Supporting Information: Sensitivity in binding free energies due to protein reorganization

Nathan M. Lim,*,† Lingle Wang,‡ Robert Abel,‡ and David L. Mobley†

†*Department of Pharmaceutical Sciences, University of California—Irvine, Irvine, California 92697, United States*

‡*Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States*

E-mail: dmobley@mobleylab.org

Simulation Interactions Diagrams

Using the 'Simulation Interactions Diagram' tool provided with Desmond, we count the number of hydrophobic contacts between the ligand and each of the surrounding protein residues at each frame. Hydrophobic contacts are categorized into three groups: π-Cation, $\pi - \pi$, and Other. The π-Cation contact is defined as an aromatic and charged group that are within 4.5Å. $\pi - \pi$ contacts are when two aromatic groups are stacked face-to-face or face-to-edge. Hydrophobic contacts that are classified as Other are when hydrophobic side-chains are within 3.6Åof a ligand's aromatic or aliphatic carbons.

Through this protein-ligand contact analysis, we can see if the ligand forms the same interactions regardless of the initial protein configuration. In Figure 1, we analyze the hydrophobic contacts for hexylbenzene in either the protein closed (purple) or open (green) MD simulations and represent the difference in red. Here, we represent the number of protein-ligand hydrophobic contacts as a percentage for the length of the MD simulations. Again, we analyze frames from 0-5ns with the default protocol and 45-55ns with the pREST

1

protocol.

Given hexylbenzene prefers the open conformation, we should expect to see Val111 contact the hexylbenzene at a percent closer to that of the protein open simulation. Using the default protocol, we show Val111 is in contact with hexylbenzene for 40.4% of the protein closed simulation and 9.6% of the protein open simulation, approximately a 30.8% difference in the interaction percentage. On the other hand, with the pREST protocol—where Val111 is included into the REST region—this difference dramatically reduces to 0.5%. Now, regardless of our initial protein starting conformation, Val111 comes in contact with the ligand 7.9% and 8.4% of the protein closed and open simulations, close to what we would have expected judging from our protein open simulation using the default protocol. Furthermore, we can compute the mean absolute difference (MAD) across all residues to yield 10.5% for the default protocol and 2.8% with the pREST protocol. Overall, the reduction in the difference in protein-ligand contacts indicates the bias from the initial protein configuration is almost entirely eliminated from using pREST.
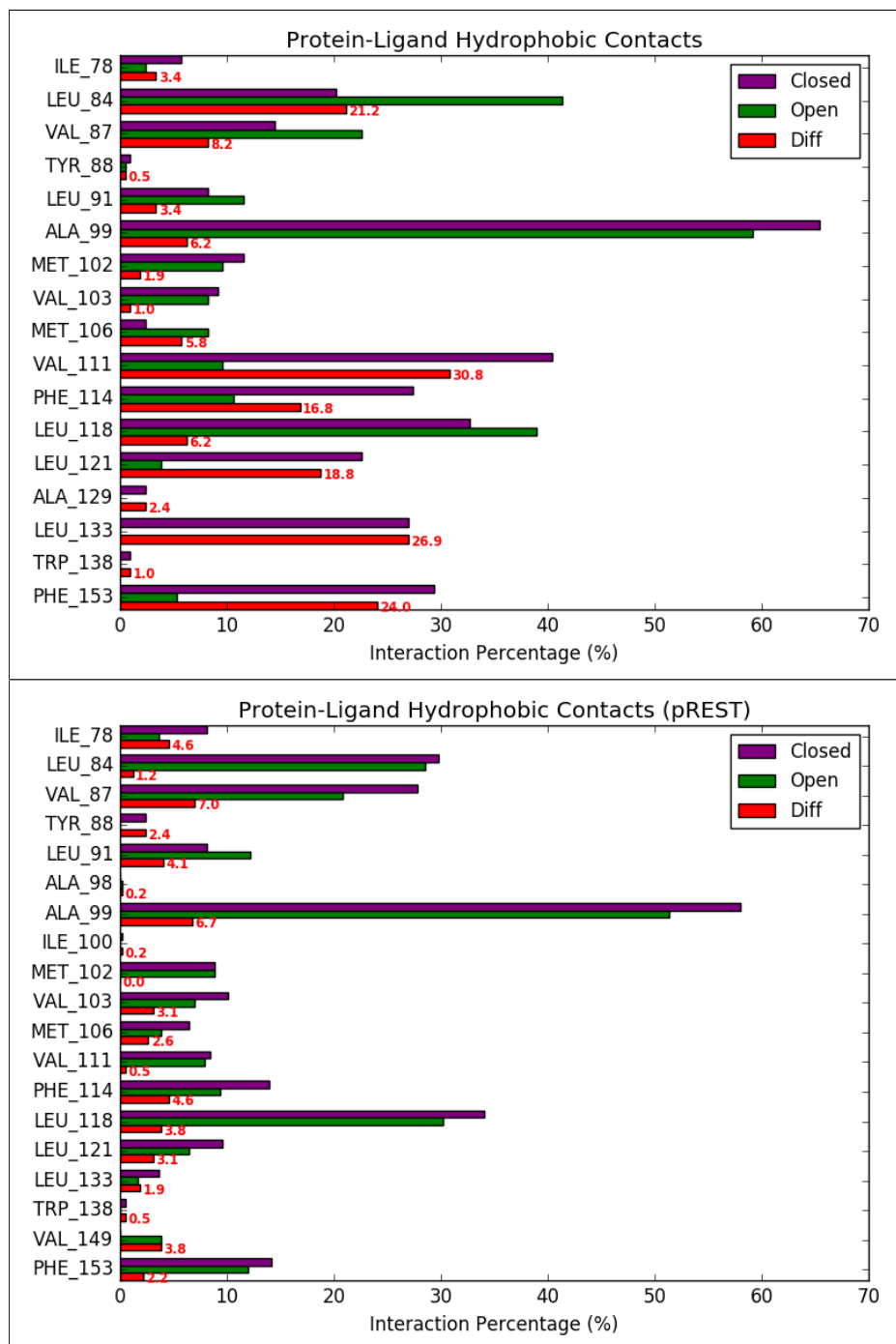
Figure 1: Bar plot of interaction percentage of protein-ligand hydrophobic contacts over the course of a trajectory. From the benzene to hexylbenzene transformation, these plots correspond to the hexylbenzene (λ_{11}) state from closed (purple) and open (green) simulations using the default versus pREST protocols. The mean absolute difference (MAD) in interaction percentage is represented in red. MAD for the default protocol is 10.5% and 2.8% for the pREST protocol.

