

Supporting Information for Nonpolar Solvation Free Energy from Proximal Distribution Functions

Shu-Ching Ou, Justin A. Drake, and B. Montgomery Pettitt*

*Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical
Branch, 301 University Blvd, Galveston, TX 77555-0304, United States*

E-mail: mpettitt@utmb.edu

Phone: (409)772-0723

Illustration of Voronoi Polyhedra

In the manuscript, we use

$$g_{\perp}^k(r) = \sum_{t=0}^T \sum_{j=1}^n \frac{\delta(\text{Inf}[|\vec{r}_{ij}(t)|]_{i=1,m} - r)}{\delta\tau(\vec{r}_j(t), k)} \quad (\text{S1})$$

to define the proximal distribution function $g_{\perp}(r)$. We use Figure S1 to illustrate this process. Assume the solute has four atoms ($i = 1 \sim 4$). The rest of the space can be divided into Voronoi polyhedra ($k = \text{I} \sim \text{IV}$) depending on the closest solute atom. If the j_{th} solvent molecule locates in k_{th} polyhedron, the distance between i_{th} solute atom and the j_{th} solvent molecule (marked as r_{ij}) is then recorded for $g_{\perp}(r)$ pair correlation calculation. In other words, the polyhedra define the sampling volume for the corresponding solute atoms. The solvent distribution is considered isotropic within each sampling space.

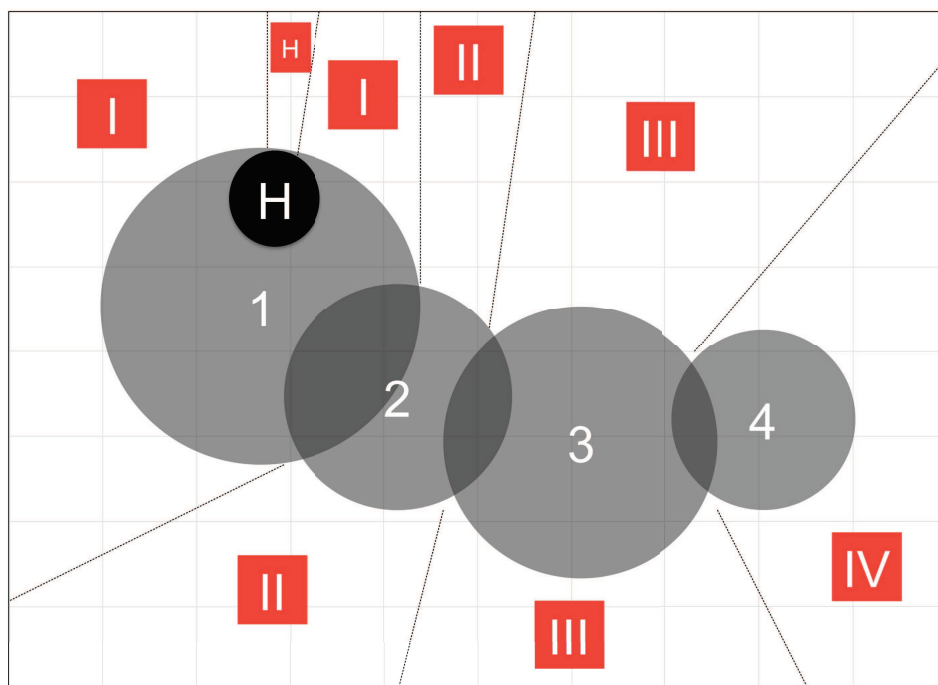


Figure S1: Illustration of Equation 1 in the manuscript. $i = 1 \sim 4$ represents the solute atoms, with $k = \text{I} \sim \text{IV}$ the corresponding Voronoi polyhedra sampling volume. The case of an atom enveloped within another is demonstrated with atom “H”. The exclusion factors are discussed below in this context.

