Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

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[2,3], and John Z.H.Zhang[2,4,5*]

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

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

[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

Details about the alchemical free energy calculation:

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[63, 82] 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,[121-130] which hinders the convergence of the simulation, we employed the nonlinear separation-shifted softcore-potential[121, 124-126, 131] 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.[1] 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.[1] According to previous experience, reliable free energy estimates can be obtained when the main diagonal and its neighbors are appreciably larger than 0.03.[1] 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-series 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 $\left.\frac{\partial H}{\partial \lambda}\right|_{\lambda=\lambda_i}$ [1-2] and the statistical inefficiency $\phi = 1 + 2\tau$. The whole dataset is then subsampled by ϕ to extract the independent data points.[3-5] Three post-processing methods including TI with trapezoid rule for numerical integration,[6-8] the bidirectional reweighting regime of BAR,[9] and multistate reweighting estimator of MBAR are used to extract the free energy estimates from alchemical free energy simulations.[10-11] According to our previous experience, these three methods are among the most reliable and efficient free energy estimators based on equilibrium dynamics.[5, 12-14]

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.[3-5] 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.[4, 15] Thus, the overall statistical uncertainties in equilibrium and nonequilibrium transformation should be similar.

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. $\Delta\Delta G_{\mathrm{MM/PBSA}}$ and $\Delta\Delta G_{\mathrm{MM/GBSA}}$ 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	$\Delta\Delta G_{\mathrm{MM/PBSA}}$	SD	$\Delta\Delta G_{\mathrm{MM/GBSA}}$	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.

