Capturing the recognition dynamics of para-sulfonato-calix[4]arenes by cytochrome c: towards a quantitative free energy assessment.

Alessio Bartocci,[†,‡,‖] Gilberto Pereira,[‡,¶] Marco Cecchini,[‡] and Elise Dumont[†,§]

†*Univ Lyon, ENS de Lyon, CNRS UMR 5182, Laboratoire de Chimie, F-69342, Lyon, France*

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

¶*Molecular Microbiology and Structural Biochemistry (MMSB, UMR 5086), CNRS & University of Lyon, 7 Passage du Vercors, 69007, Lyon, France*

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

‖*Corresponding author: abartocci@unistra.fr*

E-mail:

Abstract

Calix[n]arenes selective recognition of protein surface covers a broad range of timely applications, notably to control protein assembly and crystallization or trap partially disordered proteins. Here, the interaction of *para*-sulfonated calix-[4]-arenes with cytochrome c is investigated by all-atom, explicit water molecular dynamics simulations which allow to

characterize two binding sites in neat agreement with experimental evidences. Free energy calculations based on the MM-PBSA and the attach-pull-release (APR) methods highlight key residues implicated in the recognition process and provide binding free energy results in quantitative agreement with the isothermal titration calorimetry. Our study emphasizes the role of MD simulations to capture the "walk" of sulfonated calix-[4]-arenes on cytochrome c surface, with R13 as a pivotal interacting residue. Our MD investigation allows, through the QHMB method, to probe an allosteric reinforcement of several per-residues interactions upon calixarene binding, which suggest a more complex mode of action of these supramolecular auxiliaries.

Introduction

Calix[n]arenes have stimulated many applications in Chemistry, Biology and pharmaceutical fields[1–6], notably for chemotherapy delivery strategies[7,8]. Due to their tunable size and geometrical structure, and their unique three-dimensional surface, these synthetic macrocycles with a hydrophobic cage are prone to carry selected drug molecules (e.g. dopamine) towards a given delivery site [7,9–12]. The ease of chemical functionalization of calixarenes offers opportuities for drug-delivery[13], protein inhibition[14–16], and very recently embedding of nanoparticles for cancer therapy[17].

Among calixarenes, *para*-sulfonato-calix[4]arenes (**sclx**$_4$) present an interesting bio[2] and supramolecular chemistry[18] features, such as the ability to recognize post-translational modificated aminoacids like methylated lysines[19,20] and arginines. Like lanthanide complexes, which act as peptide/protein recognition systems [21–23] and crystallization auxiliaries[24,25], *para*-sulfonato-calix[n]arenes have been reported to function as "molecular glues" [26–29], and Crowley and coworkers use them to explore protein–protein interface chemistry, relying on rational supramolecular approaches[30] and electrostatic complementarity. Lysines and

arginines, as positively-charged residues, are the most natural anchor-points for *para*-sulfonato-calixarene-driven protein surface recognition[12,31]. [1]

To characterize para-sulfonato-calix[n]arenes binding "hot-spots" on protein surfaces[32], several experimental approaches have been proposed, from X-ray crystallography and electrospray mass spectrometry, to dynamic light scattering and atomic force microscopy[33]. Also, accurate binding free energies have been determined from isothermal calorimetry titration (ITC) measurements[28,34,35]. Para-sulfonato-calix[4]arenes (**sclx**$_4$, shown in Figure 1) were the first to be investigated in 2012 by Crowley and coworkers[35]. In particular, **sclx**$_4$ is commonly found in host-guests systems [36–39], in which the hydrophobic core plays an important role in interacting with the guest molecule. Last, these systems have attracted interest due to their high functionalization propensity[14,40], for instance with halogens to trigger specific and directional halogen bonds[41].

Para-sulfonato-calixarenes target small cationic proteins, such as hemoglobin[42], bovine serum albumin[33], lysozyme and cytochrome-c (*Cytc*)[34,35,43–46]. The latter has been particularly investigated due to its importance for respiratory and photosynthetic electron transfer. However, capturing the "walk", i.e. the diffusion of the para-sulfonato-calixarenes onto the protein surface, remains an experimental tour-de-force, which was finally achieved ten years ago by Crowley and coworkers specifically for *Cytc*[34]. It would be beneficial to hone a computational approach capable of predicting the binding sites and corresponding free energies of para-sulfonato-calixarenes onto a large range of proteins. The latter can hardly be achieved by docking[47], since it is known experimentally that there is not a single, specific binding pocket which obeys the binding-site definition. Instead, the ligand senses and scans the protein surface, where several possible anchoring points and competitive binding sites, as well as binding events, need to be assessed and quantified by binding free energy

[1] The same authors also proposed to further tune protein frameworks via auxiliary supramolecular interactions, adding a positively-charged guest as an effector to control finely the electrostatic complementarity[26].

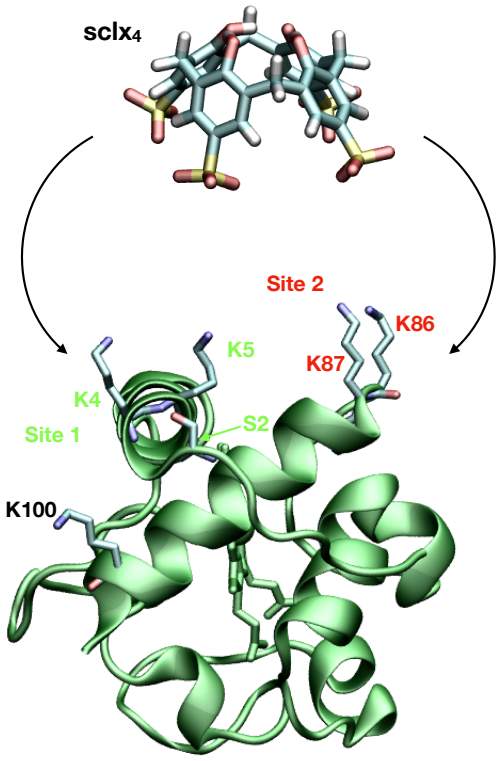

Figure 1: *Approach of the para-sulfonato-calix-[4]-arene ($sclx_4$), with Cytc (in green, PDB ID 3TYI). The two main binding sites probed experimentally, Sites 1 and 2, are labelled in green and in red throughout this work. They feature two lysines tweezers, K4 and K5, and K86 and K87.*

calculations. Also, calixarenes are macromolecules with a considerable conformational flexibility (which increases with n[26]), which is neglected in docking. Further, the response of the protein structure to ligand binding can produce conformational changes and even allosteric effects[48], that docking can only approximate by considering different protein conformations via ensemble docking protocols. Additionally, molecular docking scoring functions are highly approximated and may not be suited for calixarenes binding[49].

In this work, we employed extensive, unbiased molecular dynamics (MD) simulations followed by free energy calculations to investigate the binding of para-sulfonato-calix[4]arenes to *Cytc*, and answers the following questions: which *Cytc* residues are implied in the ligand

surface recognition process? Can one rationalize why some positively-charged residues are not experimentally implied in the binding of $sclx_4$ and clarify a possible competition with salt bridges ? Can we recover binding free energies in line with ITC measurements ?
The simulations results are consistent with both high–resolution structures from X-ray crystallographic[34] and ITC[35] data. They confirm the existence of specific binding sites targeted by the $\mathbf{sclx}_4$ ligand. We obtain an accurate estimate of the binding free energy, using the Attach-Pull-Release (APR)[50–54] method and we identify key residues implied in the protein-calixarenes recognition. By evaluating the configurational entropy loss/change of protein residues far from the calixarene binding site using the Quasi-Harmonic Multi-Basin (QHMB) approach[55], we highlight allosteric effects induced by protein-calixarene binding.

