Supporting Information:

Probing the dynamical interaction of para-sulfonato-calix[4]arene with an antifungal protein

Alessio Bartocci[*,†] and Elise Dumont[‡,¶]

†*Institut de Chimie de Strasbourg, UMR 7177, CNRS, Université de Strasbourg, Strasbourg Cedex 67083, France*

‡*Université Côte d'Azur, CNRS, Institut de Chimie de Nice, UMR 7272 - 06108 Nice, France*

¶*Institut Universitaire de France, 5 rue Descartes, 75005 Paris, France*

E-mail: abartocci@unistra.fr

Contents

List of Figures

List of Tables

Experimental binding affinity

The binding equilibrium of *para*-sulfonato-calix[4]arene (L) interacting with a small antifungal protein (*PAF*, P)

$$P + L \underset{K_D}{\overset{K_A}{\rightleftharpoons}} PL \tag{S1}$$

was experimentally characterized by Crowley and coworkers[1] by ITC experiments, with a dissociation constant K_D of 107 μM and a corresponding binding affinity equal to :

$$\Delta G^o = RT\,ln(K_D) = -5.45\ kcal \cdot mol^{-1} \tag{S2}$$

"Diffusion" Simulations

System Setup and Production Phase

We have sought to probe the interaction of the *Penicillum* antifugal protein (*PAF*) (PDB: 6HA4, 55 residues, overall charge +5) interacting with one **sclx**$_4$ molecule. In *PAF*–**sclx**$_4$ system, the **sclx**$_4$ ligand and the *PAF* protein (distance between the centers of mass) were solvated in four different initial boxes (**Box**$_1$, **Box**$_2$, **Box**$_3$ and **Box**$_4$) with TIP3P[2] water molecules. Number of water molecules and simulation box dimensions are reported in Table S1, as well as *PAF*–**sclx**$_4$ relative centers of mass distances. The ligand molecule was not placed in contact with the protein, as we aim at probing *de novo* binding site(s).

Table S1: *Number of water molecules (N_{water}), box volume (Box Vol.) and initial distances in Å between the centers of mass of PAF and* **sclx$_4$**.

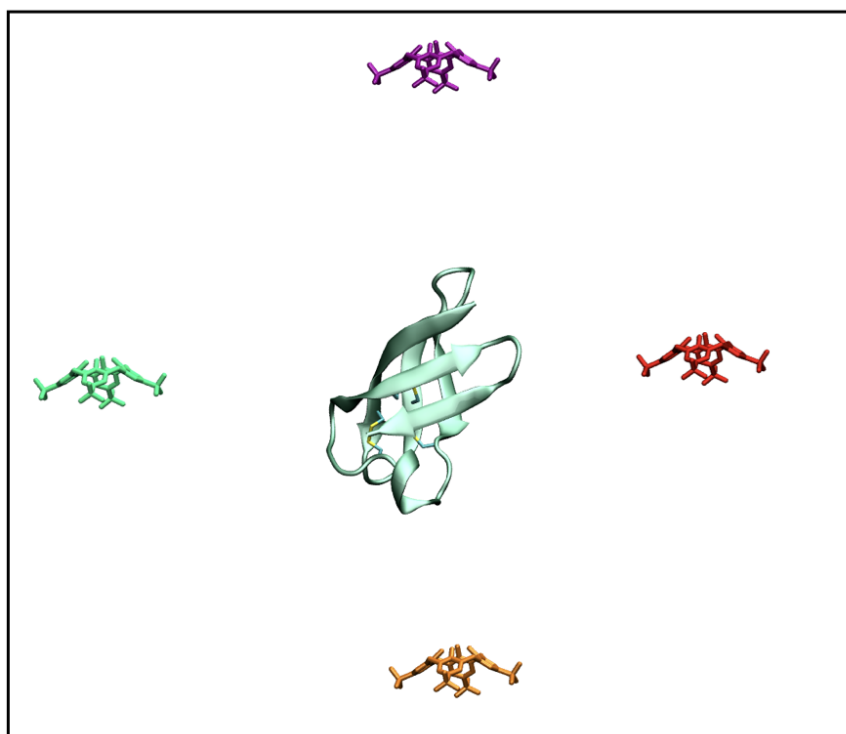| System	Box	N_{water}	Box Vol. (Å^3)	Prot–Lig
*PAF–***sclx$_4$**	**Box$_1$**	20459	90 X 90 X 90	32
	Box$_2$	25029	96 X 96 X 96	40
	Box$_3$	27546	99 X 99 X 99	37
	Box$_4$	24604	96 X 96 X 96	34

Figure S1: *PAF-***sclx$_4$** *initial systems. The protein is reported in green cartoon. For* **Box$_1$**, **Box$_2$**, **Box$_3$** *and* **Box$_4$** *the ligand is colored in orange, purple, green and red, respectively. PAF cysteines disulphur bonds are shown in lines, while water and* **sclx$_4$** *hydrogen atoms are omitted for the sake of clarity.*

The production phase for each *PAF–***sclx$_4$** simulation box (**Box$_1$**, **Box$_2$**, **Box$_3$** and **Box$_4$**) was carried out in five replicates of 300 ns each, giving 1.5 μs simulation time per box, and

in total 6 μs for the system in total.

Cluster Analysis

To obtain representative structures for the $\mathbf{sclx}_4$ molecule interacting in the different binding sites, and a good starting point for bound-state and binding free-energy simulations, cluster analysis was performed by concatenating the last 100 ns of each trajectory reported in the previous section. Clustering of the MD trajectories was carried out using the hierarchical average-linkage clustering algorithm[3] implemented in the `cpptraj` module of AmberTools18 with a random sieving frequency of 10 and a cutoff of 10 Å on the protein and ligand heavy atoms. The most representative structures of PAF-$\mathbf{sclx}_4$ are shown in Figure 2 (panels A and B) in the main article.
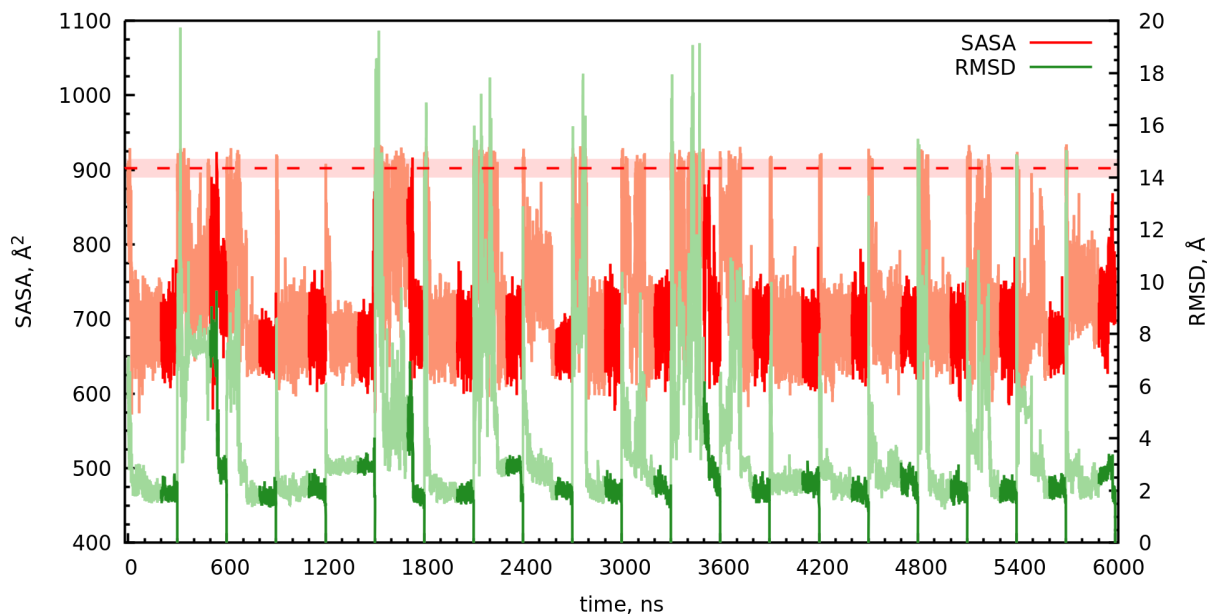
RMSD Analysis