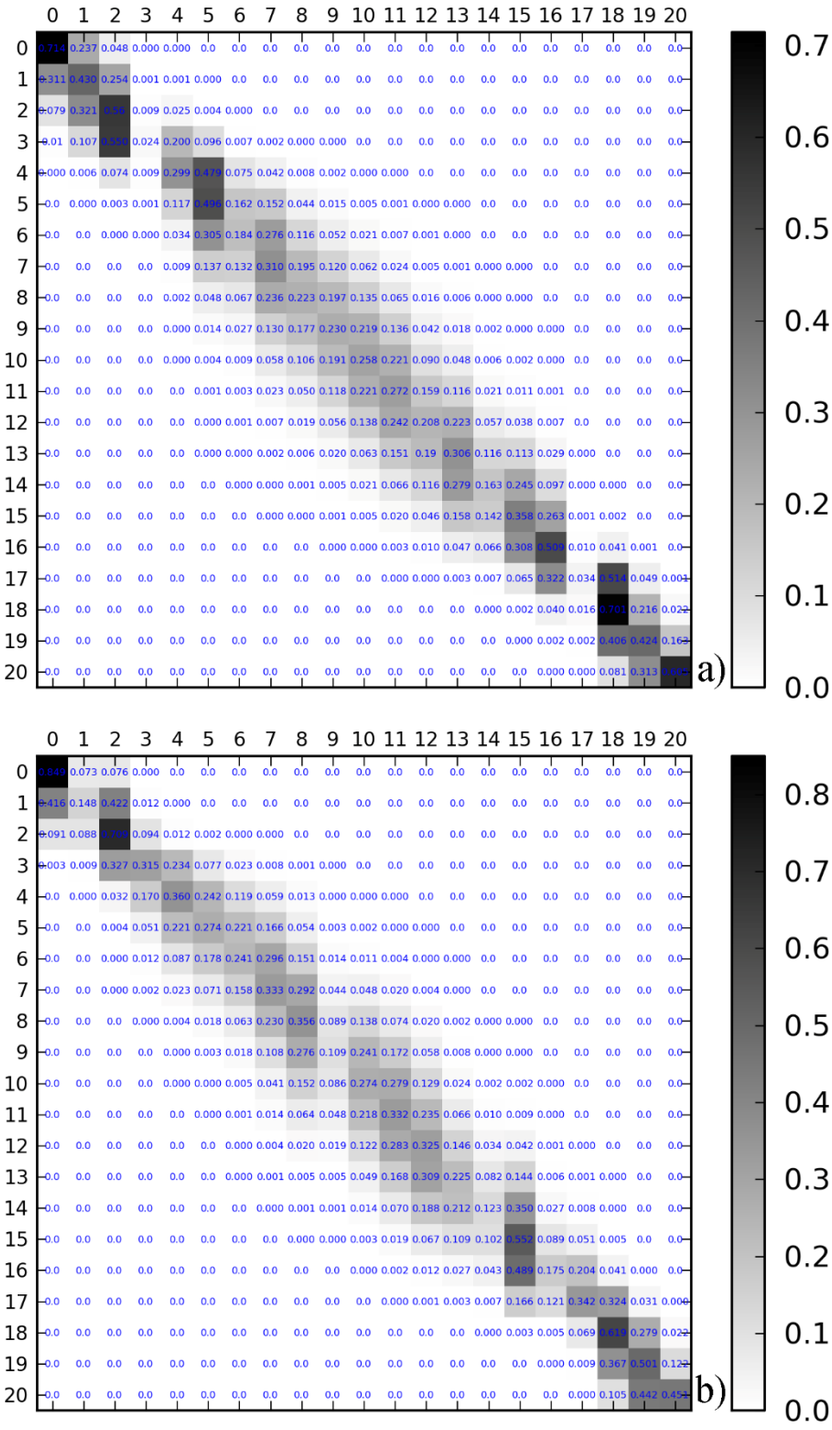
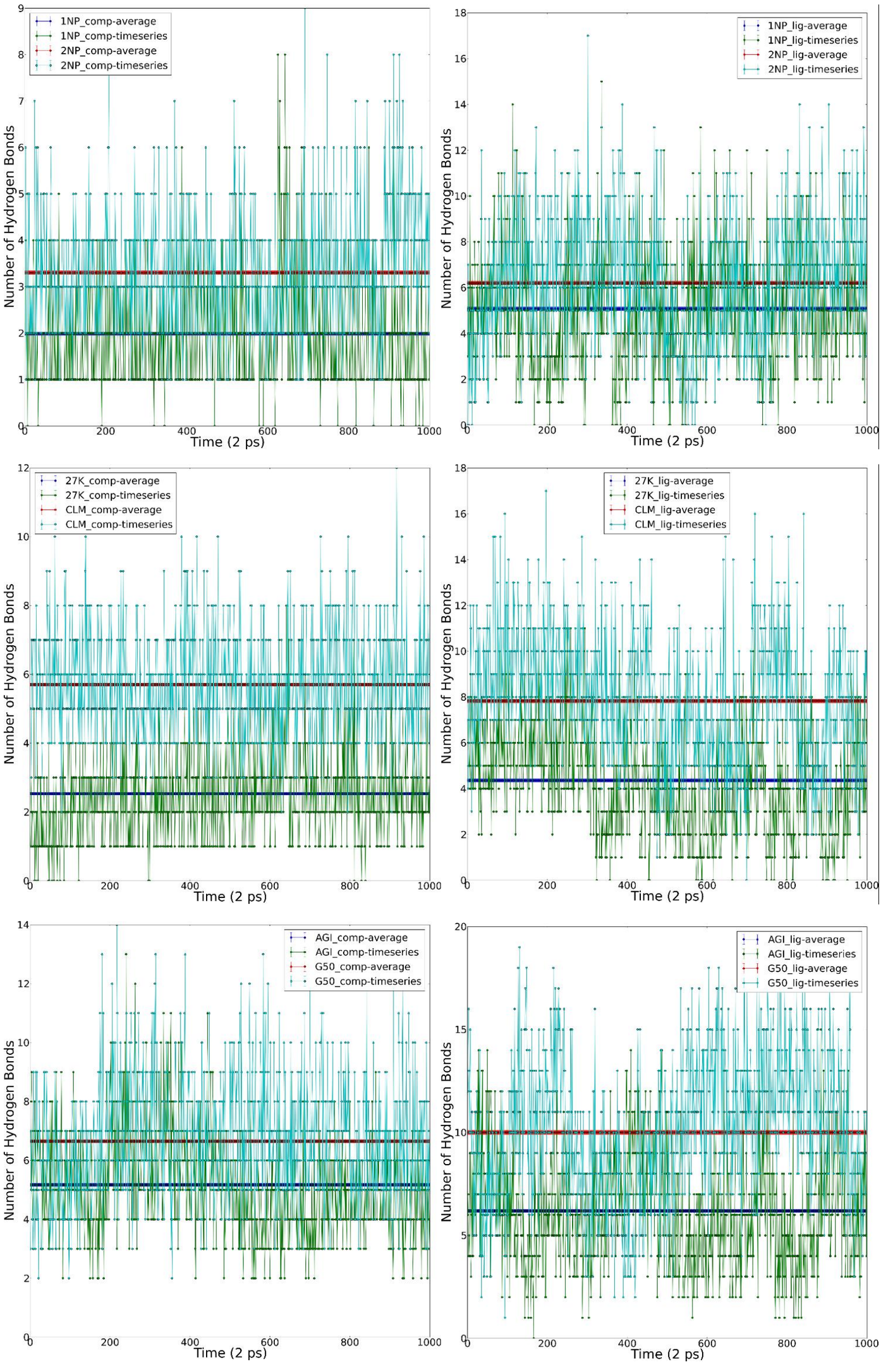
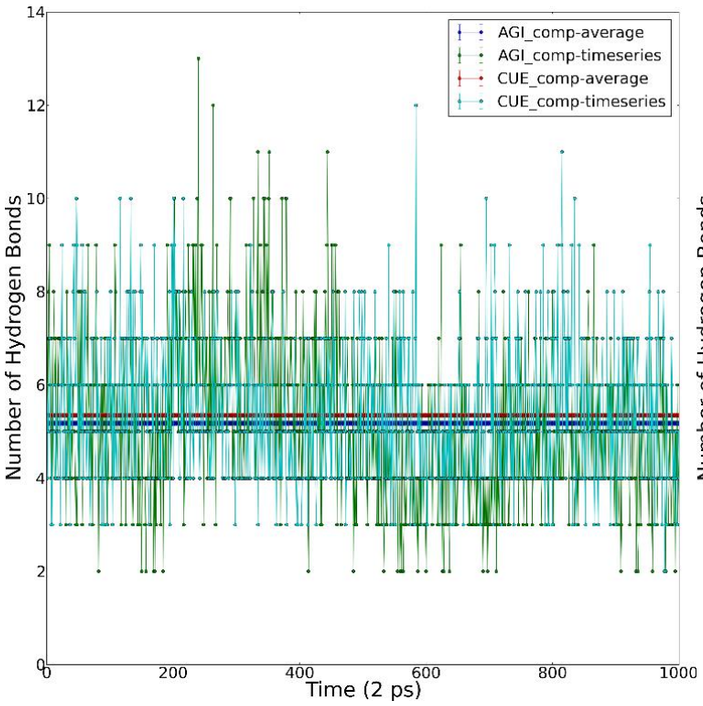