Cavity & Exclusion Factors

During the pDF-reconstructions, we used two factors δ_{cav} and δ_{excl} in order to remove the grid points that overlap with solute atoms. The total van der Waals solute-solvent interaction can be written as

$$U^{\text{vdW}} = \sum_j^n \sum_i^m U_{ij}^{\text{vdW}} \quad (\text{S2})$$

where i loops over solute atoms and j loops over the solvent molecules. m is the number of atoms in the solute and n is the number of solvent molecules. Thus,

$$\frac{\partial U^{\text{vdW}}}{\partial \lambda} = \frac{\partial \left(\sum_j^n \sum_i^m U_{ij}^{\text{vdW}} \right)}{\partial \lambda} \quad (\text{S3})$$

We can take a representative term from the above equation

$$\frac{\partial U_i^{\text{vdW}}}{\partial \lambda} = \frac{\partial \left(\sum_j^n U_{ij}^{\text{vdW}} \right)}{\partial \lambda} \quad (\text{S4})$$

For this value, we define a variable Δ_i as the difference between simulated and pDF-reconstructed of i_{th} solute atom,

$$\Delta_i = \left(\frac{\partial U_i^{\text{vdW}}}{\partial \lambda} \right)_{\text{Reconstructed}} - \left(\frac{\partial U_i^{\text{vdW}}}{\partial \lambda} \right)_{\text{Simulated}} \quad (\text{S5})$$

We can then compare Δ_i for each solute atom with both factors applied and only one of the factors applied during the pDF-reconstructions. We choose the reconstruction of ala₁ as the example. The results are presented in Figure S2.

When $\lambda = 0.1$, Δ_i are within 0.5 kcal/mol, and the results using both factors are well reproduced. The exclusion factor δ_{excl} does not seem to be crucial when the cavity volume is already well-defined by the cavity factor δ_{cav} . Without δ_{cav} , it is possible to reconstruct solvent densities for grid points laid between two different solute atoms since all $g_{\perp}(r)$ become non-zero at small r when λ is small. On the other hand, when $\lambda = 1$, both δ_{cav} or δ_{excl}

itself are required to reproduce the simulated $\langle \partial U_i^{\text{vdW}} / \partial \lambda \rangle$. Therefore, we keep both factors during our reported pDF-reconstructions.

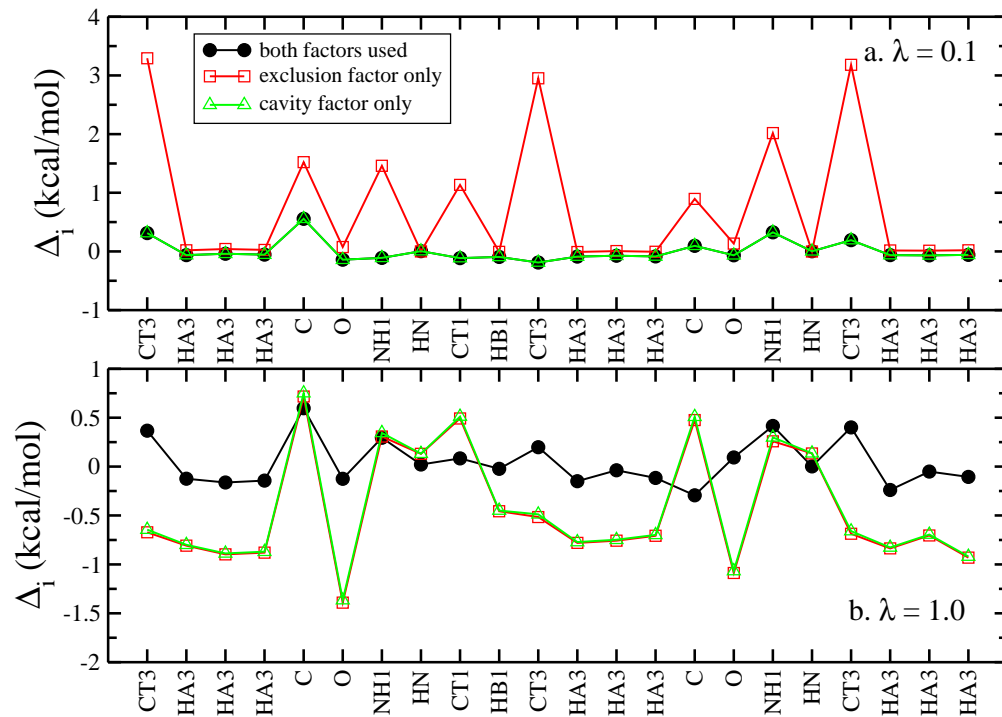
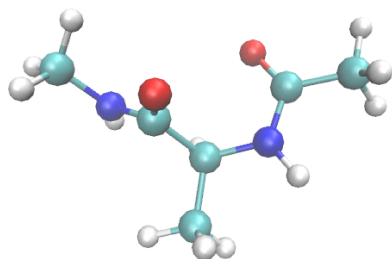