Figure S2: **sclx**$_4$ *SASA and RMSD plots in function of the simulation time for PAF-***sclx**$_4$ *system. The last frame of each replicate (bound state) was taken as RMSD reference. SASA is reported in Å*^{2} *(red lines), while RMSD in Å(green lines). The darker red and green colors, respectively, highlight the simulation part in which the post-process analysis was carried out (last 100 ns of each replicate). The unbound* **sclx**$_4$ *SASA is reported as a dashed line (902.00 $\pm$ 12.50 Å*^{2}*) as obtained in ref.*[4]*.*

MMPBSA Energy Analysis

The last 10 ns of each replicate run per system, concatenated into a 200 ns long trajectory, were analyzed using the MM-PBSA[5–7] post-processing approach. A per-residue free energy additive decomposition analysis was carried out, with the aim of identifying the residues which mediate *PAF*-**sclx**$_4$ binding.

The values of the free energy of ligand binding ΔG_{bind} were calculated according to the

MM-PBSA equation S3:

$$\Delta G_{bind} = \langle G_{compl} \rangle - \langle G_{rec} \rangle - \langle G_{lig} \rangle \qquad (S3)$$

where compl, rec, and lig stand for complex, receptor, and ligand, respectively. The free energy contribution of each species of the binding reaction was estimated as a sum of four terms as in eq. S4

$$G = \langle E_{MM} \rangle + \langle G_{psolv} \rangle + \langle G_{npsolv} \rangle - T \langle S \rangle \qquad (S4)$$

where E_{MM} is the molecular mechanics energy of the molecule expressed as the sum of the internal energy of the molecule plus the electrostatics and van der Waals interactions in gas-phase, G_{psolv} is the polar contribution to the solvation free energy of the molecule (Poisson Boltzmann PB[5]), G_{npsolv} is the non-polar contribution to the solvation free energy, T is the absolute temperature, and S is the entropy. The inclusion of the entropic contribution in eq. S4 was not considered here in the standard MM-PBSA calculations. As recommended in ref. [6], we used the default value for calculating nonpolar solvation free energies (inp=2) and for its associated variables `cavity_surften` and `cavity_offset` (0.0378 and -0.5692, respectively). A 1.4 Å probe radius (keyword `prbrad`, default value) has been used, as well as the optimized radii set (`radiopt=1`).
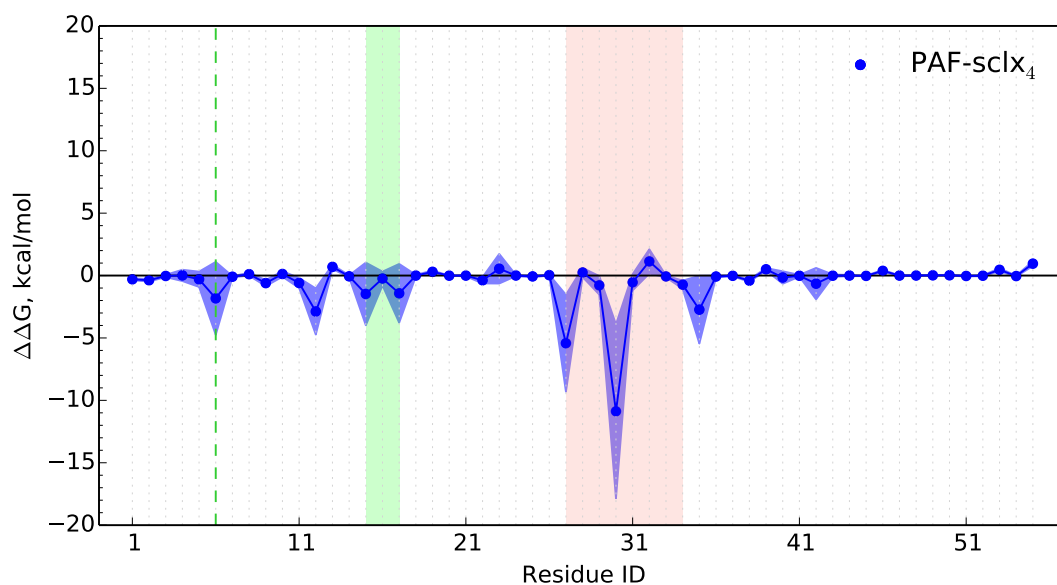
Figure S3: *Per-residue decomposition $\Delta\Delta G$ of the total interaction energy estimated by the MM-PBSA approach, in kcal.mol^{-1}, as a function of the residue number. The colored boxes in red and green correspond respectively to* **Site 1** *and* **Site 2**. *The vertical dashed green line correspond to K6, as part of* **Site 2**.
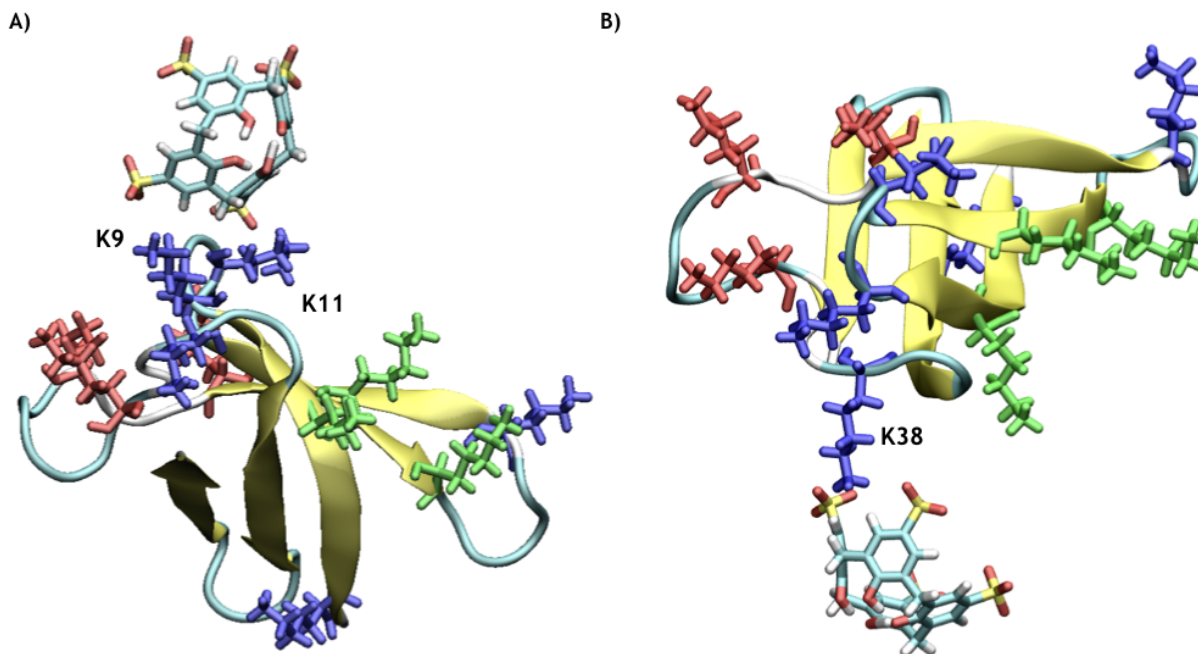
Lysine–sclx$_4$ Coordination

Figure S4: *Snapshots extracted from "Diffusion Simulations" to show supplementary lysines (K9, K11 and K38) capable of coordinating* sclx$_4$*, as well as those present in the main binding sites.*

"Bound-State" Simulations

Setup

In the two "Bound-State" representative structures, extracted from cluster-analysis and reported in Figure 2A and B in the main article, sclx$_4$ interacts in **Site 1** and **Site 2**. The two systems were solvated with 11708 and 10871 TIP3P[2] water molecules, leading to 86 X 86 X 86 and 76 X 76 X 76 Å^3 simulation boxes, respectively. The systems were called Bound$_{Site1}$ and Bound$_{Site2}$.

