Supporting information for: Standard binding free energy of a SIM SUMO complex

Alex Kötter,[*,†,‡] Henning D. Mootz,[¶] and Andreas Heuer[*,†,‡]

†*Institut für physikalische Chemie, Westfälische Wilhelms-Universität Münster, Corrensstraße 28/30, D-48149 Münster, Germany*

‡*Center for Multiscale Theory and Computation, Westfälische Wilhelms-Universität Münster, Corrensstraße 40, D-48149 Münster, Germany*

¶*Institut für Biochemie, Westfälische Wilhelms-Universität Münster, Wilhelm-Klemm-Straße 2, D-48149 Münster, Germany*

E-mail: a_koet04@wwu.de; andheuer@wwu.de

Theory

We start with a brief summary of the theoretical framework for the calculation of binding free energies, based on the work of Woo and Roux. However, we present a slightly simplified derivation. Starting point is the expression of the equilibrium constant in case of a dilute protein solution:

$$K_{eq} = \frac{N}{[L]} \frac{\int_{bound} e^{-\beta U(x)} dx}{\int_{free} e^{-\beta U(x)} dx}. \tag{1}$$

Here N and $[L]$ are number and concentration of the ligand and x are the positions of all atoms of the system and U is the potential energy. By definition $bound = \{r(x) < r_b\}$ and $free = \{r(x) \geq r_b\}$ for some value r_b. It may be chosen, such that interaction between

protein and ligand vanishes at r_b. We may then write

$$K_{eq} = \frac{N}{[L]} \frac{\int_0^{r_b} p(r)dr}{\int_{r_b}^{r_\infty} p(r)dr} \tag{2}$$

by transformation of variables and marginalizing. $p(r)$ is the probability density of the distance r and r_∞ is the largest possible value of r. The key point of Woo and Roux's method is to transform the system of interest into one, in which $p(r)$ can be sampled reliably. To this end, position, orientation and configuration of the ligand are defined using collective variables as discussed in the Methods section. Quadratic potentials are then imposed on the collective variables to restrain them to reference values in the bound state. The restraining potentials read $u(z) = k/2(z - z_0)^2$, where k is a force constant, z is the respective collective variable and z_0 is its reference value. In present case, reference values are the values of the collective variables after 100 ns of unrestrained simulation.

The total potential in the restrained system is given by $\tilde{U} = U + u_c(RMSD) + u_a(\phi, \theta) + u_o(\Theta, \Phi, \Psi)$, where the subscripts indicate restraints on configuration (c), angular part of position (a) and orientation (o). Note that here the potential on angular part of the position and orientation comprise of the restraints on θ and ϕ and Θ, Φ and Ψ respectively. I.e., $u_a(\phi, \theta) = k/2(\phi - \phi_0)^2 + k/2(\theta - \theta_0)^2$ and $u_o(\Theta, \Phi, \Psi) = k/2(\Theta - \Theta_0)^2 + k/2(\Phi - \Phi_0)^2 + k/2(\Psi - \Psi_0)^2$, while $u_c(RMSD) = k/2(RMSD)^2$. In the following discussion this reduced notation is used. Therefore contributions of the angular and orientational restraints are actually the contributions of the 2 and 3 angles mentioned above. We then define

$$\tilde{K}_{eq} = \frac{N}{[L]} \frac{\int_{bound} e^{-\beta \tilde{U}(x)} dx}{\int_{free} e^{-\beta \tilde{U}(x)} dx} = \frac{N}{[L]} \frac{\int_0^{r_b} \tilde{p}(r)dr}{\int_{r_b}^{r_\infty} \tilde{p}(r)dr} \tag{3}$$

and introduce shorthand notations

$$I_b^{-1}(u_r, W) = \frac{\int_{bound} e^{-\beta W} dx}{\int_{bound} e^{-\beta(W + u_r)} dx} \tag{4}$$

