Biophysical Journal, Volume *119*

Supplemental Information

Insights into the Microscopic Structure of RNF4-SIM-SUMO Complexes from MD Simulations

Alex Kötter, Henning D. Mootz, and Andreas Heuer

Fluctuations of the full length SIM peptide

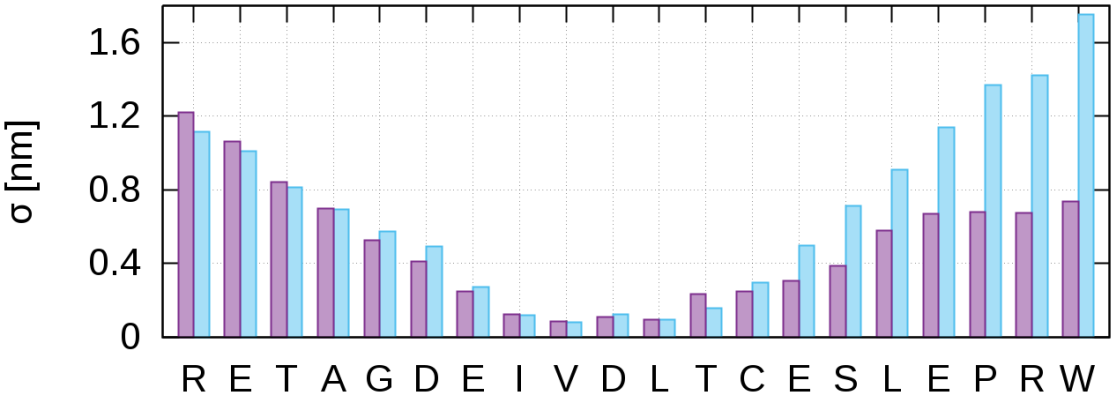

Figure 1: Standard deviations of the positions of the centers of the long SIM2 peptides residues during a 1μs simulation in parallel (purple) and antiparallel (blue) configuration. Standard deviations are computed after aligning 200 equally spaced configurations from the trajectories, using SUMO as reference for the alignment.

Free energy calculations with structures from PDB file 2mp2

The binding constant is calculated as:

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

The terms on the r.h.s. except for $\tilde{K}_{\text{bind}}$ are dimensionless correction factors. $\tilde{K}_{\text{bind}}$ itself is the binding constant in a system, in which configuration orientation and angular part of the position have been restrained to reference values. We report the contribution of the individual contributions in table 1. To be consistent with other studies using this method and the original paper by Woo and Roux, we give the logarithm of the individual correction factors. Furthermore keep in mind that $K_{b,o}$ and $K_{b,a}$ comprise of three and two components, just like their counterparts in free state (which are accessible by numerical integration however).

This means for example: $K_{b,a}^{-1} = K_{b,\theta}^{-1} \times K_{b,\phi}^{-1}$ and $\Delta G_{\phi}^{b} = -RT \log K_{b,\phi}^{-1}$.

Furthermore and also to be consistent with the notation of earlier work, we give the contributions of $\tilde{K}_{\mathrm{eq}} \times K_{f,a}$ together as $I^* S^*$.

Table 1: Contributions of the different restraints with initial configuration from PDB file 2mp2. SIM2 in parallel and SIM3 in antiparallel orientation.

contribution	SIM2 (kcal/mol)	SIM3 (kcal/mol)
ΔG_{c}^{b}	-10.35	-13.78
ΔG_{θ}^{b}	-0.28	-1.06
ΔG_{ϕ}^{b}	-0.62	-0.45
ΔG_{Θ}^{b}	-0.40	-0.47
ΔG_{Φ}^{b}	-0.60	-1.33
ΔG_{Ψ}^{b}	-0.62	-0.78
$-RTlog(I^* S^* C^0)$	-15.15	-4.02
ΔG_{o}^{f}	6.62	6.55
ΔG_{c}^{f}	16.21±0.04	14.27±0.06

We see in table 1 that the by far largest contributions are the cost of applying and releasing the restraints on the configuration (i.e. the RMSD of the peptide) and the separation of protein and peptide (denoted by I^*). I^* is calculated as

$$I^* = \exp(\beta w(r_b)) \int_0^{r_b} \exp(-\beta w(r)) dr. \tag{2}$$

Here $\beta = 1/RT$ is the Boltzmann factor, with gas constant R and temperature T. Let p_{dist} be the probability density of the distance of protein and peptide, then w is defined by the equation $w = -RT \log(p)$. w is called the potential of mean force. Finally, r_b is the distance up to which we consider protein and peptide bound. The critical part of equation (2) in terms of convergence is the potential of mean force at distance r_b, $w(r_b)$.

Figure 2 shows the convergence of the contributions of restraining the RMSD in bound and free state and $\exp(\beta w(r_b))$ against the simulation time in each umbrella window. Specifically we show $\Delta\Delta G(t) = \Delta G(t) - \Delta G(t_{\mathrm{full}})$ for the RMSD (i.e the difference between the estimate at simulation time t and full simulation time) and $\Delta w(r_b)(t) = w(r_b)(t) - w(r_b)(t_{\mathrm{full}})$ for the separation of protein and peptide.