"Bound-State" Cluster Analysis

Cluster analysis of "Bound State" simulations was performed on the entire trajectories (in total 3.5 μs) for each system, using the hierarchical average-linkage clustering algorithm via the `cpptraj` module of `AmberTools18`, with a random sieving frequency of 10 and a cutoff of 10 Å on the protein and ligand heavy atoms. The most representative structures of Bound$_{Site1}$ and Bound$_{Site2}$ are reported in Figure 3 in the main article and Figure S5.
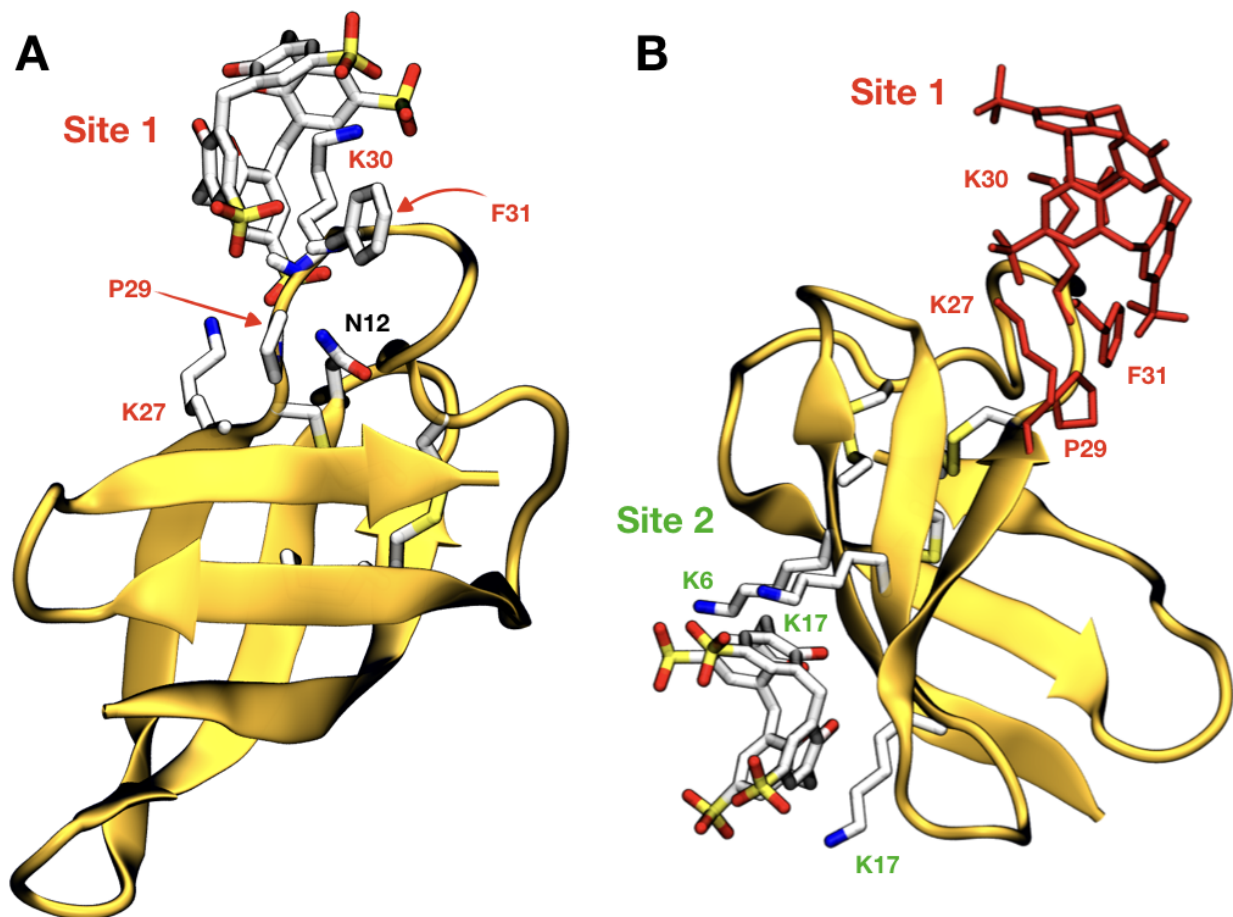
Figure S5: **A**: *most representative structure (80 %) of* **sclx₄** *interacting with PAF in* **Site 1** *obtained from "Bound State" simulations (Bound$_{Site1}$, shown in Figure 2A in the main article).*
B: *superimposition of the first (Cluster 1,* **sclx₄** *in* **Site 2***, 51.4 %) and second (Cluster 2,* **sclx₄** *in* **Site 1***, 45 %) most representative structures obtained from cluster analysis of Bound$_{Site2}$ system simulations. PAF is colored in orange cartoon;* **sclx₄** *in Cluster 1, and the binding site residues K6, K15 and K17 in white licorice tubes, while* **sclx₄** *in Cluster 2, and the binding site residues K27, P29, K30 and F31 in red lines.*
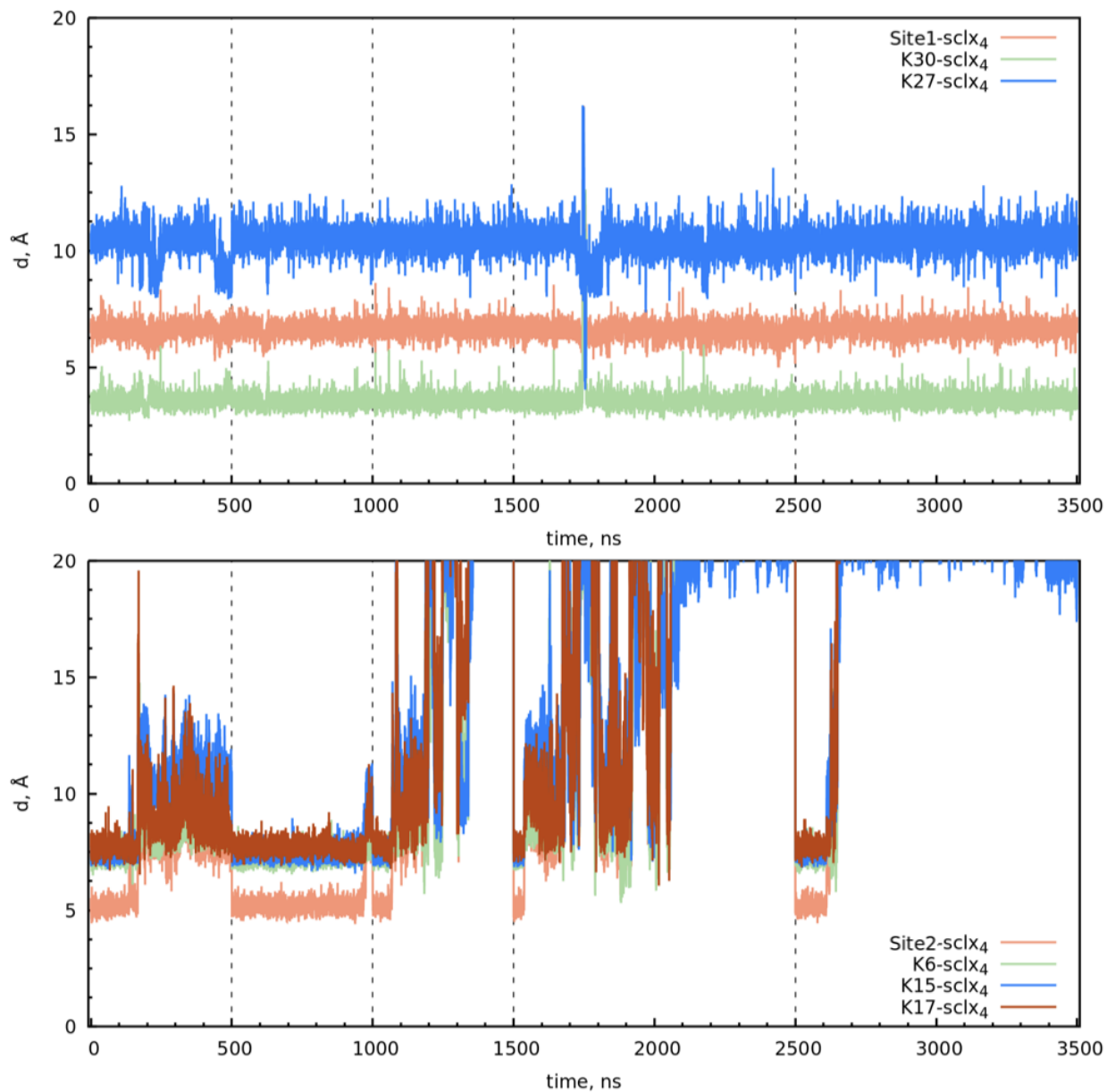
12