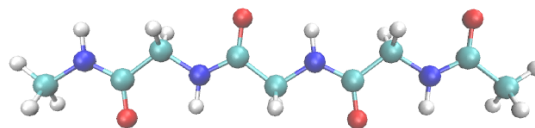
Figure S2: Comparison of Δ_i using both factors or only one factor during the pDF-reconstruction, for $\lambda = 0.1$ and 1.

Solute Configurations

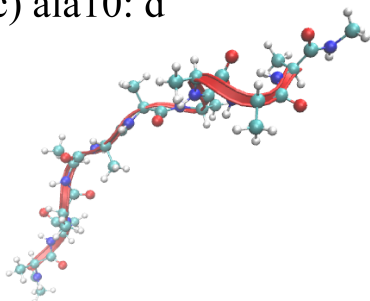
(a) ala1



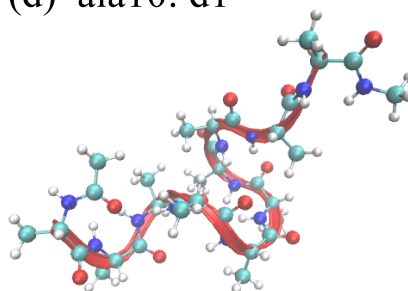
(b) gly3



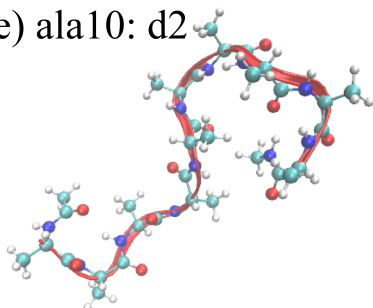
(c) ala10: d



(d) ala10: d1



(e) ala10: d2



(f) ala10: d3

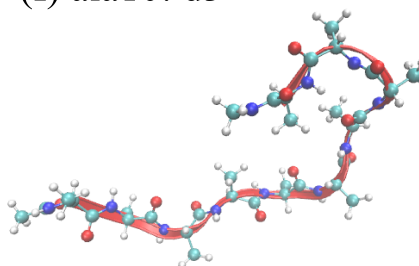


Figure S3: Solute configurations used in fixed simulations. For ala₁₀, peptide backbones are represented with red ribbons. For convenience, we designate d, d1, d2 and d3 for the configuration of deca-alanines respectively. Configurations are visualized using VMD.¹