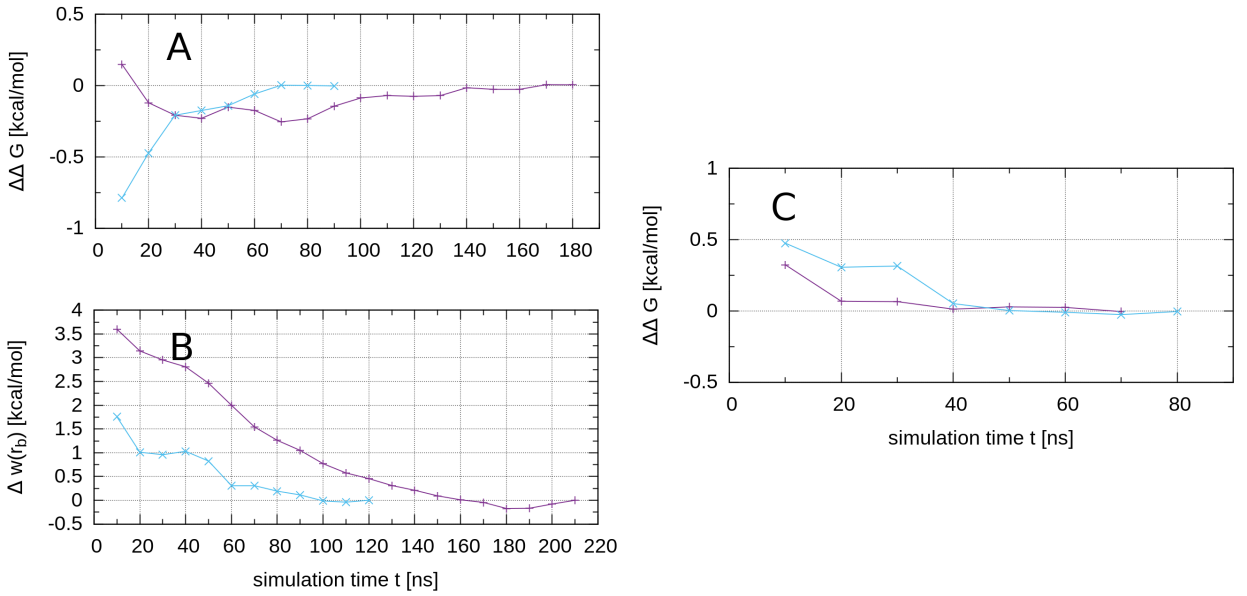

Figure 2: A: Convergence of the estimate for ΔG_c^f, B: convergence of the estimate for $w(r_b)$, C: convergence of the estimate for ΔG_c^b. Purple curves: parallel orientation, blue curves: antiparallel orientation.

The parameters of the REMD-US simulations are given in table 2:

Table 2: Overview of the umbrella sampling parameters. These are the parameters used for all simulations. The simulation times are lower bounds for each window. In case of the RMSD, additional simulations with a higher force constant of 5000 kJ/(mol·nm^2) were used for the windows at 0 nm, 0.05 nm and 0.1 nm. In case of the distance, we used a spacing of 0.1 nm between windows in the regions where we expected a flat potential of mean force. For the other windows a spacing of 0.5 nm was used. Furthermore, after preliminary simulation times of 20 ns for all windows, we significantly extended the simulation time for critical windows. These are the windows in the region starting where the potential of mean force starts to increase strongly and ending where it becomes flat (see also reference 18 in the main text).

collective variable	window spacing	force constant	simulation time
RMSD in bound state	0.05 nm	1000 (5000) kJ/(mol·nm^2)	70-320ns
Θ in bound state	0.1 rad	1000 kJ/(mol·rad^2)	15 ns
Φ in bound state	0.1 rad	1000 kJ/mol	15 ns
Ψ in bound state	0.1 rad	1000 kJ/mol	15 ns
θ in bound state	0.1 rad	1000 kJ/(mol·rad^2)	15 ns
ϕ in bound state	0.1 rad	1000 kJ/mol	15 ns
distance r	0.1 (0.05) nm	1000 kJ/(mol·nm^2)	20 ns (80-290 ns)
RMSD in free state	0.05 nm	1000 (5000) kJ/(mol·nm^2)	90-260 ns

Free energy calculations with docked structures

The tables show the contribution of the individual steps and the figures show the convergence of the critical contributions in the same way as for the simulations with NMR input structures.

Table 3: Contributions of the different restraints for the SIM2 peptide. Simulation with docked initial configurations.

contribution	parallel (kcal/mol)	antiparallel (kcal/mol)
ΔG_c^b	-12.32	-11.41
ΔG_θ^b	-0.31	-0.31
ΔG_ϕ^b	-0.25	-0.54
ΔG_Θ^b	-0.32	-0.46
ΔG_Φ^b	-0.69	-0.61
ΔG_Ψ^b	-0.52	-0.67
$-RTlog(I^*S^*C^0)$	-13.45	-11.99
ΔG_o^f	6.58	6.56
ΔG_c^f	15.46	13.90
ΔG	-5.81	-5.57

Table 4: Contributions of the different restraints for the SIM3 peptide. Simulation with docked initial configurations.

contribution	parallel (kcal/mol)	antiparallel (kcal/mol)
ΔG_c^b	-10.74	-12.62
ΔG_θ^b	-0.50	-0.19
ΔG_ϕ^b	-0.37	-0.32
ΔG_Θ^b	-0.34	-0.35
ΔG_Φ^b	-0.63	-0.68
ΔG_Ψ^b	-0.65	-0.58
$-RTlog(I^*S^*C^0)$	-15.83	-13.87
ΔG_o^f	6.70	6.55
ΔG_c^f	16.83	17.17
ΔG	-5.53	-4.88