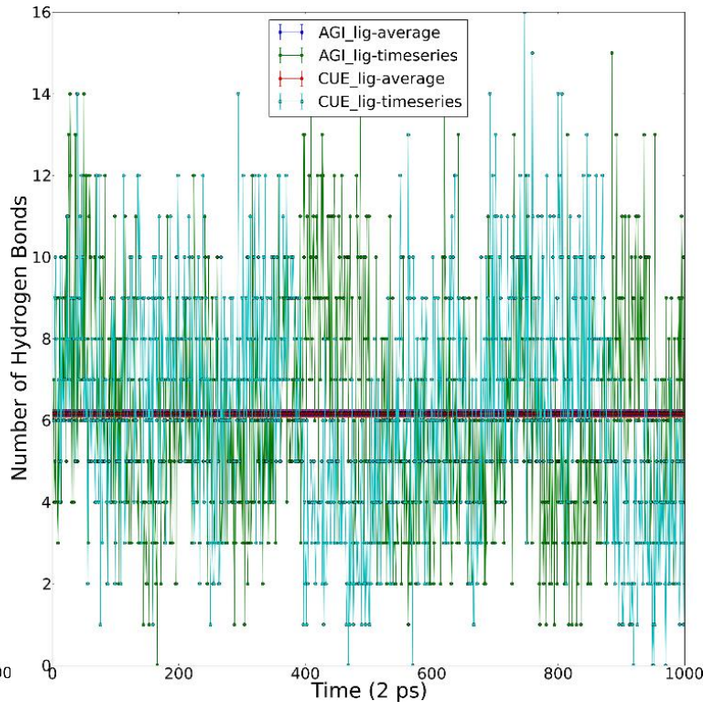
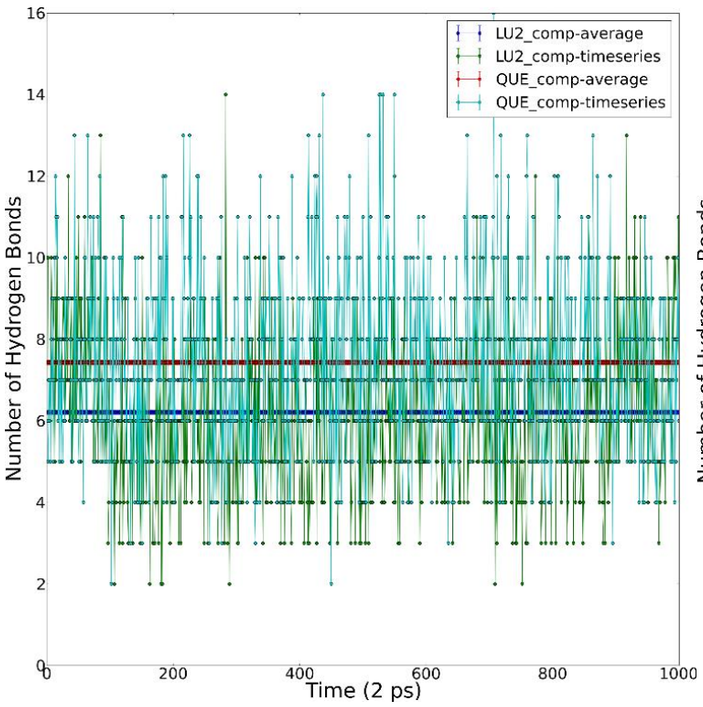


