# Supporting Information

Bin W. Zhang, Di Cui, Nobuyuki Matubayasi, and Ronald M. Levy*

E-mail: ronlevy@temple.edu

## Solvation Free Energy Estimated from Two $\gamma$-states: Additional Data

The easy positions are showed with white background in Table.1. For the positions in this group, the solvation free energy of the tagged water molecule can be estimated by only using the first $10\,ns$ data points at the $\gamma = 0$ and $\gamma = 1$ states. The positions that belongs to the difficult group are marked with gray background in Table.1. For the positions in this group, when only using the $10\,ns$ data points at the $\gamma = 0$ and $\gamma = 1$ states, UWHAM analysis either fails to converge or the estimate of the solvation free energy of the tagged water molecule significantly differs from the benchmark. However, if the data set which includes $100\,ns$ of sampling at the $\gamma = 0$ states is used, the UWHAM estimates of the solvation free energy of the tagged water molecule agree with the benchmark within 95% confidence interval.

Extending the simulation at the $\gamma = 0$ state from $10\,ns$ to $100\,ns$ is essential for obtaining accurate estimates of the solvation free energy for the water positions in the difficult group. The data in Table.1 show that it also decreases the uncertainty of an estimate $2 \sim 3$ times smaller for water positions in the easy group. Next we examine the influence of changing the number of input data points at the $\gamma = 1$ state. The data in the column $\Delta F^{**}$(2 states) show the estimates of the solvation free energy when $0.1\,ns$ simulation data at the $\gamma = 1$ state and $100\,ns$ simulation data at the $\gamma = 0$ state are UWHAMMed. Compared with the

$\Delta F$(2 states) column, the number of the input data points at the $\gamma = 1$ state is 100 times less. However, despite the uncertainties increase $2 \sim 3$ times larger, the UWHAM estimates still match the benchmark within 95% confidence interval. Apparently, the convergence of endpoint calculations does not require long simulations at the $\gamma = 1$ state.

Table 1: Solvation free energy ($kcal/mol$) for growing a water molecule in a solution containing a protein molecule (FXa) and 15951 water molecules. The data in the $d_{min}$ column show the minimum of the distances ($nm$) between the oxygen atom of the tagged water molecule and the atoms of FXa. The data in the $\Delta F$(32 states) column are the benchmarks. The $\Delta F$(2 states) column shows the estimates when the $10\,ns$ simulation data at the $\gamma = 1$ state and the $100\,ns$ simulation data at the $\gamma = 0$ state are UWHAMMed. The $\Delta F^*$(2 states) column shows the estimates when the first $10\,ns$ simulation data at both the $\gamma = 0$ and $\gamma = 1$ states are UWHAMMed. The $\Delta F^{**}$(2 states) column shows the estimates when $0.1\,ns$ simulation data at the $\gamma = 1$ state and the $100\,ns$ simulation data at the $\gamma = 0$ state are UWHAMMed. The uncertainties marked with star are the lower limits of the uncertainty.