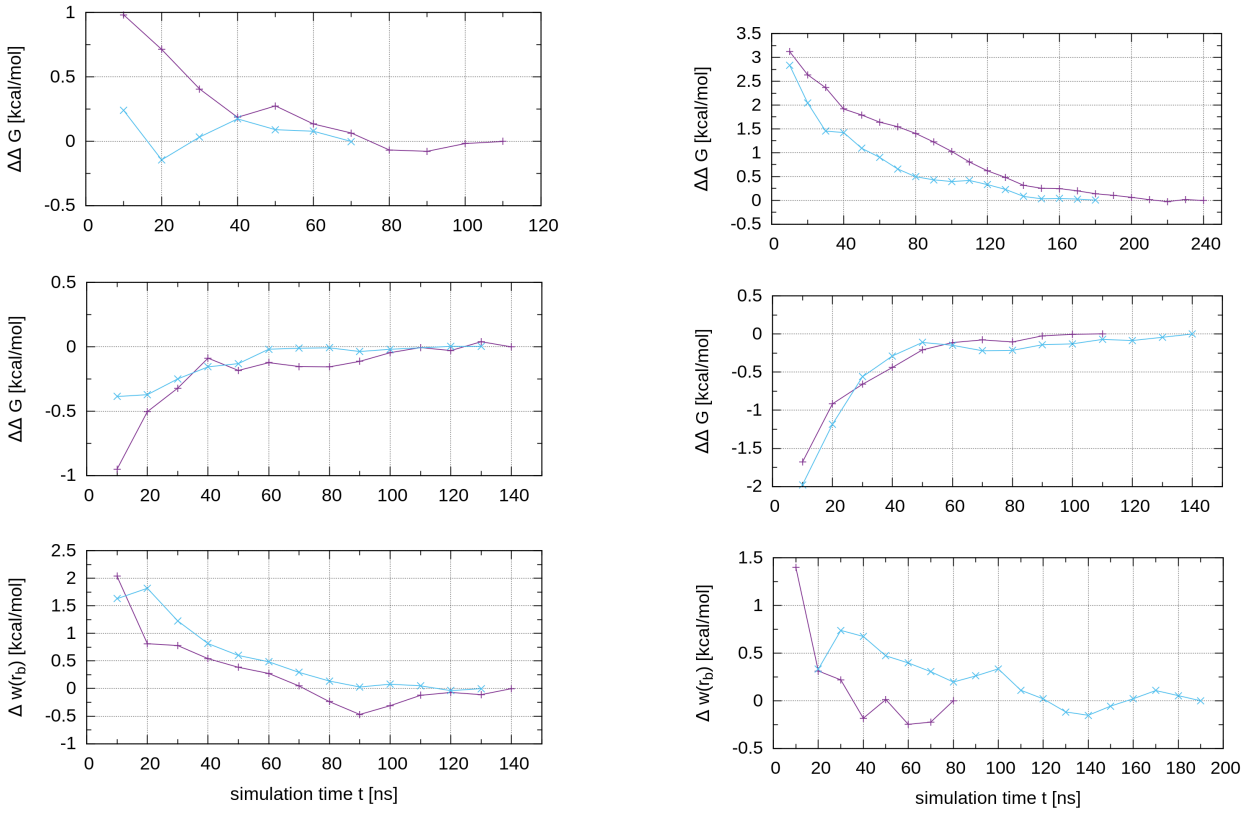

Figure 3: Top: Convergence of the estimate for ΔG_c^b, middle: convergence of the estimate for ΔG_c^f, bottom: convergence of the estimate for $w(r_b)$. Left column: SIM2, right column: SIM3. Purple curves: parallel orientation, blue curves: antiparallel orientation.

Structural properties

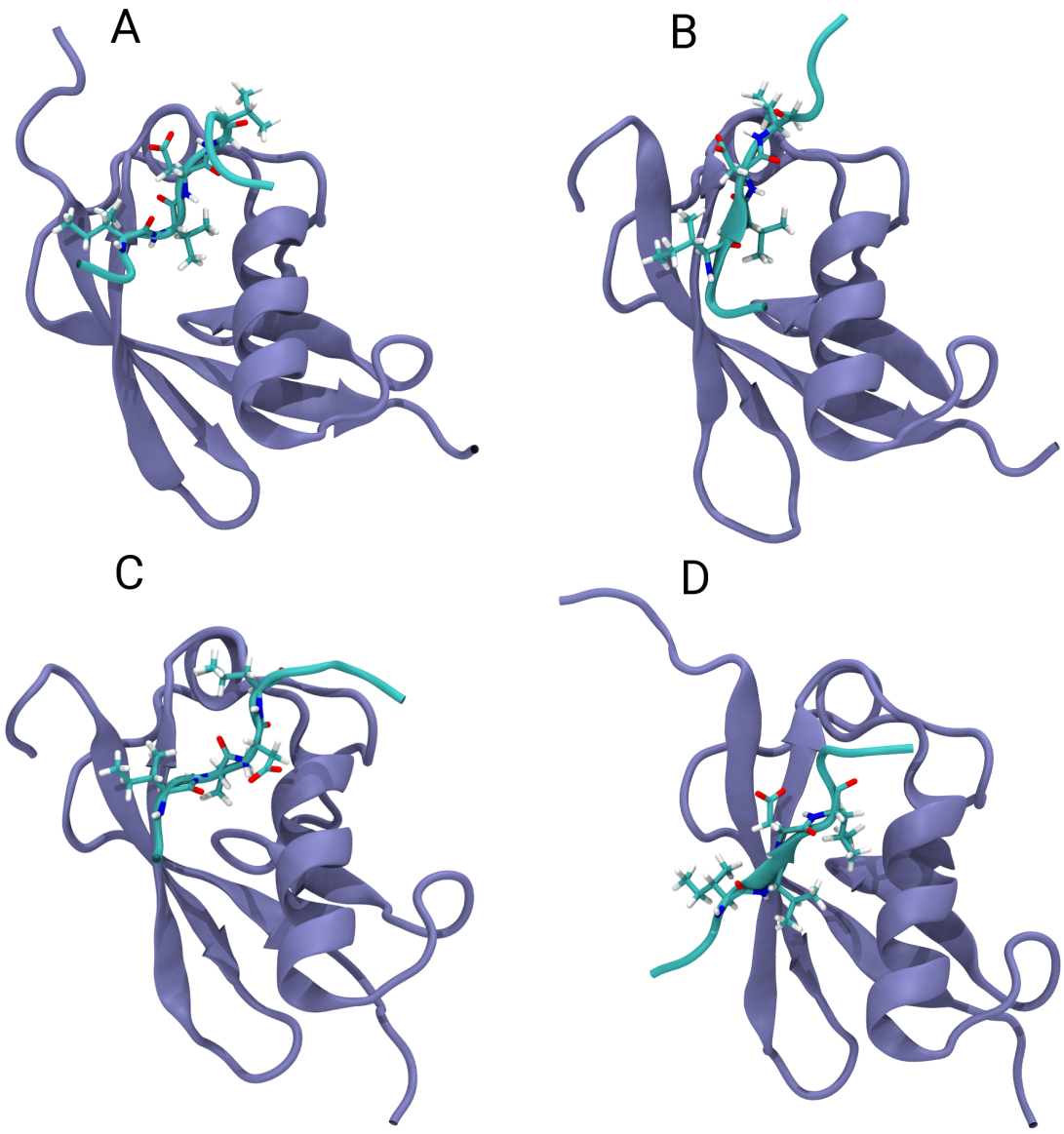

Figure 4: Snapshots of the simulation of the SIM2 peptide in parallel orientation starting in the structure from PDB file 2mp2. A: initial structure, B: 500 ns, C: 1000 ns, D: docked structure after 1000 ns for comparison.