pDFs for Solute Atoms with Water Hydrogen

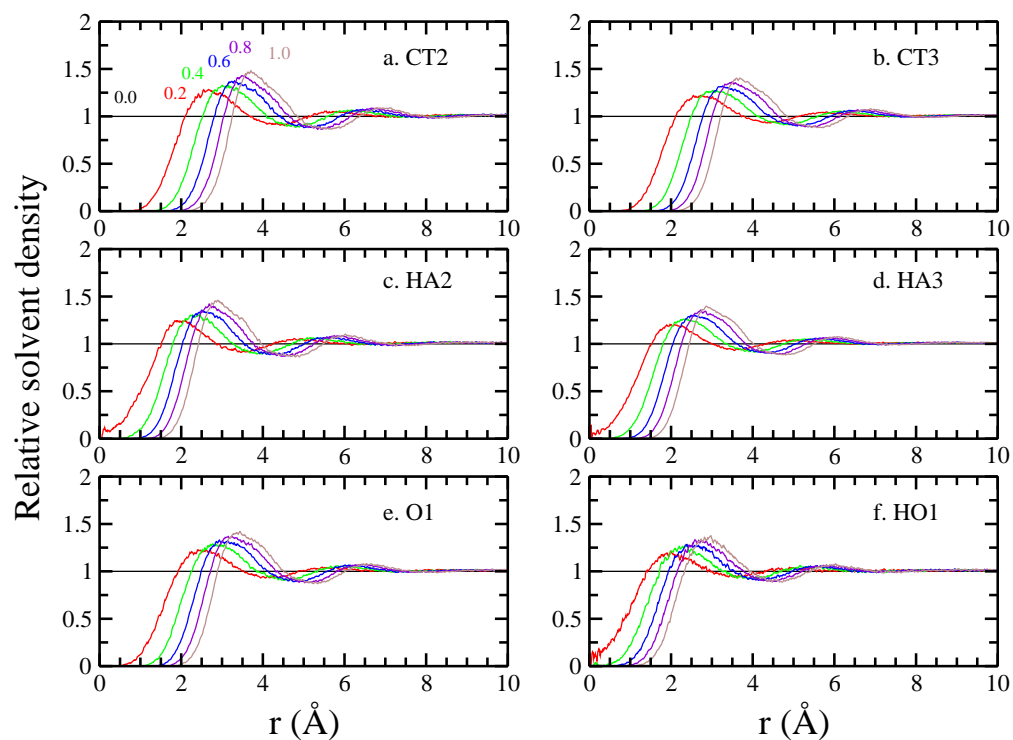


Figure S4: pDF of butane/propanol atoms to water hydrogen atoms at different λ . Numbers above each pDF indicate the corresponding λ of this atom type.

