibrium hard-disk fluid: Explicit size effects. *J. Chem. Phys.* **1997**, *107* (12), 4635-4641.

28.   Lebowitz, J.; Percus, J., Long-range correlations in a closed system with applications to nonuniform fluids. *Phys. Rev.* **1961**, *122* (6), 1675.

29.   Salacuse, J.; Denton, A.; Egelstaff, P., Finite-size effects in molecular dynamics simulations: Static structure factor and compressibility. I. Theoretical method. *Physical Review E* **1996**, *53* (3), 2382.

30.   Liu, W.; Sakane, S.; And, R. H. W.; Doren, D. J., The Hydration Free Energy of Aqueous Na+ and Cl- at High Temperatures Predicted by ab Initio/Classical Free Energy Perturbation: 973 K with 0.535 g/cm3 and 573 K with 0.725 g/cm3. *J.phys.chem.a* **2002**, *106* (7), 1409-1418.

31.   Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A., Hybrid ab initio quantum mechanics/molecular mechanics calculations of free energy surfaces for enzymatic reactions: the nucleophilic attack in subtilisin. *The Journal of Physical Chemistry B* **1998**, *102* (12), 2293-2301.

32.   Caveayland, C.; Skylaris, C. K.; Essex, J. W., Direct Validation of the Single Step Classical to Quantum Free Energy Perturbation. *J. Phys. Chem. B* **2014**, *119* (3), 1017-25.

33.   Lameira, J. S.; Kupchencko, I.; Warshel, A., Enhancing Paradynamics for QM/MM Sampling of Enzymatic Reactions. *J. Phys. Chem. B* **2016**, *120* (9), 2155.

34.   Moran, J.; Smith, A. G.; Carris, R. M.; Johnson, J. S.; Krische, M. J., A simple QM/MM approach for capturing polarization effects in protein-ligand binding free energy calculations. *J. Phys. Chem. B* **2011**, *115* (17), 4911-26.

35.   Plotnikov, N.; Kamerlin, S. C. L.; Warshel, A., ParaDynamics: An Effective and Reliable Model for Ab Initio QM/MM Free Energy Calculations and Related Tasks. *J. Phys. Chem. B* **2011**, *115* (24), 7950-62.

36.   Plotnikov, N. V.; Warshel, A., Exploring, refining, and validating the paradynamics QM/MM sampling. *J. Phys. Chem. B* **2012**, *116* (34), 10342-10356.

37.   Parenti, M. D.; Rastelli, G., Advances and applications of binding affinity prediction methods in drug discovery. *Biotechnol Adv* **2012**, *30* (1), 244-50.

38.   Gohlke, H.; Case, D. A., Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf. *J. Comput. Chem.* **2004**, *25* (2), 238-250.

39.   Olsson, M. A.; Söderhjelm, P.; Ryde, U., Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level. *J. Comput. Chem.* **2016**, *37* (17), 1589-1600.

40.   Monroe, J. I.; Shirts, M. R., Converging free energies of binding in cucurbit[7]uril and octa-acid host–guest systems from SAMPL4 using expanded ensemble simulations. *J. Comput. Aided Mol. Des.* **2014**, *28* (4), 401-15.

41.   Sun, Z.; Wang, X.; Song, J., Extensive Assessment of Various Computational Methods for Aspartate's pKa Shift. *J. Chem. Inf. Model.* **2017**, *57* (7), 1621-1639.

42.   Mezei, M., Adaptive Umbrella Sampling: Self-consistent Determination of the Non-Boltzmann Bias. *J. Comput. Phys.* **1987**, *68* (1), 237-248.

43.   Hooft, R. W.; van Eijck, B. P.; Kroon, J., An Adaptive Umbrella Sampling Procedure in Conformational Analysis using Molecular Dynamics and Its Application to Glycol. *J. Chem. Phys.* **1992**, *97* (9), 6690-6694.

44.   Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1* (6), 932-942.

45.   Sun, Z.; Wang, X.; Zhang, J. Z. H., Protonation-dependent Base Flipping in The Catalytic Triad of A Small RNA. *Chem. Phys. Lett.* **2017**, *684*, 239-244.

46.   Wang, X.; Deng, B.; Sun, Z., Thermodynamics of helix formation in small peptides of varying length in vacuo, in implicit solvent, and in explicit solvent. *J. Mol. Model.* **2019**, *25* (1), 3.

47.   Wang, X.; He, Q.; Sun, Z., BAR-Based Multi-Dimensional Nonequilibrium Pulling for Indirect Construction of a QM/MM Free Energy Landscape. *Phys. Chem. Chem. Phys.* **2019**, *21* (12), 6672-6688

48.   Wang, X.; Xingzhao, T.; Boming, D.; John Z. H., Z.; Sun, Z., BAR-based Optimum Adaptive Steered MD for Configurational Sampling. *J. Comput. Chem.* **2019**, *9999*, 1-20.

49.   Wang, X.; Sun, Z., Determination of Base Flipping Free Energy Landscapes from Nonequilibrium Stratification. *J. Chem. Inf. Model.* **2019**.

50.   Sun, Z.; Wang, X.; Zhang, J. Z. H.; He, Q., Sulfur-substitution-induced base flipping in the DNA duplex. *Phys. Chem. Chem. Phys.* **2019**.

51.   Fukunishi, H.; Watanabe, O.; Takada, S., On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116* (116), 9058-9067.

52.   Itoh, S. G.; Damjanovic, A.; Brooks, B. R., pH replica-exchange method based on discrete protonation states. *Proteins* **2011**, *79* (12), 3420-36.

53.   Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C., Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model. *J. Chem. Theory Comput.* **2006**, *2* (2), 420.

54.   Sabri, D. D.; Roitberg, A. E., Optimization of Umbrella Sampling Replica Exchange Molecular Dynamics by Replica Positioning. *J. Chem. Theory Comput.* **2013**, *9* (9), 4692–4699.

55.   Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1-2), 141-151.

56.   Sugita, Y.; Kitao, A.; Okamoto, Y., Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113* (15), 6042-6051.

57.   Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P., Molecular mechanics methods for predicting protein–ligand binding. *Physical Chemistry Chemical Physics Pccp* **2006**, *8* (44), 5166-77.

58.   Genheden, S.; Ryde, U.; Söderhjelm, P., Binding affinities by alchemical perturbation using QM/MM with a large QM system and polarizable MM model. *J. Comput. Chem.* **2015**, 2114-2124.

59.   Wang, Y.; Liu, J.; Zhang, L.; He, X.; Zhang, J. Z., Computational search for aflatoxin binding proteins. *Chem. Phys. Lett.* **2017**, *685*, 1-8.

60.   Sun, Z. X.; Wang, X. H.; Zhang, J. Z. H., BAR-based Optimum Adaptive Sampling Regime for Variance Minimization in Alchemical Transformation. *Phys. Chem. Chem. Phys.* **2017**, *19* (23), 15005-15020.

61.   Giovannelli, E.; Procacci, P.; Cardini, G.; Pagliai, M.; Volkov, V.; Chelli, R., Binding free energies of host-guest systems by nonequilibrium alchemical simulations with constrained dynamics: Theoretical framework. *J. Chem. Theory Comput.* **2017**.

62.   Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc* **2015**, *137* (7), 2695-703.

63.   Wang, X.; Sun, Z., Understanding PIM-1 kinase inhibitor interactions with free energy simulation. *Phys. Chem. Chem. Phys.* **2019**, *21*, 7544-7558.

64.   Sun, Z.; Wang, X.; Zhang, J. Z., Determination of Binding Affinities of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Inhibitors from Free Energy calculation. *Chem. Phys. Lett.* **2019**, *723*, 1-10.

65.   Davies, M. N.; Toseland, C. P.; Moss, D. S.; Flower, D. R., Benchmarking pKa prediction. *Bmc Biochemistry* **2006**, *7* (1), 1-12.

66.   Stanton, C. L.; Houk, K. N., Benchmarking pKa Prediction Methods for Residues in Proteins. *J. Chem. Theory Comput.* **2008**, *4* (6), 951-966.

67.   Archontis, G.; Simonson, T., Proton binding to proteins: a free-energy component analysis using a dielectric continuum model. *Biophys. J.* **2005**, *88* (6), 3888-904.

68.   Wang, X.; Tu, X.; Zhang, J. Z. H.; Sun, Z., BAR-based Optimum Adaptive Sampling Regime for Variance Minimization in Alchemical Transformation: The Nonequilibrium Stratification. *Phys. Chem. Chem. Phys.* **2018**, *20* (3), 2009-2021.

69.   Gerhard Hummer, †, ‡; And, L. R. P., §; García†, A. E., Free Energy of Ionic Hydration. *J. Phys. Chem.* **2010**, *100* (4), 1206-1215.

70.   Hummer, G.; Pratt, L. R.; Garcia, A. E., Hydration free energy of water. *J. Phys. Chem.* **1995**, *99* (38), 14188-14194.

71.   Shirts, M. R.; Pande, V. S., Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J Chem Phys* **2005**, *122* (13), 134508.

72.   Gohlke, H.; Kiel, C.; Case, D. A., Insights into protein–protein binding by binding free energy calculation and free energy

decomposition for the Ras–Raf and Ras–RalGDS complexes. *Journal of molecular biology* **2003,** *330* (4), 891-913.

73. Kilburg, D.; Gallicchio, E., Assessment of a Single Decoupling Alchemical Approach for the Calculation of the Absolute Binding Free Energies of Protein-Peptide Complexes. *Frontiers in molecular biosciences* **2018,** *5*, 22.

74. Bai, H. J., Protein-Protein Interactions:Interface Analysis,Binding Free Energy Calculation and Interaction Design. *ACTA PHYSICO-CHIMICA SINICA* **2010,** *26* (7), 1988-1997(10).

75. Sun, Z.; Yan, Y. N.; Yang, M.; Zhang, J. Z., Interaction Entropy for Protein-Protein Binding. *J. Chem. Phys.* **2017,** *146* (12), 124124.

76. Gapsys, V.; de Groot, B. L., Alchemical Free Energy Calculations for Nucleotide Mutations in Protein–DNA Complexes. *J. Chem. Theory Comput.* **2017,** *13* (12), 6275-6289.

77. Wang, X.; Sun, Z., A Theoretical Interpretation of Variance-based Convergence Citeria in Perturbation-based Theories. *arXiv preprint arXiv:1803.03123* **2018**.

78. Swope, W. C., A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982,** *76* (1), 637.

79. Resat, H.; Mezei, M., Studies on free energy calculations. I. Thermodynamic integration using a polynomial path. *J. Chem. Phys.* **1993,** *99* (8), 6052-6061.

80. Resat, H.; Mezei, M., Studies on free energy calculations. II. A theoretical approach to molecular solvation. *J. Chem. Phys.* **1994,** *101* (7), 6126-6140.

81. Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammont, J. A., Improving the Efficiency of Free Energy Calculations in the Amber Molecular Dynamics Package. *J Chem Theory Comput* **2013,** *9* (9), 4131-4139.

82. Zwanzig, R. W., High Temperature Equation of State by A Perturbation Method. *I.nonpolar Gases J.chem.phys* **1954,** *22* (8), 1420-1426.

83. Fenwick, M. K.; Escobedo, F. A., On the use of Bennett's acceptance ratio method in multi-canonical-type simulations. *J Chem Phys* **2004,** *120* (7), 3066-74.

84. Tan, Z., On a Likelihood Approach for Monte Carlo Integration. *J. Am. Stat. Assoc.* **2004,** *99* (December), 1027-1036.

85. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008,** *129* (12), 124105.

86. Jarzynski, C., A Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997,** *78* (14), 2690-2693.

87. Jarzynski, C., Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E* **1997,** *56* (5), 5018-5035.

88. Crooks, G. E., Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *J. Stat. Phys.* **1998,** *90* (5), 1481-1487.

89. Miller, B. R.; Mcgee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: an efficient program for end-state free energy calculations. J Chem Theory Comput 8:3314-3321. *J. Chem. Theory Comput.* **2012,** *8* (9), 3314-3321.

90. Ramos, J. L.; Duque, E.; Gallegos, M.-T.; Godoy, P.; Ramos-Gonzalez, M. I.; Rojas, A.; Terán, W.; Segura, A., Mechanisms of solvent tolerance in gram-negative bacteria. *Annual Reviews in Microbiology* **2002,** *56* (1), 743-768.

91. Nikaido, H. In *Preventing drug access to targets: cell surface permeability barriers and active efflux in bacteria*, Semin. Cell Dev. Biol., Elsevier: 2001; pp 215-223.

92. Newberry, K. J.; Huffman, J. L.; Miller, M. C.; Vazquez-Laslop, N.; Neyfakh, A. A.; Brennan, R. G., Structures of BmrR-drug complexes reveal a rigid multidrug binding pocket and transcription activation through tyrosine expulsion. *J. Biol. Chem.* **2008,** *283* (39), 26795-26804.