Figure S2. Hydrogen bonds formed between the ligand and its surroundings in protein-ligand system (left) and ligand-only system (right). The widths of the average lines are the standard error of the mean of the number of hydrogen bonds.




AGI_comp-average
AGI_comp-timeseries
CUE_comp-average
CUE_comp-timeseries
Number of Hydrogen Bonds
Time (2 ps)


AGI_lig-average
AGI_lig-timeseries
CUE_lig-average
CUE_lig-timeseries
Number of Hydrogen Bonds
Time (2 ps)


LU2_comp-average
LU2_comp-timeseries
QUE_comp-average
QUE_comp-timeseries
Number of Hydrogen Bonds
Time (2 ps)

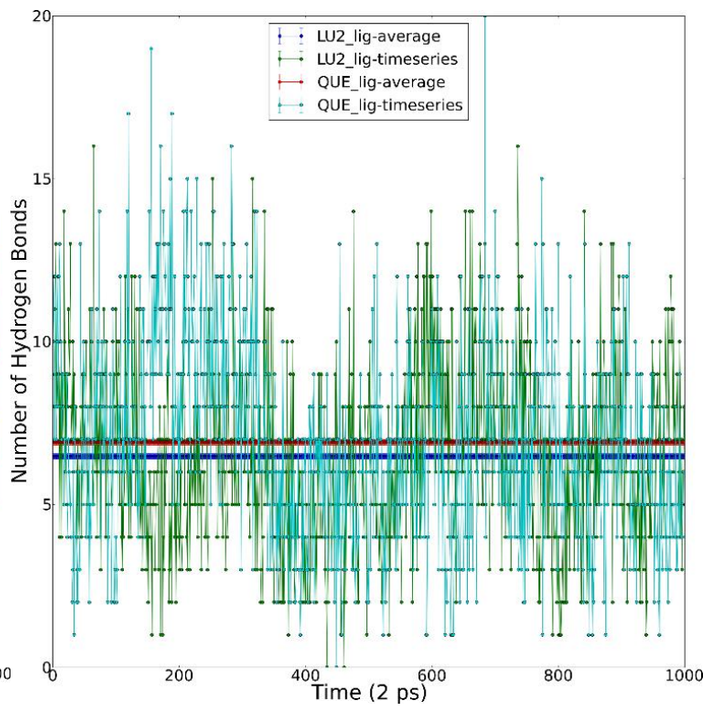
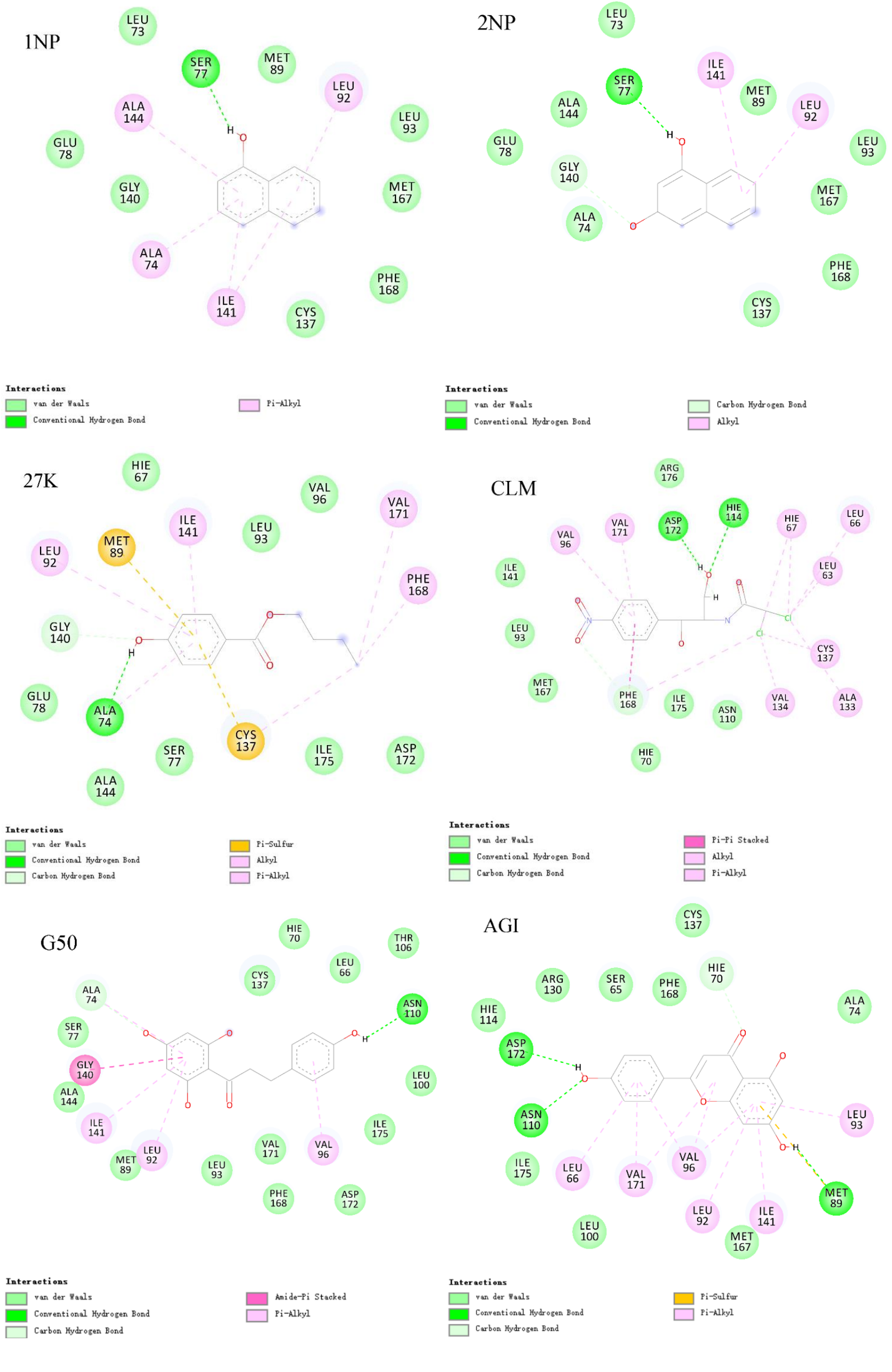
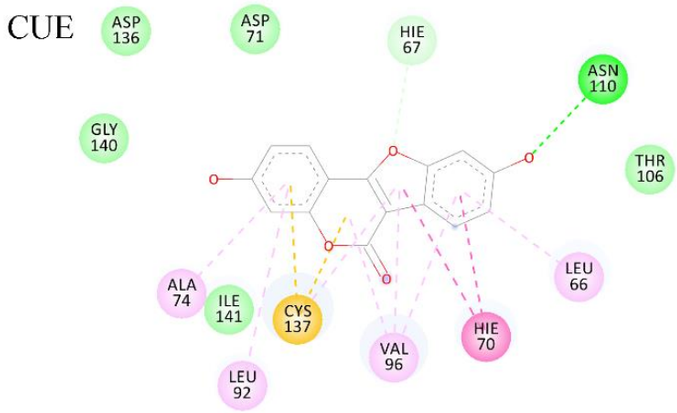