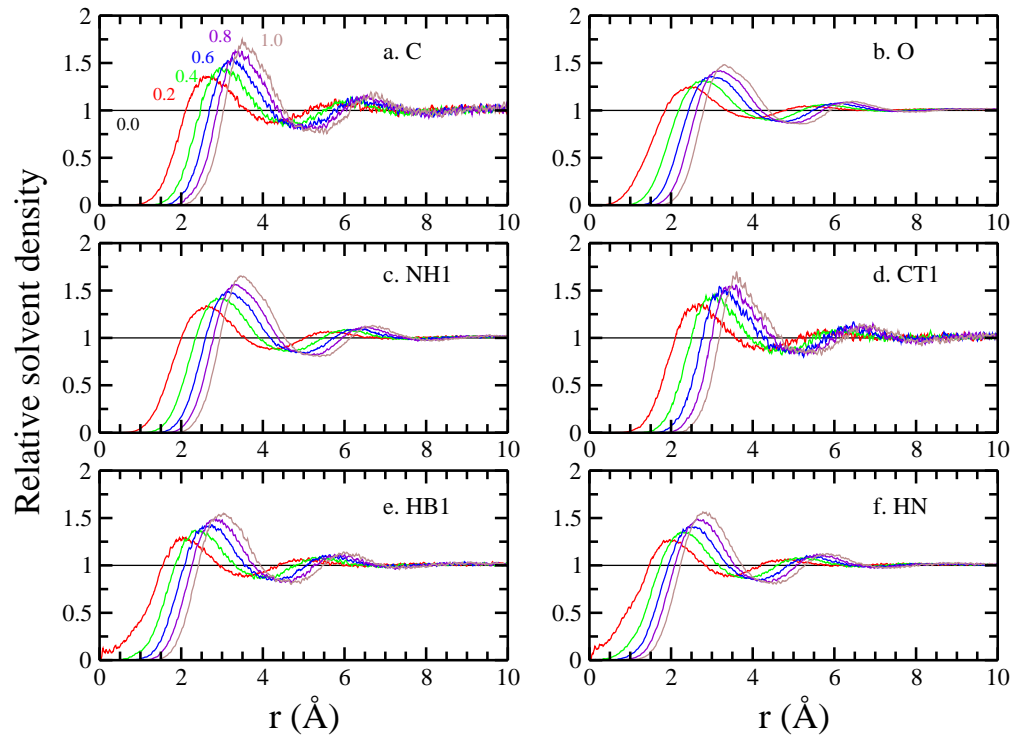


Figure S5: pDF of alanine/glycine atoms to water hydrogen atoms at different λ . Numbers above each pDF indicate the corresponding λ of this atom type.

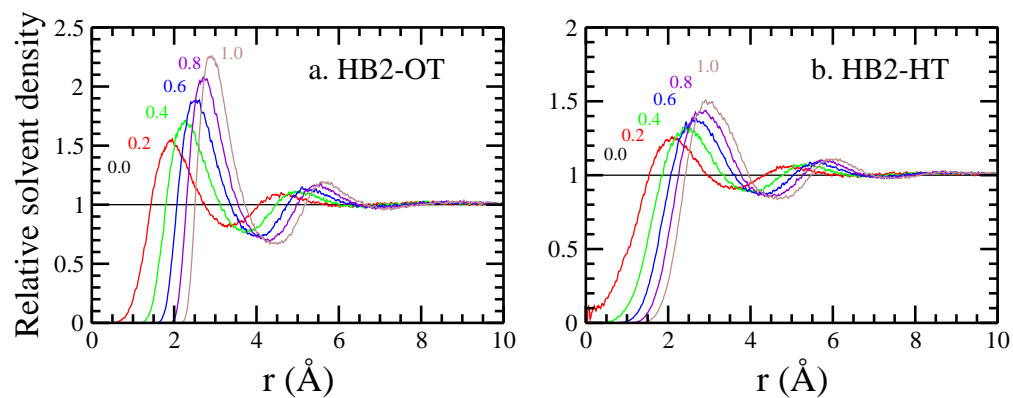


Figure S6: pDF of HB2 atoms to water oxygen/hydrogen atoms at different λ . Numbers above each pDF indicate the corresponding λ of this atom type.

Convergence: $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ and λ Spacing

In this section we address the convergence issues. Figure S7 shows the $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ as a function of sampling time at different λ windows, for the case of ala₁₀ (configuration d), since the deca-alanine is the largest solute throughout this study. In all cases, $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ is certainly converged within 2 ns. One might argue that the uncertainties for small λ cases (~ 1) are still large relative the uncertainties for larger λ (~ 0.2), but we should realize that these uncertainties are already within 1% of the reported $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$.