$$I_f(u_r, W) = \frac{\int_{free} e^{-\beta(W+u_r)} dx}{\int_{free} e^{-\beta W} dx} \tag{5}$$

for some potential W and $r \in \{a, o, c\}$. It is then easy to see that for the relationship between K_{eq} and $\tilde{K}_{eq}$ one obtains

$$\begin{aligned} K_{eq} = I_b^{-1}(u_c, U) \times I_b^{-1}(u_o, U + u_c) \times I_b^{-1}(u_a, U + u_c + u_o) \times \\ \tilde{K}_{eq} \times \underline{I_f(u_a, U + u_c + u_o)} \times \underline{I_f(u_o, U + u_c)} \times I_f(u_c, U). \end{aligned} \tag{6}$$

Here the underlined terms are accessible by numerical integration (see below). For reasons of brevity we rewrite this as

$$K_{eq} = K_{b,c}^{-1} \times K_{b,o}^{-1} \times K_{b,a}^{-1} \times \tilde{K}_{eq} \times \underline{K_{f,a}} \times \underline{K_{f,o}} \times K_{f,c}. \tag{7}$$

In this sense, introducing the restraints is related to an additional binding constant or free energy change $\Delta G_r^f = -RT \log(K_{f,r})$ and $\Delta G_r^b = -RT \log(K_{b,r}^{-1})$ respectively. We start by evaluating $\tilde{K}_{eq}$. For $r \geq r_b$ we may write

$$\tilde{p}(r) = \frac{\tilde{p}(r_b)}{r_b^2} r^2. \tag{8}$$

Plugging this in into the expression for $\tilde{K}_{eq}$ and using $N/[L] = V$, where V is the system volume we obtain:

$$\tilde{K}_{eq} = V \frac{\int_0^{r_b} \tilde{p}(r) dr}{\int_{r_b}^{r_\infty} \tilde{p}(r) dr} = \frac{V r_b^2}{\tilde{p}(r_b)} \frac{\int_0^{r_b} \tilde{p}(r) dr}{\int_{r_b}^{r_\infty} r^2 dr} = \frac{4\pi r_b^2}{\tilde{p}(r_b)} \int_0^{r_b} \tilde{p}(r) dr \tag{9}$$

We proceed with the remaining factors. As an example, consider the term $I_f(u_c, U)$.

$$I_f(u_c, U) = \frac{\int_{free} e^{-\beta(U(x)+u_c(RMSD(x))} dx}{\int_{free} e^{-\beta U(x)} dx} = \int_{free} e^{-\beta u_r(RMSD)} p_{free}(RMSD) dRMSD, \tag{10}$$

where $p(RMSD)$ is the probability density of the $RMSD$ in the free state. We remark that it is common to work with the PMF of a reaction coordinate z, $w(z) = -RT\log(p(z))$, instead of working directly with the probability density. In general the probability densities of the collective variables have to be estimated from simulation data. For the contributions of the angles in the free state however, no simulation is needed, since it is natural to assume that $\phi(x), \theta(x), \Theta(x), \Phi(x)$ and $\Psi(x)$ are uniformly distributed in the free state. Change of variables then results in the probability densities

$$p_{free}(\theta, \phi) = \frac{1}{4\pi}\sin(\theta),\ p_{free}(\Theta, \Phi, \Psi) = \frac{1}{8\pi^2}\sin(\Theta). \tag{11}$$

The contribution of restraining these angles is then accessible via numerical integration, for example:

$$K_{f,a} = \frac{\int_{free} e^{-\beta(U+u_c+u_0+u_a)}dx}{\int_{free} e^{-\beta(U+u_c+u_0)}dx} = \int_0^{\pi}\int_0^{2\pi} e^{-\beta u_a(\theta,\phi)}\frac{1}{4\pi}\sin(\theta)d\phi d\theta \tag{12}$$