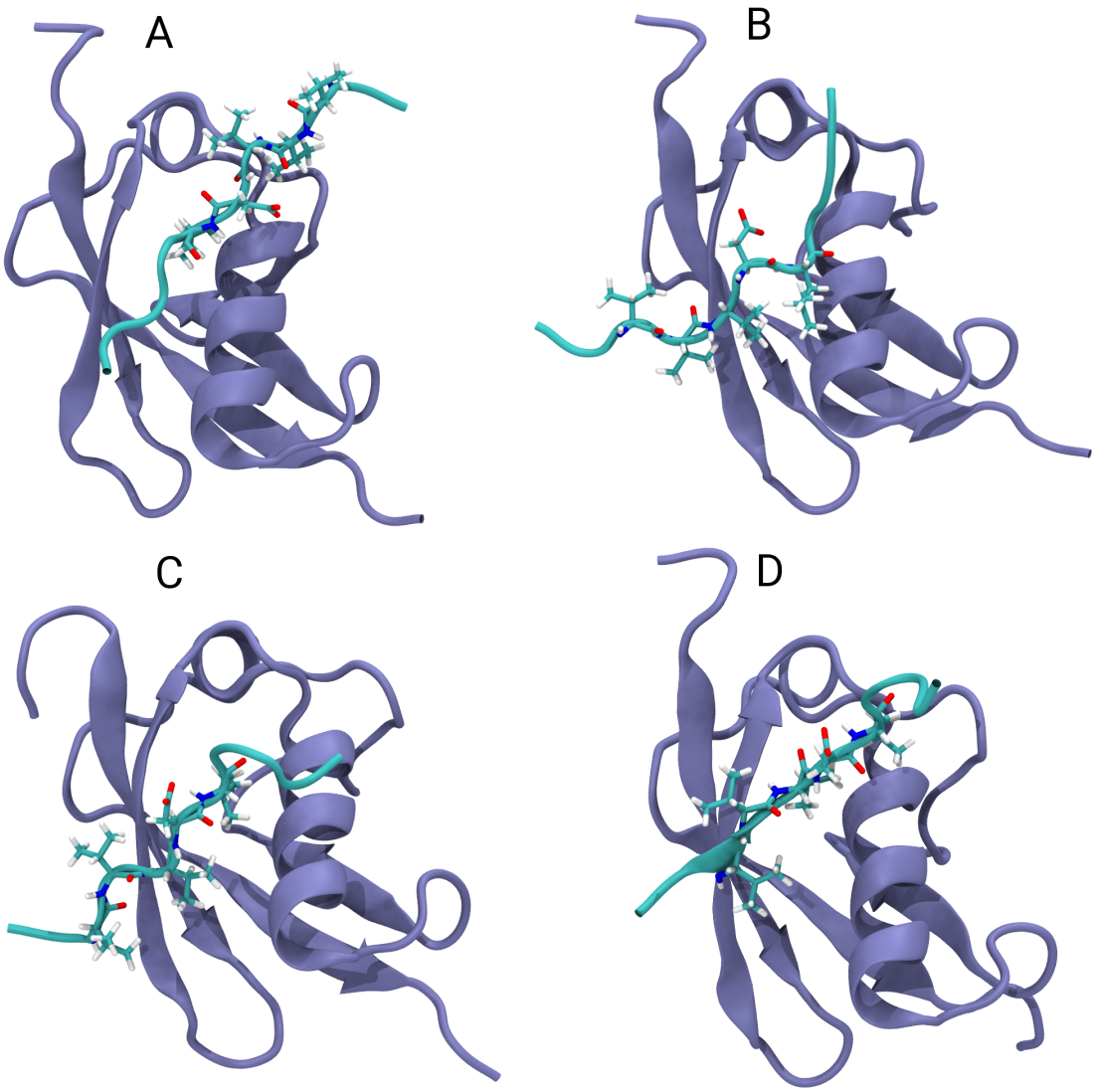

Figure 5: Snapshots of the simulation of the SIM3 peptide in antiparallel orientation starting in the structure from PDB file 2mp2. Note that the orientation changes during the simulation. A: initial structure, B: 500 ns, C: 1000 ns, D: docked structure after 1000 ns for comparison.