93. Sun, J.; Deng, Z.; Yan, A., Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. *Biochem. Biophys. Res. Commun.* **2014,** *453* (2), 254-267.

94. Moraleda-Muñoz, A.; Pérez, J.; Extremera, A. L.; Muñoz-Dorado, J., Differential regulation of six heavy metal efflux systems in the response of Myxococcus xanthus to copper. *Appl. Environ. Microbiol.* **2010,** *76* (18), 6069-6076.

95. Tseng, T.-T.; Gratwick, K. S.; Kollman, J.; Park, D.; Nies, D. H.; Goffeau, A.; Saier Jr, M. H., The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *J. Mol. Microbiol. Biotechnol.* **1999,** *1* (1), 107-125.

96. Terán, W.; Felipe, A.; Segura, A.; Rojas, A.; Ramos, J.-L.; Gallegos, M.-T., Antibiotic-dependent induction of Pseudomonas putida DOT-T1E TtgABC efflux pump is mediated by the drug binding repressor TtgR. *Antimicrob. Agents*

Chemother. **2003,** *47* (10), 3067-3072.

97.   María-Eugenia, G.; María-Trinidad, G.; Ramos, J. L.; Tino, K., Different modes of binding of mono- and biaromatic effectors to the transcriptional regulator TTGV: role in differential derepression from its cognate operator. *J. Biol. Chem.* **2007,** *282* (22), 16308.

98.   Ramos, J. L.; Duque, E.; Huertas, M.-J.; HaïDour, A., Isolation and expansion of the catabolic potential of a Pseudomonas putida strain able to grow in the presence of high concentrations of aromatic hydrocarbons. *J. Bacteriol.* **1995,** *177* (14), 3911-3916.

99.   Rojas, A.; Duque, E.; Mosqueda, G.; Golden, G.; Hurtado, A.; Ramos, J. L.; Segura, A., Three Efflux Pumps Are Required To Provide Efficient Tolerance to Toluene in Pseudomonas putidaDOT-T1E. *J. Bacteriol.* **2001,** *183* (13), 3967-3973.

100. Itou, H.; Watanabe, N.; Yao, M.; Shirakihara, Y.; Tanaka, I., Crystal structures of the multidrug binding repressor Corynebacterium glutamicum CgmR in complex with inducers and with an operator. *Journal of molecular biology* **2010,** *403* (2), 174-184.

101. Willems, A.; Tahlan, K.; Taguchi, T.; Zhang, K.; Lee, Z.; Ichinose, K.; Junop, M.; Nodwell, J., Crystal structures of the Streptomyces coelicolor TetR-like protein ActR alone and in complex with actinorhodin or the actinorhodin biosynthetic precursor (S)-DNPA. *Journal of molecular biology* **2008,** *376* (5), 1377-1387.

102. Manjasetty, B. A.; Halavaty, A. S.; Luan, C.-H.; Osipiuk, J.; Mulligan, R.; Kwon, K.; Anderson, W. F.; Joachimiak, A., Loop-to-helix transition in the structure of multidrug regulator AcrR at the entrance of the drug-binding cavity. *Journal of structural biology* **2016,** *194* (1), 18-28.

103. Le, T. B.; Stevenson, C. E.; Fiedler, H.-P.; Maxwell, A.; Lawson, D. M.; Buttner, M. J., Structures of the TetR-like simocyclinone efflux pump repressor, SimR, and the mechanism of ligand-mediated derepression. *Journal of molecular biology* **2011,** *408* (1), 40-56.

104. Ramos, J. L.; Martínez-Bueno, M.; Molina-Henares, A. J.; Terán, W.; Watanabe, K.; Zhang, X.; Gallegos, M. T.; Brennan, R.; Tobes, R., The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* **2005,** *69* (2), 326-356.

105. Schumacher, M. A.; Brennan, R. G., Deciphering the molecular basis of multidrug recognition: crystal structures of the Staphylococcus aureus multidrug binding transcription regulator QacR. *Res. Microbiol.* **2003,** *154* (2), 69-77.

106. Yu, Z.; Reichheld, S. E.; Savchenko, A.; Parkinson, J.; Davidson, A. R., A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *Journal of molecular biology* **2010,** *400* (4), 847-864.

107. Krell, T.; Terán, W.; Mayorga, O. L.; Rivas, G.; Jiménez, M.; Daniels, C.; Molina-Henares, A. J.; Martínez-Bueno, M.; Gallegos, M. T.; Ramos, J. L., Optimization of the Palindromic Order of the TtgR Operator Enhances Binding Cooperativity. *Journal of Molecular Biology* **2007,** *369* (5), 1188-1199.

108. Guazzaroni, M.-E.; Terán, W.; Zhang, X.; Gallegos, M.-T.; Ramos, J. L., TtgV bound to a complex operator site represses transcription of the promoter for the multidrug and solvent extrusion TtgGHI pump. *J. Bacteriol.* **2004,** *186* (10), 2921-2927.

109. Duque, E.; Segura, A.; Mosqueda, G.; Ramos, J. L., Global and cognate regulators control the expression of the organic solvent efflux pumps TtgABC and TtgDEF of Pseudomonas putida. *Mol. Microbiol.* **2001,** *39* (4), 1100-1106.

110. Ahmed, S.; Ning, J.; Cheng, G.; Ahmad, I.; Li, J.; Liu, M.; Wei, Q.; Iqbal, M.; Shabbir, M. A. B.; Yuan, Z., Receptor-based screening assays for the detection of antibiotics residues – A review. *Talanta* **2017,** *166*, 176-186.

111. Alguel, Y.; Meng, C.; Terán, W.; Krell, T.; Ramos, J. L.; Gallegos, M.-T.; Zhang, X., Crystal structures of multidrug binding protein TtgR in complex with antibiotics and plant antimicrobials. *Journal of molecular biology* **2007,** *369* (3), 829-840.

112. Fiser, A.; Do, R. K. G., Modeling of loops in protein structures. *Protein Sci.* **2000,** *9* (9), 1753-1773.

113. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002,** *23* (16), 1623-41.

114. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25*, 1157–1173.

115. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

116. Price, D. J.; Brooks III, C. L., A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004,** *121* (20), 10096-10103.

117. Joung, I. S.; Cheatham III, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The journal of physical chemistry B* **2008,** *112* (30), 9020-9041.

118. Joung, I. S.; Cheatham, T. E., Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *The Journal of Physical Chemistry B* **2009,** *113* (40), 13279-13290.

119. Pearlman, D. A.; Kollman, P. A., The lag between the Hamiltonian and the system configuration in free energy perturbation calculations. *Journal of Chemical Physics* **1989,** *91* (12), 7831-7839.

120. Ravishanker, G.; Mezei, M.; Beveridge, D. L., Conformational stability and flexibility of the ala dipeptide in free space and water: Monte Carlo computer simulation studies †. *J. Comput. Chem.* **1986,** *7* (3), 345–348.

121. Cross, A. J., Influence of hamiltonian parameterization on convergence of kirkwood free energy calculations. *Chem. Phys. Lett.* **1986,** *128* (2), 198-202.

122. Steinbrecher, T.; Mobley, D. L.; Case, D. A., Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J Chem Phys* **2007,** *127* (21), 214108.

123. Zacharias, M.; Straatsma, T. P.; Mccammon, J. A., Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994,** *100* (12), 9025-9031.

124. Beutler, T. C.; Mark, A. E.; Schaik, R. C. V.; Gerber, P. R.; Gunsteren, W. F. V., Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994,** *222* (6), 529-539.

125. Pitera, J. W.; van Gunsteren, W. F., A Comparison of Non-Bonded Scaling Approaches for Free Energy Calculations. *Molecular Simulation* **2002,** *28* (1-2), 45-65.

126. Bitetti, R., Generalized ensembles serve to improve the convergence of free energy simulations. *Chem. Phys. Lett.* **2003,** *377* (5-6), 633-641.

127. Chipot, C.; Rozanska, X.; Dixit, S. B., Can free energy calculations be fast and accurate at the same time? Binding of low-affinity, non-peptide inhibitors to the SH2 domain of the src protein. *J. Comput. Aided Mol. Des.* **2005,** *19* (11), 765-70.

128. Fowler, P. W.; Jha, S.; Coveney, P. V., Grid-based steered thermodynamic integration accelerates the calculation of binding free energies. *Philosophical Transactions of the Royal Society A Mathematical Physical & Engineering Sciences* **2005,** *363* (1833), 1999-2015.

129. Levitt, M., Protein folding by restrained energy minimization and molecular dynamics. *Journal of Molecular Biology* **1983,** *170* (3), 723-64.

130. Klimovich, P. V.; Shirts, M. R.; Mobley, D. L., Guidelines for the Analysis of Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2015,** *29* (5), 397-411.

131. Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A., Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2007,** *3* (1), 26--41.

132. Bruckner, S.; Boresch, S., Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **2011,** *32* (7), 1320–1333.

133. Bennett, C. H., Efficient Estimation of Free Energy Differences from Monte Carlo data. *J. Comput. Phys.* **1976,** *22* (2), 245-268.

134. Paliwal, H.; Shirts, M. R., A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods. *J Chem Theory Comput* **2011,** *7* (12), 4115-34.

135. Shirts, M. R.; Pande, V. S., Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.* **2005,** *122* (14), 144107-144107.

136. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins Structure Function & Bioinformatics* **2004,** *55* (55), 383-94.

137. Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A., Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.* **2004,** *25* (2), 265-84.

138. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of The Cartesian Equations of Motion of A System with Constraints: Molecular Dynamics of n -alkanes. *J. Comput. Phys.* **1977,** *23* (3), 327-341.

139. Miyamoto, S.; Kollman, P. A., Settle: An Analytical Version of The SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992,** *13* (8), 952-962.

140. Pastor, R. W.; Brooks, B. R.; Szabo, A., An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Molecular Physics* **1988,** *65* (6), 1409-1419.

141. York, D. M.; Darden, T. A.; Pedersen, L. G., The Effect of Long-range Electrostatic Interactions in Simulations of

Macromolecular Crystals: A Comparison of The Ewald and Truncated List Methods. *J. Chem. Phys.* **1993**, *99* (10), 8345-8348.

142. Lafont, V.; Armstrong, A. A.; Ohtaka, H.; Kiso, Y.; Mario Amzel, L.; Freire, E., Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chemical biology & drug design* **2007,** *69* (6), 413-422.

143. Brandt, T.; Holzmann, N.; Muley, L.; Khayat, M.; Wegscheid-Gerlach, C.; Baum, B.; Heine, A.; Hangauer, D.; Klebe, G., Congeneric but still distinct: how closely related trypsin ligands exhibit different thermodynamic and structural properties. *Journal of molecular biology* **2011,** *405* (5), 1170-1187.

144. Talhout, R.; Villa, A.; Mark, A. E.; Engberts, J. B., Understanding binding affinity: a combined isothermal titration calorimetry/molecular dynamics study of the binding of a series of hydrophobically modified benzamidinium chloride inhibitors to trypsin. *J. Am. Chem. Soc.* **2003,** *125* (35), 10570-10579.

145. Makhatadze, G. I.; Privalov, P. L., Hydration effects in protein unfolding. *Biophys. Chem.* **1994,** *51* (2-3), 291-309.

146. Cornish-Bowden, A., Enthalpy—entropy compensation: a phantom phenomenon. *J. Biosci. (Bangalore)* **2002,** *27* (2), 121-126.

147. Cooper, A., Microcalorimetry of heat capacity and volumetric changes in biomolecular interactions—the link to solvation? *J. Therm. Anal. Calorim.* **2011,** *104* (1), 69-73.

148. Krug, R.; Hunter, W.; Grieger, R., Statistical interpretation of enthalpy–entropy compensation. *Nature* **1976,** *261* (5561), 566.

149. Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nöllmann, M., Heat does not come in different colours: entropy–enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys. Chem.* **2001,** *93* (2-3), 215-230.

150. Lumry, R.; Rajender, S., Enthalpy–entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous properly of water. *Biopolymers: Original Research on Biomolecules* **1970,** *9* (10), 1125-1227.

151. Grunwald, E.; Steel, C., Solvent reorganization and thermodynamic enthalpy-entropy compensation. *J. Am. Chem. Soc.* **1995,** *117* (21), 5687-5692.

152. Lee, B.; Graziano, G., A two-state model of hydrophobic hydration that produces compensating enthalpy and entropy changes. *J. Am. Chem. Soc.* **1996,** *118* (22), 5163-5168.

153. Breiten, B.; Lockett, M. R.; Sherman, W.; Fujita, S.; Al-Sayah, M.; Lange, H.; Bowers, C. M.; Heroux, A.; Krilov, G.; Whitesides, G. M., Water networks contribute to enthalpy/entropy compensation in protein–ligand binding. *J. Am. Chem. Soc.* **2013,** *135* (41), 15579-15584.