Distance Analysis

Figure S6: *Distance (d, Å) between the centers of mass of* **sclx$_4$** *interacting in* **Site 1** *(upper panel, Bound$_{Site1}$ system) and in* **Site 2** *(lower panel, Bound$_{Site2}$ system) and the coordinating partners in function of the simulation time. For the entire binding site (Site(i)-* **sclx$_4$** *light red line) the center of mass is defined by the coupled residues K27, P29, K30 and F31 for* **Site 1** *and K6, K15 and K17 for* **Site 2**, *while it corresponds to the center of mass of the specific lysine in the other cases.*
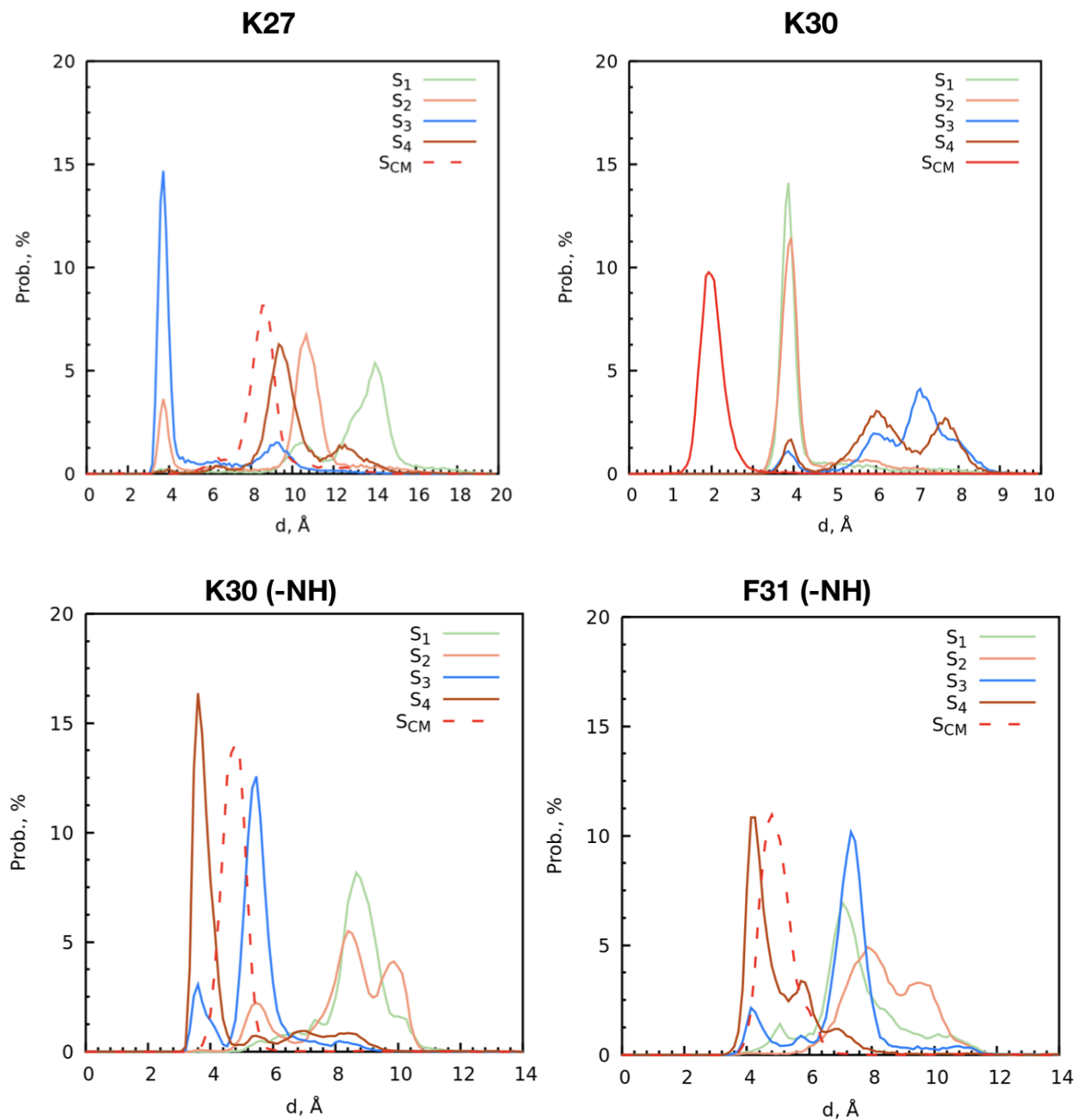
Figure S7: *Distance distribution (Prob, %) between* **sclx**$_4$*'s sulphur atoms and K27 (left panel) and K30 (right panel) -NH$_3^+$ groups (Upper panels) and between* **sclx**$_4$*'s sulphur atoms and K30 (left panel) and F31 (right panel) -NH amide groups (Lower panels), extracted from the "Bound State" simulations of* **sclx**$_4$ *interacting in* **Site 1** *(Bound$_{Site1}$ system).*

Table S2: *Average distances, reported as probability distributions in Figure S7 and extracted from the "Bound State" simulations of* **sclx**$_4$ *interacting in* **Site 1** *(Bound$_{Site1}$ system), between the nitrogen atom of the -NH$_3^+$ group and the* **sclx**$_4$ *sulphur atoms (d$_{S-NZ}$) and between the amide nitrogen atom and the* **sclx**$_4$ *sulphur atoms (d$_{S-NH}$). S$_{CM}$ refers to consider the center of mass of the four sulphur atoms instead of a single sulphur. Distance values are given in Å.*

Res.		$\mathbf{S_1}$	$\mathbf{S_2}$	$\mathbf{S_3}$	$\mathbf{S_4}$	$\mathbf{S_{CM}}$
K27	d$_{S-NZ}$	13.10 ± 2.27	9.71 ± 3.00	5.34 ± 2.46	10.13 ± 1.85	8.71 ± 1.26
K30	d$_{S-NZ}$	4.27 ± 1.02	4.28 ± 0.84	6.76 ± 1.15	6.44 ± 1.25	2.10 ± 0.44
	d$_{S-NH}$	8.68 ± 1.04	8.38 ± 1.51	5.48 ± 1.03	4.61 ± 1.62	4.82 ± 0.46
F31	d$_{S-NH}$	7.63 ± 1.46	8.51 ± 1.20	7.20 ± 1.40	5.10 ± 1.21	5.10 ± 0.66