Material and Methods

Force field parameters

Classical all-atom molecular dynamics (MD) simulations have been performed using the AMBER18 package[56]. A single *Cytc* monomer structure was used (108 residues), taken from Cytochrome c PDB ID 3TYI *Saccharomyces cerevisae* structure[34]. The protonation state of titratable residues was determined via H++ server (http://biophysics.cs.vt.edu/H++)[57] at the experimental pH of 6.0[34,35]. The heme group was connected to the H18 and M80 residues using the "Metal Center Parameter Builder" (MCPB) module in AmberTools18[58]. The atomic charges for the heme were obtained from RESP calculations[59] on a geometry optimized at the DFT-B3LYP/6-31G(d) level of theory, following a previous parametrization in our group[48]. A Stuttgart-Dresden SDD pseudopotential was used for the iron center of the heme cofactor. All DFT calculations were performed using the Gaussian 16 revision B.01 series of programs[60]. We note that the heme is not directly involved in the surface interaction between *Cytc* and $\mathbf{sclx}_4$.

Force field parameters for the **sclx**$_4$ ligand were taken from GAFF2[61], while the protein was described with the AMBER/ff14SB[62] force field. The atomic charges for **sclx**$_4$ were assessed from the RESP procedure[59] at the DFT-B3LYP-D3BJ/6-31+G** level of theory. The **sclx**$_4$ is present with one deprotonated phenolic hydroxyl (net charge of -5), as reported in ref.[34].

MD Simulations Protocol

MD simulations were carried out using coordinates from a dimeric co-crystallized *Cytc*-**sclx**$_4$ X-ray structure (PDB ID code: 3TYI), from which one *Cytc* monomer was considered. The ligand **sclx**$_4$ was removed and then placed far from the protein (at more than 10 Å between the centers of mass, see Figure S1), to probe molecular association with no bias. The binding of a second **sclx**$_4$ molecule (*Cytc*-2 **sclx**$_4$ system) was also simulated with independent trajectories (see Figure S2). All the details about relative positions and distances are reported in Table S1. Independent replica of 200 ns (see Table S2) were used to ensure ligand diffusion on the protein surface and sample stable binding events. The last 100 ns of the production trajectories were used to capture relevant features of the main binding modes.

Per-residue energy decomposition was carried out using the molecular mechanics Poisson-Boltzmann surface area method (MM-PBSA)[63,64], which has the key advantage to afford an additive per-residue decomposition with individual contributions $\Delta\Delta G$. Binding free energy calculations were carried out using the attach-pull-release (APR) method[50–54]. Within APR, starting from the initially bound complex, the ligand is forcibly detached from the protein surface towards bulk solvent, and the distance between the two partners is selected as the reaction coordinate along which the potential of mean force (PMF) is constructed (see ESI for further details).

The conformational entropy loss of the ligand as well as protein residue in proximity of the ligand binding site was quantified using the Multibasin Quasi-Harmonic (QHMB)[55]

approach. Details of the method are given in ESI. All pictures were rendered using VMD[65].

MD Simulations Details

The systems were 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 $T = 300\,K$). Pressure control in NPT simulations was achieved using the Berendsen barostat. For both systems, different NPT replica production runs (total simulation time reported in Table S2) with different initial velocities have followed, and they have been carried out with the `pmemd.cuda` module of AMBER18, through the use of graphics processing units (GPUs). In the equilibration and production runs, a cutoff of 10 Å was applied for the van der Waals, for electrostatic interactions and for the real space of the electrostatic interaction. Long-range electrostatic interactions were computed using Particle Mesh Ewald (PME) algorithm[66,67].

Results and Discussion

Cytc is overall positively-charged (+10), which reflects the imperfect balance between positively-charged and negatively-charged residues along the sequence of 108 residues. *Cytc* encompasses 16 lysines (K-2, K4, K5, K11, K22, K27, K54, K55, K72, K73, K79, K86, K87, K89, K99, and K100 — hence with six of them acting as tweezers), 3 arginines (R13, R38 and R91) and 4 histidines (H18, H26, H33, and H39). Negatively-charged residues correspond to 4 aspartates (D50, D60, D90 and D93) and 7 glutamates (E-4, E21, E44, E61, E66, E88 and E103). Ten residues are involved in salt bridges K5-D90, D50-K54, K89-D93, K99-D103 and

K55-E66 (the latter one being an interchain salt bridge in the crystallographic structure). This contributes to reduce the number of stable interaction sites between $\mathbf{sclx}_4$ and the five lysines (K5, K54, K55, K89 and K99) already implicated in salt bridges, even if the competition between salt bridges and $\mathbf{sclx}_4$ remains to be investigated and is one of the scopes of this study.

Structural Characterization

Cytc is a competitive system for $\mathbf{sclx}_4$ binding, yet existing salt bridges interactions reduce the number of possible, stable binding sites $\mathbf{sclx}_4$. Two stable binding sites were identified by visual inspection of our 15 independent MD trajectories of 200 ns each and corroborated by clustering analysis (see ESI). The first, stable interaction between $\mathbf{sclx}_4$ and *Cytc* involves K4 and K5 (Figure 2-a). K4's side chain adopts an inchworm conformation, as the ammonium terminal group lies in the plane of the sulfonato substituents sticking out of the $\mathbf{sclx}_4$ cage. It interacts through a hydrogen bond with the oxygens of sulfonate groups, forming a contact at an average distance of 2.8 Å as the -NH_3^+ terminus rotates with a lasso motion in the plane of the four sulfonato moities. Analogous inchworm conformations have been found for host-guest systems between spermine, spermidine and other supramolecular cages[68]. Meanwhile, K5 interacts with one sulfonato group and remains outside of the cavity. Furthermore, HBs between the sulfonate oxygens and the hydroxyl group of serine S2 (2.6 Å) are also observed. Such configuration has been also observed by X-ray crystallography (3.2 and 2.8 Å respectively, for contacts with K4 and S2), revealing the importance of these residues in calixarene recognition[34].

In the second representative structure (Figure 2-b) a similar binding geometry is observed, with K86's side chain rotating inside the $\mathbf{sclx}_4$ cavity, as characterized by an average interatomic distance between the ammonium terminal nitrogen and the sulfonato sulfurs of 2.9 Å. K87 also interacts with the outer surface of $\mathbf{sclx}_4$. R13's side chain additionally also

forms a HB with one sulfonate group (2.8 Å). Interestingly, Crowley and coworkers[34,35] suggested that the binding process in solution may involve two binding sites centered around K4-K5 and K86-K87, in full agreement with our MD simulations started from fully non-interacting conformations. These two binding sites co-exist, thus raising the possibility that two $\mathbf{sclx}_4$ molecules interact simultaneously with *Cytc*[34,35]. This stands as a pre-requisite for using *para*-sulfonato-calixarenes as molecules glues to consolidate protein-protein interface.

Figure 2-c shows the most representative structure for two calixarenes interacting with the *Cytc* protein at Sites 1 and 2. As already noticed for *Cytc*-$\mathbf{sclx}_4$ system, two lysines adopt an inchworm conformation, whereas vicinal lysines interact outside the calixarene surface with one sulfonato moiety to stabilize binding. Other residues help anchoring the $\mathbf{sclx}_4$ ligand. At Site 1, a sulfonato moiety establishes an additional hydrogen bond with the hydroxyl group of S2. The role of R13 that gives rise to a HB with the sulfonate group of $\mathbf{sclx}_4$ close to Site 2 is also highlighted. In the case of *Cytc*–2 $\mathbf{sclx}_8$, on the other hand, R13 acts as a bridge between the two binding sites, interacting simultaneously with the two ligands because of the higher surface area covered by ligand molecule[48].

Figure 2-d provides a superimposition of $\mathbf{sclx}_4$ interacting at Site 1 in the X-ray reference structure. The relative RMSD values with respect to the reference structure, obtained considering the whole protein backbone atoms, S2 and K4 side-chains and the $\mathbf{sclx}_4$ molecule, are 1.00 Å, 0.90 Å, 0.96 Å and 0.96 Å for site 1, while 1.00 Å, 0.66 Å, 0.60 Å and 0.93 Å for site 2. These low RMSD values show that the MD-captured structure of $\mathbf{sclx}_4$ in Site 1 is in excellent agreement with the crystallographic structure.