Table 1. Free energy differences computed via various methods (ΔG_{TI} , ΔG_{BAR} , ΔG_{MBAR} , $\Delta G_{MM/PBSA}$, $\Delta G_{MM/GBSA}$) as well as the corresponding experimental values (ΔG_{EXP} , $\Delta\Delta G_{EXP}$). All values are in kcal/mol. The mean signed error (MSE), the mean absolute error (MAE), the root-mean-squared error (RMSE), and Pearlman's predictive index (PI) serve as quality measurements. The ligand of AGI is used as reference and thus its values are all zeros.

Ligand	ΔG_{EXP}	$\Delta\Delta G_{EXP}$	$\Delta\Delta G_{TI}$	SD	$\Delta\Delta G_{BAR}$	SD	$\Delta\Delta G_{MBAR}$	SD	$\Delta\Delta G_{CE}$	SD	$\Delta\Delta G_{MM/PBSA}$	SD	$\Delta\Delta G_{MM/GBSA}$	SD
1NP	-5.3	1.4	4.0	3.3	1.8	0.7	1.5	0.8	1.9	0.8	-0.6	0.2	11.6	0.2
2NP	-5.8	0.8	2.0	3.5	-0.0	0.8	-0.1	0.9	0.8	0.9	2.0	0.2	12.5	0.2
27K	-6.5	0.2	2.0	3.1	0.4	0.7	0.5	0.8	0.3	0.7	-1.3	0.3	5.6	0.3
CLM	-6.9	-0.2	1.0	3.4	-0.1	0.7	-0.1	0.9	-0.7	0.8	9.7	0.3	5.4	0.3
G50	-6.9	-0.2	1.6	2.2	0.1	0.5	0.2	0.5	0.1	0.4	-2.6	0.3	-2.4	0.3
AGI	-6.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CUE	-6.6	0.1	-0.6	1.3	-0.7	0.3	-1.0	0.3	-0.6	0.3	6.4	0.3	5.3	0.2
LU2	-7.1	-0.5	1.0	1.2	0.3	0.2	0.0	0.3	0.4	0.4	9.3	0.3	7.0	0.2
QUE	-8.1	-1.4	-0.2	2.1	-1.1	0.5	-1.0	0.5	-1.0	0.6	11.2	0.3	5.1	0.2
RMSE			1.5		0.5		0.5		0.5		6.7		6.9	
MSE			-1.2		-0.1		0.0		-0.1		-3.8		-5.5	
MAE			1.3		0.4		0.4		0.4		5.1		6.0	
PI			0.7		0.7		0.6		0.8		-0.6		0.6	

Table 2. Free energy decomposition into enthalpic and entropic contributions from MBAR reweighting for transformations from the reference ligand AGI to another ligand in ligand-only and protein-ligand systems. The difference between the enthalpy change for mutations in protein-ligand complex and ligand-only system, $\Delta\Delta H = \Delta H_{complex} - \Delta H_{ligands}$, gives the enthalpic contribution to the difference between binding affinities, and the entropic contribution $-T\Delta\Delta S = -T\Delta S_{complex} + T\Delta S_{ligands}$ can be obtained similarly. The free energy difference in each transformation in Table S1 can be calculated with the enthalpic and entropic components in this table via $\Delta G = \Delta H - T\Delta S$. The relationship between the data in this table and those in Table 1 is $\Delta\Delta H - T\Delta\Delta S = \Delta\Delta G$. All values are in kcal/mol. The reference ligand is AGI and thus all its values are zeros.