| # | $\epsilon$ | $d_{min}$ | $\Delta F$ (32 states) | $\Delta F$(2 states) | $\Delta F^*$ (2 states) | $\Delta F^{**}$(2 states) |
|---|---|---|---|---|---|---|
| 1 | $-22.684$ | 0.221 | $-13.84 \pm 0.09$ | $-14.0 \pm 0.5^*$ | - | $-13.9 \pm 0.9^*$ |
| 2 | $-21.011$ | 0.258 | $-12.7 \pm 0.17^*$ | $-13.3 \pm 0.4^*$ | - | $-13.3 \pm 1.0^*$ |
| 3 | $-19.093$ | 0.183 | $-11.0 \pm 0.4^*$ | $-11.8 \pm 0.4^*$ | - | $-12.1 \pm 0.4^*$ |
| 4 | $-17.471$ | 0.263 | $-12.7 \pm 0.13^*$ | - | - | - |
| 5 | $-16.157$ | 0.204 | $-11.40 \pm 0.07$ | $-11.9 \pm 0.4^*$ | - | $-11.4 \pm 0.6^*$ |
| 6 | $-14.444$ | 0.177 | $-10.36 \pm 0.03$ | $-10.51 \pm 0.11$ | $-10.8 \pm 0.4^*$ | $-10.5 \pm 0.2$ |
| 7 | $-12.760$ | 0.180 | $-11.43 \pm 0.08$ | $-10.7 \pm 0.6^*$ | - | $-10.2 \pm 0.7^*$ |
| 8 | $-11.111$ | 0.274 | $-9.033 \pm 0.014$ | $-9.00 \pm 0.15$ | $-9.0 \pm 0.3$ | $-9.1 \pm 0.4$ |
| 9 | $-9.627$ | 0.207 | $-8.782 \pm 0.016$ | $-8.87 \pm 0.12$ | $-8.7 \pm 0.2$ | $-8.9 \pm 0.4$ |
| 10 | $-7.834$ | 0.263 | $-9.095 \pm 0.013$ | $-9.06 \pm 0.07$ | $-8.92 \pm 0.12$ | $-9.1 \pm 0.2$ |
| 12 | $-6.40$ | 0.190 | $-7.774 \pm 0.017$ | $-7.67 \pm 0.13$ | $-7.0 \pm 0.3$ | $-7.7 \pm 0.4$ |
| 11 | $-5.777$ | 0.269 | $-8.895 \pm 0.012$ | $-8.92 \pm 0.09$ | $-8.8 \pm 0.2$ | $-8.9 \pm 0.3$ |
| 14 | $-4.74$ | 0.301 | $-5.765 \pm 0.011$ | $-5.85 \pm 0.08$ | $-5.7 \pm 0.3$ | $-5.7 \pm 0.3$ |
| 13 | $-4.189$ | 0.283 | $-5.66 \pm 0.02$ | $-5.82 \pm 0.14$ | $-5.4 \pm 0.5$ | $-6.0 \pm 0.4$ |
| 15 | $-2.080$ | 0.260 | $-9.33 \pm 0.06$ | $-9.26 \pm 0.09$ | $-9.5 \pm 0.2^*$ | $-9.2 \pm 0.2$ |
| 16 | $-0.015$ | 0.632 | $-6.013 \pm 0.010$ | $-6.02 \pm 0.08$ | $-5.8 \pm 0.2$ | $-6.2 \pm 0.3$ |
| 17 | 1.044 | 0.408 | $-4.731 \pm 0.015$ | $-4.61 \pm 0.12$ | $-4.2 \pm 0.3$ | $-4.8 \pm 0.4$ |
| 18 | 3.142 | 0.318 | $-6.200 \pm 0.013$ | $-6.19 \pm 0.13$ | $-6.3 \pm 0.3$ | $-6.5 \pm 0.4$ |
| 19 | 5.286 | 0.268 | $-4.578 \pm 0.014$ | $-4.46 \pm 0.08$ | $-4.43 \pm 0.15$ | $-4.6 \pm 0.3$ |
| 20 | 7.040 | 0.277 | $-11.0 \pm 0.12^*$ | - | - | - |
| 21 | 8.935 | 0.261 | $-8.60 \pm 0.14$ | $-7.7 \pm 0.6^*$ | - | $-8.2 \pm 1.0^*$ |

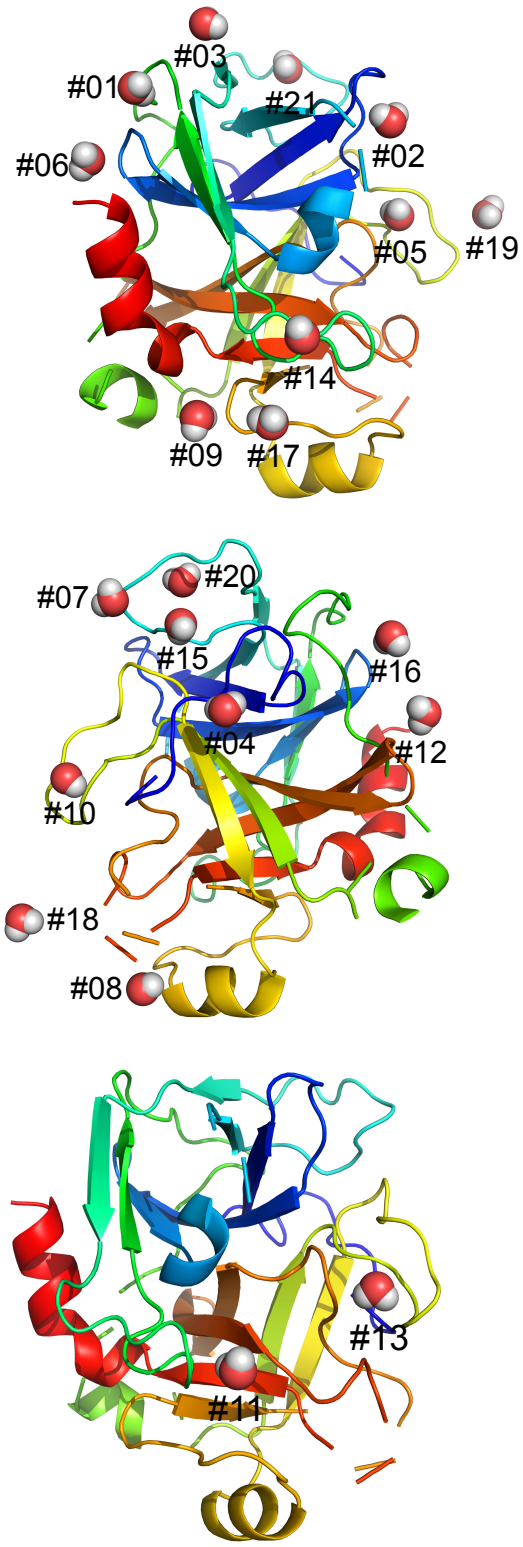
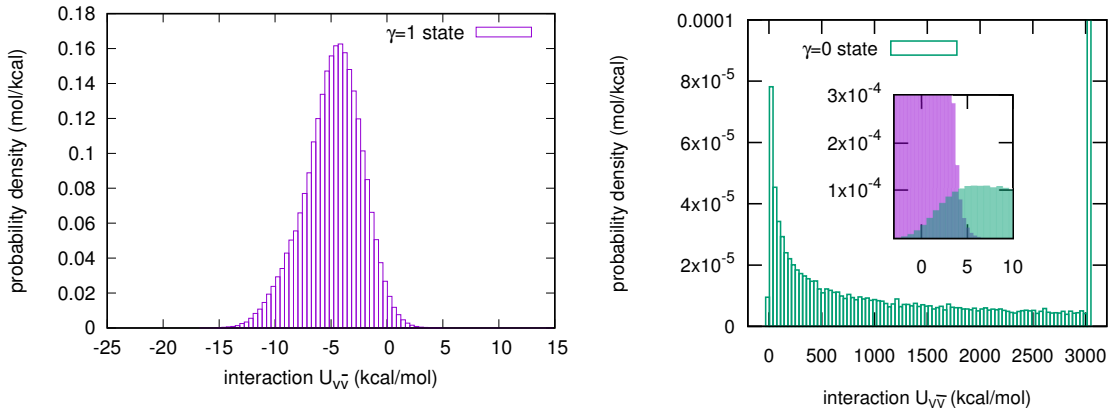