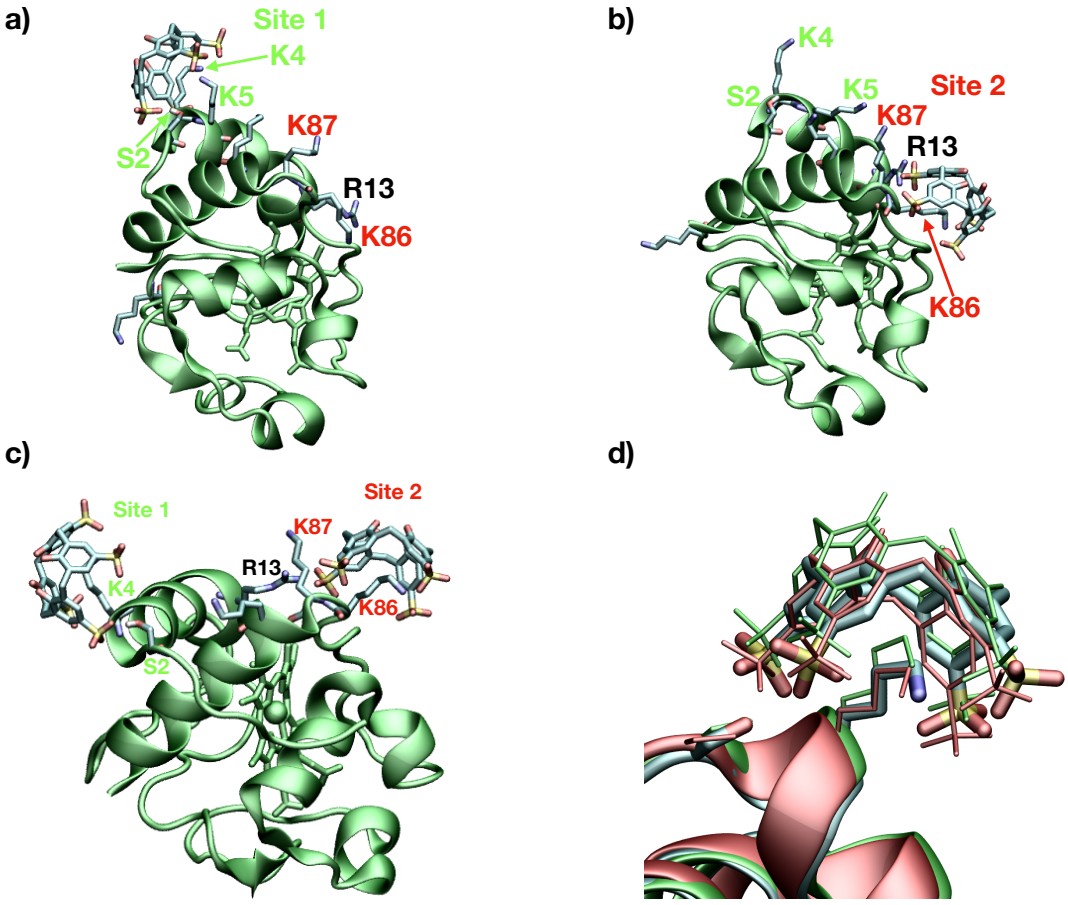

Figure 2: *Upper panel: two most representative structures for Cytc-*$\mathbf{sclx_4}$*, with the ligand sclx4 interacting in a) Site 1 and in b) Site 2, obtained from cluster analysis of the trajectories. c) Most representative structure for Cytc-2* $\mathbf{sclx_4}$ *system, with the two ligands interacting in Sites 1 and 2. R13 residue is also labelled in black. Cytc is reported in cartoon representation colored of green, while the* $\mathbf{sclx_4}$ *and the heme cofactor of Cytc are shown in licorice tubes. d) Inset of the superimposition of the X-ray structure (Cytc in cyan cartoons,* $\mathbf{sclx_4}$ *molecule, S2 and K4 residues in licorice tubes) with structures of panel a (Cytc in red cartoons,* $\mathbf{sclx_4}$*, S2 and K4 residues in red lines) and c (Cytc in green cartoons,* $\mathbf{sclx_4}$*, S2 and K4 residues in green lines). Hydrogen atoms are not shown for sake of clarity. K4 adopts an inchworm conformation.*

Per-residue Decomposition

The interaction between the surface residues of *Cytc* and $\mathbf{sclx}_4$ was analysed by decomposing the free energy of binding, evaluated by the Molecular Mechanics/Poisson-Boltzmann surface area (MM-PBSA) approach[69] on the production MD trajectories (see Table S2 of ESI), into per residue contributions.

Although MM-PBSA usually strongly overshoots the absolute binding affinity of $\mathbf{sclx}_4$ for *Cytc*, this analysis provides a readable "footprint" of the protein-ligand interaction (Figure 3). The most sizeable contributions are listed in Table 1. The key role of charged residues like lysines and arginines single out in Figure 3 and corroborate the visual inspection of the stable binding sites. For the *Cytc*–$\mathbf{sclx}_4$ system, it is observed that the calixarene interacts with protein residues K4 and K5 at Site 1, with an interaction energy of -3.4 ± 1.3 and -1.4 ± 0.9 kcal·mol^{-1}, respectively. Similarly, residues K86 and K87 play a central role to bind the calixarene $\mathbf{sclx}_4$ at Site 2, with a per-residue contribution of -7.7 ± 1.2 and -7.3 ± 1.3 kcal·mol^{-1}, respectively. The presence of the second ligand slightly modifies the interaction pattern. In this case, the contribution of K4 is enhanced (-6.7 ± 1.2 kcal·mol^{-1}), and it is closer to that of K86 (-5.0 ± 1.6 kcal·mol^{-1}) and K87 (-4.5 ± 1.8 kcal·mol^{-1}). Another non negligible contribution is introduced by R13, which is similar in both systems (-4.3 ± 1.5 and -3.4 ± 2.0 kcal·mol^{-1}). The R13 role, effectively, is to stabilize the calixarene binding at Site 2.

Interestingly, a third tweezer K72-K73 singles out from monitoring the per-residue contributions (see Figure 3, lower panel), lower panel, with a per residue contribution of -1.1 ± 0.8 and -0.7 ± 0.2 kcal·mol^{-1}.

Specific interactions of $\mathbf{sclx}_4$ with the small-size *Cytc* involve residues at Sites 1 and 2, with no other significant contribution, as shown by the color map provided in Figure 4. Per-type contributions from positively and negatively-charged residues are reported at the bottom panel of Figure 4. The most important contributions include lysines residues (-26.6 ± 2.8 and