Transformation	ligands				complex				complex-ligands			
	$\Delta H_{ligands}$	SD	$-T\Delta S_{ligands}$	SD	$\Delta H_{complex}$	SD	$-T\Delta S_{complex}$	SD	$\Delta\Delta H$	SD	$-T\Delta\Delta S$	SD
27K-> 1NP	3.2	1.9	-2.9	1.9	33.9	12.1	-32.4	12.1	30.6	12.2	-29.5	12.2
1NP->2NP	-3.3	1.9	-0.8	1.9	0.0	19.9	-5.8	19.9	3.3	20.0	-4.9	20.0
27K->G50	-29.6	9.1	22.0	9.1	16.8	29.6	-24.6	29.6	46.3	30.9	-46.6	30.9
27K->CLM	-25.7	3.6	14.6	3.6	30.2	34.2	-41.8	34.2	55.9	34.3	-56.4	34.3
AGI->G50	-25.8	9.3	12.2	9.3	48.9	24.0	-62.3	24.0	74.7	25.7	-74.5	25.7
AGI->CUE	-22.3	2.1	-3.2	2.1	17.6	12.1	-44.6	12.1	39.9	12.2	-41.3	12.3
AGI->LU2	-13.1	3.7	5.0	3.7	-9.2	12.2	1.1	12.2	3.9	12.8	-3.9	12.8
LU2->QUE	-12.0	5.2	13.1	5.2	-49.9	24.5	50.2	24.5	-37.9	25.1	37.1	25.1
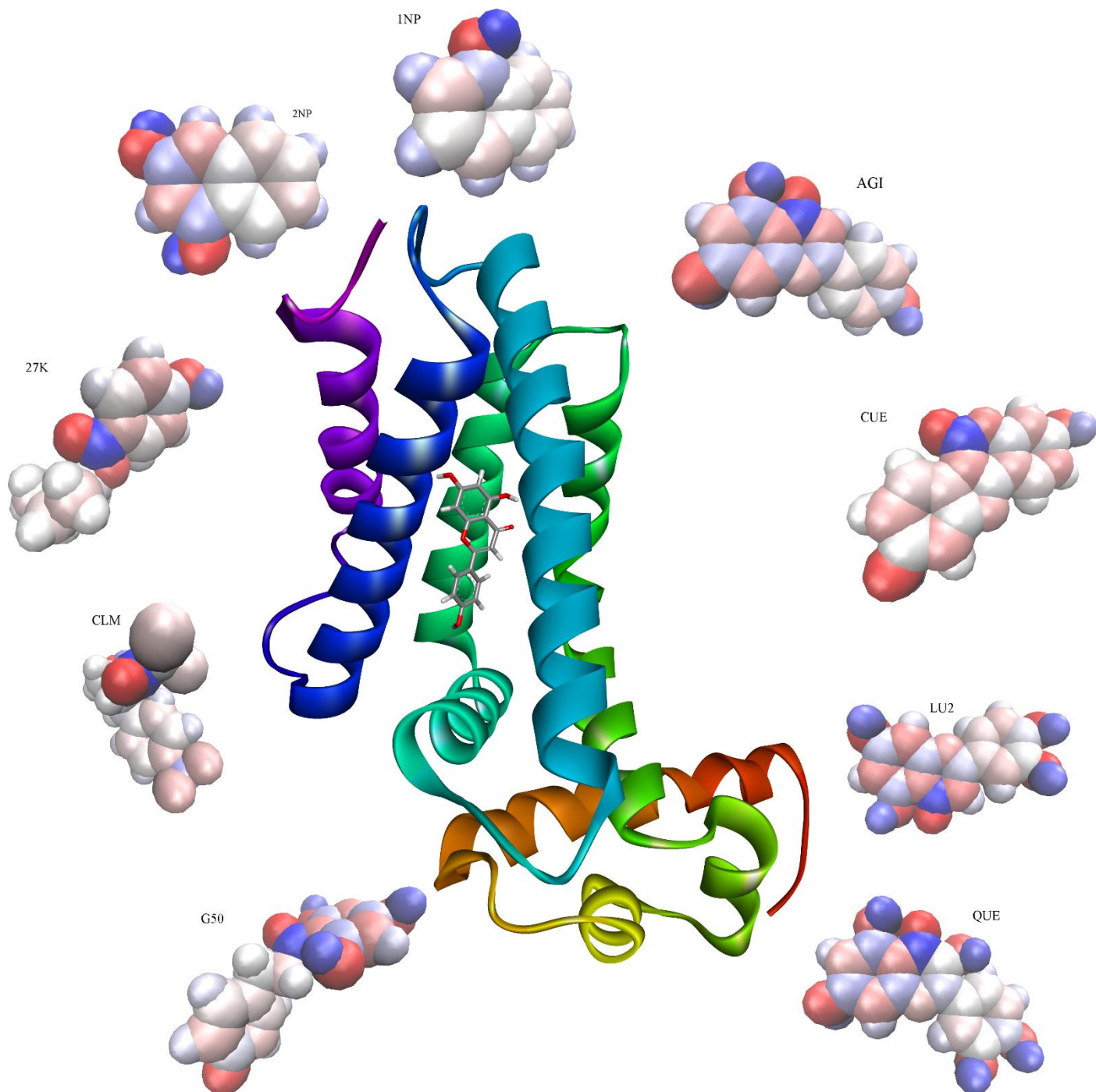
Figure 1. An illustration of ligand binding to transcriptional regulator TtgR is given at the center. The ligand AGI is used to generate the picture. The charge maps of the 9 ligands binding to TtgR are shown around the complex. Their 2D chemical structures are shown in Figure 2.
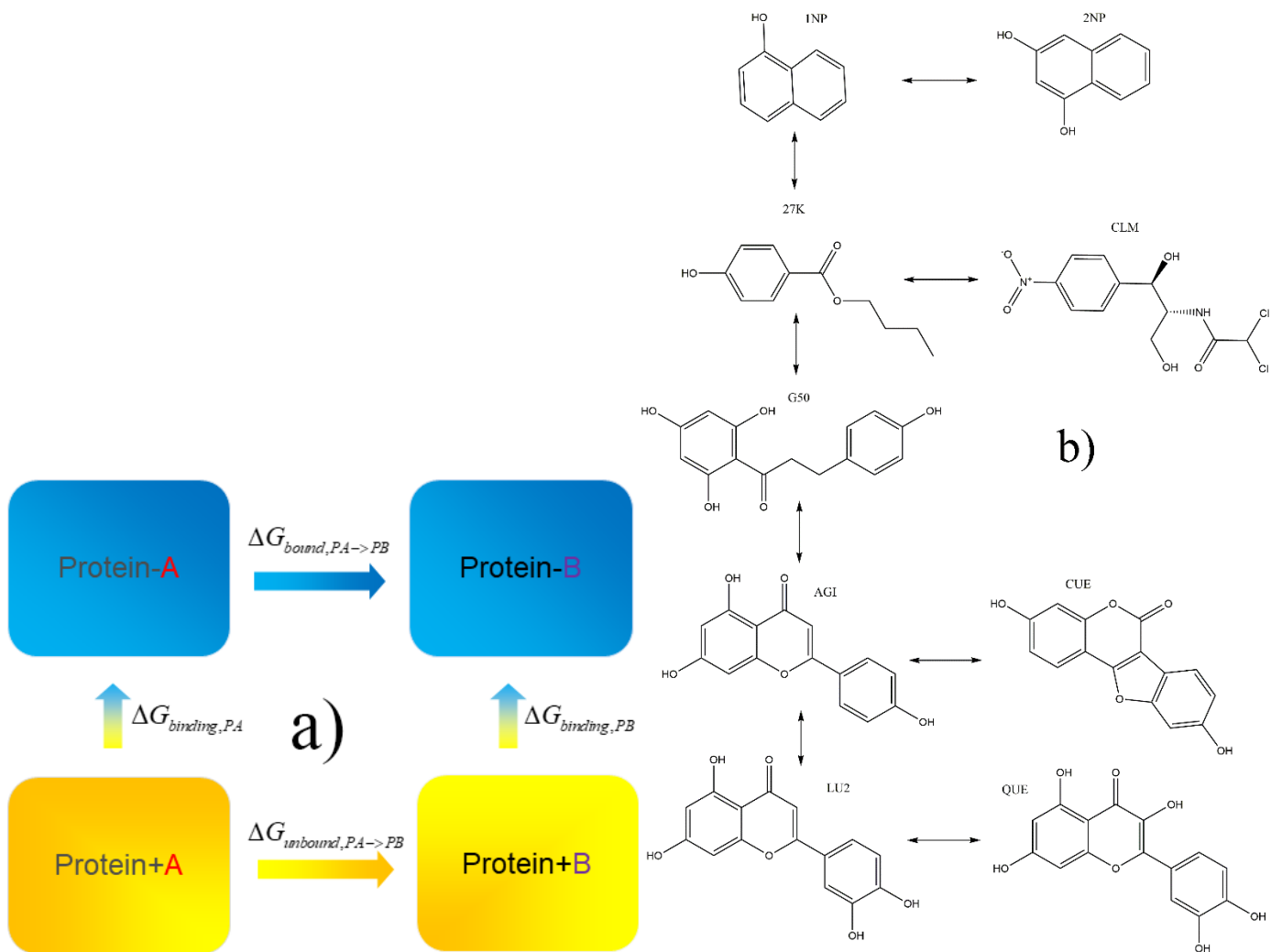
Figure 2. a) Thermodynamic cycle used to compute the relative binding affinities between different ligands. A and B denote two ligands binding at the same site, Protein-A is the protein-ligand complex and Protein+A refers to unbound protein and ligand. $\Delta G_{binding,PA}$ and $\Delta G_{binding,PB}$ denote the binding affinity between protein and ligand A and that between protein and ligand B, respectively. $\Delta G_{unbound,PA->PB}$ is the free energy difference between unbound protein and B in solution and unbound protein and A, and $\Delta G_{bound,PA->PB}$ represents the free energy difference between solvated protein-B complex and solvated protein-A complex. The relative binding affinity between protein-B and protein-A complexes can be obtained from the double free energy difference $\Delta G_{bound,PA->PB} - \Delta G_{unbound,PA->PB} = \Delta G_{binding,PB} - \Delta G_{binding,PA}$, according to the closure of the thermodynamic cycle. b) The transformation map used to calculate the relative binding affinities between different ligands. The transformations in the map are defined according to similarities in the structures. Such setting minimizes the magnitude of alchemical perturbation and thus enhances the convergence.
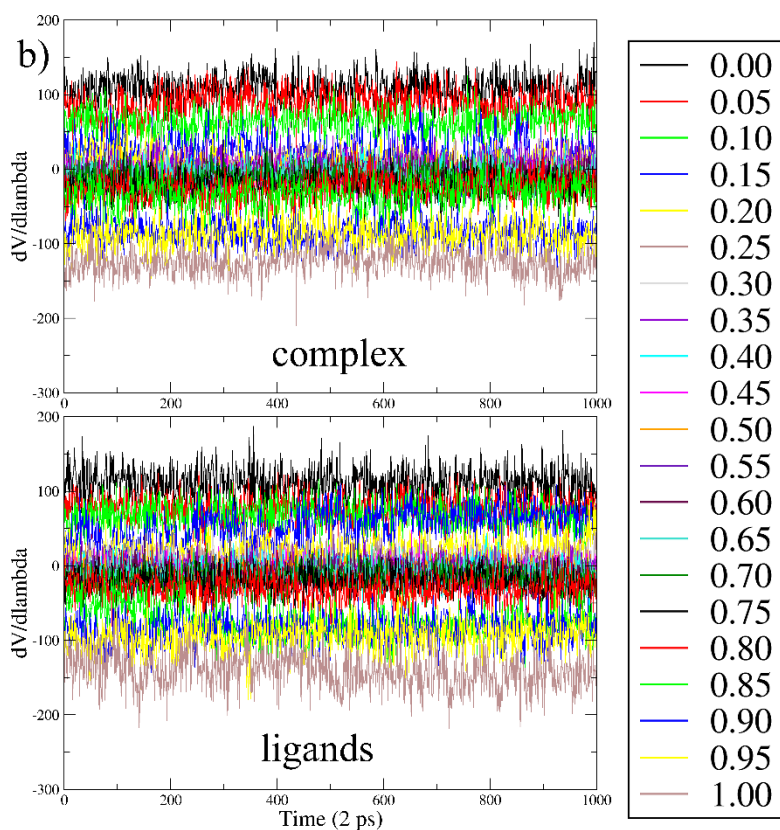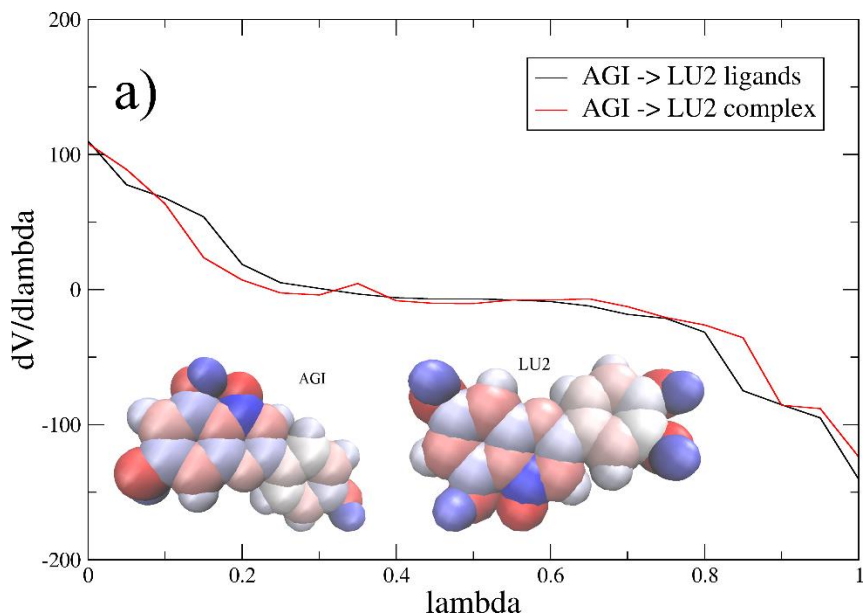
Figure 3. a) $\left\langle \dfrac{\partial U}{\partial \lambda} \right\rangle_{\lambda_i}$ as a function of λ in protein-ligand system and ligand-only system for transformations from AGI to LU2. The distribution of the atomic charges of the ligands are also given. b) Timeseries of $\left. \dfrac{\partial U}{\partial \lambda} \right|_{\lambda_i}$. The time-evolution of $\left. \dfrac{\partial U}{\partial \lambda} \right|_{\lambda_i}$ does not show much fluctuation and no multistate behavior is observed. The unit of $\left\langle \dfrac{\partial U}{\partial \lambda} \right\rangle_{\lambda_i}$ and $\left. \dfrac{\partial U}{\partial \lambda} \right|_{\lambda_i}$ is kcal/mol.
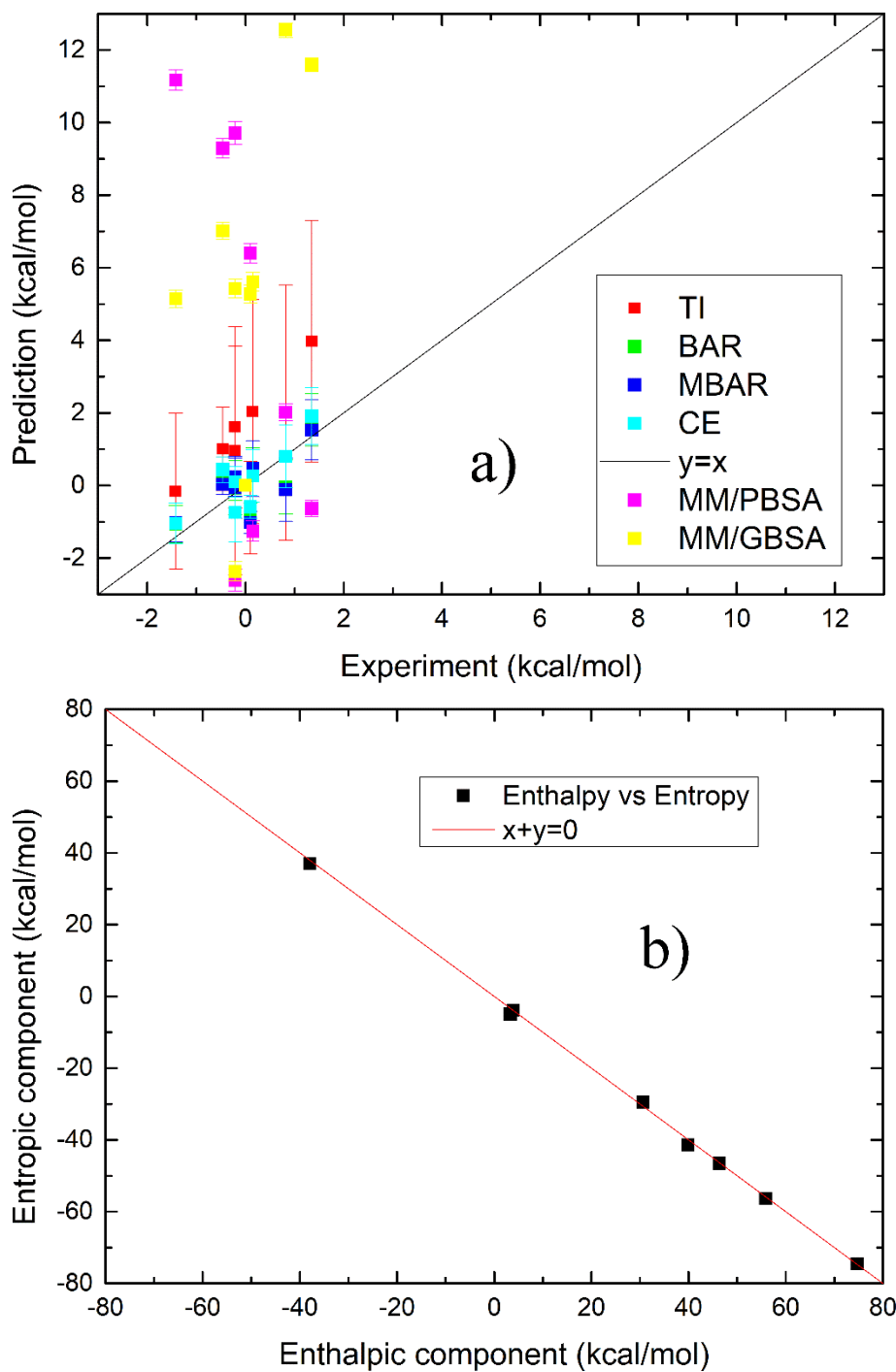
Figure 4. a) Comparison between calculated binding affinities and experimental values. We notice that both the statistical error and the bias of TI are relatively large compared with BAR and MBAR. The agreement between BAR/MBAR predictions and the experimental values is good. Compared with the alchemical method, MM/PBSA and MM/GBSA give worse predictions. b) Correlation between the enthalpic and entropic contributions of difference between binding affinities obtained from MBAR reweighting. The mutation induced enthalpic change depends linearly on the entropic term and their sums are almost zeros in all mutation cases, namely $x + y \approx 0$, indicating the existence of the entropy-enthalpy compensation.
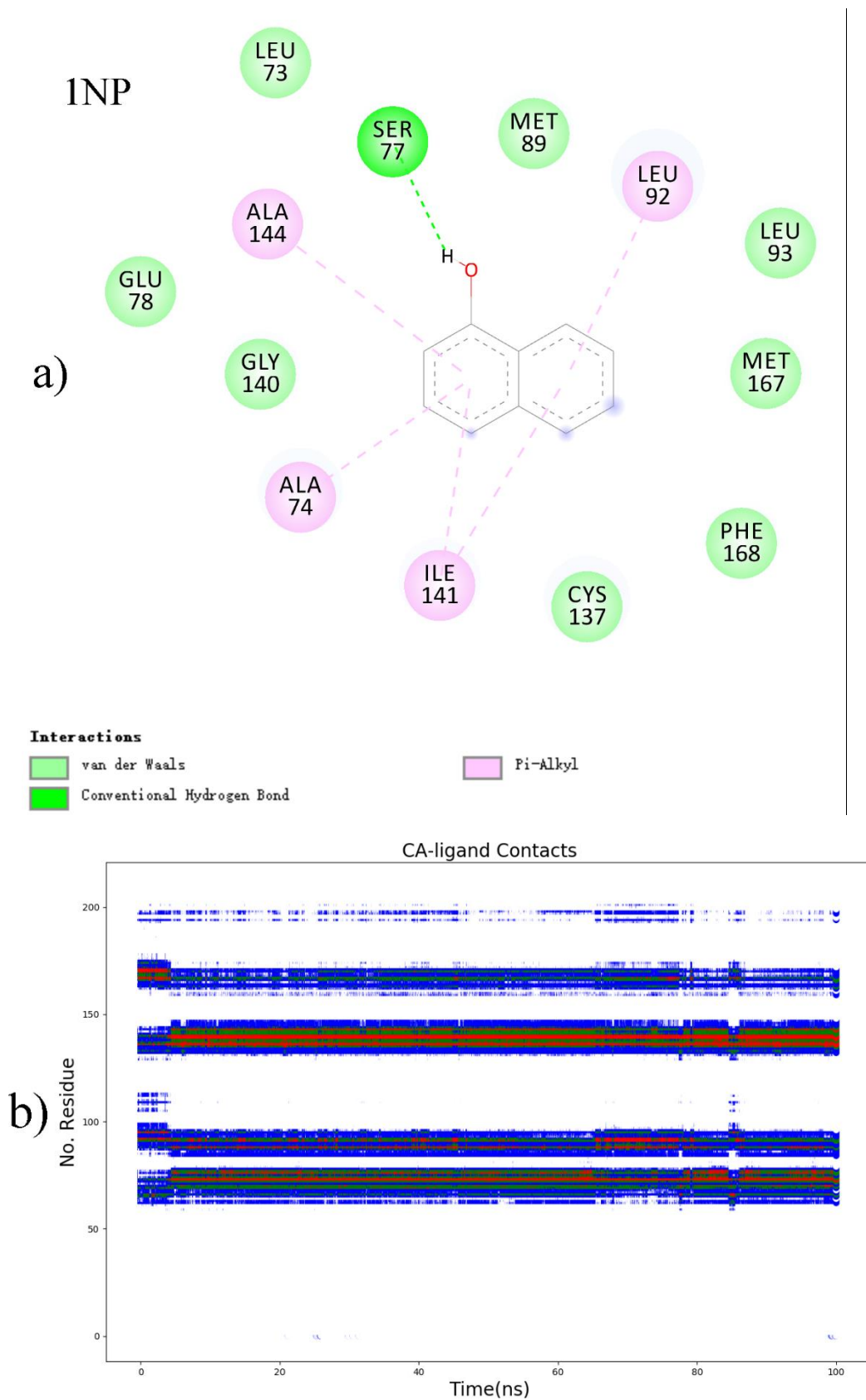
Figure 5. a) Interaction maps for protein-1NP complex obtained from equilibrated structures. b) Time evolution of number of contacts between alpha-C atoms of protein and the ligand. Red dots denote contacts larger than 10, green dots represent contact number between 5 and 10, blue ones are those larger than 1, and the other are represented by white dots.
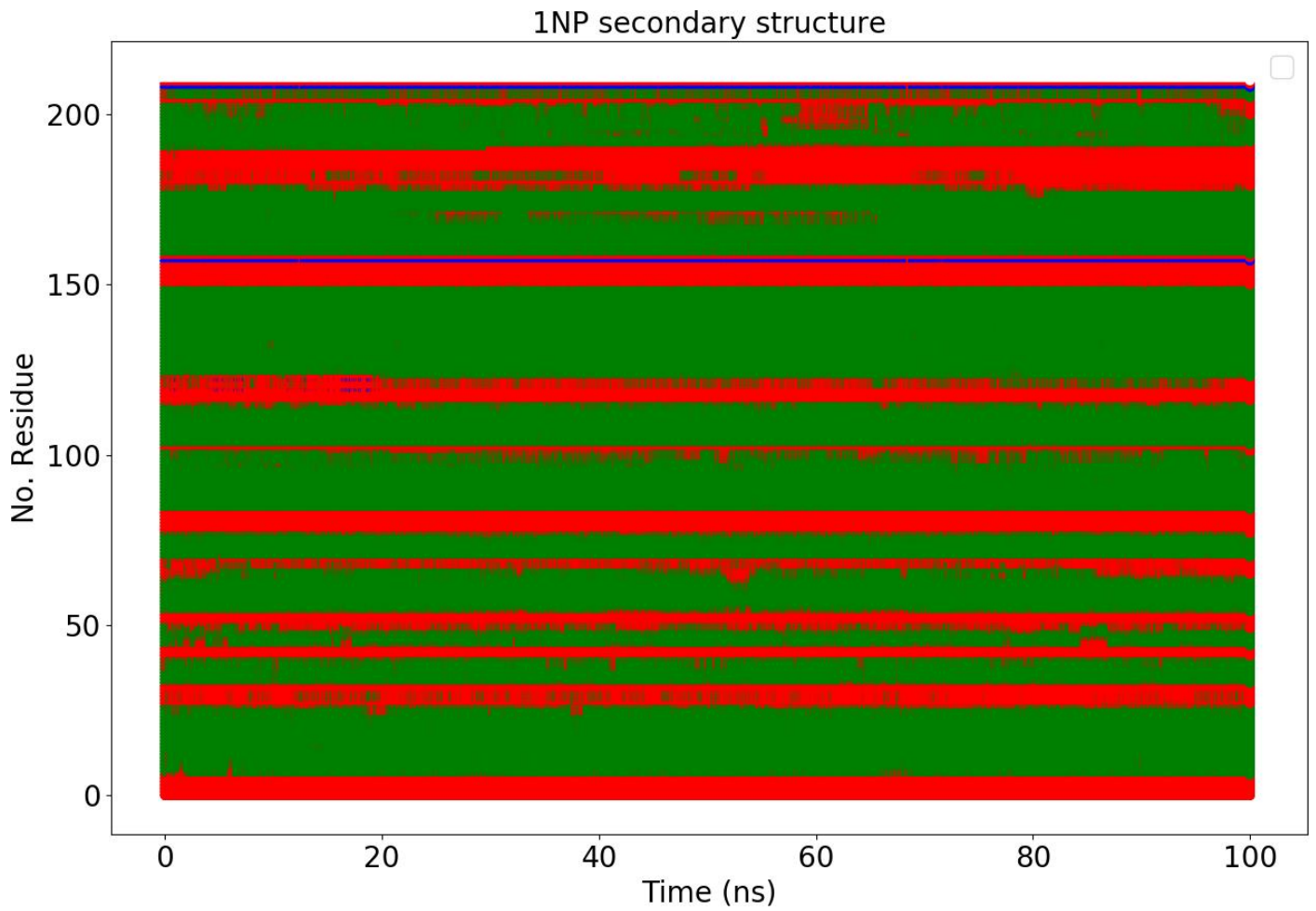
Figure 6. Time evolution of secondary structures of protein-1NP complex. In the plot of secondary structures, the helical structure is plotted with green dots, the beta component is plotted with red dots, and the coil region is represented by blue dots.
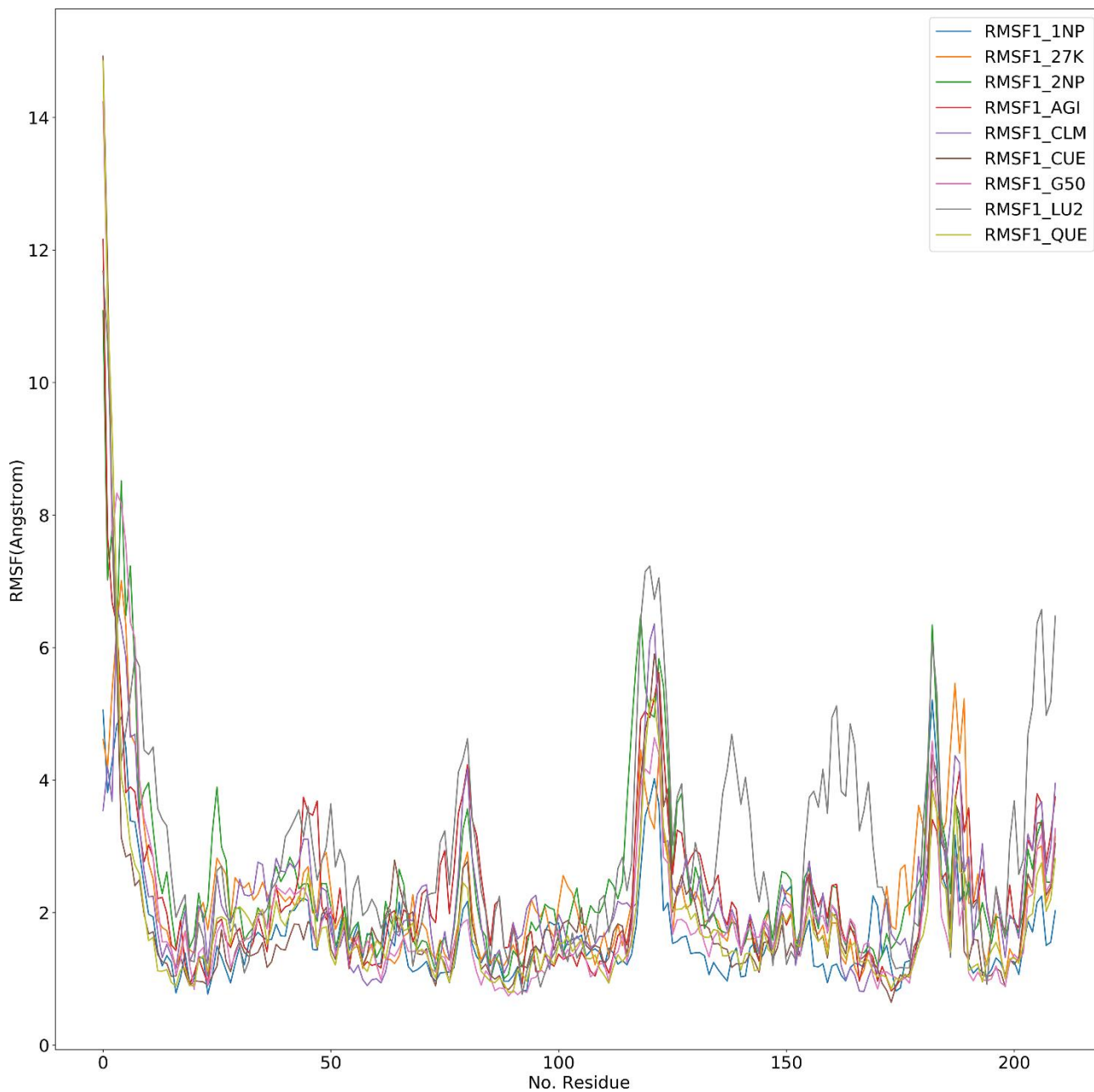
Figure 7. RMSF of C-α atoms of protein during conventional MD simulations.