LU2_lig-average
LU2_lig-timeseries
QUE_lig-average
QUE_lig-timeseries
Number of Hydrogen Bonds
Time (2 ps)

Figure S3. Interaction maps for protein-ligand complexes obtained from equilibrated structures.



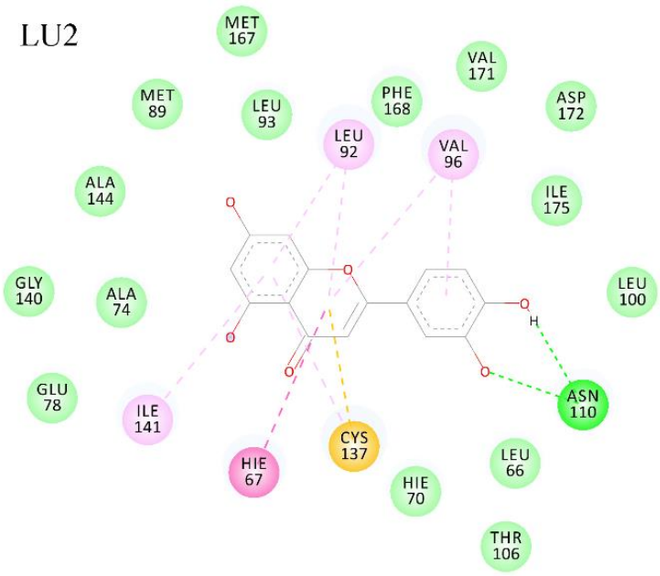
CUE



Interactions

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

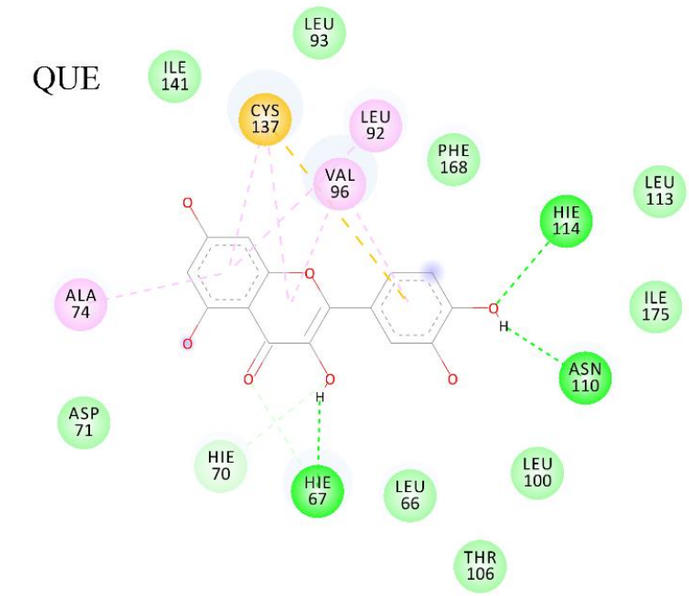
LU2



Interactions