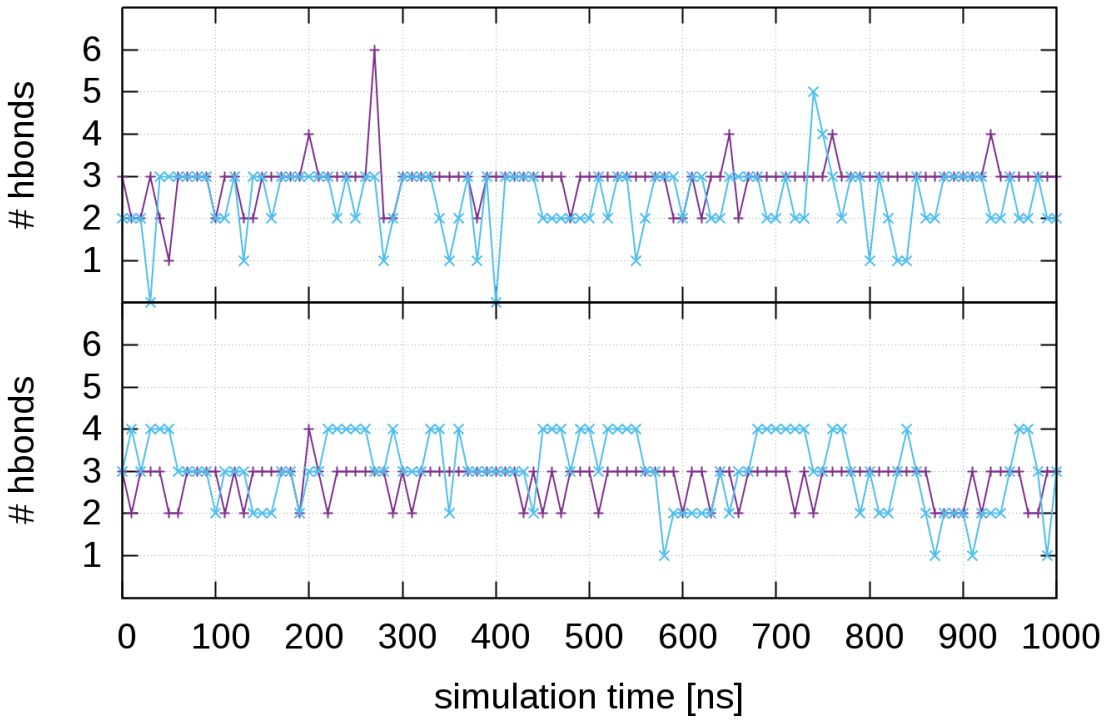

Figure 6: Number of hydrogen bonds between backbone of SIM peptide and SUMO. Top SIM2, bottom: SIM3, purple: parallel orientation, blue: antiparallel orientation.

Table 5: Average number of hydrogen bonds between backbone atoms of the second beta sheet of SUMO and SIM peptide.

SIM2 parallel	2.9
SIM2 antiparallel	2.5
SIM3 parallel	2.8
SIM3 antiparallel	3.0

Figures 7 and 8 show a more detailed comparison of the position of the SIM residues in the last frames of our simulations and typical examples from literature, which serve as reference. We compare the parallel structures of SIM2 and SIM3 with the interfaces found in PDB files 2rpq and 5d2m and the antiparallel structures of SIM2 and SIM3 with the interfaces found in PDB files 3uin and 5d2m.

For the parallel configurations, the reference SIMs consist of the residues (YIDL) and

(VIDL) and thus are very similar to SIM2 and SIM3, in particular they are also SIMs of the type (hhXh). We see that in the parallel configurations, SIM2 and SIM3 assume positions very close to those of the SIMs from literature. Notably, in all cases two hydrophobic side chains lie close to the first and second helix turn of SUMO. We remark, however, that the side chains of Val-3 and Val-5 and not the side chains of Val-5 and Val-7 occupy the binding groove.

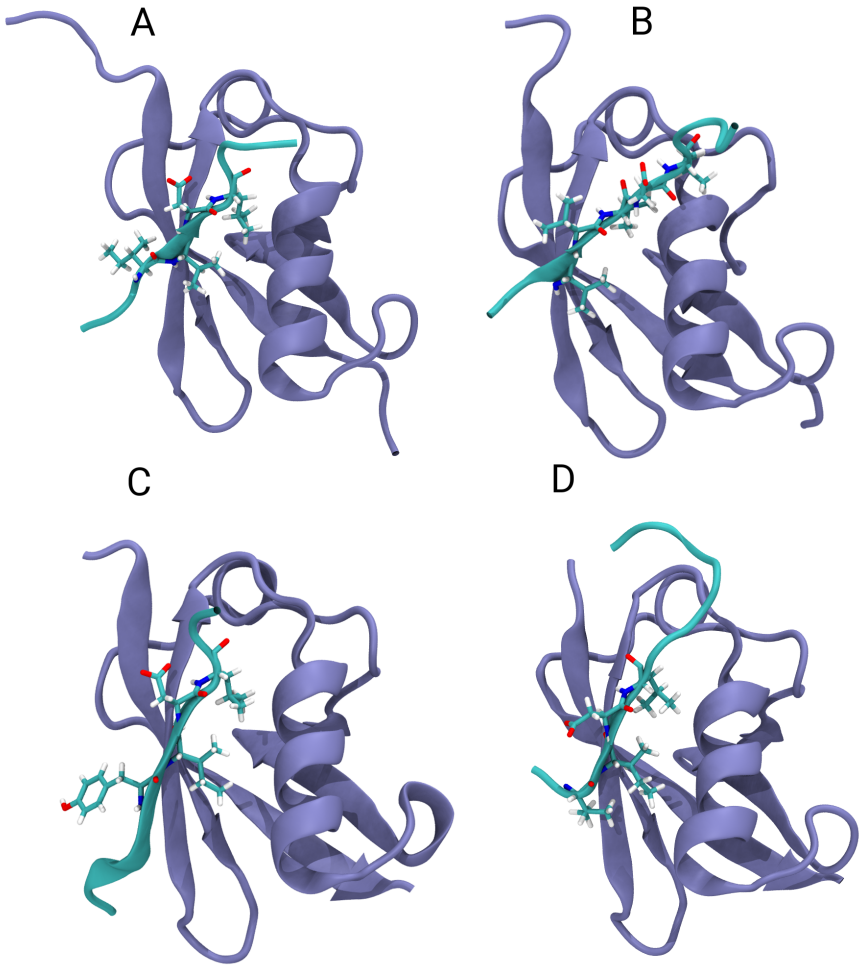

Figure 7: Comparison of final structures of SIM2 and SIM3 peptides in parallel orientation. A: final structure of the SIM2 peptide, B: final structure of the SIM3 peptide, C: structure from PDB file 5d2m, D: structure from PDB file 2rpq.

For the antiparallel configurations, the reference SIMs consist of the residues (VLIV) and (IQFV) (PDB file 5d2m contains the structure of a diSIM peptide in complex with two

SUMOs, thus we obtained one antiparallel and one parallel structure from PDB file 5d2m). These differ from SIM2 and SIM3 in that the charged residue on position 3 is replaced by a hydrophobic one and in case of the second reference SIM, the hydrophobic residue on position 2 is replaced by the polar glutamine. Still, the configuration adopted by residues of SIM2 in complex is close to what we see in the structures of the reference SIMs. Naturally, the position in the SUMO groove close to the second helix turn is occupied by the charged aspartic acid instead of a hydrophobic residue, as in case of the reference SIMs. While in the reference SIMs the residues on position 1 and 3 occupy the hydrophobic groove of SUMO, in case of SIM3 the residues on position 2 and 4 occupy these spots. This seems beneficial for binding since these residues are hydrophobic.