Supporting Information: Theoretical understanding of the thermodynamics and interactions in transcriptional regulator TtgR-inhibitor binding from free energy simulation

Zhaoxi Sun[1,2*], Xiaohui Wang[1,3], and John Z.H.Zhang[1,4,5*]

[1]*State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

[2]*Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich, Jülich 52425, Germany*

[3]*Institute of Computational Science, Università della Svizzera italiana (USI), Via Giuseppe Buffi 13, CH-6900, Lugano, Ticino, Switzerland*

[4]*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

[5]*Department of Chemistry, New York University, NY, NY 10003, USA*

*To whom correspondence should be addressed: proszx@163.com, john.zhang@nyu.edu

Table S1. Detailed free energy differences from MM/PBSA and MM/GBSA. ΔH_{gas} is the gas-phase enthalpy change upon binding or the protein-ligand interaction energy. ΔG_{sol} is the solvation free energy and the subscripts of PB and GB denote the implicit solvent model used to calculate the solvation free energy. $\Delta\Delta G_{MM/PBSA}$ and $\Delta\Delta G_{MM/GBSA}$ are the relative binding affinity of ligands with AGI as the reference.

Ligand	ΔH_{gas}	SD	$\Delta G_{sol,PBSA}$	SD	$\Delta G_{sol,GBSA}$	SD	MM/PBSA	SD	MM/GBSA	SD	$\Delta\Delta G_{MM/PBSA}$	SD	$\Delta\Delta G_{MM/GBSA}$	SD
1NP	-35.51	0.07	27.91	0.06	13.09	0.05	-7.60	0.10	-22.41	0.09	-0.64	0.22	11.59	0.19
2NP	-34.27	0.08	29.32	0.08	12.81	0.06	-4.95	0.12	-21.46	0.10	2.01	0.23	12.55	0.20
27K	-41.46	0.16	33.24	0.12	13.07	0.11	-8.22	0.21	-28.40	0.20	-1.26	0.29	5.61	0.26
CLM	-47.17	0.16	49.92	0.19	18.59	0.10	2.75	0.25	-28.58	0.19	9.71	0.32	5.42	0.25
G50	-63.17	0.18	53.59	0.13	26.79	0.09	-9.58	0.23	-36.37	0.20	-2.62	0.30	-2.37	0.26
AGI	-50.80	0.14	43.83	0.14	16.79	0.08	-6.96	0.20	-34.00	0.17	0.00	0.28	0.00	0.24
CUE	-39.29	0.14	38.72	0.10	10.56	0.11	-0.56	0.18	-28.73	0.18	6.40	0.27	5.27	0.24
LU2	-47.88	0.12	50.21	0.12	20.88	0.10	2.33	0.18	-26.99	0.16	9.29	0.27	7.01	0.23
QUE	-51.71	0.14	55.93	0.15	22.85	0.11	4.21	0.21	-28.86	0.18	11.17	0.29	5.14	0.24

Table S2. Number of hydrogen bonds formed between the ligand and its surroundings in ligand-only and protein-ligand systems. N_{hbond} denotes the number of hydrogen bond and SD is the standard deviation of the mean of N_{hbond}.