Tables

Table 1: Closed-Intermediate Transformations[a]

Ligand 1	Ligand 2	$\Delta\Delta G_C$	σ_C	$\Delta\Delta G_O$	σ_O	$\Delta\Delta G_\varepsilon$
benzene	butylbenzene	0.58	0.07	-0.59	0.08	1.17
toluene	butylbenzene	-0.28	0.06	-1.27	0.09	0.99
ethylbenzene	butylbenzene	0.24	0.06	-0.23	0.07	0.47
propylbenzene	butylbenzene	0.99	0.06	0.63	0.04	0.36
benzene	sec-butylbenzene	2.36	0.09	2.14	0.11	0.22
toluene	sec-butylbenzene	1.47	0.07	1.14	0.08	0.33
ethylbenzene	sec-butylbenzene	1.90	0.08	1.77	0.08	0.13
propylbenzene	sec-butylbenzene	2.86	0.06	2.67	0.05	0.19

[a] Calculated free energies and their uncertainties in 'closed-intermediate' alchemical transformations from simulations that were started from the protein closed state($\Delta\Delta G_C$) versus the protein open state($\Delta\Delta G_O$). Free energies are in units of kcal/mol. The inconsistency($\Delta\Delta G_\varepsilon$) is obtained from the difference in the relative free energies between protein closed and open simulations. Inconsistency is highlighted red or green if above or below 1 kcal/mol, respectively. RMSI is 0.6 kcal/mol. Table is represented in Fig 5c.