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

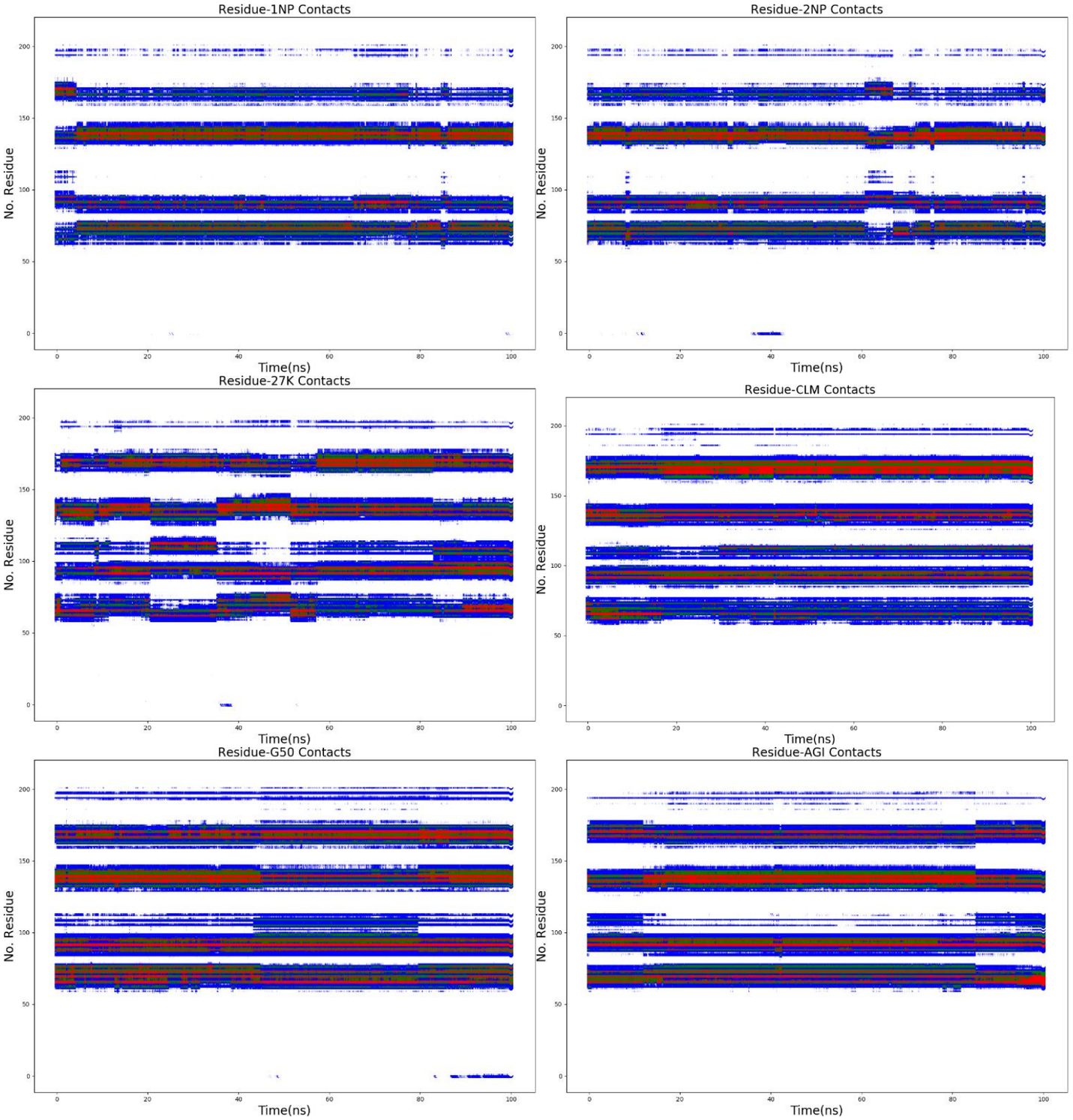
QUE

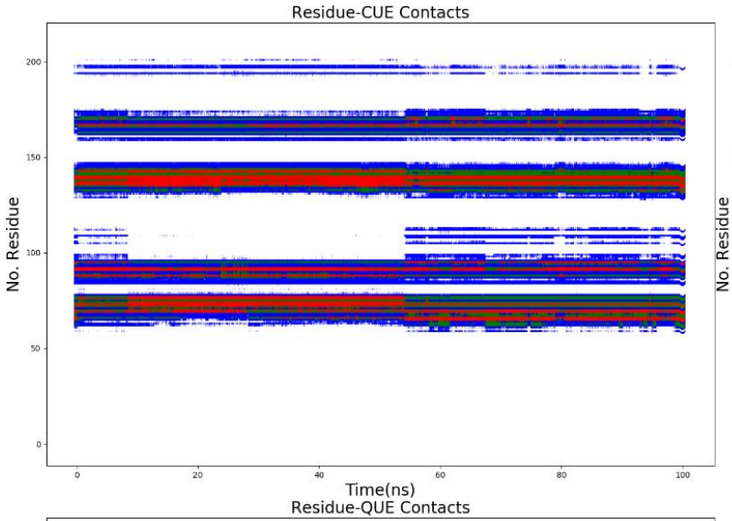


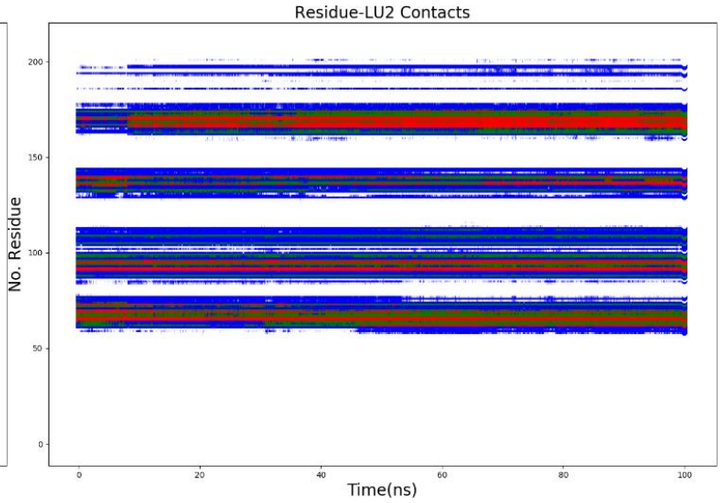
Interactions

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

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-CUE Contacts
No. Residue
Time(ns)
0
50
100
150
200
0
20
40
60
80
100