Ligand	ligand		complex	
	N_{hbond}	SD	N_{hbond}	SD
1NP	5.08	0.08	1.99	0.04
2NP	6.20	0.08	3.30	0.04
27K	4.36	0.07	2.54	0.04
CLM	7.83	0.09	5.70	0.05
G50	10.01	0.11	6.65	0.06
AGI	6.19	0.08	5.18	0.05
CUE	6.17	0.09	5.35	0.05
LU2	6.48	0.09	6.21	0.06
QUE	6.92	0.10	7.43	0.07

Figure S1. The overlap matrices of alchemical transformation from AGI to LU2 in a) ligand-only system and b) protein-ligand system for overlap check. The plot clearly shows that the elements are large enough for reliable reweighting.
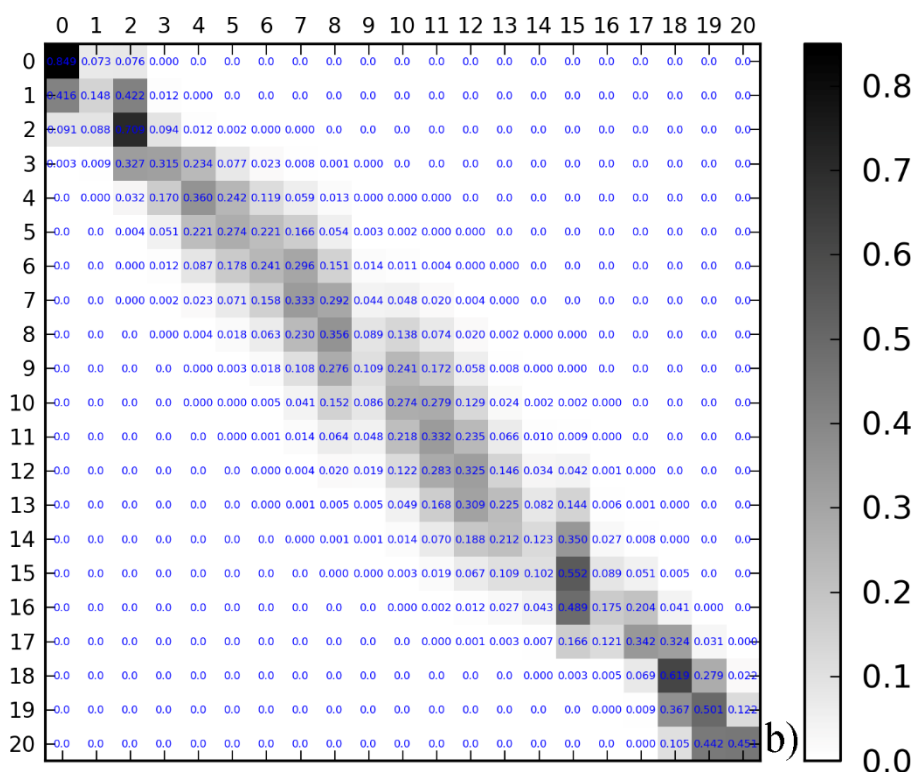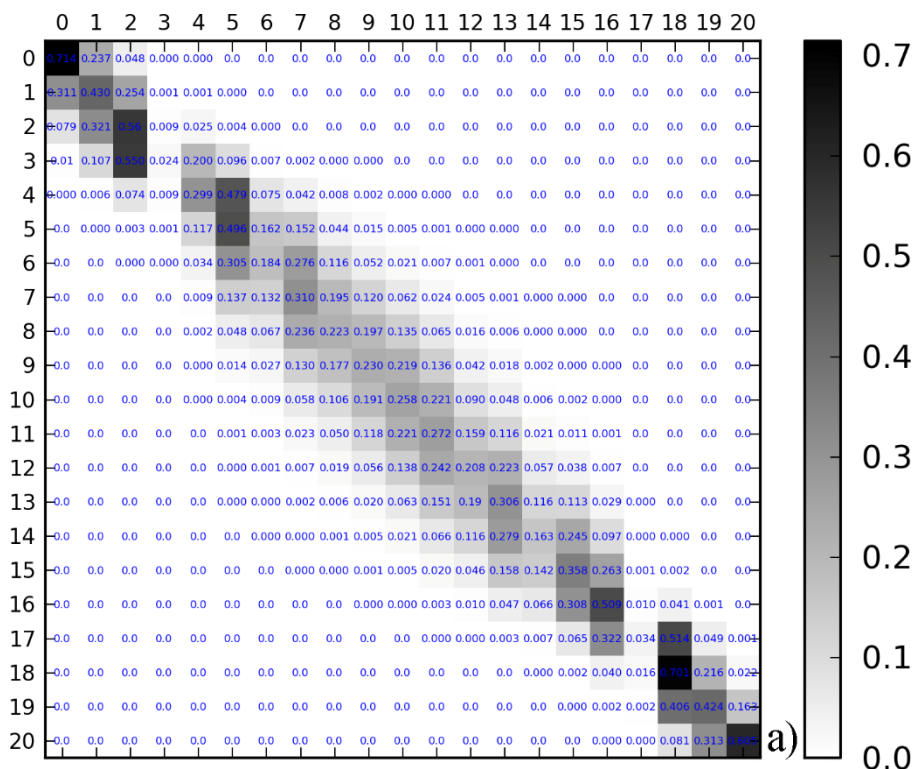
Figure S2. Hydrogen bonds formed between the ligand and its surroundings in protein-ligand system (left) and ligand-only system (right). The widths of the average lines are the standard error of the mean of the number of hydrogen bonds.
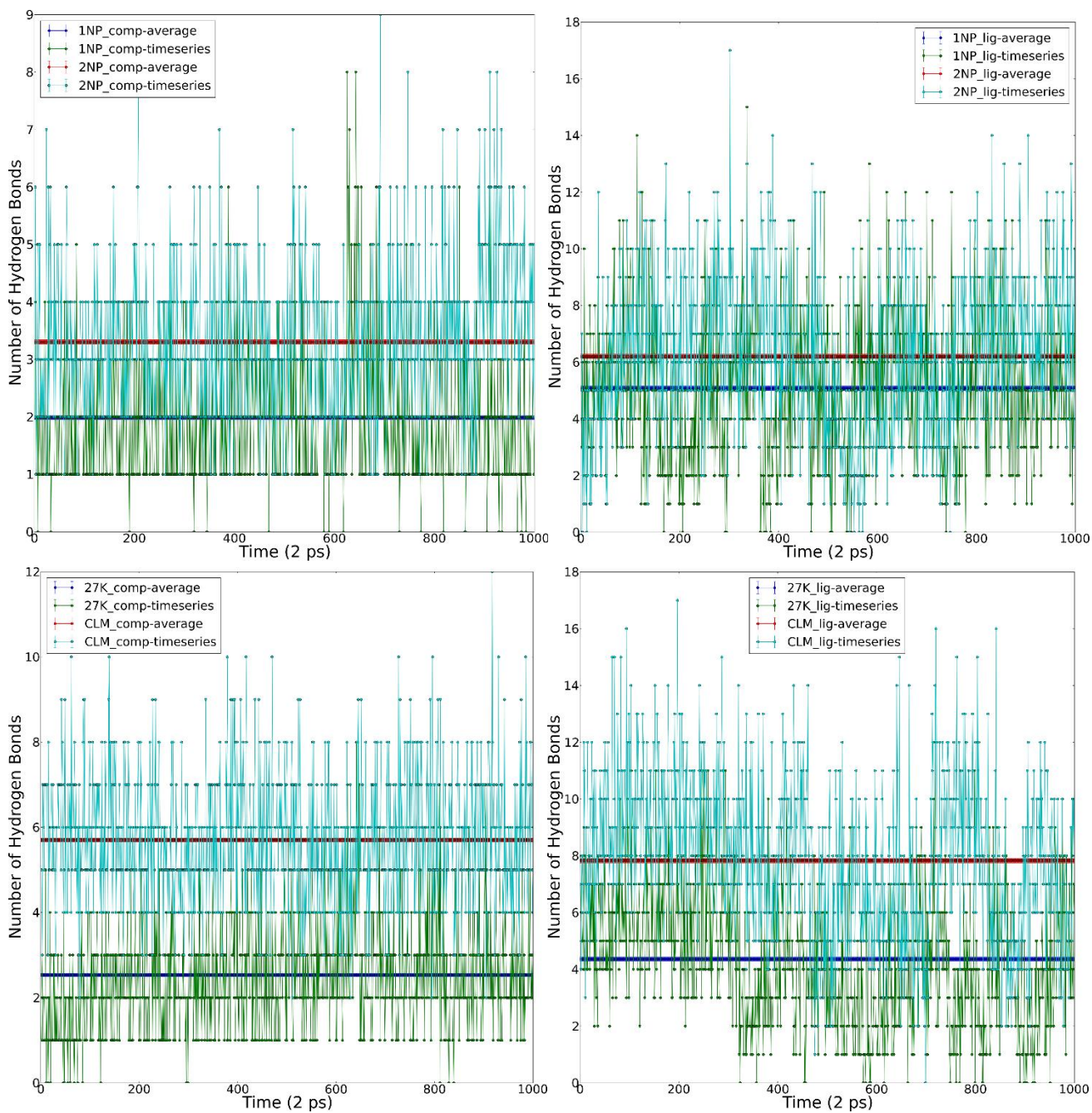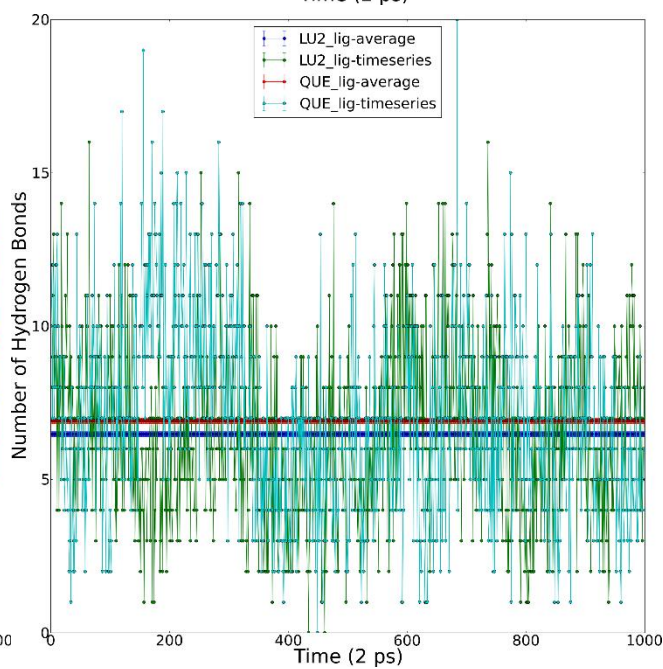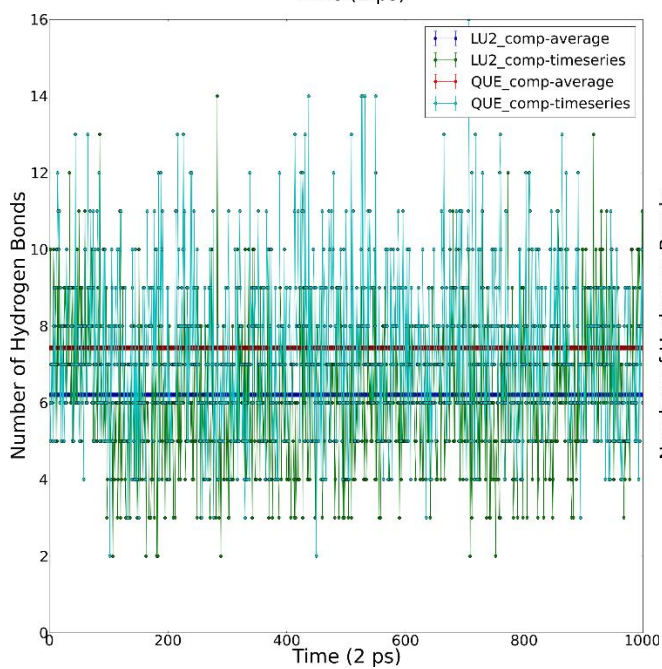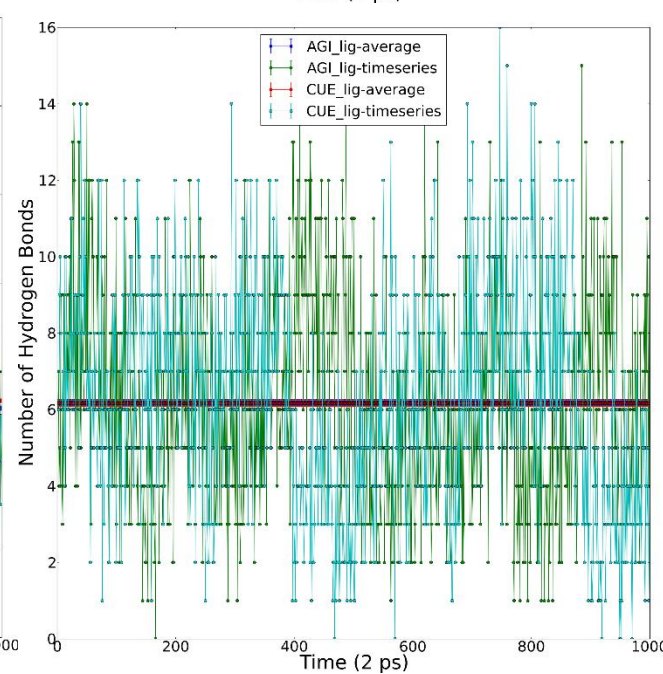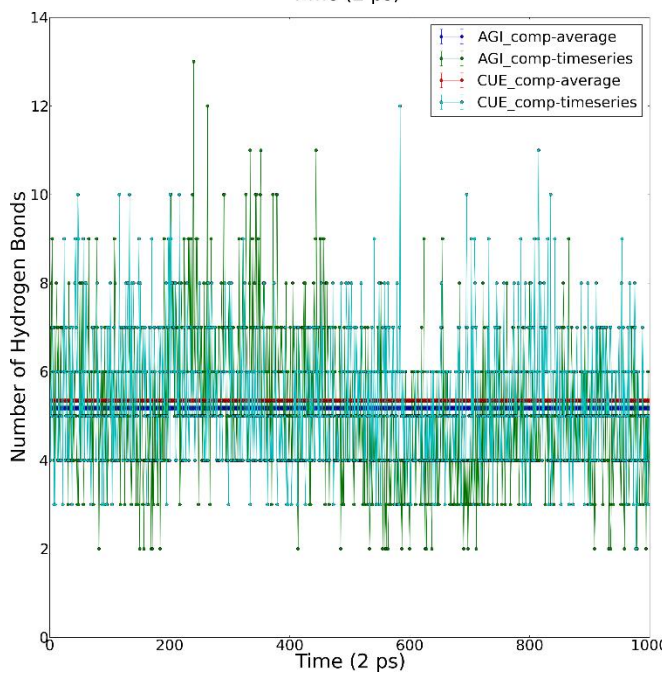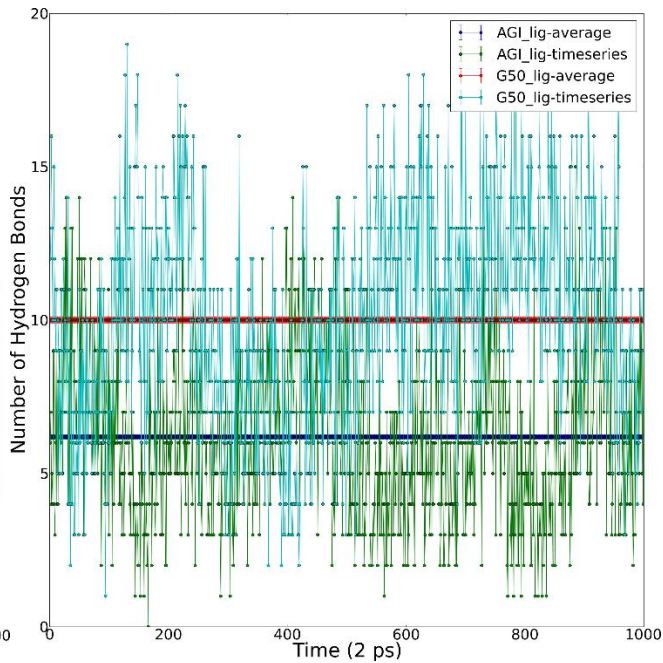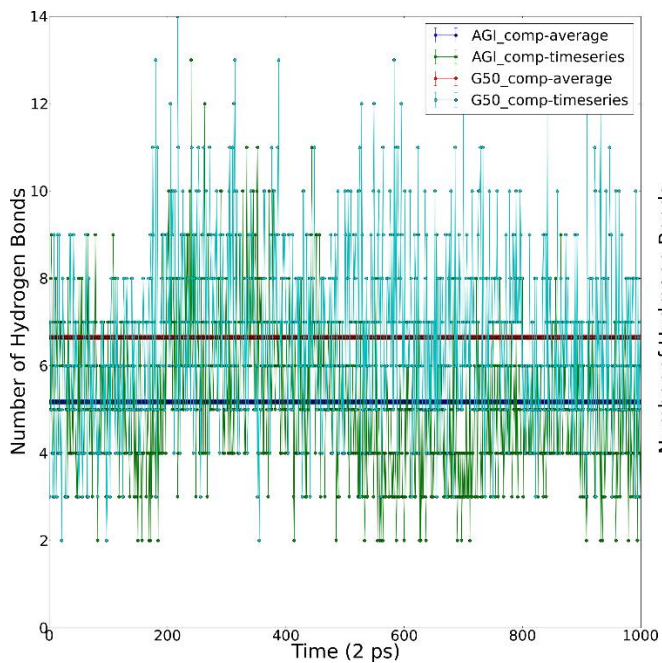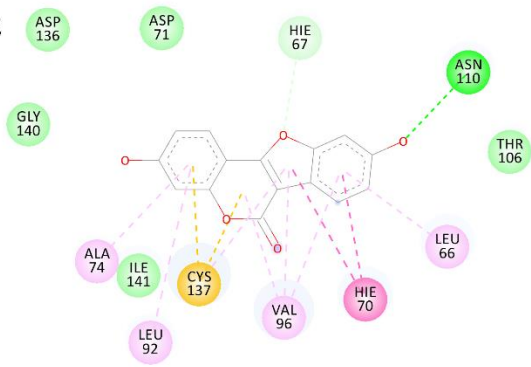
Figure S3. Interaction maps for protein-ligand complexes obtained from equilibrated structures.
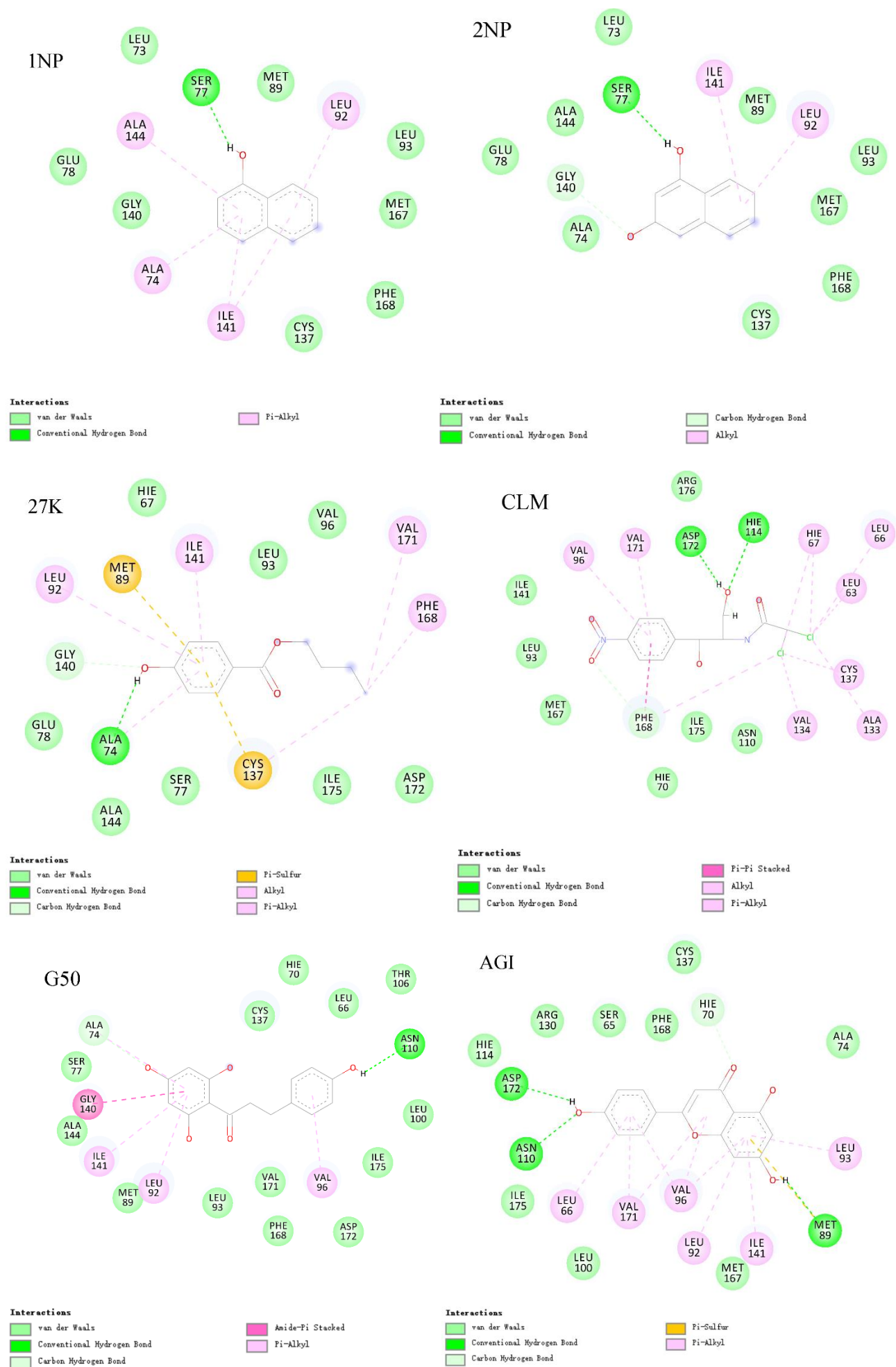
CUE