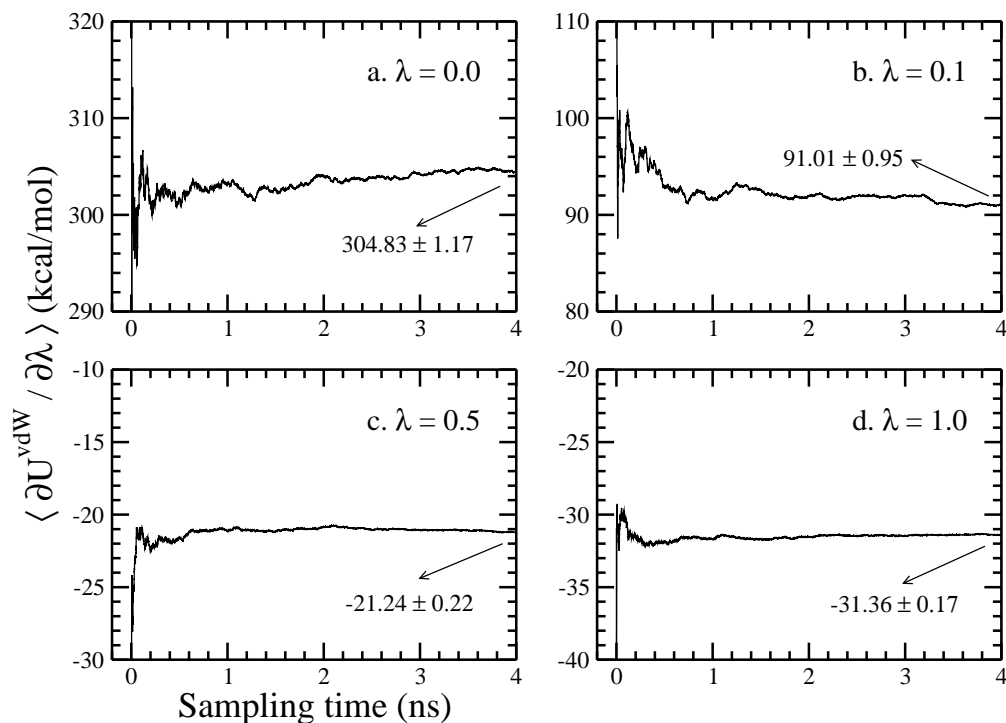


Figure S7: $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ as a function of sampling time. The values with uncertainties indicate the simulated $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ reported in the main text.

Due to the nature of simulating the soft-core potential at $\lambda = 0$ we simply zero the value of the correlation for our reported studies. Here we present the results of simulated $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ and ΔG_{vdW} using different $\Delta \lambda$ in Figure S8. We use error propagation to estimate the uncertainties of ΔG_{vdW} , thus the uncertainty reduces as the number of λ de-

creases. Compared with the $\Delta\lambda = 0.05$ case, ΔG_{vdW} becomes more than doubled when $\Delta\lambda = 0.2$. When $\Delta\lambda = 0.1$, the ΔG_{vdW} is overestimated with 1.82 kcal/mol, which is already within the uncertainty of ΔG_{vdW} obtained from $\Delta\lambda = 0.05$ case. Therefore, to simplify the illustration of our results and focus on the method itself, we chose the case of $\Delta\lambda = 0.1$ in our main text.