We remark that Woo and Roux denote $\tilde{K}_{eq}K_{f,a}$ by I^*S^*. If we now partition the bound state into two regions, one that we assign to the parallel (p) and one that we assign to the antiparallel (a) orientation, we may write

$$K_{eq} = \frac{N}{[L]}\frac{\int_{bound} e^{-\beta U(x)}dx}{\int_{free} e^{-\beta U(x)}dx} = \frac{N}{[L]}\left(\frac{\int_{bound,p} e^{-\beta U(x)}dx}{\int_{free} e^{-\beta U(x)}dx} + \frac{\int_{bound,a} e^{-\beta U(x)}dx}{\int_{free} e^{-\beta U(x)}dx}\right). \tag{13}$$

We could now proceed with the forgoing analysis for both summands and get the result shown in equation (6). The difference would be that instead of defining $I_b(u_r, U_{ee})$ we would now define $I_{b,p}(u_r, U_{ee})$ and $I_{b,a}(u_r, U_{ee})$, i.e. we would condition on the set bound $\cap$ parallel (p) and bound $\cap$ antiparallel (a).

Supporting figures

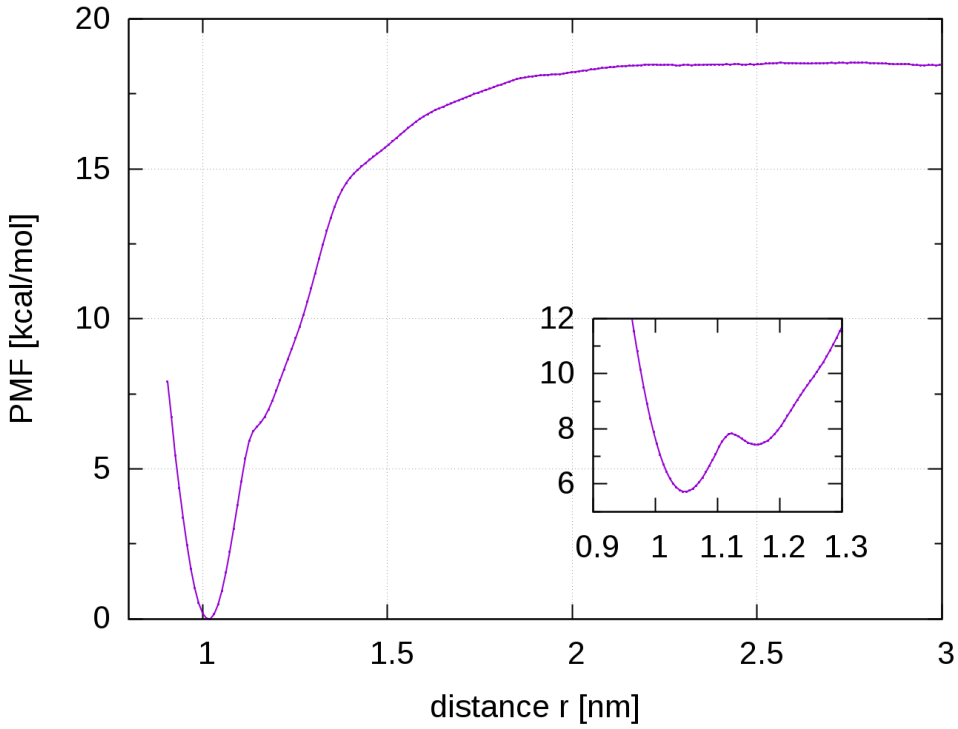

Figure 1: The potential of mean force of the distance of protein and ligand in parallel orientation in units of $kcal/mol$. The inset shows the resulting effective potential in the window at 1.25 nm also in units of $kcal/mol$.