Table 2: Closed-Intermediate Transformations pREST[a]

Ligand 1	Ligand 2	$\Delta\Delta G_C$	σ_C	$\Delta\Delta G_O$	σ_O	$\Delta\Delta G_\varepsilon$
benzene	butylbenzene	-0.07	0.10	-0.72	0.12	0.65
toluene	butylbenzene	0.89	0.10	-0.42	0.08	1.31
ethylbenzene	butylbenzene	-0.20	0.09	-0.43	0.08	0.23
propylbenzene	butylbenzene	1.02	0.06	0.49	0.06	0.53
benzene	sec-butylbenzene	0.45	0.08	1.34	0.12	0.89
toluene	sec-butylbenzene	0.60	0.12	0.60	0.10	0.00
ethylbenzene	sec-butylbenzene	1.09	0.10	1.69	0.09	0.60
propylbenzene	sec-butylbenzene	1.88	0.07	3.05	0.07	1.17

[a] Calculated free energies and their uncertainties in 'closed-intermediate' alchemical transformations from 'pREST' simulations that were started from the protein closed state($\Delta\Delta G_C$) versus the protein open state($\Delta\Delta G_O$). Free energies were calculated using frames from 0 to 5ns. Free energies are in units of kcal/mol. The inconsistency($\Delta\Delta G_\varepsilon$) is obtained from the difference in the relative free energies between protein closed and open simulations. Inconsistency is highlighted red or green if above or below 1 kcal/mol, respectively. RMSI is 0.79 kcal/mol. Table is represnted in Fig 5d.

Table 3: Closed-Intermediate Transformations pREST Extended[a],[b]

Ligand 1	Ligand 2	$\Delta\Delta G_C$	σ_C	$\Delta\Delta G_O$	σ_O	$\Delta\Delta G_\varepsilon$
benzene	butylbenzene	-1.14[b]	0.09	-0.72	0.12	0.42
toluene	butylbenzene	-0.62[b]	0.09	-0.96[a]	0.09	0.34
ethylbenzene	butylbenzene	-0.20	0.09	-0.43	0.08	0.23
propylbenzene	butylbenzene	0.51[b]	0.06	0.49	0.06	0.02
benzene	sec-butylbenzene	0.45	0.08	1.34	0.12	0.89
toluene	sec-butylbenzene	0.60	0.12	0.60	0.10	0.00
ethylbenzene	sec-butylbenzene	1.09	0.10	1.69	0.09	0.60
propylbenzene	sec-butylbenzene	1.88	0.07	1.84[b]	0.06	0.04

[a] Calculated free energies and their uncertainties in 'closed-intermediate' alchemical transformations from 'pREST' simulations that were started from the protein closed state($\Delta\Delta G_C$) versus the protein open state($\Delta\Delta G_O$). Free energies were calculated using frames from 15 to 25ns. Free energies are in units of kcal/mol. The inconsistency($\Delta\Delta G_\varepsilon$) is obtained from the difference in the relative free energies between protein closed and open simulations. Inconsistency is highlighted red or green if above or below 1 kcal/mol, respectively. RMSI is 0.43 kcal/mol. Table is represnted in Fig 5d.
[b] Only these alchemical transformations were extended up to 25ns of simulation time in order to lower the inconsistency and error with experiment.

Table 4: Closed-Open Transformations[a]

Ligand 1	Ligand 2	$\Delta\Delta G_C$	σ_C	$\Delta\Delta G_O$	σ_O	$\Delta\Delta G_\varepsilon$
benzene	pentylbenzene	2.36	0.12	-1.33	0.10	3.69
toluene	pentylbenzene	1.77	0.09	0.34	0.10	1.43
ethylbenzene	pentylbenzene	2.45	0.08	0.46	0.09	1.99
propylbenzene	pentylbenzene	3.46	0.08	-0.22	0.07	3.68
benzene	hexylbenzene	4.13	0.16	-0.61	0.15	4.74
toluene	hexylbenzene	2.90	0.13	-1.63	0.08	4.53
ethylbenzene	hexylbenzene	3.63	0.11	-0.76	0.09	4.39
propylbenzene	hexylbenzene	5.85	0.10	0.13	0.06	5.72

[a] Calculated free energies and their uncertainties in 'closed-open' alchemical transformations from simulations that were started from the protein closed state($\Delta\Delta G_C$) versus the protein open state($\Delta\Delta G_O$). Free energies were calculated using frames from 0 to 5ns. Free energies are in units of kcal/mol. The inconsistency($\Delta\Delta G_\varepsilon$) is obtained from the difference in the relative free energies between protein closed and open simulations. Inconsistency is highlighted red or green if above or below 1 kcal/mol, respectively. RMSI is 4.0 kcal/mol. Table is represented in Fig 5a.