Interacting Surface of sclx$_4$

The angle Θ is formed between the centers of mass of the protein (P$_{CM}$), **sclx**$_4$ sulfonate's sulphur (S$_{CM}$) and hydroxyl oxygen atoms (O$_{CM}$). The normalized distribution, extracted from the "Bound State" simulations, is reported in Figure S8. The main *exo* and *endo* conformations have an average $\Theta < 50°$ and $\simeq 125°$, respectively.
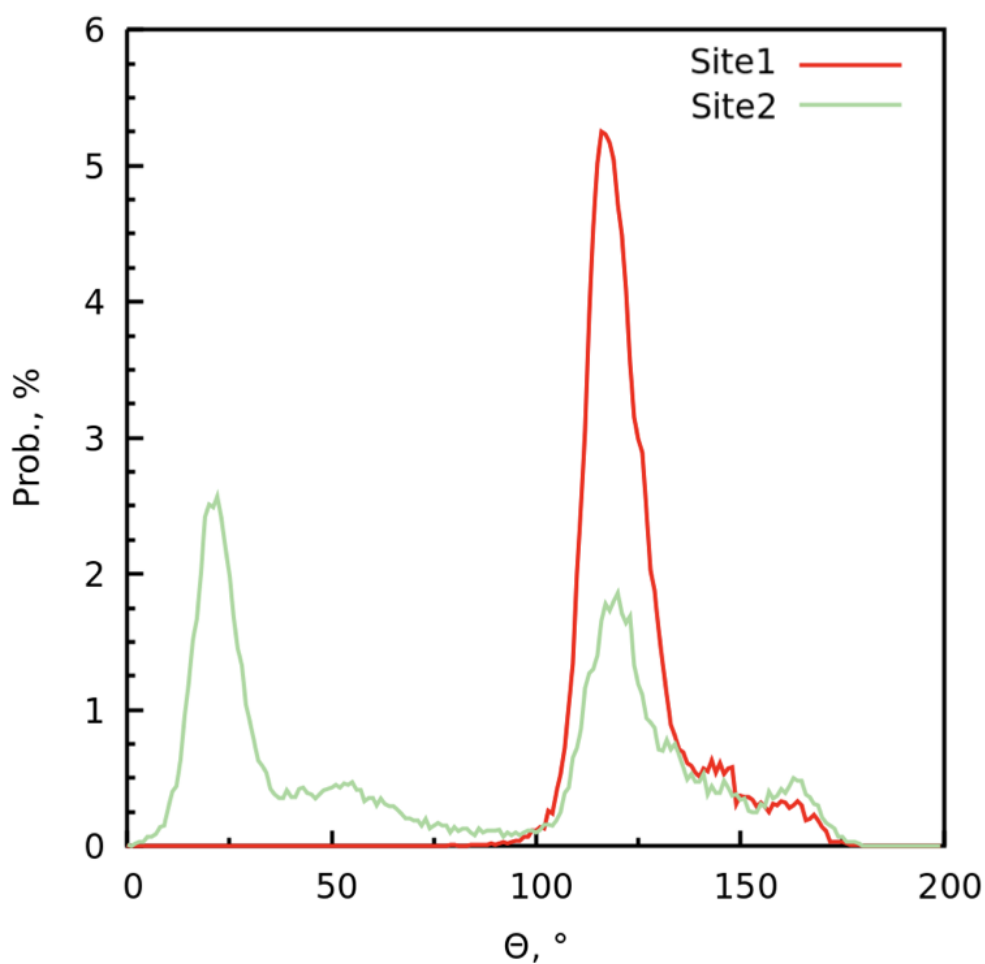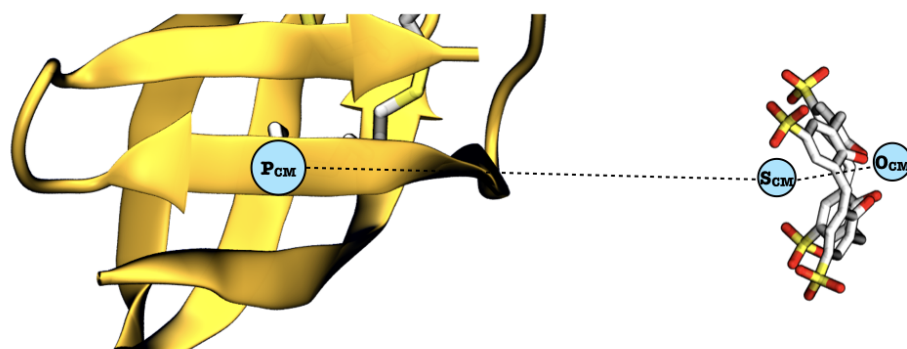
Figure S8: *A schematic visualization of* Θ *angle is reported in the upper panel. In the lower panel, normalized distributions of* Θ *angle (in degrees °), formed between the centers of mass of the protein,* **sclx**$_4$ *sulfonate's sulphur and hydroxyl oxygen atoms.* **sclx**$_4$ *interacting in* **Site 1** *(Bound$_{Site1}$ system) and in* **Site 2** *(Bound$_{Site2}$ system) are reported as red and green lines, respectively.*

16

Table S3: *Distances (in Å) between the nitrogen atom of the -NH$_3^+$ group and the closest* **sclx**$_4$ *sulphur atom (d_{S-NZ}) and between the amide nitrogen atom and the closest* **sclx**$_4$ *sulphur atom (d_{S-NH}), and Θ angle (in degrees °), formed between the centers of mass of the protein, the sulfonate's sulphur and hydroxyl oxygen atoms (see section "Interacting Surface of* **sclx**$_4$*"). X-ray refers to the value in the crystallographic structure (asymmetric unit in Figure 1 in the main article), MD to the average value obtained from the "Bound State" simulations, while Diff$_1$, Bound$^1_{Site2}$ and Bound$^2_{Site2}$ to the values extracted from the structures shown in Figures 2A and 3B in the main article.*

Res.		X-ray[1]	MD	Bound$^1_{Site2}$	Bound$^2_{Site2}$	Diff$_1$
K27	d_{S-NZ}	5.10	5.34 ± 2.46	3.90	3.91	4.20
	d_{C-NZ}	6.81	6.28 ± 0.43	6.81	6.40	6.23
K30	d_{S-NZ}	3.90	4.27 ± 1.02	3.80	3.90	3.44
	d_{S-NH}	2.70	4.61 ± 1.62	3.60	4.13	4.10
	d_{C-NZ}	4.87	5.00 ± 0.82	4.90	4.80	5.64
F31	d_{S-NH}	3.34	5.10 ± 1.21	4.00	5.86	4.25
sclx$_4$	Θ	140.68	123.90 ± 13.10	118.00	126.25	114.65

Binding Affinity

Within the APR (attach-pull-release)[8–11] protocol, the standard (absolute) binding free energy between the protein (*PAF*) and a ligand (**sclx**$_4$) is computed as the result of the reversible work of transferring the ligand from the binding site to solution (unbound state). The free energy difference is calculated as a sum of the work W required to attach restraints on the ligand (W_{attach}), to pull the ligand away from the binding site (W_{pull}), and then to release the attached restraints and to place the ligand at the standard concentration ($W_{release-std}$).

$$\Delta G^{\circ}_{bind} = -\left(W_{attach} + W_{pull} + W_{release-std} \right) \tag{S5}$$

For further details on the protocol implementation, we refer the reader to the references[8–11].