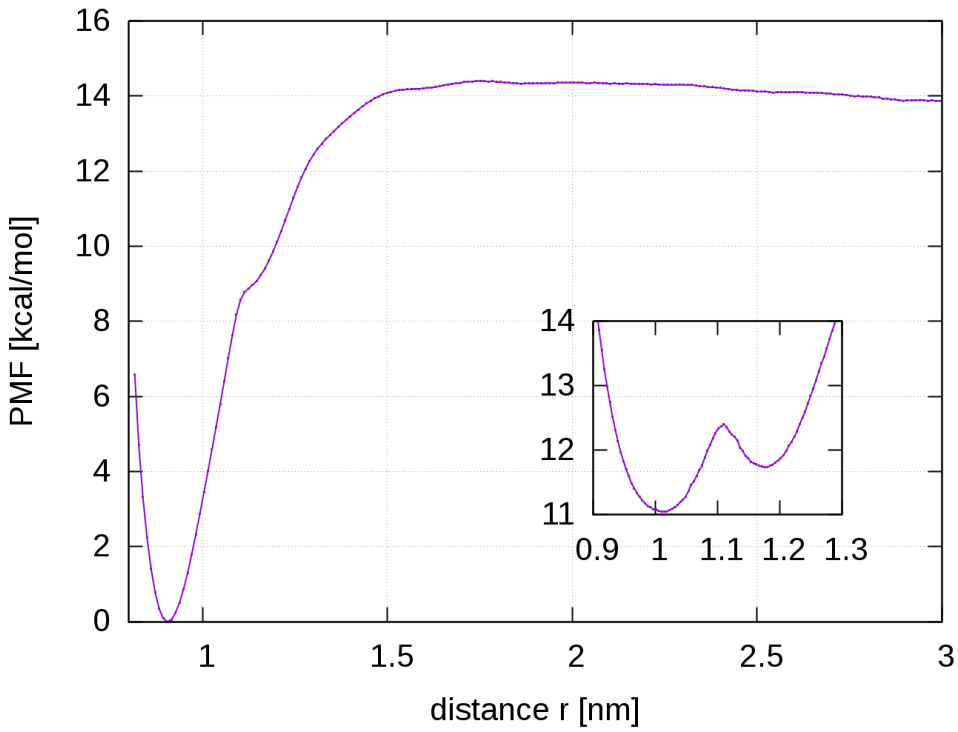

Figure 2: The potential of mean force of the distance of protein and ligand in antiparallel orientation in units of $kcal/mol$. The inset shows the resulting effective potential in the window at 1.25 nm also in units of $kcal/mol$.