ASP 136　ASP 71　HIE 67　ASN 110

GLY 140　THR 106

ALA 74　ILE 141　CYS 137　VAL 96　HIE 70　LEU 66

LEU 92

Interactions

van der Waals	Pi-Sulfur
Conventional Hydrogen Bond	Pi-Pi T-shaped
Carbon Hydrogen Bond	Pi-Alkyl

LU2

MET 167　VAL 171

MET 89　LEU 93　PHE 168　ASP 172

ALA 144　LEU 92　VAL 96　ILE 175

GLY 140　ALA 74　LEU 100

GLU 78　ILE 141　CYS 137　LEU 66

HIE 67　HIE 70

THR 106　ASN 110

Interactions

van der Waals	Pi-Pi T-shaped
Conventional Hydrogen Bond	Pi-Alkyl
Pi-Sulfur	

QUE

LEU 93

ILE 141　CYS 137　LEU 92

VAL 96　PHE 168　LEU 113

ALA 74　HIE 114

ILE 175

ASP 71　ASN 110

HIE 70　HIE 67　LEU 66　LEU 100

THR 106
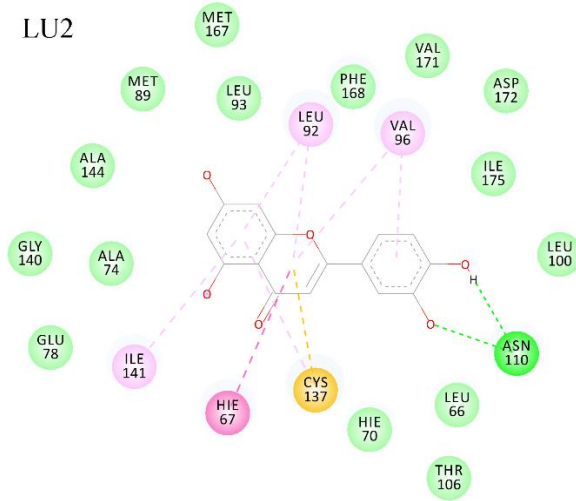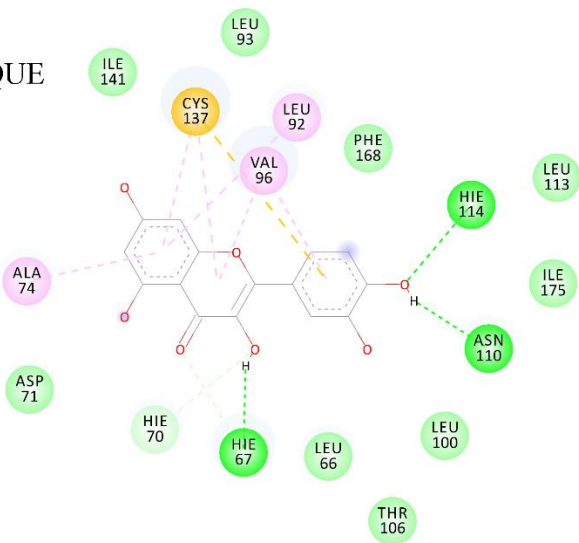
Interactions

van der Waals	Pi-Sulfur
Conventional Hydrogen Bond	Pi-Alkyl
Carbon Hydrogen Bond	

Figure S4. Residue-specific numbers of contacts between alpha-C atoms and the ligand in protein-ligand complexes. Red dots denote contacts larger than 10, green dots represent contact number between 5 and 10, blue ones are those larger than 1, and the other are represented by white dots.