Simulations Setup

As starting *PAF*–**sclx**$_4$ snapshot, the most representative structure obtained from cluster analysis of the "Diffusion" simulations, reported in Figure 2A in the main article and in which **sclx**$_4$ interacts in **Site 1**, was used. *PAF*–**sclx**$_4$ complex was solvated with 10000 TIP3P[2] water molecules.

The protein is kept fixed and the ligand's distance and angle restraints set up by the use of three non-interacting anchor dummy particles with zero charge, zero LJ radius and well-depth, mass of 220 Da and subject to positional restraints of 50 kcal/(mol Å^2) (see Figure S9). The force constants of the distance and angle restraints, applied between the anchor particles and *PAF* and between the anchor particles and **sclx**$_4$ were, respectively, 5 kcal/(mol Å^2) and 100 kcal/(mol rad^2). The reaction coordinate adopted for the pulling phase is represented by the distance between a dummy particle (P1 in Figure S9) and the **sclx**$_4$ atom C18, and the pull force constant is 5 kcal/(mol Å^2).
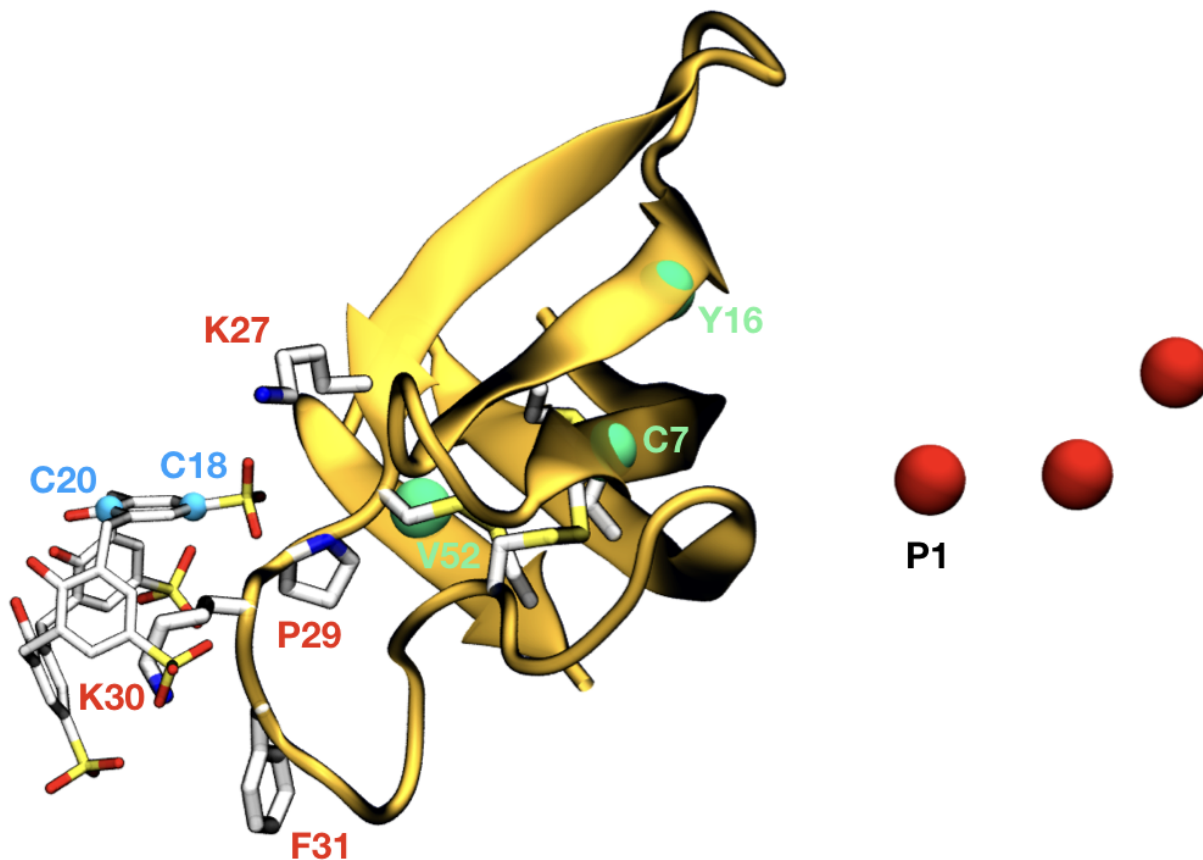
Figure S9: *APR starting structure for PAF–**sclx**$_4$ system, with **sclx**$_4$ interacting in **Site 1**. PAF is represented as orange cartoon, while the ligand molecule and the binding site protein residues in white licorice tubes. Dummy atoms (P1, P2 and P3) are reported as red spheres, while the protein (C7, Y16 and V52 Cα carbons) and the ligand anchor's atoms (C18 and C20) as green and cyan spheres, respectively.*
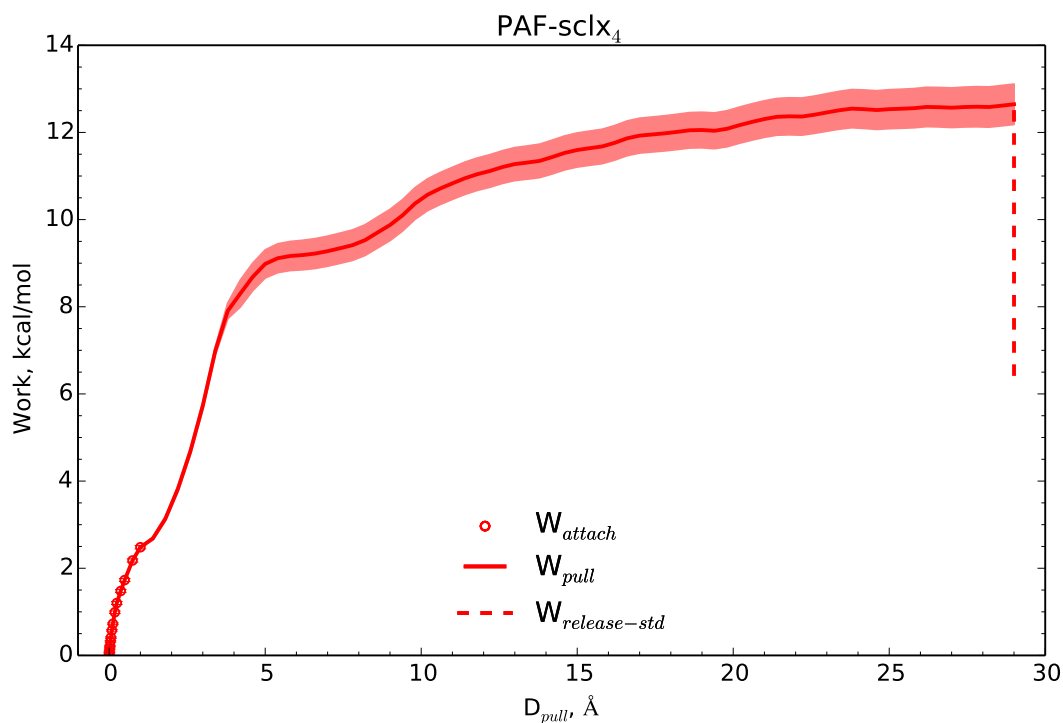
Figure S10: W_{pull} is represented by the potential of mean force (PMF) along the coordinate reaction, while $W_{release-std}$ corresponds to the work of releasing the ligand at the standard concentration (1M) and it is evaluated semi-analytically. The x-axis is defined as the pulling reaction, but the pull is shown starting at 1 Å. The points from 0 to 1 Å must be considered as the attaching phase W_{attach} values (increasing of the constraints values).