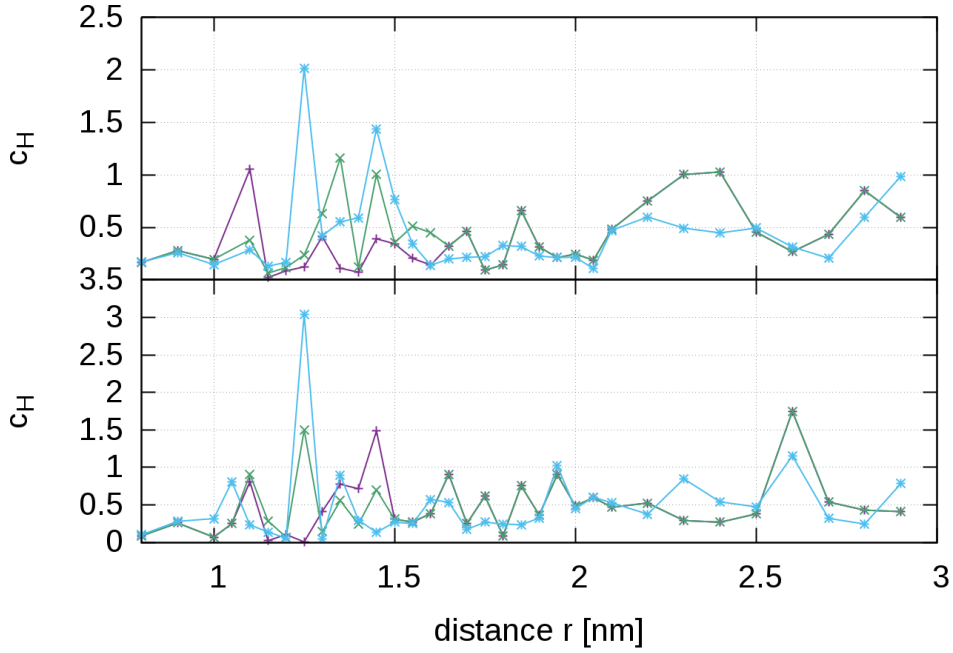

Figure 3: The consistency coefficient for the umbrella sampling of the distance of protein and ligand for parallel orientation (upper figure) and antiparallel orientation (bottom figure) for different simulation times. Blue: 4 ns per window. Green: 40 ns for the critical windows and 20 ns for the remaining windows. Purple: Full trajectories. The coefficient measures the consistency between adjacent windows, the point at 0.8 nm refers to windows at 0.8 and 0.9 nm and so on.