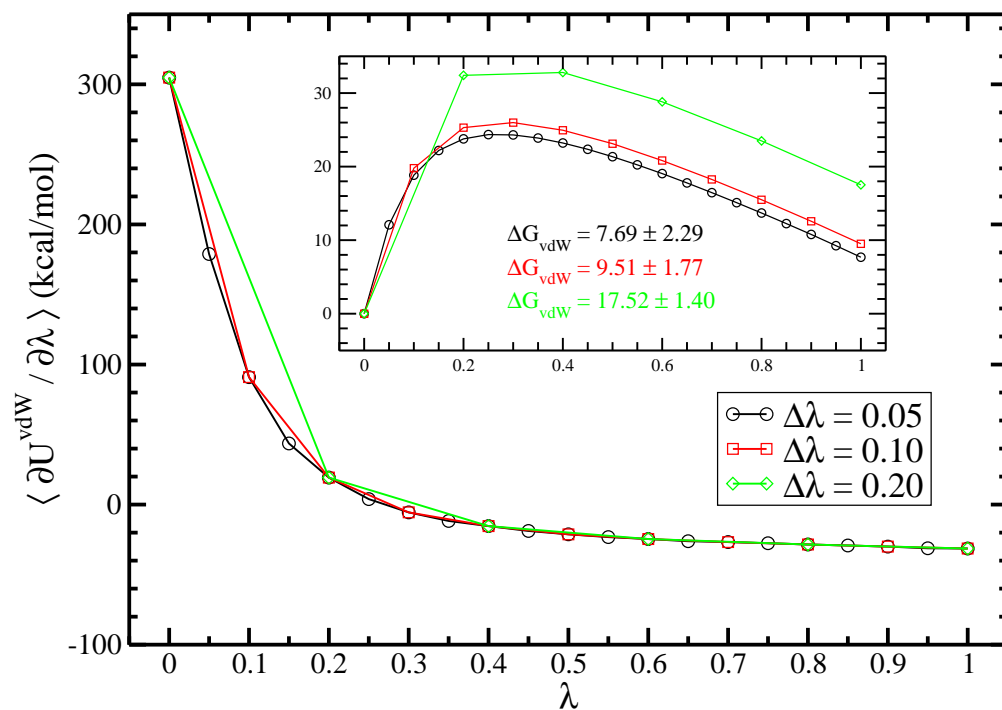


Figure S8: Simulated $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ using the λ spacings of 0.05, 0.10, and 0.20. The inset shows the corresponding cumulative ΔG_{vdW} . Uncertainties are estimated using the propagation of block standard errors at the λ windows.

pDF-Reconstructions for Ala₁₀

Simulated and pDF-reconstructed $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ at each λ window for different configuration of ala₁₀.

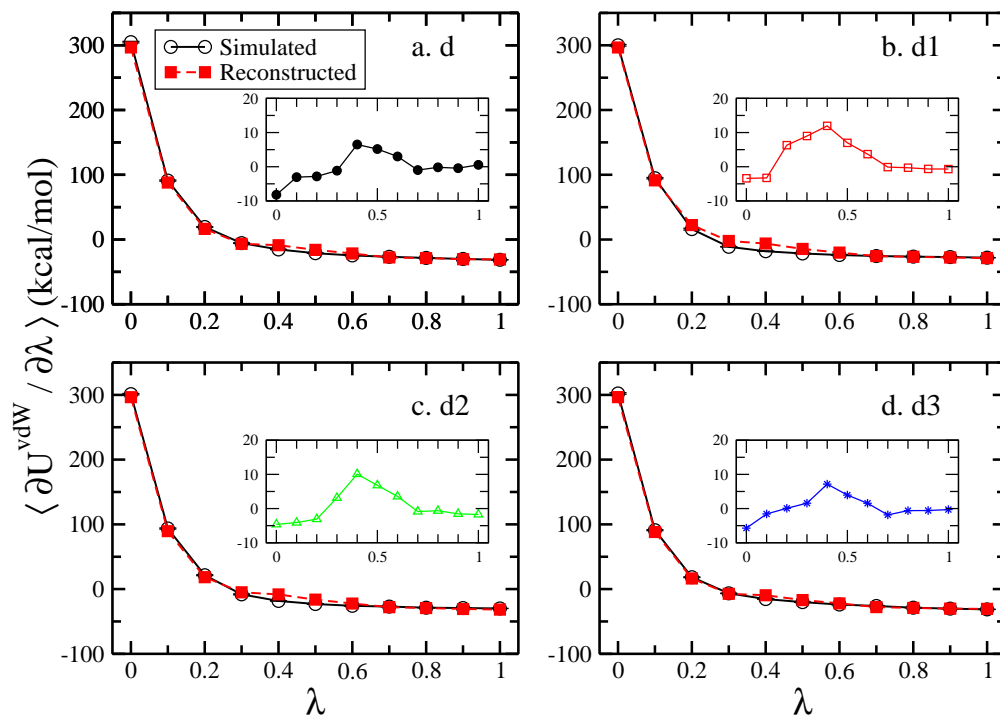


Figure S9: Simulated and pDF-reconstructed $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ at each λ window for different configuration of ala₁₀. The inset of each panel shows the difference between the simulated and pDF-reconstructed $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$.

References

- (1) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.