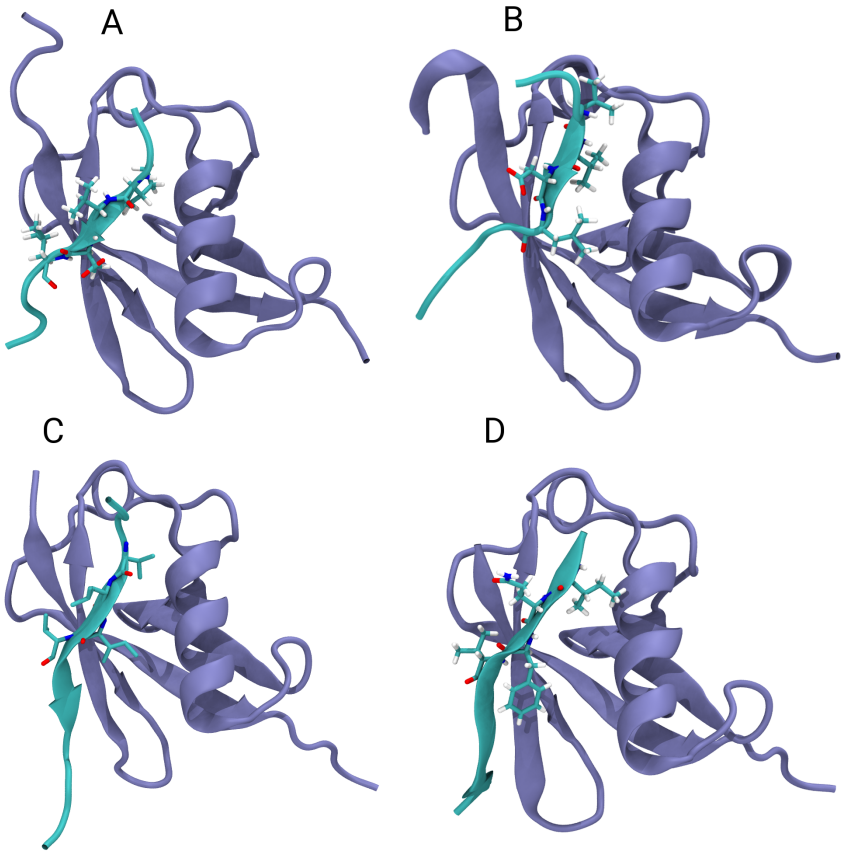

Figure 8: Comparison of final structures of SIM2 and SIM3 peptides in antiparallel orientation. A: final structure of the SIM2 peptide, B: final structure of the SIM3 peptide, C: structure from PDB file 3uin, D: structure from PDB file 5d2m.

Role of flanking acidic residues

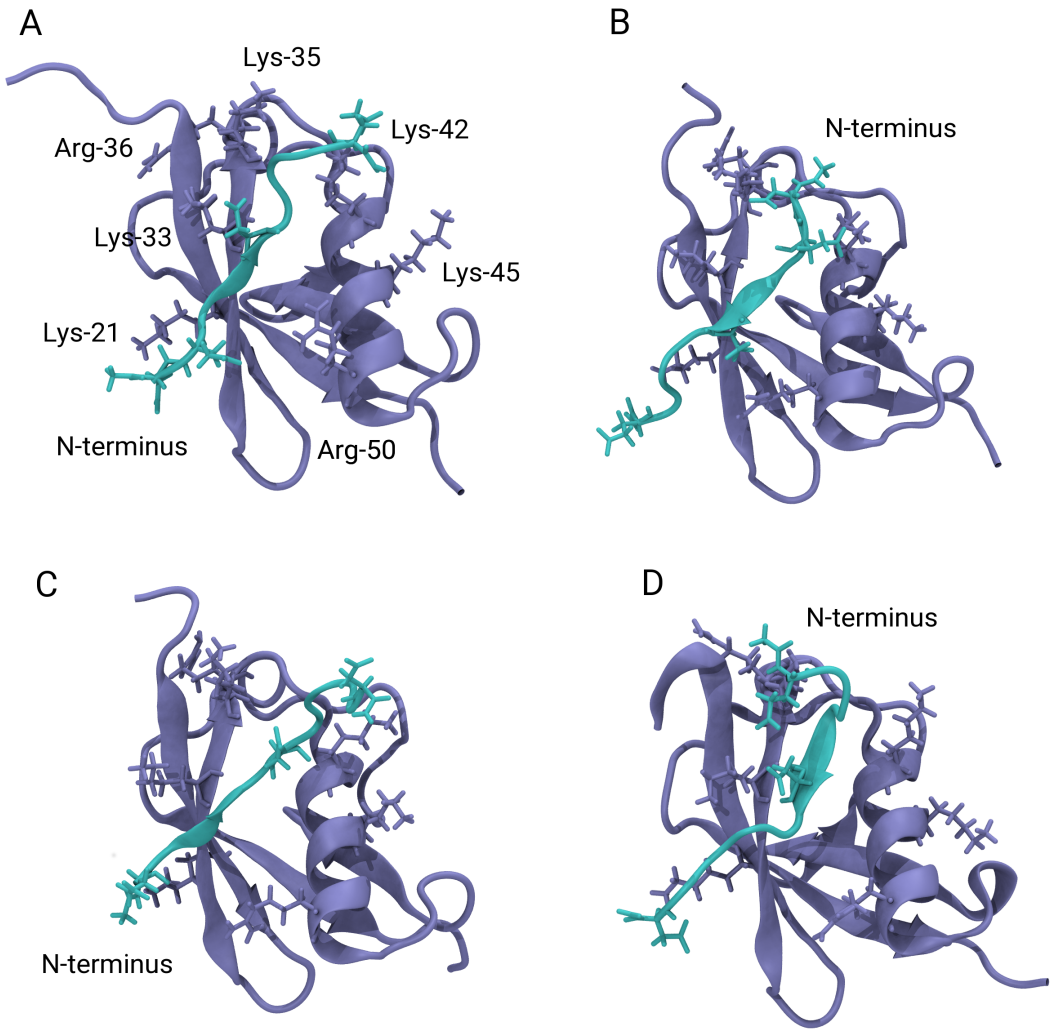

Figure 9: Structures of the SIM-SUMO complexes with charged residues highlighted. A: SIM2 in parallel orientation, B: SIM2 in antiparallel orientation, C: SIM3 in parallel orientation, D: SIM3 in antiparallel orientation,