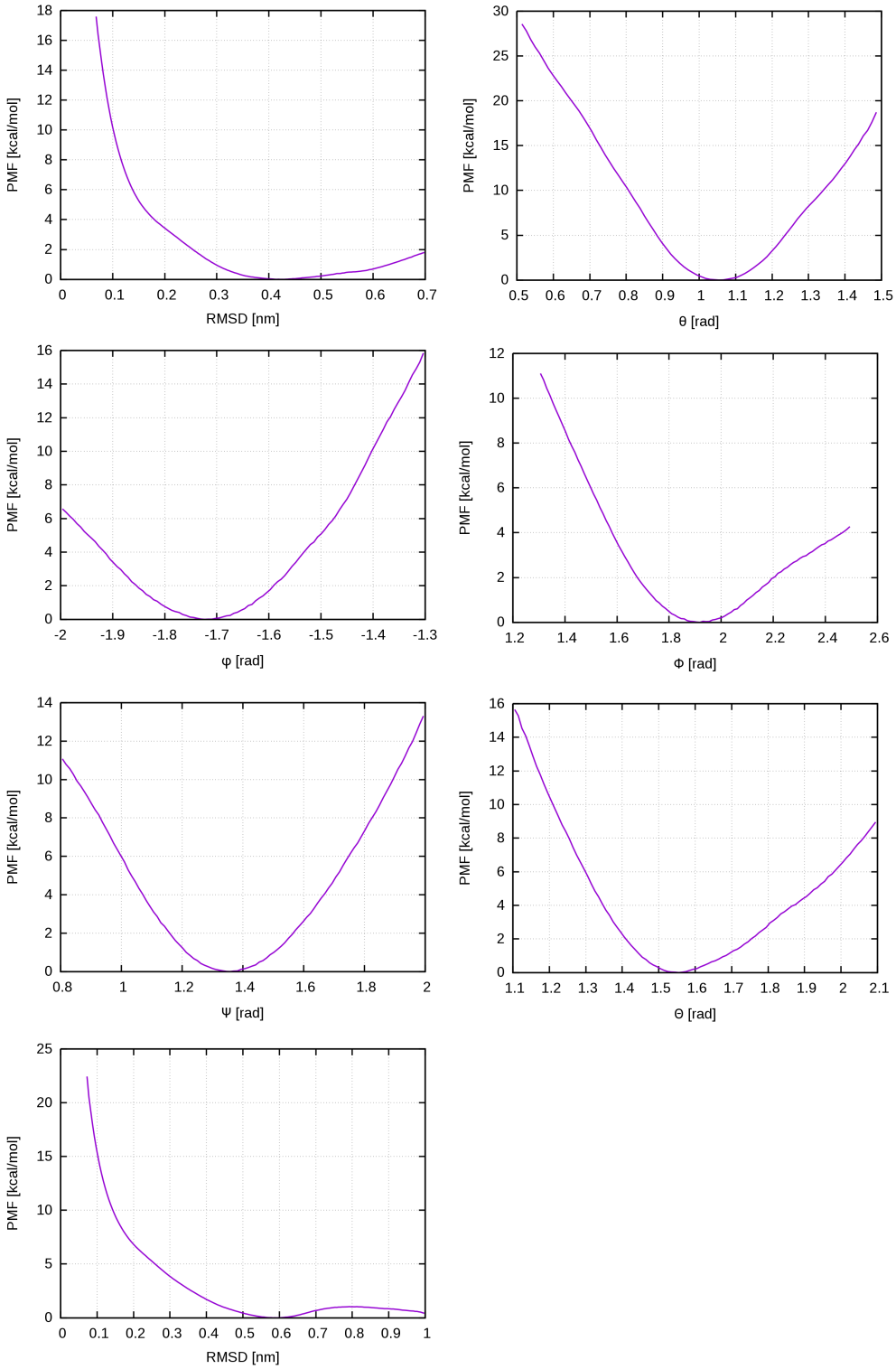

Figure 4: Potentials of mean force of the collevtive variables in parallel orientation. From top left to bottom: $RMSD$ in the bound state, position angles θ and ϕ, orientation angles Φ, Ψ and Θ and $RMSD$ in the free state.