Salt Bridge, RMSF and SASA Analysis

PAF unbound was simulated in a box of 72 X 72 X 72 Å^3, solvated by 9778 TIP3P water molecules. Cl− counterions were added to neutralize the system total charge. The system was initially minimized for 10000 steps (5000 of steepest descent and 5000 of conjugate gradient), and then heated up from $0\,K$ to $300\,K$ (with an integration time step t_{step} of 1.0 fs) for a total of 30 ps using the Langevin thermostat ($\gamma_{coll} = 1\,ps^{-1}$) in the isothermal-isochoric ensemble (NVT). Equilibration was carried out for 1 ns, using an integration timestep of 2 fs in the isothermal-isobaric (NPT) ensemble ($P = 1\,atm$ and $T = 300\,K$). During these phases, position restraints were applied on the protein heavy atoms, with a force constant of 5 kcal/mol·Å^2 during the minimization and heating processes and of 2.5 kcal/mol·Å^2 during the equilibration. A subsequent equilibration of 1ns followed with no position restraints applied. The production run was carried out for 1 μs, while for the analysis the first 100 were discarded.
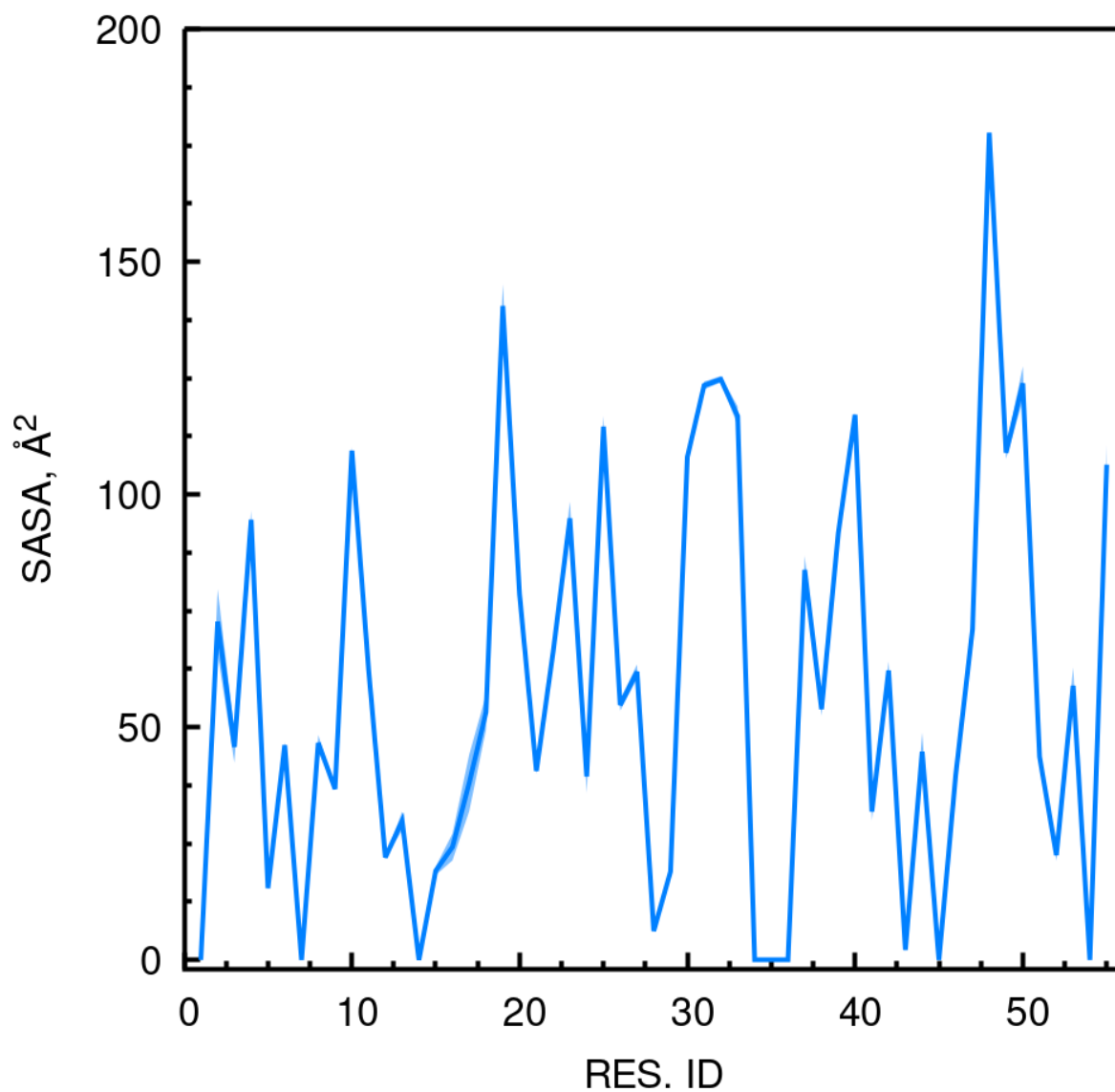
Figure S11: *SASA (in $\mathring{A}^2$) obtained from PAF unbound simulation. The lines indicate the average obtained by block-averaging the simulations: PAF unbound was decomposed in 3 blocks of 300ns each, while the first 100 ns were discarded. The filled blue light area indicate the corresponding standard deviations.*
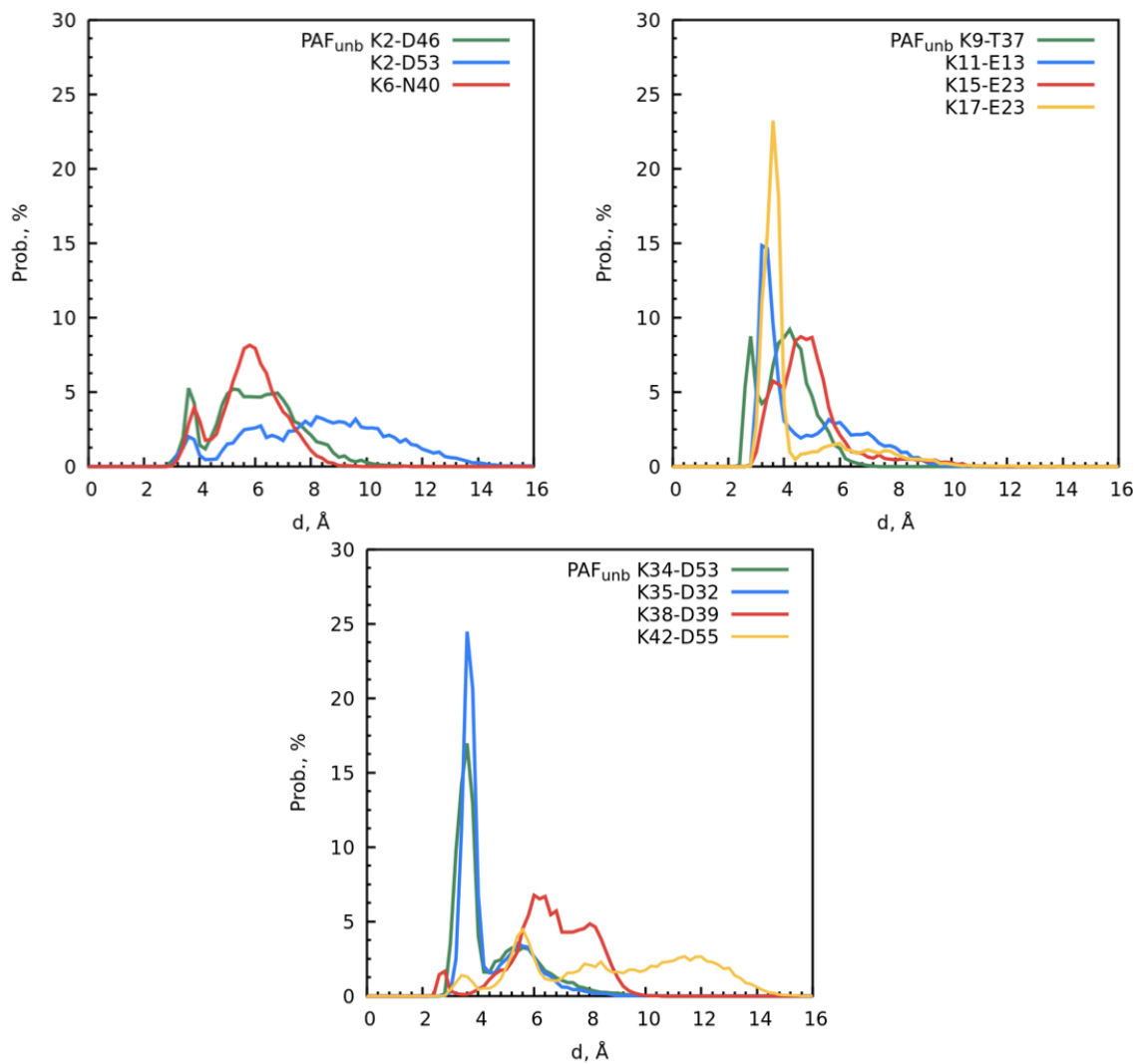
Figure S12: *Lysine salt-bridge distributions (Prob, %) extracted from PAF unbound simulation (see text for details).*
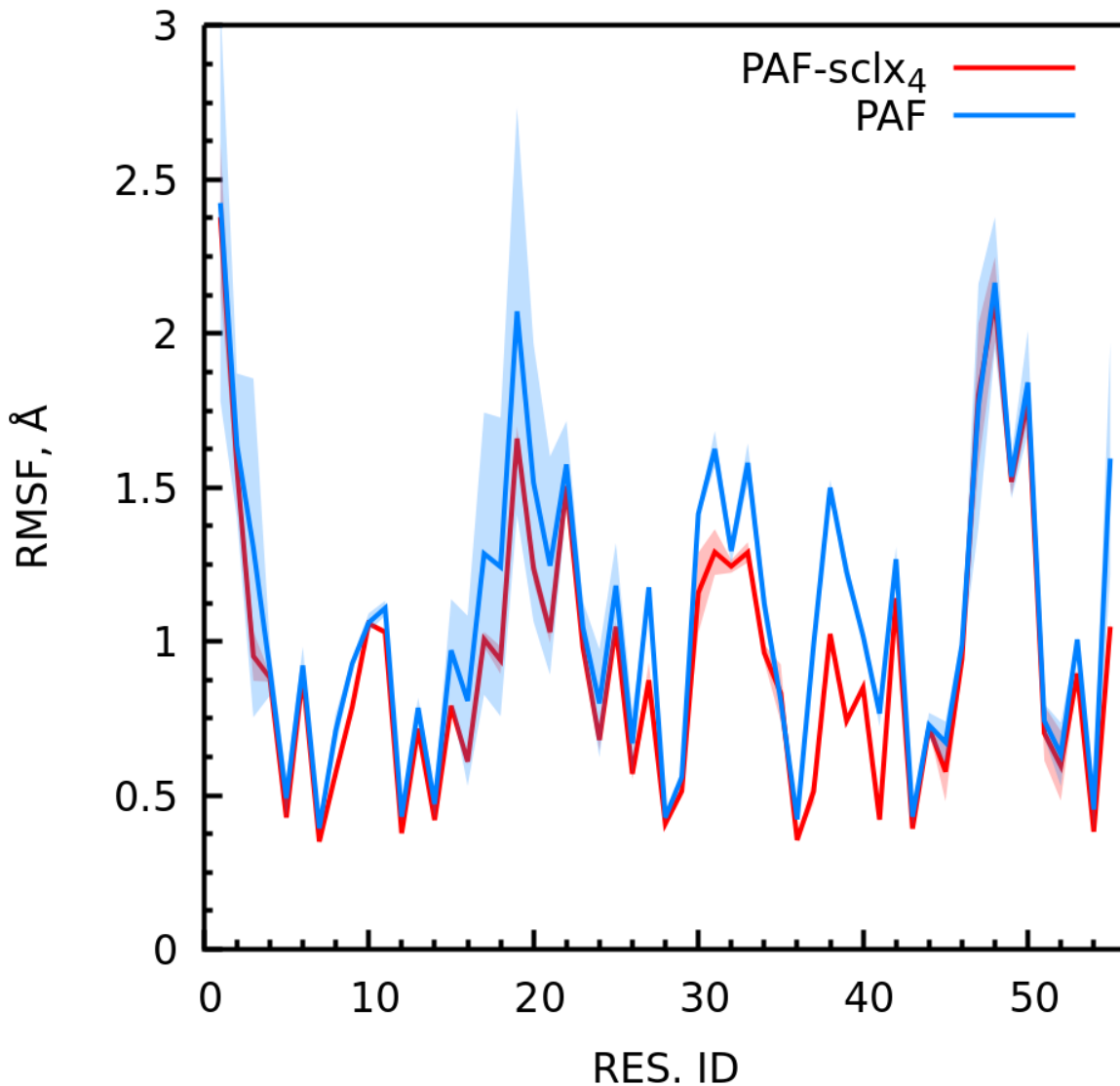
Figure S13: *RMSF (in Å) obtained from PAF unbound simulation (PAF, blue line) and PAF–sclx₄ system (red line), with sclx₄ interacting in* **Site 1***. The lines indicate the average obtained by block-averaging the simulations: PAF unbound was decomposed in 3 blocks of 300ns each, while PAF–sclx₄ in 3 blocks of 500 ns and 2 blocks of 1 μs. The filled blue and red light areas indicate the corresponding standard deviations.*