Residue-LU2 Contacts

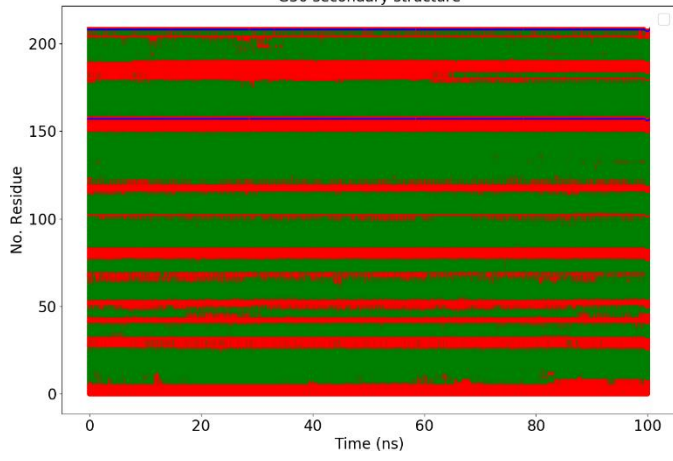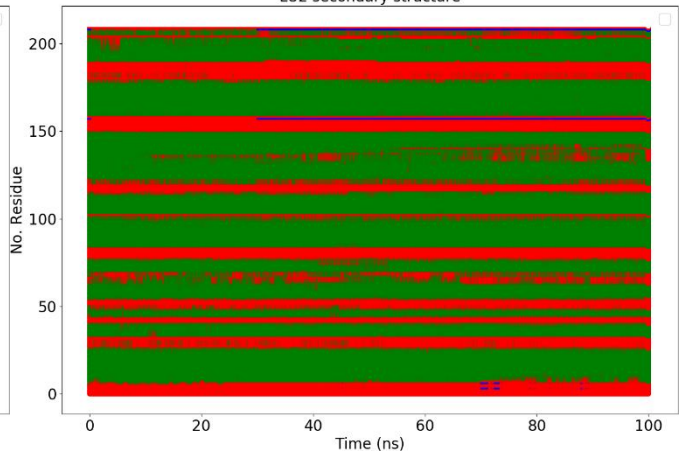Residue-NAR Contacts

Residue-QUE Contacts
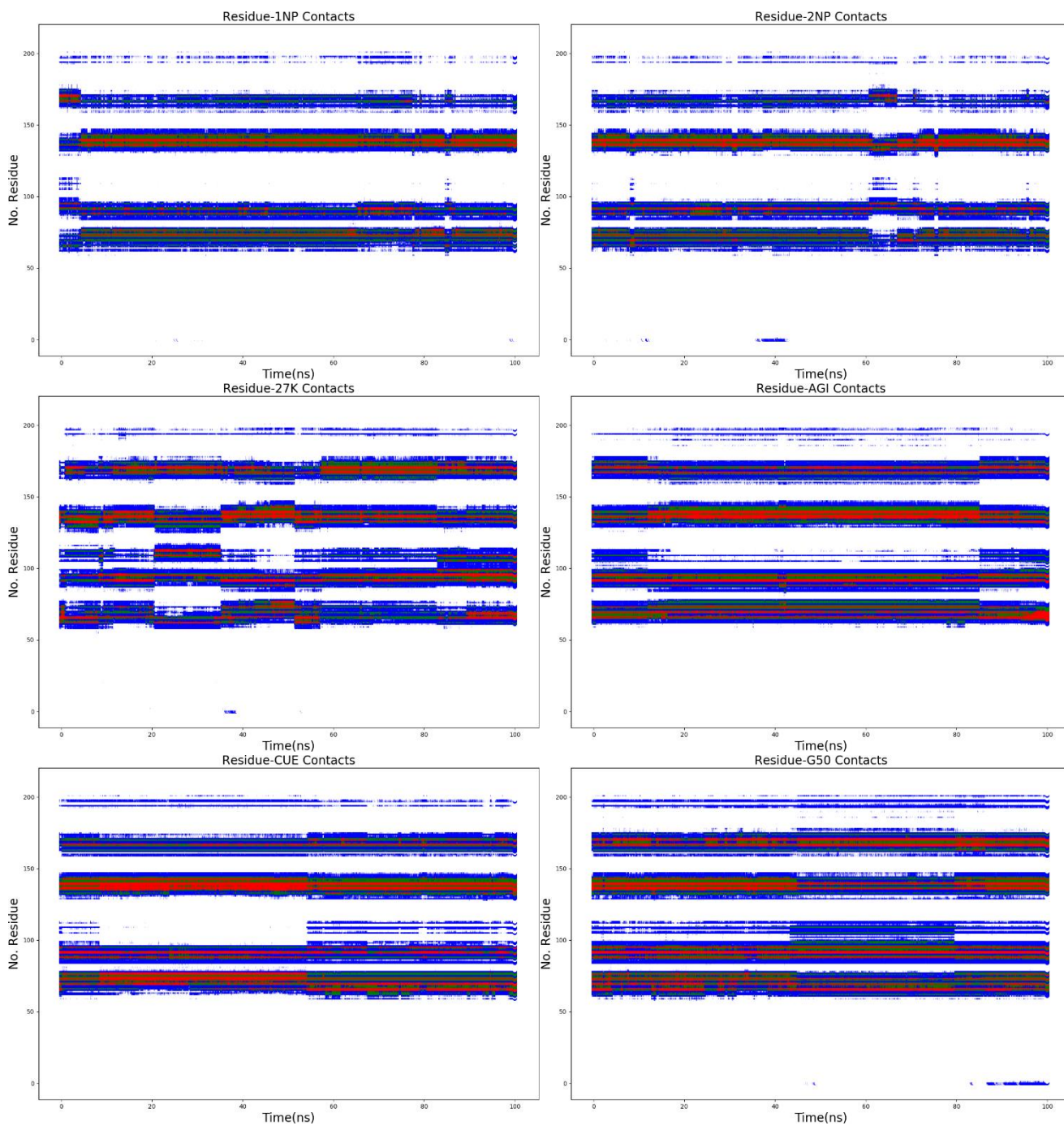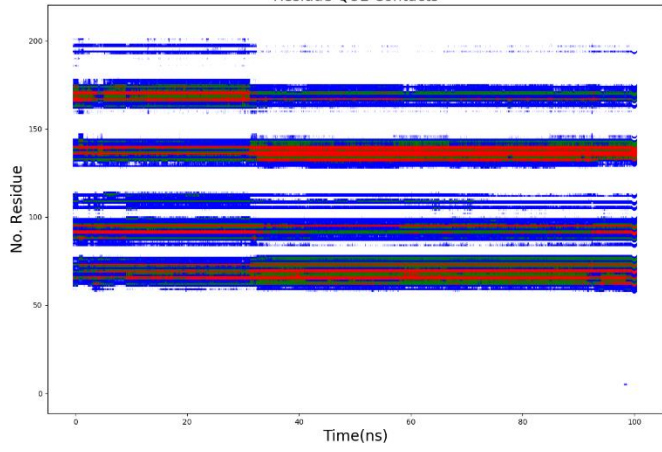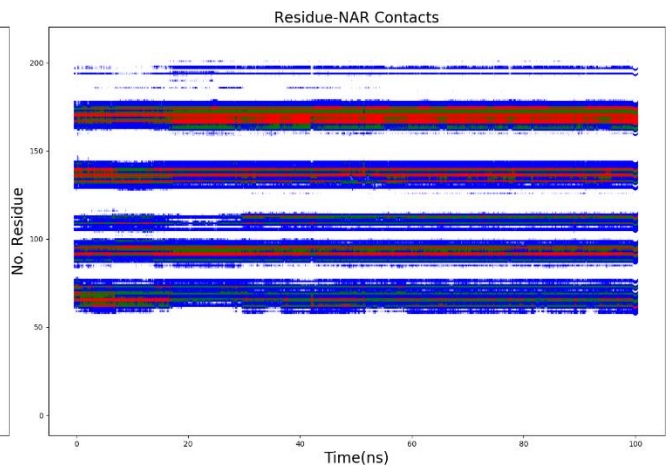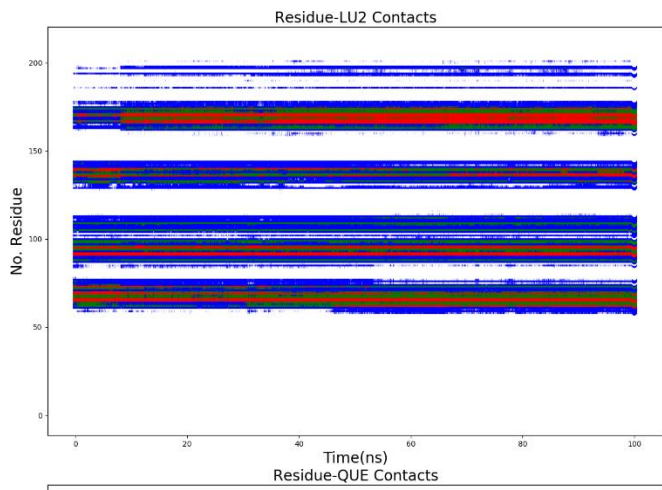
Figure S5. Time evolution of secondary structures of protein-ligand complexes. The helical structure is plotted with green dots, the beta component is plotted with red dots, and the coil region is represented by blue dots.
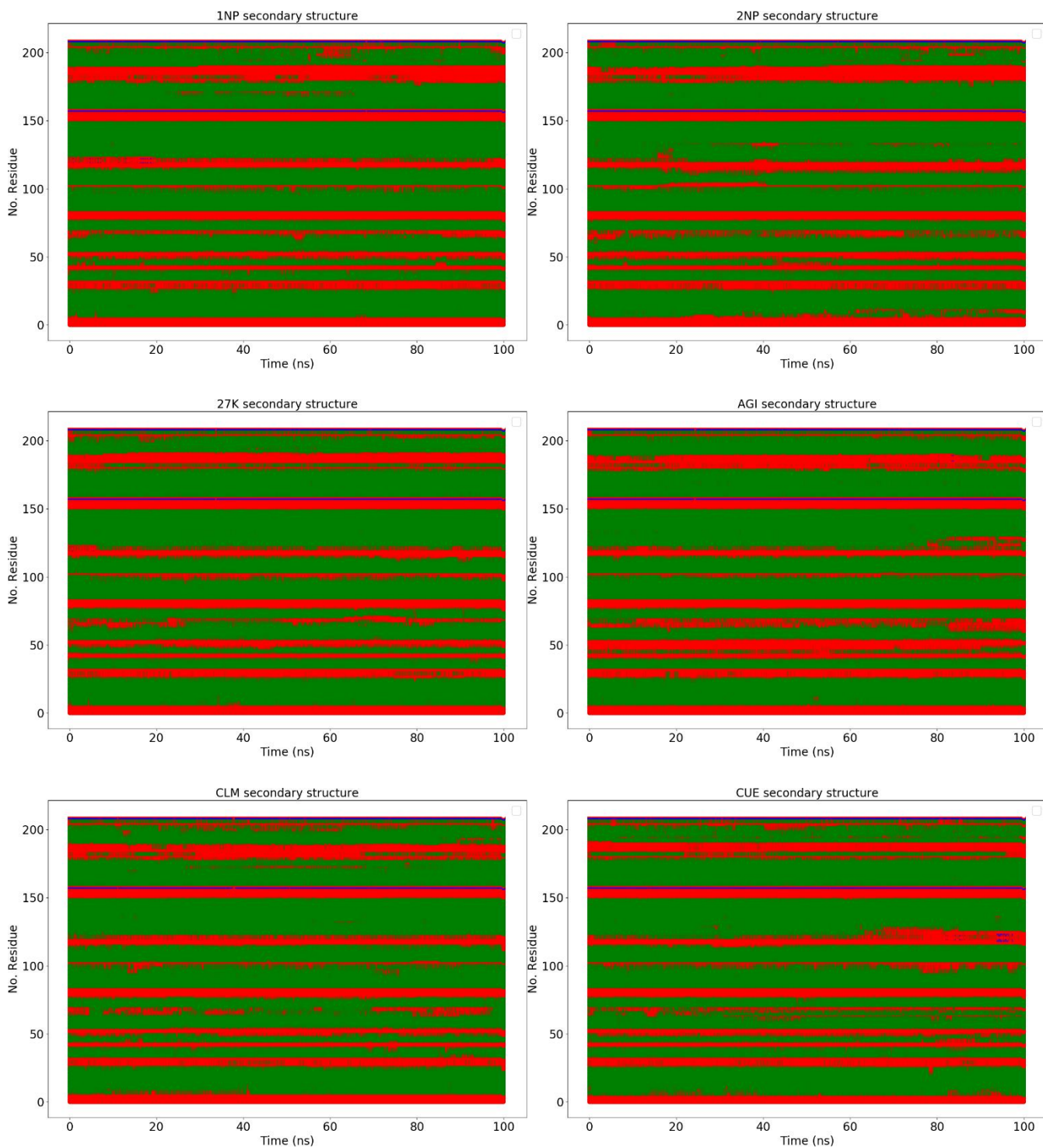
G50 secondary structure

LU2 secondary structure

QUE secondary structure