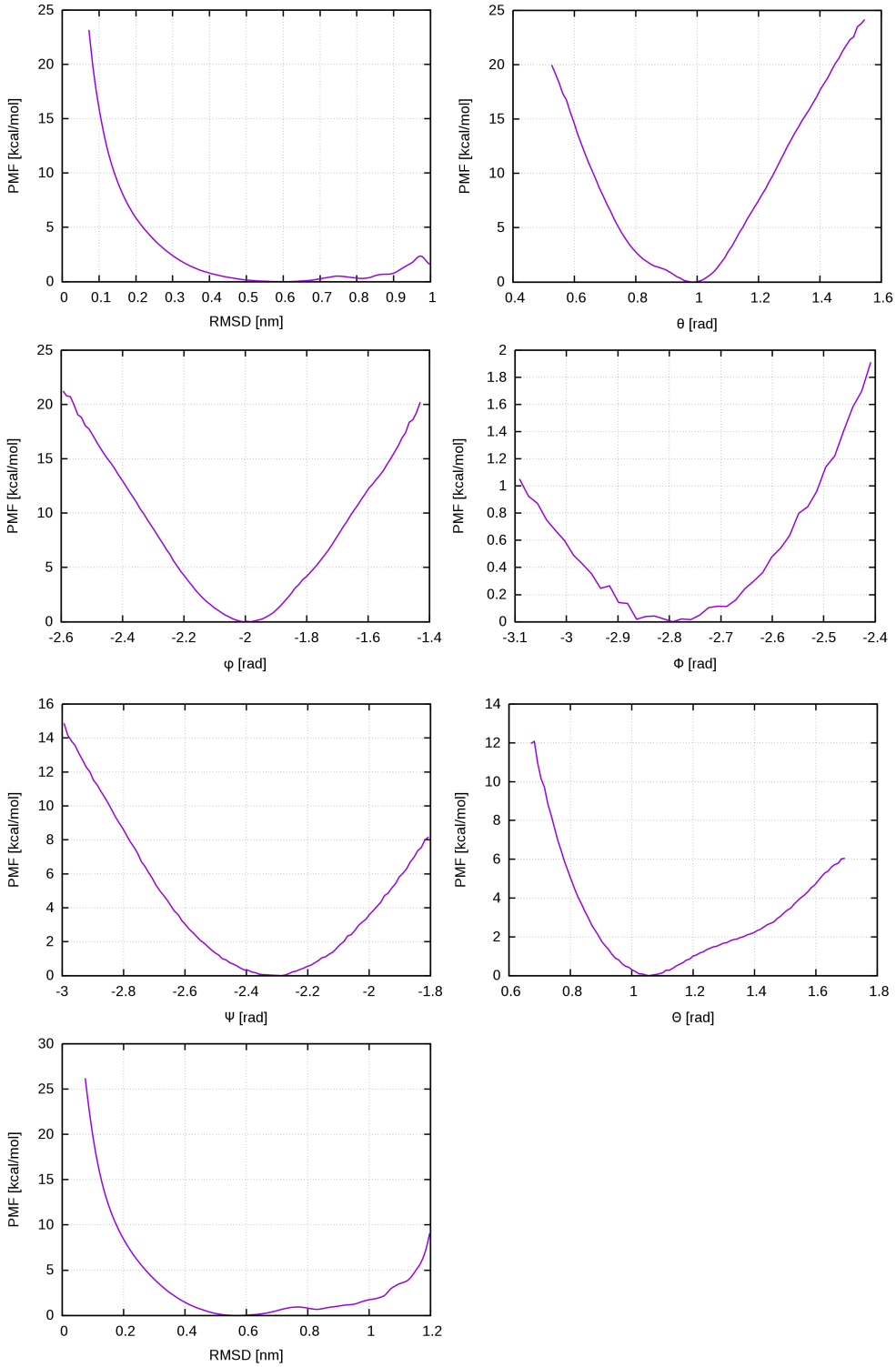

Figure 5: Potentials of mean force of the collevtive variables in antiparallel orientation. From top left to bottom: $RMSD$ in the bound state, position angles θ and ϕ, orientation angles Φ, Ψ and Θ and $RMSD$ in the free state.

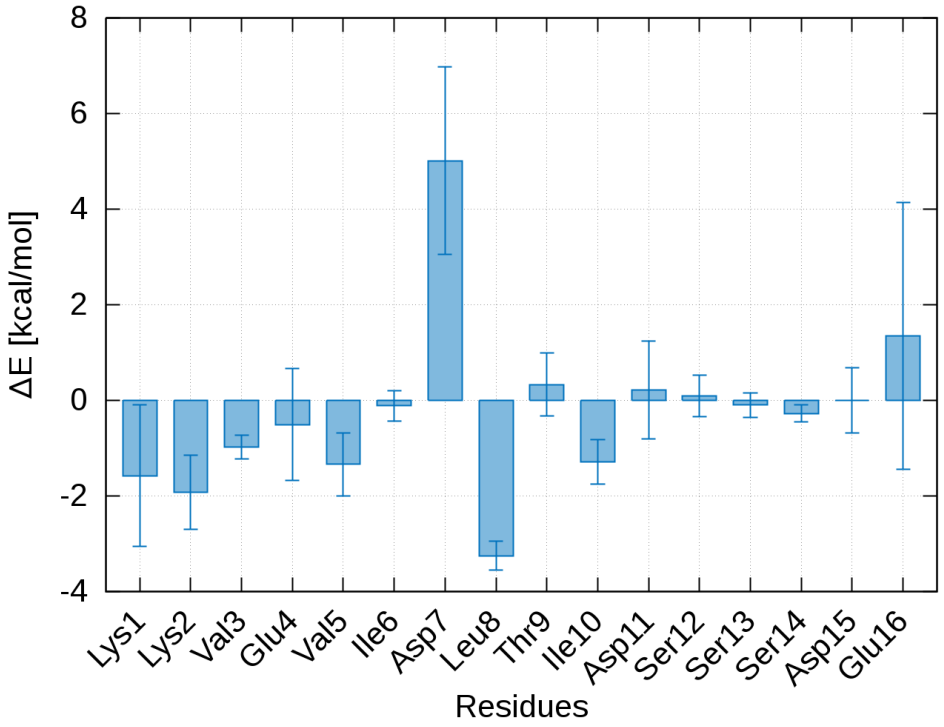

Figure 6: The potential energy differences for each residue of the PIAS1 between bound and free state in antiparallel orientation. We include non bonded interactions to SUMO3, PIAS1 residues and solute.