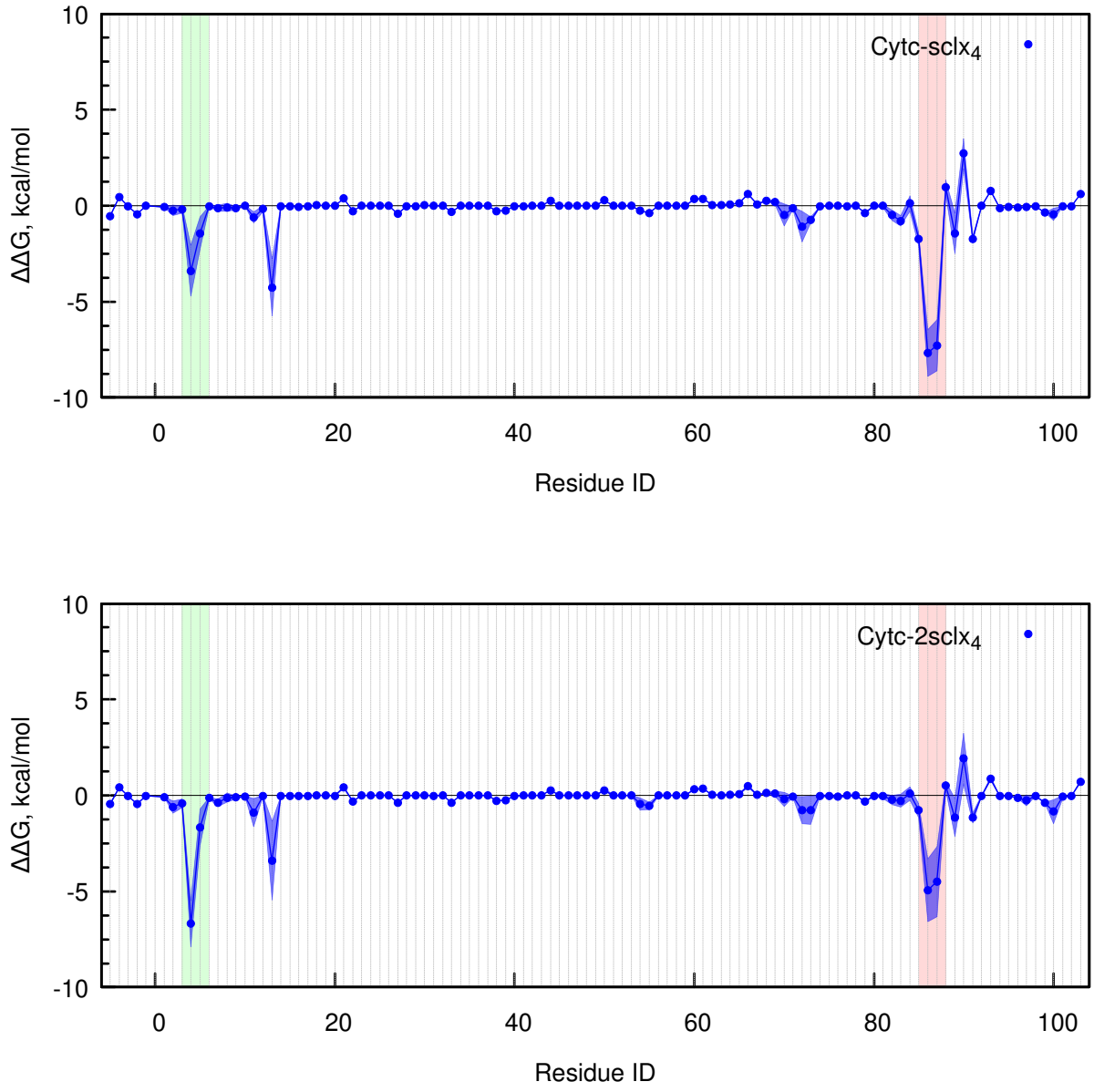

Figure 3: *Per-residue decomposition $\Delta\Delta G$ of the total interaction energy estimated by the MM-PBSA approach, in kcal·mol^{-1} as a function of residue number for the system Cytc-* **sclx**$_4$ *(upper panel), and for the system Cytc-2* **sclx**$_4$ *(lower panel). The colorboxes in green and red correspond respectively to Site 1 and Site 2.*

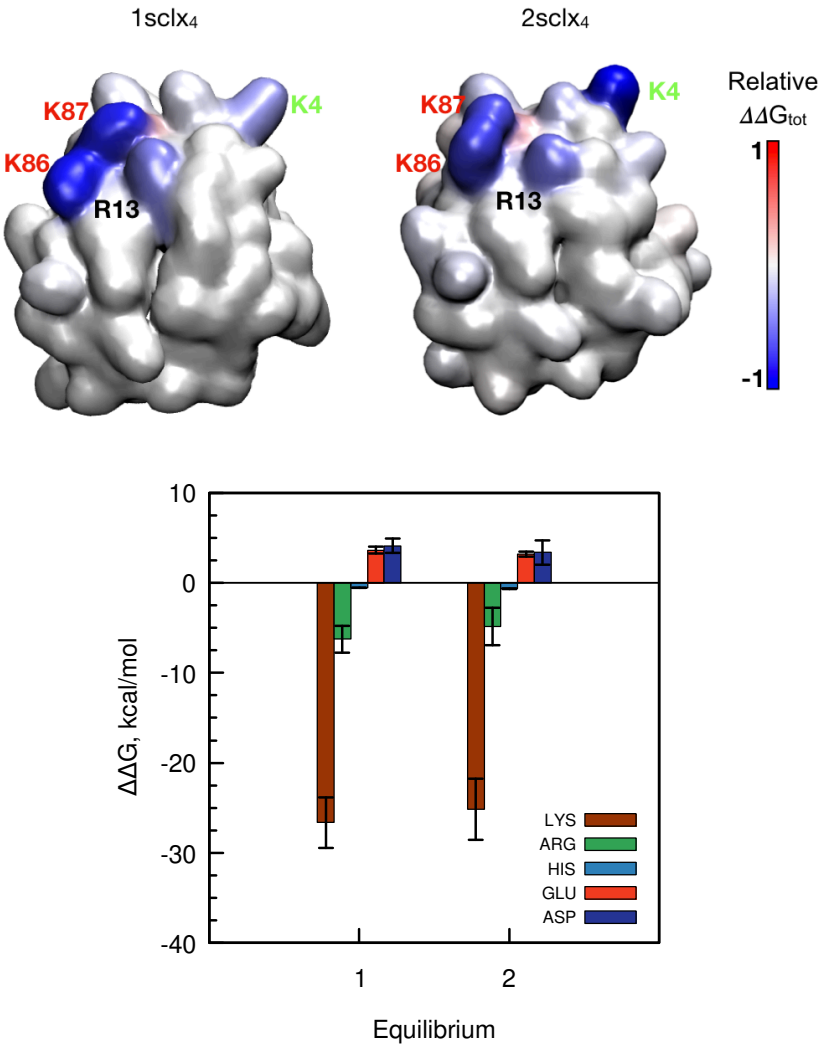

Figure 4: *Upper panel: color map of per-residue decomposition* $\Delta\Delta G_1$ *(leftside) or* $\Delta\Delta G_2$ *(rightside) for Cytc-***sclx**$_4$ *and Cytc-2* **sclx**$_4$ *systems, respectively. The color scale is defined by dividing each contribution by the maximal absolute one (*$\Delta\Delta G_1$*(K86) and* $\Delta\Delta G_2$*(K4)) and using the color codes blue for attractive interactions and red for repulsive interactions. Lower panel: per-type decomposition over charged residues: lysine (LYS), arginine (ARG), histidine (HIS), glutamate (GLU), and aspartate (ASP) for Cytc-***sclx**$_4$ *and Cytc-2* **sclx**$_4$ *systems, respectively.*

-25.1 ± 3.4 kcal·mol^{-1} for equilibrium 1 and 2, respectively). Additionally, arginines show a non-negligible attractive contribution (-6.3 ± 1.5 and -4.9 ± 2.1 kcal·mol^{-1}). Similar to *Cytc*–**sclx**$_8$ [48] recognition, the ability of calixarenes to interact strongly with charged residues makes them adaptable to the protein surface area, corroborating the binding process as electrostatically-driven. It is well known experimentally that protein-calixarene binding sites are most often lysine-rich regions [70]. Binding implies sulfonato groups of **sclx**$_4$ molecule which interact through electrostatics and hydrogen bonds (HB) (see upper panel of Figure 2). Note that since the fourth K44-K55 tweezer, which belongs to the Ω loop, is already implicated in two salt bridge interactions, and hence cannot serve as a potential anchor point for **sclx**$_4$ binding.