Figure 1: Positions of tagged water molecules. The last picture shows the positions of the two tagged water molecules in the binding site.

# Probability Density of $U_{v\bar{v}}$ for the Inhomogeneous Liquid



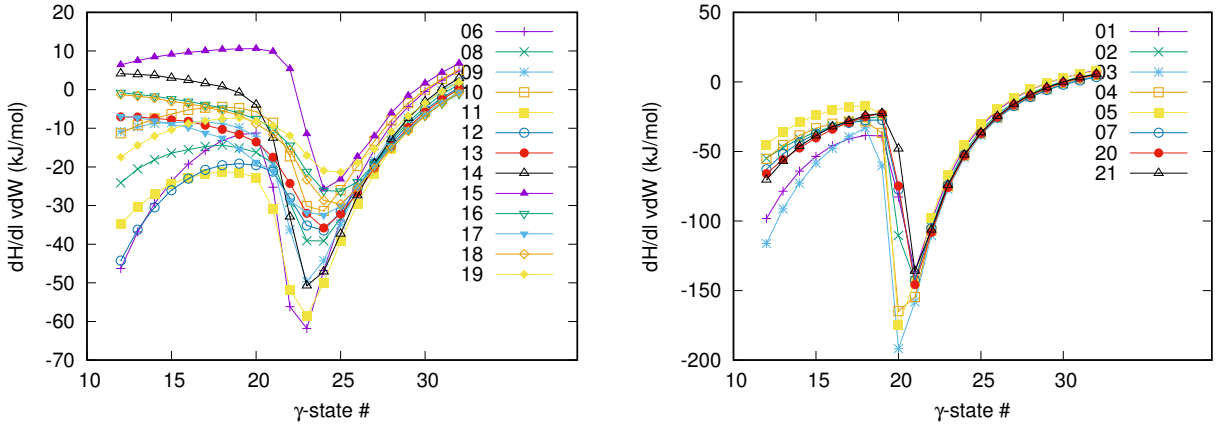

(a) $\gamma = 1$ state (b) $\gamma = 0$ state

Figure 2: Probability density of $U_{v\bar{v}}$ at the $\gamma = 0$ state and the $\gamma = 1$ state for the inhomogeneous liquid (water #6 in Table.1).