References

(1) Alex, J. M.; Rennie, M. L.; Engilberge, S.; Lehoczki, G.; Dorottya, H.; Fizil, A.; Batta, G.; Crowley, P. B. Calixarene-mediated assembly of a small antifungal protein. *IUCrJ* **2019**, *6*, 238–247.

(2) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, *79*, 926–935.

(3) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *Journal of chemical theory and computation* **2007**, *3*, 2312–2334.

(4) Bartocci, A.; Pereira, G.; Cecchini, M.; Dumont, E. Capturing the Recognition Dynamics of para-Sulfonato-calix[4]arenes by Cytochrome c: Toward a Quantitative Free Energy Assessment. *Journal of Chemical Information and Modeling* **2022**, *62*, 6739–6748.

(5) Homeyer, N.; Gohlke, H. Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. *Molecular Informatics* **2012**, *31*, 114–122.

(6) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **2012**, *8*, 3314–3321.

(7) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z.; Hou, T. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chemical reviews* **2019**, *119*, 9478–9508.

(8) Velez-Vega, C.; Gilson, M. K. Overcoming dissipation in the calculation of standard binding free energies by ligand extraction. *Journal of computational chemistry* **2013**, *34*, 2360–2371.

(9) Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. Computational calorimetry: high-precision calculation of host–guest binding thermodynamics. *Journal of chemical theory and computation* **2015**, *11*, 4377–4394.

(10) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Gilson, M. K. The SAMPL5 host–guest challenge: computing binding free energies and enthalpies from explicit solvent simulations by the attach-pull-release (APR) method. *Journal of computer-aided molecular design* **2017**, *31*, 133–145.

(11) Heinzelmann, G.; Henriksen, N. M.; Gilson, M. K. Attach-pull-release calculations of ligand binding and conformational changes on the first BRD4 bromodomain. *Journal of chemical theory and computation* **2017**, *13*, 3260–3275.