Table 1: *Interacting residues for Sites 1 and 2 and auxiliaries* **sclx**$_4$ *coordinating residues onto the Cytc surface, and their per-residue decomposition contributions of the first Cytc–*$\mathbf{sclx}_4(\Delta\Delta G_1)$ *and of the second Cytc–2* $\mathbf{sclx}_4(\Delta\Delta G_2)$ *equilibrium. The values of the interface area (Int. Area) and* Δ^i_{SASA} *(see Eq. S6 and S7) are also reported, while residues identified in the experimental crystal structure are boldfaced*[34].

	RES	$\Delta\Delta G_1$ (kcal.mol^{-1})	$\Delta\Delta G_2$ (kcal.mol^{-1})	Int. Area (Å^2) *endo*	Int. Area (Å^2) *exo*	Δ^i_{SASA} (Å^2) *endo*	Δ^i_{SASA} (Å^2) *exo*
Site 1	**S2**	-0.2 ± 0.2	-0.6 ± 0.3	216 ± 9.0 (**230**)	210.6 ± 3.0 (**200**)	21.1 ± 2.0	-
	K4	-3.4 ± 1.3	-6.7 ± 1.2			106.6 ± 3.8	65.0 ± 4.2
	K5	-1.4 ± 0.9	-1.7 ± 1.0			-	21.5 ± 6.0
Site 2	**K86**	-7.7 ± 1.2	-5.0 ± 1.6	272.4 ± 36.1	202 ± 68.4	90.6 ± 30.0	45.6 ± 25.0
	K87	-7.3 ± 1.3	-4.5 ± 1.8			40.8 ± 10.0	45.5 ± 9.8
Other	R13	-4.3 ± 1.5	-3.4 ± 2.0	-		33.1 ± 20	
	K72	-1.1 ± 0.8	-0.8 ± 0.7	-	145.6 ± 5.1	-	40.5 ± 3.7
	K73	-0.7 ± 0.2	-0.8 ± 0.7			-	34.4 ± 24.0

Ligand Conformations

Our MD simulations also shed light on the conformations adopted by the ligand as it approaches the protein surface, which can be directly compared with experimental data from X-ray structures. Indeed, Crowley and co-workers have obtained contact area results at the protein-ligand interface[34]. Table 1 collates the values of the interface area and Δ^{i}_{SASA}, defined as the contribution of the *i*-th residue to the total interface area (see Eq. S6 and S7). These values provide a direct measure of the residue surface offered to the calixarene while binding, and are in good agreement with the experimental findings (marked in bold in Table 1). Inspection of these data corroborate the dynamic nature of the interaction and suggests how the ligand might explore the protein surface, adopting different conformations, as already pointed out by the analysis of the crystal packing[34]. X-ray structures reveal that $\mathbf{sclx}_4$ can also interact at the *exo* surface (lower rim), in which case the alkyl portion of lysine protein residues is important. In order to be able to catch these conformations, an analysis of the approaching angle Θ, defined as the angle formed between the centers of mass of the protein, the ligand sulfur and hydroxyl oxygen atoms, has been performed while the ligand interacts onto the protein surface. A normalized distribution is reported in Figure S3, pointing out that the *exo* and the *endo* (upper rim) conformations have an average Θ angle of $\simeq 50\ ^{\circ}$ and $\simeq 140\ ^{\circ}$, respectively. The *exo* conformations, thus, have a role in allowing *Cytc* to anchor the $\mathbf{sclx}_4$ to from multiple sides, as shown by its broader distributions. According to these results, $\mathbf{sclx}_4$ spans different conformations while probing the surface. The *exo* conformation corresponds, as shown in Figure S4, to an alternative binding mode where a lysine tweezer coordinates $\mathbf{sclx}_4$ and the alkyl side chain of lysine form CH$\cdots\pi$ interaction with the phenyl ring of the calixarene.

Protein-$\mathbf{sclx_4}$ Association: Binding Free-Energy and Allostery

Our agreement with experimental findings also stands from the thermodynamic point of view. The experimental dissociation constant K_d for the *Cytc*-$\mathbf{sclx_4}$ system was obtained via ITC experiments[35], and corresponds to a binding affinity of -6.3 ± 0.2 kcal·mol^{-1}. Table 2 provides a comparison between this value and computed ΔG°_{bind} by APR approach.

Table 2: *Experimental and calculated (APR method)* ΔG°_{bind} *binding affinity for Cytc-*$\mathbf{sclx_4}$ *association.*

System	**Exp.**[35]	**Method**	**Site 1**	**Site 2**	**Total**
Cytc–$\mathbf{sclx_4}$	-6.3 ± 0.2	APR	-7.5 ± 0.5	-6.4 ± 0.6	-7.6 ± 0.4

The calculated binding affinities through the APR method for the *Cytc*–$\mathbf{sclx_4}$ system lie in very good agreement with the experimental result ($\Delta\Delta G$ = 1.3 kcal·mol^{-1}), while the MM-PBSA (-31.72 ± 0.13 kcal·mol^{-1}) method strongly overestimates the value. Per-site binding affinities differ by ca. 1.1 kcal·mol^{-1}, pointing to nearly equivalent binding sites. Our computational approach also allows to identify protein allosteric events upon $\mathbf{sclx_4}$ binding, and how they influence the flexibility of near and distant residues from the binding site. The configurational entropy loss upon $\mathbf{sclx_4}$ binding was evaluated through the QHMB method, from bound and unbound states. The binding process also entails a freezing of the motions of some residues in the binding site, as well as a global rearrangement of the protein salt bridges network all around. Their configurational entropy loss upon binding for a selection of proximal residues in both sites (see Figure S5 for details) has been calculated following Eq. S8, and range between 8.5 and 11 kcal·mol^{-1}(see values in Table S4). The final values for both equilibria are reported in Figure 5.

Entropic contributions evidence a significant change in the flexibility of the ligand and

the protein residues involved in the binding. The second ligand can trigger conformational modifications to the other binding site as shown by $\Delta\left(T\Delta S^{conf}_{b-unb}\right)$ between Site 1 and 2, that passes from 0.44 to -1.50 kcal·mol^{-1}(Figure 5). This feature denotes a cooperative contribution of the binding even if enthalpic and solvation contributions may also play an effective role in the per-site binding mechanism.

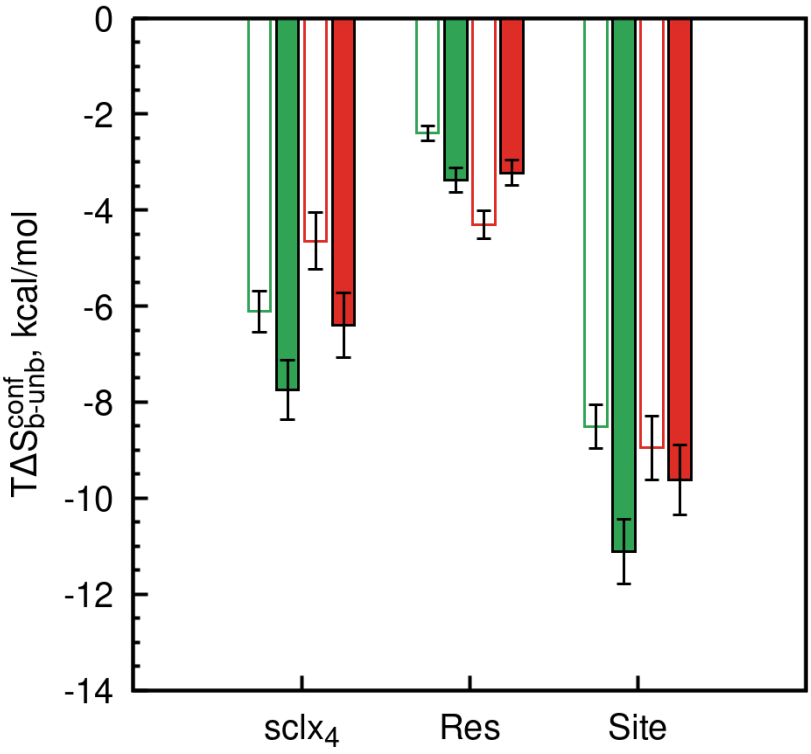

Figure 5: *Entropic contributions($T\Delta S^{conf}_{b-unb}$) in kcal·mol^{-1} for the first (Cytc–***sclx**$_4$*, empty histograms) and the second (Cytc–2* **sclx**$_4$*, filled histograms) associations for binding sites 1 and 2 (in green and red), through the QHMB approach. The corresponding values are reported in Table S4.*