# Solvation Free Energy Estimated from Three $\gamma$-states

Table 2: Solvation free energy ($kcal/mol$) for growing a water molecule in a solution containing a protein molecule (FXa) and 15951 water molecules. The data in the $\Delta F(32$ states) column are the benchmarks. The $\Delta F(3$ states) column shows the estimates when the $10\,ns$ simulation data at the $\gamma = 1$ state, the $10\,ns$ simulation data at the intermediate state ($\gamma_e = 1.0$ and $\gamma_v = 0.4$) and the $100\,ns$ simulation data at the $\gamma = 0$ state are UWHAMMed. The uncertainties marked with star are the lower limits of the uncertainty.

| # | $\epsilon$ | $\Delta F$ (32 states) | $\Delta F(3$ states) |
|---|---|---|---|
| 1 | $-22.684$ | $-13.84 \pm 0.09$ | $-13.88 \pm 0.15$ |
| 2 | $-21.011$ | $-12.7 \pm 0.17^*$ | $-12.6 \pm 0.2$ |
| 3 | $-19.093$ | $-11.0 \pm 0.4^*$ | $-11.3 \pm 0.4^*$ |
| 4 | $-17.471$ | $-12.7 \pm 0.13^*$ | $-12.56 \pm 0.15$ |
| 5 | $-16.157$ | $-11.40 \pm 0.07$ | $-11.30 \pm 0.09$ |
| 6 | $-14.444$ | $-10.36 \pm 0.03$ | $-10.39 \pm 0.06$ |
| 7 | $-12.760$ | $-11.43 \pm 0.08$ | $-11.47 \pm 0.12$ |
| 8 | $-11.111$ | $-9.033 \pm 0.014$ | $-8.96 \pm 0.05$ |
| 9 | $-9.627$ | $-8.782 \pm 0.016$ | $-8.80 \pm 0.05$ |
| 10 | –7.834 | $-9.095 \pm 0.013$ | $-9.07 \pm 0.03$ |
| 11 | $-6.40$ | $-7.774 \pm 0.017$ | $-7.80 \pm 0.05$ |
| 12 | $-5.777$ | $-8.895 \pm 0.012$ | $-8.89 \pm 0.05$ |
| 13 | $-4.74$ | $-5.765 \pm 0.011$ | $-5.77 \pm 0.05$ |
| 14 | $-4.189$ | $-5.66 \pm 0.02$ | $-5.63 \pm 0.05$ |
| 15 | $-2.080$ | $-9.33 \pm 0.06$ | $-9.55 \pm 0.09$ |
| 16 | $-0.015$ | $-6.013 \pm 0.010$ | $-6.04 \pm 0.05$ |
| 17 | 1.044 | $-4.731 \pm 0.015$ | $-4.60 \pm 0.06$ |
| 18 | 3.142 | $-6.200 \pm 0.013$ | $-6.12 \pm 0.06$ |
| 19 | 5.286 | $-4.578 \pm 0.014$ | $-4.56 \pm 0.06$ |
| 20 | 7.040 | $-11.0 \pm 0.12^*$ | $-10.98 \pm 0.16$ |
| 21 | 8.935 | $-8.60 \pm 0.14$ | $-8.7 \pm 0.3$ |




(a) easy (b) difficult and challenging

Figure 3: $\frac{dH}{d\gamma}$ of each $\gamma$-state for thermodynamic integration.

In Fig.3, we plot $\frac{dH}{d\gamma}$ versus the number of the $\gamma$-state index. The values of $\frac{dH}{d\gamma}$ are provided by GROMACS for thermodynamic integration (TI) calculations during the free energy perturbation (FEP) simulations. Although we did not do TI calculations in this study, the values of $\frac{dH}{d\gamma}$ offer useful suggestion to choose the third $\gamma$-states for three-states FEP calculations. As can be seen in Fig.3, $\frac{dH}{d\gamma}$ has the largest magnitude at the 20th ($\gamma_e = 1.0$ and $\gamma_v = 0.4$) or the 21th ($\gamma_e = 1.0$ and $\gamma_v = 0.45$) $\gamma$-states for the water positions in the difficult or challenging category. Note that in GROMACS, $\gamma_e = 1$ and $\gamma_v = 1$ corresponds to the $\gamma$-state at which the interactions between the tagged water molecule and the other molecules are totally turned off.

## Uncertainty Estimates

In this study, the uncertainties of the estimates of solvation free energy were estimated by the block bootstrap method. The raw data were divided into blocks before performing bootstrap analyses and each block contains $5,000$ data points ($0.5\,ns$ simulation). When estimating the uncertainty of $\Delta F^{**}$, which are the estimates of solvation free energy when the $0.1\,ns$ simulation data at the $\gamma = 1$ state and the $100\,ns$ simulation data at the $\gamma = 0$ state are UWHAMMed, the input data at the $\gamma = 1$ state were blocks (continuous $0.1\,ns$ simulation data) randomly chosen from the original $10\,ns$ simulation data at the $\gamma = 1$ state.

However, for some water positions in the difficult and challenging categories, although UWHAM converged by using the original simulation data, UWHAM failed to estimate the solvation free energy by using the input data constructed by bootstrapping, namely re-sampling. In such cases, UWHAM either cannot converge or provides an estimate which extremely differs from the one by using the original simulation data. The estimates based on those input data constructed by resampling have not been included for the calculation of uncertainty. Therefore, the corresponding uncertainty estimates are the lower limits of the true uncertainty, and they are marked with stars in the tables.