Definition of the distance: Let C_t be the atomic positions of the acid carbons of acidic residues flanking the SIM and N_t the atomic positions of the amine nitrogen in case of lysines of SUMO and the guanidinium carbon in case of argininesof SUMO at time t. We define D_t as the set of distances $d(c, n)$ for $c \in C_t$ and $n \in N_t$ (i.e. $D_t = \{d(c, t)|c \in C_t, n \in N_t\}$). Figure 4 of the main text shows the first and second order statistic of the set D_t against the

simulation time.

Convergence of the binding free energy estimates: Protonated systems

As before, the tables contain the individual contributions to the affinities and the figures show the convergence of the critical contributions.

Table 6: Contributions of the different restraints for the SIM2 peptide. Simulation with docked initial configurations.

contribution	parallel (kcal/mol)	antiparallel (kcal/mol)
ΔG_c^b	-12.50	-13.41
ΔG_θ^b	-0.18	-0.20
ΔG_ϕ^b	-0.37	-0.94
ΔG_Θ^b	-1.00	-0.45
ΔG_Φ^b	-0.47	-0.79
ΔG_Ψ^b	-0.45	-0.53
$-RTlog(I^*S^*C^0)$	-8.98	-7.60
ΔG_o^f	6.57	6.59
ΔG_c^f	13.48	14.65
ΔG	-3.91	-2.66

Table 7: Contributions of the different restraints for the SIM3 peptide. Simulation with docked initial configurations.

contribution	parallel (kcal/mol)	antiparallel (kcal/mol)
ΔG_c^b	-13.66	-11.38
ΔG_θ^b	-0.59	-0.23
ΔG_ϕ^b	-0.33	-0.29
ΔG_Θ^b	-0.35	-0.33
ΔG_Φ^b	-0.65	-0.71
ΔG_Ψ^b	-0.56	-0.67
$-RTlog(I^*S^*C^0)$	-14.11	-14.34
ΔG_o^f	6.69	6.58
ΔG_c^f	18.31	16.80
ΔG	-5.26	-4.55

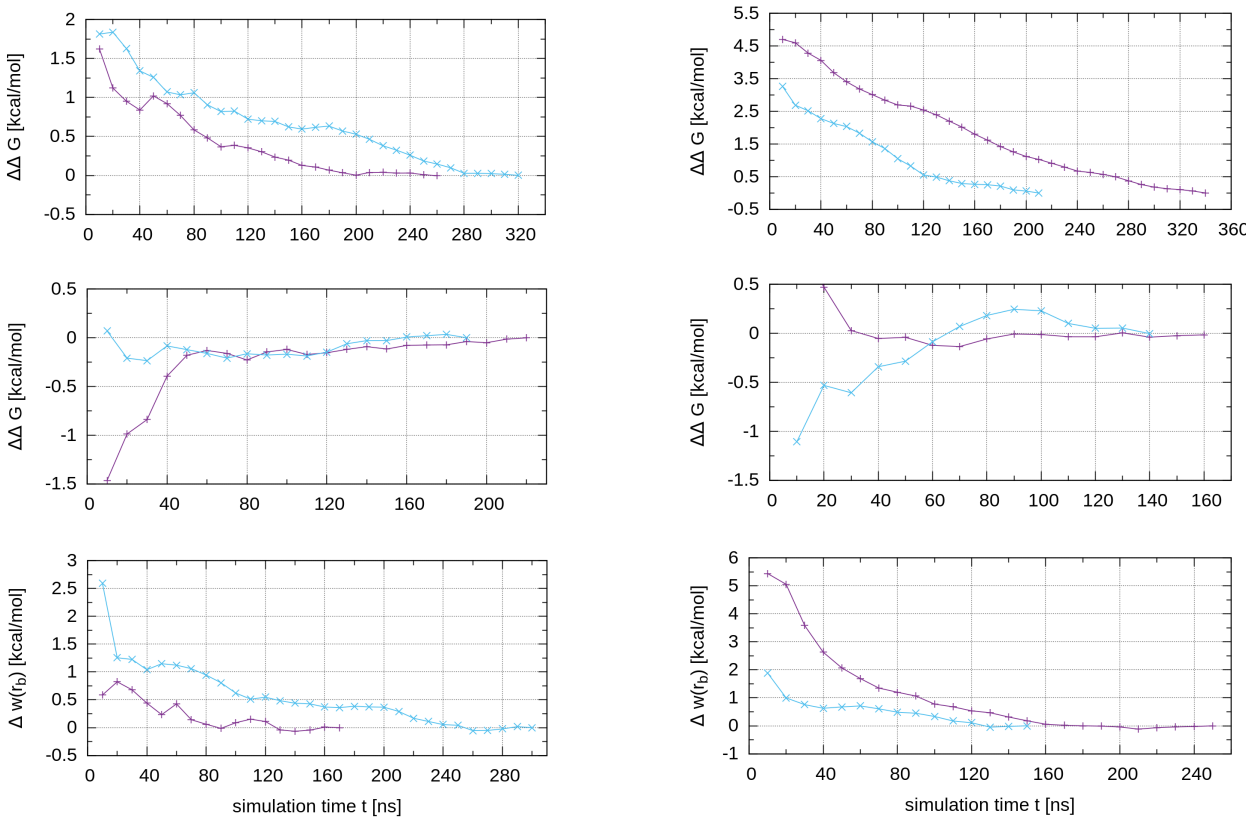

Figure 10: Top: Convergence of the estimate for ΔG_c^b, middle: convergence of the estimate for ΔG_c^f, bottom: convergence of the estimate for $w(r_b)$. Left column: SIM2, right column: SIM3. Purple curves: parallel orientation, blue curves: antiparallel orientation.