By enlarging the configurational entropic analysis to all protein residues, we highlight allosteric effects upon ligand binding. Figure 6 provides a color map for $T\Delta S^{conf}_{b-unb}$, where red colored residues exhibit more confined movements in the presence of the ligand(s) (bound

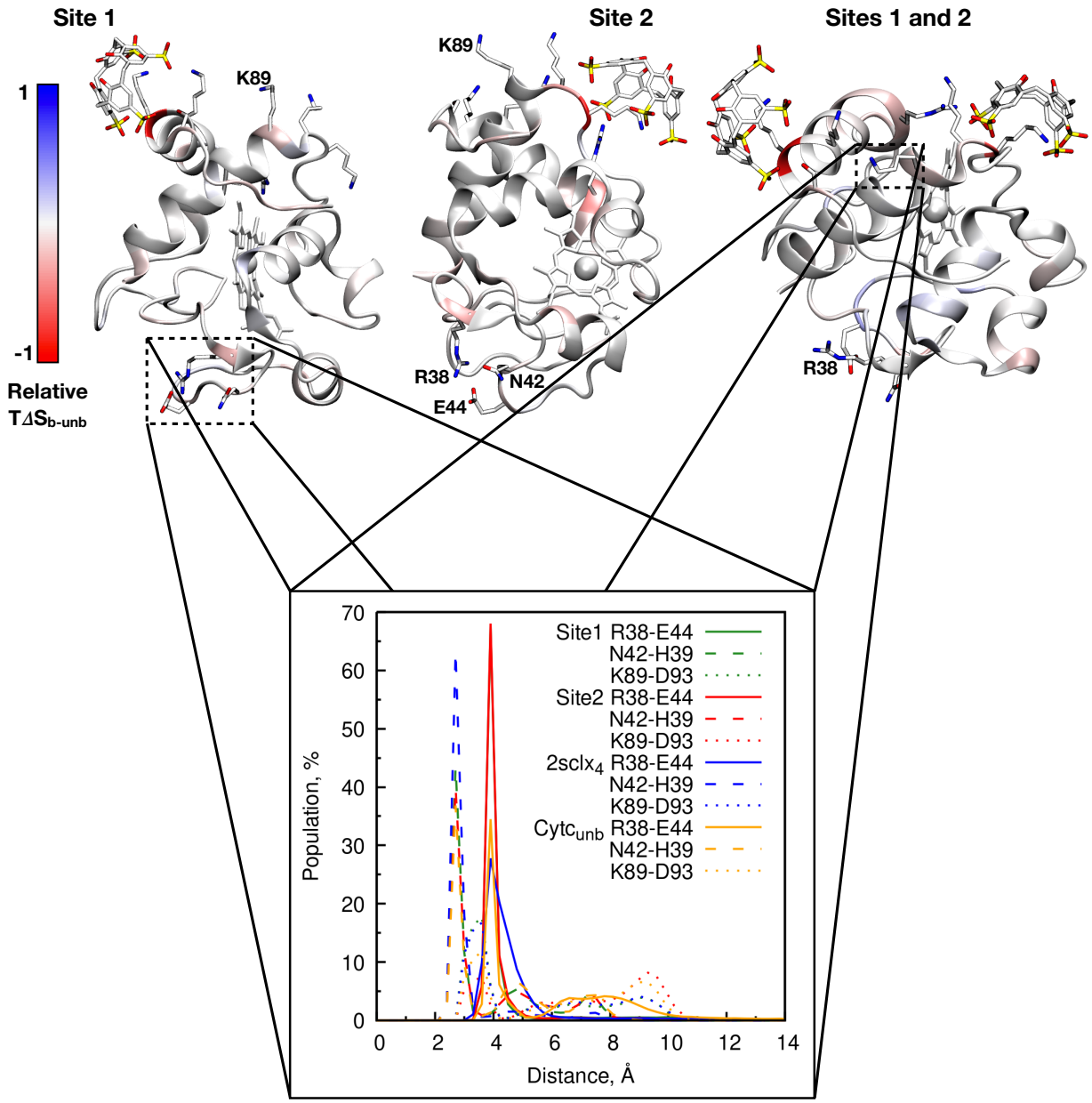

Figure 6: *Upper panels: color maps of* $T\Delta S^{Res_i}_{b-unb}$ *for* $\mathbf{sclx}_4$ *ligand interacting in Site 1 (left panel), Site 2 (central panel) and in Sites 1 and 2 simultaneously (right panel), respectively. The color scale is defined by dividing each contribution by the maximal absolute one* ($T\Delta S^{Res_i}_{b-unb}$ *(K4), (K86) (K4), respectively, for the three panels) and using the color codes red for negative values and blue for positive values. Residues R38, N42, E44 and K89 are reported as directly affected by the analysis*
Lower panel: salt-bridge distributions for R38-E44, N42-H39 and K89-D93 pair interactions.

state) than without (unbound state), which arise from the tightening of the interaction network with surrounding residues. R38, N42 and E44, belonging to the Ω-loop opposite to Sites 1 and 2 (Figure 6), undergo a dynamical modification of salt bridge network in which they are involved. The presence of one $\mathbf{sclx}_4$ molecule strenghtens R38-E44 salt bridge, as shown by the distance distribution reported as inset in Figure 6. This induces less fluctuations for R38 and E44. N42-H39 interaction is also reinforced upon binding of two $\mathbf{sclx}_4$, compared to the binding of a unique ligand. On the other hand, K89, that lies between sites 1 and 2, establishes salt bridges with D90 and D93 negatively-charged residues. K89's accessibility in solution is highly influenced by the dynamical salt bridge formation with D93[34], such that the binding occurs at K86-K87 rather than K89. Such interaction is maintained mostly with one ligand bound in Site 1, stressing out the cooperativity effects upon binding of a second $\mathbf{sclx}_4$ molecule.

Conclusions

The association between $\mathbf{sclx}_4$, proposed as molecular glues, and cytochrome c is investigated by MD simulations and free energy calculations. Computational results provide a detailed structural and thermodynamic description of *para*-sulfonated calix-[4]-arenes-protein binding consistent with experimental evidences by Crowley and coworkers. Two calixarene recognition sites are identified *de novo* by MD simulations and lie in neat agreement with previous studies. Analysis of the per-residue contributions to binding highlights the importance of residues like K4, K5, K86, K87, in agreement with experimental NMR and X-ray crystallographic data[34], and R13.

For such electrostatically-driven protein-ligand interaction, we show the superiority of the APR method, which yields results in quantitative agreement with ITC measurements[35]. This method is set up as a reliable computational approach to explore the energetics of

protein-calixarene binding and may be used for the rational design of ad-hoc functionalized protein crystallization auxiliaries.

MD simulations also allow to probe an accurate protocol which offers insights into prominent protein systems. The evaluation of the configurational entropic penalty, which affects more strongly large and flexible ligands as $\mathbf{sclx}_4$, with the QHMB method has enabled a detailed analysis of the dynamical rearrangement of protein-ligand system during the binding. It has revealed, in fact, that the protein salt bridge network can be influenced by the presence of the ligand. Specific surface regions may become more rigid, highlighting a reduction of surface entropy, and the ligand can thus promote protein-protein contacts[71].