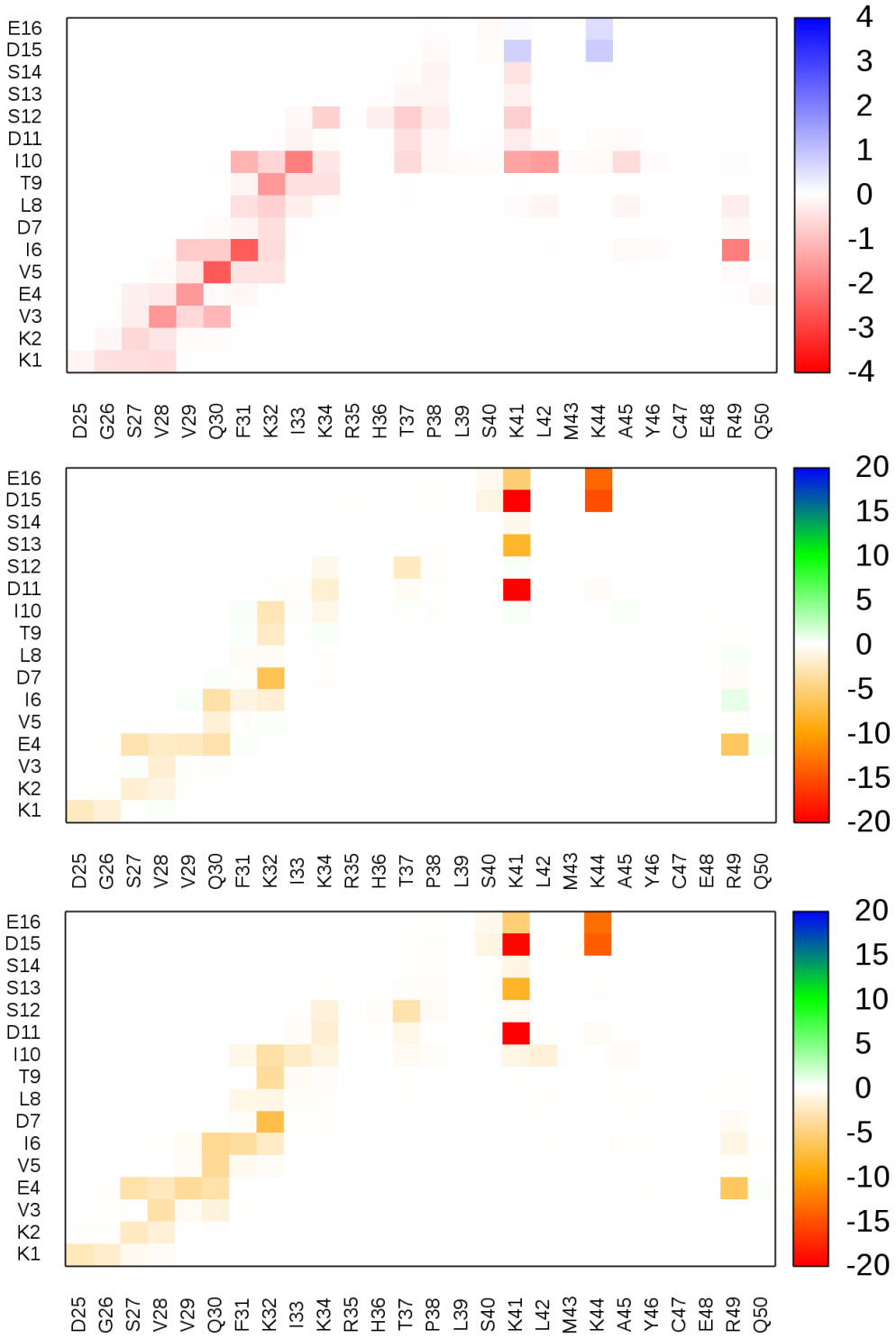

Figure 7: Lennard jones interaction (top), coloumb interaction (middle) and the sum of the former (bottom) between PIAS1 residues (y axis) and SUMO3 residues (x axis) in $kcal/mol$.