Residue-LU2 Contacts
No. Residue
Time(ns)
0
50
100
150
200
0
20
40
60
80
100

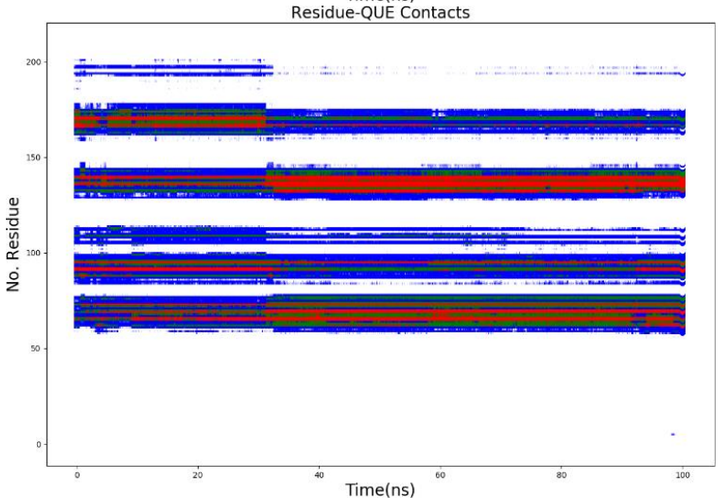
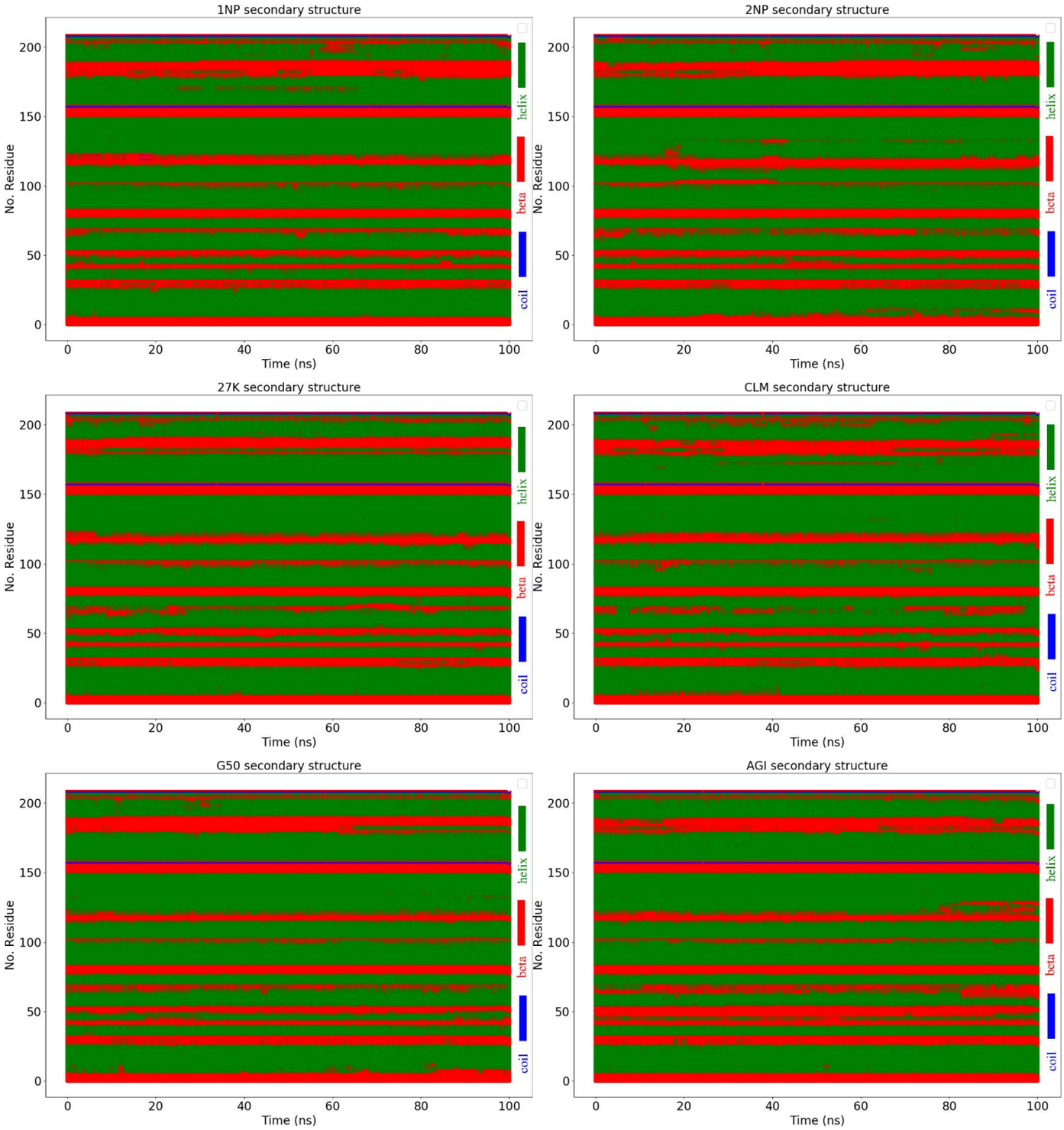
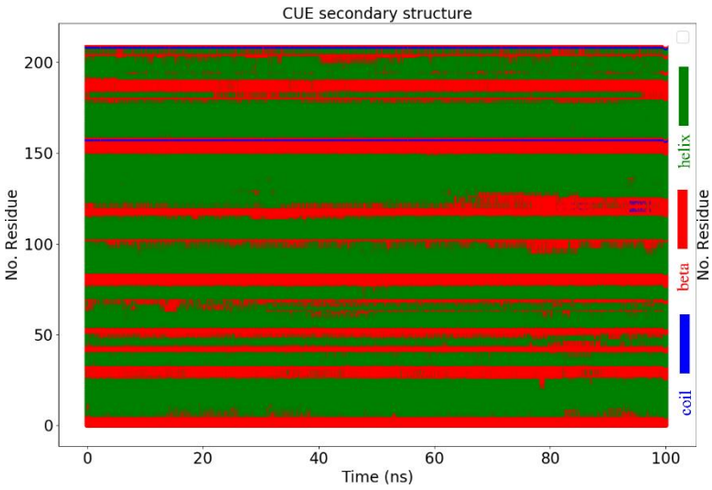
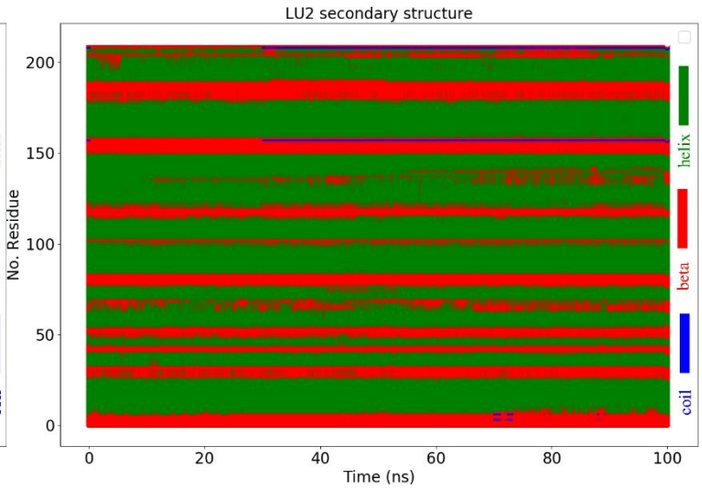
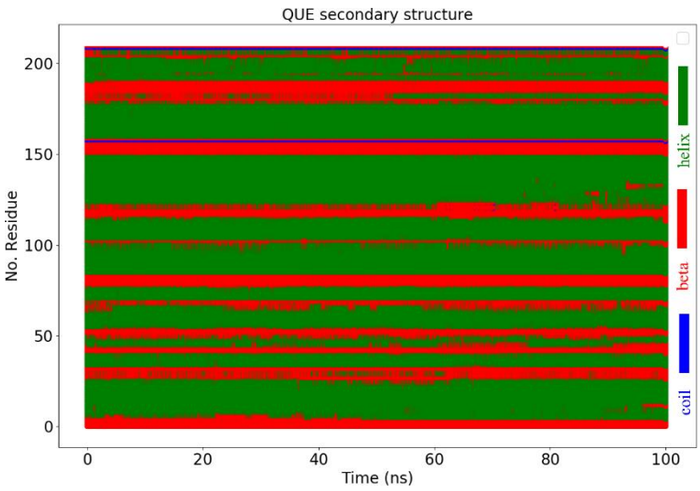