Acknowledgement

A.B. is grateful for a post-doctoral fellowship by the Fundation "Maison de la Chimie". The authors thank the SYSPROD project and AXELERA Pôle de Compétitivité for financial support (PSMN Data Center). G.P. is grateful to the Doctoral School of Chemical Sciences of the University of Strasbourg (ED222) for financial support. Computational HPC resources from the Mesocentre of the University of Strasbourg and the Mesocentre of the University of Reims, ROMEO, are also gratefully acknowledged.

References

(1) Giuliani, M.; Morbioli, I.; Sansone, F.; Casnati, A. Moulding calixarenes for biomacromolecule targeting. *Chem. Commun.* **2015**, *51*, 14140–14159.

(2) Perret, F.; Lazar, A. N.; Coleman, A. W. Biochemistry of the para-sulfonato-calix [n] arenes. *Chemical Communications* **2006**, 2425–2438.

(3) Español, E. S.; Villamil, M. M. Calixarenes: Generalities and Their Role in Improving

the Solubility, Biocompatibility, Stability, Bioavailability, Detection, and Transport of Biomolecules. *Biomolecules* **2019**, *9*, 90.

(4) Nimse, S. B.; Kim, T. Biological applications of functionalized calixarenes. *Chem. Soc. Rev.* **2013**, *42*, 366–386.

(5) Gutsche, C. D.; Bauer, L. J. Calixarenes. 13. The conformational properties of calix [4] arenes, calix [6] arenes, calix [8] arenes, and oxacalixarenes. *Journal of the American Chemical Society* **1985**, *107*, 6052–6059.

(6) Danil de Namor, A. F.; Cleverley, R. M.; Zapata-Ormachea, M. L. Thermodynamics of calixarene chemistry. *Chemical reviews* **1998**, *98*, 2495–2526.

(7) Basilotta, R.; Mannino, D.; Filippone, A.; Casili, G.; Prestifilippo, A.; Colarossi, L.; Raciti, G.; Esposito, E.; Campolo, M. Role of Calixarene in Chemotherapy Delivery Strategies. *Molecules* **2021**, *26*, 3963.

(8) Fahmy, S. A.; Ponte, F.; Fawzy, I. M.; Sicilia, E.; Azzazy, H. M. E.-S. Betaine host–guest complexation with a calixarene receptor: enhanced in vitro anticancer effect. *RSC Advances* **2021**, *11*, 24673–24680.

(9) Wang, L.; Li, L.-l.; Fan, Y.-s.; Wang, H. Host–guest supramolecular nanosystems for cancer diagnostics and therapeutics. *Advanced Materials* **2013**, *25*, 3888–3898.

(10) Rahimi, M.; Karimian, R.; Noruzi, E. B.; Ganbarov, K.; Zarei, M.; Kamounah, F. S.; Yousefi, B.; Bastami, M.; Yousefi, M.; Kafil, H. S. Needle-shaped amphoteric calix [4] arene as a magnetic nanocarrier for simultaneous delivery of anticancer drugs to the breast cancer cells. *International journal of nanomedicine* **2019**, *14*, 2619.

(11) Pan, Y.-C.; Hu, X.-Y.; Guo, D.-S. Biomedical applications of calixarenes: state of the art and perspectives. *Angewandte Chemie International Edition* **2021**, *60*, 2768–2794.

(12) Douteau-Guével, N.; Coleman, A. W.; Morel, J.-P.; Morel-Desrosiers, N. Complexation of basic amino acids by water-soluble calixarene sulphonates as a study of the possiblemechanisms of recognition of calixarene sulphonates by proteins. *Journal of physical organic chemistry* **1998**, *11*, 693–696.

(13) Athar, M.; Lone, M. Y.; Jha, P. C. Designing of calixarene based drug carrier for dasatinib, lapatinib and nilotinib using multilevel molecular docking and dynamics simulations. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **2018**, *90*, 157–169.

(14) Buldenko, V. M.; Trush, V. V.; Kobzar, O. L.; Drapailo, A. B.; Kalchenko, V. I.; Vovk, A. I. Calixarene-based phosphinic acids as inhibitors of protein tyrosine phosphatases. *Bioorganic & medicinal chemistry letters* **2019**, *29*, 797–801.

(15) Vovk, A. I.; Kalchenko, V. I.; Cherenok, S. A.; Kukhar, V. P.; Muzychka, O. V.; Lozynsky, M. O. Calix [4] arene methylenebisphosphonic acids as calf intestine alkaline phosphatase inhibitors. *Organic & biomolecular chemistry* **2004**, *2*, 3162–3166.

(16) Chini, M. G.; Terracciano, S.; Riccio, R.; Bifulco, G.; Ciao, R.; Gaeta, C.; Troisi, F.; Neri, P. Conformationally locked calixarene-based histone deacetylase inhibitors. *Organic letters* **2010**, *12*, 5382–5385.

(17) Liu, Q.; Zhang, T.-X.; Zheng, Y.; Wang, C.; Kang, Z.; Zhao, Y.; Chai, J.; Li, H.-B.; Guo, D.-S.; Liu, Y., et al. Calixarene-Embedded Nanoparticles for Interference-Free Gene–Drug Combination Cancer Therapy. *Small* **2021**, *17*, 2006223.

(18) Guo, D.-S.; Liu, Y. Supramolecular chemistry of p-sulfonatocalix [n] arenes and its biological applications. *Accounts of chemical research* **2014**, *47*, 1925–1934.

(19) Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F. A simple calixarene recognizes post-translationally methylated lysine. *ChemBioChem* **2010**, *11*, 63–66.

(20) Deng, D.; Yang, X.; An, J.; Zhang, K.; Lin, S.; Dong, X. Sulfonated calix [4] arene functionalized SiO2@ TiO2 for recognition of lysine methylation. *Talanta* **2021**, *224*, 121819.

(21) Denis-Quanquin, S.; Riobé, F.; Delsuc, M.-A.; Maury, O.; Giraud, N. Paramagnetic DOSY: An Accurate Tool for the Analysis of the Supramolecular Interactions between Lanthanide Complexes and Proteins. *Chemistry – A European Journal* **2016**, *22*, 18123–18131.

(22) Denis-Quanquin, S.; Bartocci, A.; Szczepaniak, F.; Riobé, F.; Maury, O.; Dumont, E.; Giraud, N. Capturing the dynamic association between a tris-dipicolinate lanthanide complex and a decapeptide: a combined paramagnetic NMR and molecular dynamics exploration. *Physical Chemistry Chemical Physics* **2021**, *23*, 11224–11232.

(23) Dumont, E.; Pompidor, G.; d'Aléo, A.; Vicat, J.; Toupet, L.; Kahn, R.; Girard, E.; Maury, O.; Giraud, N. Exploration of the supramolecular interactions involving tris-dipicolinate lanthanide complexes in protein crystals by a combined biostructural, computational and NMR study. *Physical Chemistry Chemical Physics* **2013**, *15*, 18235–18242.

(24) Pompidor, G.; d'Aleo, A.; Vicat, J.; Toupet, L.; Giraud, N.; Kahn, R.; Maury, O. Protein crystallography through supramolecular interactions between a lanthanide complex and arginine. *Angewandte Chemie* **2008**, *120*, 3436–3439.

(25) Roux, A.; Talon, R.; Alsalman, Z.; Engilberge, S.; d'Aléo, A.; Di Pietro, S.; Robin, A.; Bartocci, A.; Pilet, G.; Dumont, E.; T., W.; Seigo, S.; Riobé, F.; Girard, E.; Maury, O. Influence of divalent cations in the protein crystallization process assisted by lanthanide-based additives. *Inorganic Chemistry* **2021**, *60*, 15208–15214.

(26) Engilberge, S.; Rennie, M. L.; Dumont, E.; Crowley, P. B. Tuning Protein Frameworks via Auxiliary Supramolecular Interactions. *ACS Nano* **2019**, *13*, 10343–10350.