Comparison with experimental results on the bound orientation

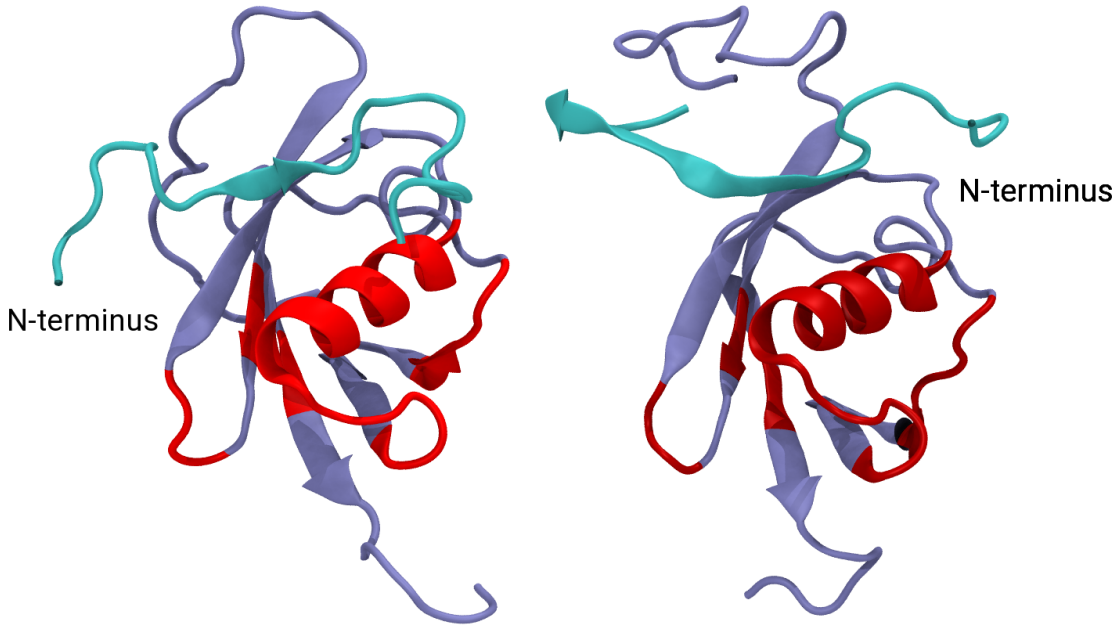

Figure 11: Structure of SUMO with regions highlighted in red, which are mostly affected by the spin label at the N-terminus of the SIM3 peptide. Data by Kung et al was used to make this figure. The peptide shown is a peptide containing SIM2 and should only serve as guide to the eye.

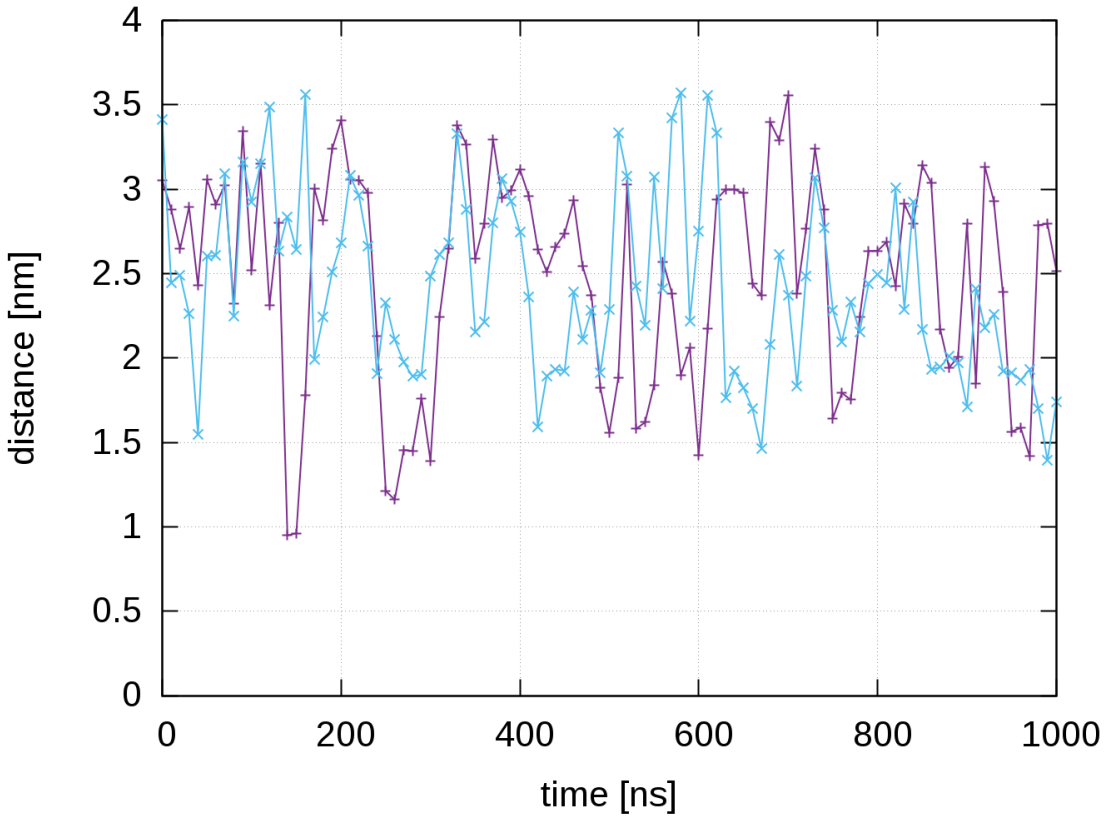

Figure 12: Distance between C-terminus of SUMO and N-terminus of the second beta sheet of SUMO (purple) and N-terminus of the alpha helix of SUMO (blue) respectively.