Table 5: Closed-Open Transformations pREST[a]

Ligand 1	Ligand 2	$\Delta\Delta G_C$	σ_C	$\Delta\Delta G_O$	σ_O	$\Delta\Delta G_\varepsilon$
benzene	pentylbenzene	1.49	0.13	0.07	0.10	1.42
toluene	pentylbenzene	1.41	0.12	0.79	0.13	0.62
ethylbenzene	pentylbenzene	2.90	0.10	1.23	0.09	1.67
propylbenzene	pentylbenzene	4.25	0.09	1.01	0.07	3.24
benzene	hexylbenzene	2.75	0.14	1.25	0.11	1.50
toluene	hexylbenzene	3.22	0.11	-1.15	0.11	4.37
ethylbenzene	hexylbenzene	3.34	0.12	-0.22	0.11	3.56
propylbenzene	hexylbenzene	4.93	0.12	1.21	0.11	3.72

[a] Calculated free energies and their uncertainties in 'closed-open' alchemical transformations from 'pREST' simulations that were started from the protein closed state($\Delta\Delta G_C$) versus the protein open state($\Delta\Delta G_O$). Free energies are in units of kcal/mol. The inconsistency($\Delta\Delta G_\varepsilon$) is obtained from the difference in the relative free energies between protein closed and open simulations. The inconsistency is highlighted red or green if above or below 1 kcal/mol, respectively. RMSI is 2.82 kcal/mol.

Table 6: Closed-Open Transformations pREST Extended[a]

Ligand 1	Ligand 2	$\Delta\Delta G_C$	σ_C	$\Delta\Delta G_O$	σ_O	$\Delta\Delta G_\varepsilon$
benzene	pentylbenzene	1.37	0.08	1.41	0.08	0.04
toluene	pentylbenzene	1.08	0.08	0.81	0.07	0.27
ethylbenzene	pentylbenzene	1.71	0.08	1.50	0.07	0.21
propylbenzene	pentylbenzene	3.20	0.06	2.34	0.05	0.86
benzene	hexylbenzene	2.15	0.09	1.14	0.08	1.01
toluene	hexylbenzene	0.32	0.10	1.10	0.07	0.78
ethylbenzene	hexylbenzene	1.96	0.08	2.17	0.07	0.21
propylbenzene	hexylbenzene	3.38	0.07	3.12	0.06	0.26

[a] Calculated free energies and their uncertainties in 'closed-open' alchemical transformations from extended 'pREST' simulations that were started from the protein closed state($\Delta\Delta G_C$) versus the protein open state($\Delta\Delta G_O$). Free energies are in units of kcal/mol. Free energies were calculated using frames from 40 to 55ns. The inconsistency($\Delta\Delta G_\varepsilon$) is obtained from the difference in the relative free energies between protein closed and open simulations. Inconsistency is highlighted red or green if above or below 1 kcal/mol, respectively. RMSI is 0.57 kcal/mol. Table is represented in Fig 5b.

Table 7: Experimental Ligand Transformations[a]

Ligand 1	Ligand 2	ΔG_{exp}	σ_{exp}	$\Delta\Delta G_C$	σ_C	$\Delta\Delta G_O$	σ_O
benzene	toluene	-0.33	0.16	0.71	0.04	0.59	0.05
benzene	ethylbenzene	-0.19	0.17	0.15	0.05	-0.09	0.06
benzene	propylbenzene	-1.36	0.16	-0.82	0.06	-0.69	0.08
benzene	butylbenzene	-1.51	0.16	0.58	0.07	-0.59	0.08
toluene	ethylbenzene	-0.24	0.08	-0.59	0.04	-0.40	0.05
toluene	propylbenzene	-1.03	0.04	-2.21	0.05	-0.89	0.05
toluene	butylbenzene	-1.18	0.04	-0.28	0.06	-1.27	0.09
ethylbenzene	propylbenzene	-0.79	0.07	-0.86	0.05	-0.80	0.03
ethylbenzene	butylbenzene	-0.94	0.07	0.24	0.06	-0.23	0.07
propylbenzene	butylbenzene	-0.15	0.03	0.99	0.06	0.63	0.04