Residue-QUE Contacts
No. Residue
Time(ns)
0
50
100
150
200
0
20
40
60
80
100

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.




CUE secondary structure
No. Residue
Time (ns)
helix
beta
coil


LU2 secondary structure
No. Residue
Time (ns)
helix
beta
coil


QUE secondary structure
No. Residue
Time (ns)
helix
beta
coil

References

1. 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.
2. 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.
3. 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.
4. 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.
5. 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.
6. Bruckner, S.; Boresch, S., Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **2011,** *32* (7), 1320–1333.
7. Resat, H.; Mezei, M., Studies on free energy calculations. I. Thermodynamic integration using a polynomial path. *J. Chem. Phys.* **1993,** *99* (8), 6052-6061.
8. Resat, H.; Mezei, M., Studies on free energy calculations. II. A theoretical approach to molecular solvation. *J. Chem. Phys.* **1994,** *101* (7), 6126-6140.
9. Bennett, C. H., Efficient Estimation of Free Energy Differences from Monte Carlo data. *J. Comput. Phys.* **1976,** *22* (2), 245-268.
10. Tan, Z., On a Likelihood Approach for Monte Carlo Integration. *J. Am. Stat. Assoc.* **2004,** *99* (December), 1027-1036.
11. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008,** *129* (12), 124105.
12. 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.
13. 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.
14. 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.
15. 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,** *40* (12), 1270-1289.