(27) Ramberg, K. O.; Engilberge, S.; Skorek, T.; Crowley, P. B. Facile Fabrication of Protein–Macrocycle Frameworks. *Journal of the American Chemical Society* **2021**, *143*, 1896–1907.

(28) Rennie, M. L.; Doolan, A. M.; Raston, C. L.; Crowley, P. B. Protein dimerization on a phosphonated calix [6] arene disc. *Angewandte Chemie International Edition* **2017**, *56*, 5517–5521.

(29) Selkti, M.; Coleman, A. W.; Nicolis, I.; Douteau-Guével, N.; Villain, F.; Tomas, A.; de Rango, C. The first example of a substrate spanning the calix [4] arene bilayer: the solid state complex of p-sulfonatocalix [4] arene with l-lysine. *Chemical Communications* **2000**, 161–162.

(30) Alex, J. M.; McArdle, P.; Crowley, P. B. Supramolecular stacking in a high Z? calix [8] arene–porphyrin assembly. *CrystEngComm* **2020**, *22*, 14–17.

(31) McGovern, R. E.; McCarthy, A. A.; Crowley, P. B. Protein assembly mediated by sulfonatocalix [4] arene. *Chemical Communications* **2014**, *50*, 10412–10415.

(32) Martins, J. N.; Lima, J. C.; Basílio, N. Selective recognition of amino acids and peptides by small supramolecular receptors. *Molecules* **2021**, *26*, 106.

(33) Memmi, L.; Lazar, A.; Brioude, A.; Ball, V.; Coleman, A. W. Protein–calixarene interactions: complexation of bovine serum albumin by sulfonatocalix [n] arenes. *Chemical Communications* **2001**, 2474–2475.

(34) McGovern, R. E.; Fernandes, H.; Khan, A. R.; Power, N. P.; Crowley, P. B. Protein camouflage in cytochrome c–calixarene complexes. *Nature chemistry* **2012**, *4*, 527.

(35) Doolan, A. M.; Rennie, M. L.; Crowley, P. B. Protein recognition by functionalized sulfonatocalix [4] arenes. *Chemistry–A European Journal* **2018**, *24*, 984–991.

(36) Rebek Jr, J. Host–guest chemistry of calixarene capsules. *Chemical Communications* **2000**, 637–643.

(37) Garnier, L.; Bonal, C.; Malfreyt, P. Thermodynamics of Supramolecular Associations with Macrocyclic Water-Soluble Hosts. *ACS omega* **2019**, *4*, 16899–16905.

(38) Kim, H. J.; Lee, M. H.; Mutihac, L.; Vicens, J.; Kim, J. S. Host–guest sensing by calixarenes on the surfaces. *Chemical Society Reviews* **2012**, *41*, 1173–1190.

(39) Tian, H.-W.; Chang, Y.-X.; Hu, X.-Y.; Shah, M. R.; Li, H.-B.; Guo, D.-S. Supramolecular imaging of spermine in cancer cells. *Nanoscale* **2021**,

(40) Trush, V. V.; Cherenok, S. O.; Tanchuk, V. Y.; Kukhar, V. P.; Kalchenko, V. I.; Vovk, A. I. Calix [4] arene methylenebisphosphonic acids as inhibitors of protein tyrosine phosphatase 1B. *Bioorganic & medicinal chemistry letters* **2013**, *23*, 5619–5623.

(41) Voth, A. R.; Khuu, P.; Oishi, K.; Ho, P. S. Halogen bonds as orthogonal molecular interactions to hydrogen bonds. *Nature Chemistry* **2009**, *1*, 74–79.

(42) Ahire, V. K.; Malkhede, D. D. Interaction studies of haemoglobin with p-sulfonatocalix [8] arene by spectrophotometric methods. *Chemical Physics Letters* **2019**, *731*, 136597.

(43) Louie, G. V.; Brayer, G. D. High-resolution refinement of yeast iso-1-cytochrome c and comparisons with other eukaryotic cytochromes c. *Journal of molecular biology* **1990**, *214*, 527–555.

(44) Li, A.; Acevedo-Rocha, C. G.; D'Amore, L.; Chen, J.; Peng, Y.; Garcia-Borràs, M.; Gao, C.; Zhu, J.; Rickerby, H.; Osuna, S., et al. Regio-and stereoselective steroid hy-

droxylation at C7 by cytochrome P450 monooxygenase mutants. *Angewandte Chemie International Edition* **2020**, *59*, 12499–12505.

(45) Rennie, M. L.; Fox, G. C.; Pérez, J.; Crowley, P. B. Auto-regulated Protein Assembly on a Supramolecular Scaffold. *Angewandte Chemie* **2018**, *130*, 13960–13965.

(46) Rennie, M. L.; Crowley, P. B. A Thermodynamic Model of Auto-regulated Protein Assembly by a Supramolecular Scaffold. *Chem. Phys. Chem.* **2019**, *20*, 1011–1017.

(47) Khairutdinov, B.; Ermakova, E.; Sitnitsky, A.; Stoikov, I.; Zuev, Y. Supramolecular complex formed by DNA oligonucleotide and thiacalix [4] arene. NMR-spectroscopy and molecular docking. *Journal of Molecular Structure* **2014**, *1074*, 126–133.

(48) Bartocci, A.; Gillet, N.; Jiang, T.; Szczepaniak, F.; Dumont, E. Molecular Dynamics Approach for Capturing Calixarene–Protein Interactions: The Case of Cytochrome C. *The Journal of Physical Chemistry B* **2020**,

(49) Śledź, P.; Caflisch, A. Protein structure-based drug design: from docking to molecular dynamics. *Current Opinion in Structural Biology* **2018**, *48*, 93–102.

(50) 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.

(51) 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.

(52) 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 simu-

lations by the attach-pull-release (APR) method. *Journal of computer-aided molecular design* **2017**, *31*, 133–145.

(53) 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.

(54) Heinzelmann, G.; Gilson, M. K. Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Scientific Reports* **2021**, *11*, 1–18.

(55) Pereira, G. P.; Cecchini, M. Multibasin Quasi-Harmonic Approach for the Calculation of the Configurational Entropy of Small Molecules in Solution. *Journal of Chemical Theory and Computation* **2021**, *17*, 1133–1142.

(56) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M., et al. AMBER 2018; 2018. *University of California, San Francisco*

(57) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic acids research* **2012**, *40*, W537–W541.

(58) Li, P.; Merz Jr, K. M. MCPB. py: A python based metal center parameter builder. 2016.

(59) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of computational chemistry* **2000**, *21*, 1049–1074.

(60) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; et al., H. N. Gaussian 16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.

(61) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(62) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **2015**, *11*, 3696–3713.

(63) Wang, C.; Greene, D.; Xiao, L.; Qi, R.; Luo, R. Recent Developments and Applications of the MMPBSA Method. *Frontiers in Molecular Biosciences* **2018**, *4*, 87.

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

(65) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(66) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.

(67) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593.

(68) Jeamet, E.; Septavaux, J.; Héloin, A.; Donnier-Maréchal, M.; Dumartin, M.; Ourri, B.; Mandal, P.; Huc, I.; Bignon, E.; Dumont, E.; Morell, C.; Francoia, J.-P.; Perret, F.;

Vial, L.; Leclaire, J. Wetting the lock and key enthalpically favours polyelectrolyte binding. *Chem. Sci.* **2019**, *10*, 277–283.

(69) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; 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.

(70) Perret, F.; Coleman, A. W. Biochemistry of anionic calix [n] arenes. *Chemical Communications* **2011**, *47*, 7303–7319.

(71) Cooper, D. R.; Boczek, T.; Grelewska, K.; Pinkowska, M.; Sikorska, M.; Zawadzki, M.; Derewenda, Z. Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallographica Section D: Biological Crystallography* **2007**, *63*, 636–645.