[a] Experimental and calculated free energies with their uncertainties in alchemical transformations involving ligands with available experimental affinities. Calculated free energies are from simulations that were started from the protein closed state($\Delta\Delta G_C$) versus the protein open state($\Delta\Delta G_O$). Free energies are in units of kcal/mol. Here, we compare $\Delta\Delta G_C$ and $\Delta\Delta G_O$ against ΔG_{exp} to compute the error with experiment. RMSE is 1.0 kcal/mol and 0.58 kcal/mol for closed and open simulations, respectively. RMSI is 0.68 kcal/mol. Table is represented in Fig 6a.

Table 8: Experimental Ligand Transformations pREST[a]

Ligand 1	Ligand 2	ΔG_{exp}	σ_{exp}	$\Delta\Delta G_C$	σ_C	$\Delta\Delta G_O$	σ_O
benzene	toluene	-0.33	0.16	0.10	0.08	0.31	0.06
benzene	ethylbenzene	-0.19	0.17	-0.20	0.08	-0.62	0.08
benzene	propylbenzene	-1.36	0.16	-1.57	0.09	-1.26	0.09
benzene	butylbenzene	-1.51	0.16	-1.14	0.11	-0.72	0.12
toluene	ethylbenzene	-0.24	0.08	-0.84	0.07	-0.83	0.07
toluene	propylbenzene	-1.03	0.04	-1.91	0.09	-1.44	0.08
toluene	butylbenzene	-1.18	0.04	-0.62	0.09	-0.96	0.09
ethylbenzene	propylbenzene	-0.79	0.07	-1.08	0.07	-0.80	0.06
ethylbenzene	butylbenzene	-0.94	0.07	-0.20	0.09	-0.43	0.08
propylbenzene	butylbenzene	-0.15	0.03	0.51	0.06	0.49	0.06

[a] Experimental and calculated free energies with their uncertainties in alchemical transformations involving ligands with available experimental affinities. Calculated free energies are from 'pREST' simulations that were started from the protein closed state($\Delta\Delta G_C$) versus the protein open state($\Delta\Delta G_O$). Free energies are in units of kcal/mol. Here, we compare $\Delta\Delta G_C$ and $\Delta\Delta G_O$ against ΔG_{exp} to compute the error with experiment. RMSE is 0.54 kcal/mol and 0.50 kcal/mol for closed and open simulations, respectively. RMSI is 0.31 kcal/mol. Table is represented in Fig 6b.

Data Sets and Analysis Tool Download

All related data sets and analysis tools are available online and described below. The files can be downloaded at http://n2t.net/ark:/b7280/d1js3b (doi:10.7280/D1JS3B).

Prepared protein-ligand crystal structures: `xtal_prepped.tar.gz`

Contains the reference protein-ligand crystal structures that have been aligned and prepared through the 'Protein Preparation Wizard' tool as `.mae` files.

RMSD analysis: `scripts.tar.gz`

Contains the scripts `RMSD-analysis.tcl` and `plot-RMSD.py` used for RMSD analysis. A `README.txt` file has been provided which details how to generate the RMSD/time and Colormap plots found in the paper.

FEP data: `C-I_default.tar.gz`, `C-I_pPREST.tar.gz`, `C-O_default.tar.gz`, `C-O_pREST.tar.gz`, `EXP_default.tar.gz,` and `EXP_pREST.tar.gz`

Contains archived data from the FEP calculations for each alchemical transformation set, where within each archived set are directories for individual FEP calculations. These contain input `.mae` files containing the protein/ligand structures, the simulation configuration `.msj` files, energy files, and a `README.txt` that summarizes information about each FEP calculation. Complete trajectories are included.

This information is available free of charge via the Internet at http://pubs.acs.org