ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

# Capturing the supramolecular association of calixarenes onto proteins relying on molecular dynamics simulations: the case of cytochrome c.†

Alessio Bartocci,[a] Florence Szczepaniak,[a‡], Tao Jiang,[a] Natacha Gillet,[a] and Elise Dumont*,[a],[b]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Functionalized supramolecular cages are of growing importance in biology and biochemistry. They have recently been proposed as efficient auxiliaries to obtain high-resolution co-crystallized proteins. Here, we propose a molecular dynamics investigation of the supramolecular association of sulfonated calix-[8]-arenes to cytochrome c. The binding sites prone to interactions with sulfonated calixarenes can be identified without prior knowledge of the X-ray structure, and the binding free energies estimated by molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) post-analysis are found to be in neat agreement with the isothermal titration calorimetry (ITC) measurements (-12.0 and -10.8 kcal.mol$^{-1}$, [Rennie *et al., Chem. Phys. Chem.*, 2019, 20, 1011]). The per-residue decomposition reveals the detailed picture of this electrostatically-driven association and notably the role of the arginine R13 as a bridge residue between the two main anchoring sites. In addition, the analysis of the residue behavior by means of a supervised machine learning protocol unveils the formation of an hydrogen bond network far from the binding sites, increasing the rigidity of the protein. This study paves the way towards an automated procedure for predicting computationally supramolecular protein–cages association, with the possibility of a computational screening of new promising derivatives for controlled protein assembly and protein recognition surfaces processes.

## 1 Introduction

Regulated protein assembly and disassembly is fundamental in biology[1]. These mechanisms, as many biophysical and biochemical processes[2–5], lead to supramolecular aggregates and macromolecules. Chemistry at protein-protein interfaces has been widely investigated, mostly to develop inhibitors[6]. Conversely, and more recently, small auxiliary molecules have been proposed to consolidate protein-protein interfaces in view of chemically-assisted protein crystallization[7–12]. The modulation of protein-protein interfaces by ligands whose morphology and electrostatic potential offer tremendous perspectives from high-resolution X-ray structural determination to controlled self-assembly[13]. In that realm, functionalized calixarenes, notably para-sulfonato-calix-[$n$]-arenes with high potential in biochemistry[9,14,15].

The use of "molecular glues" has been pioneered in 2012[8], with the obtention of a series of high-resolution structures characterizing the role of surface residues notably lysines (K) and arginines (R) targetting functionalized calixarenes, whose role as biosensors is well recognized[16]. Other residues, prone to more opportunistic non-covalent interactions, may also contribute to the recognition process of calixarenes at protein-protein interfaces,

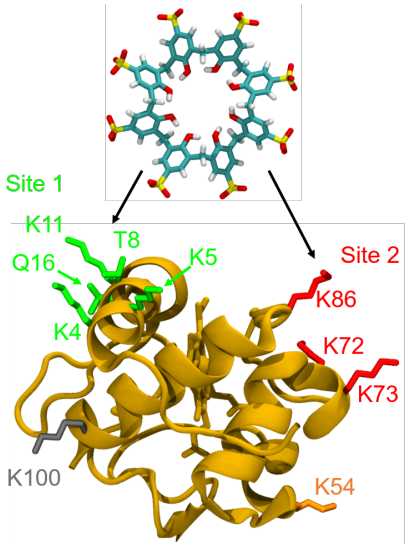


Fig. 1 *Interaction of the sulfonated calix-[8]-arene (**sclx**$_8$, with the cytochrome c (Cyt c PDB ID 6GD6). Two main sites (Site 1 in green, Site 2 in red) are identified based on X-ray (PDB ID 6GD6), which features essentially lysine residues, some of them vicinal ("tweezers").*

[a] *Univ Lyon, ENS de Lyon, CNRS UMR 5182, Université Claude Bernard Lyon 1, Laboratoire de Chimie, F-69342, Lyon, France. Fax: +33 4 72 72 88 60; Tel: +33 4 72 72 80 10; E-mail: elise.dumont@ens-lyon.fr*

[b] *Institut Universitaire de France, 5 rue Descartes, 75005 Paris, France.*

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/cXCP00000x/

which suggests that electrostatic interactions may not be the exclusive driving force of this binding.

Recent isothermal titration calorimetry (ITC) measurements by Crowley and coworkers[17] allow to characterize the free energies for the binding of a first sulfonated calix-[8]-arene (**sclx**$_8$ see Figure 1 top), on the surface of a small heme protein, cytochrome c[18] (*Cyt c*, see Figure 1), followed by the fixation of a second ligand. Yet such experimental measurements for protein–ligand association remain challenging[19]. Molecular dynamics (MD) simulations offer an alternative way to evaluate binding free energies, with the key possibility of per-residue decomposition to dissect the overall binding free energy and pinpoint crucial residues or cooperative network. A computational approach comes with a predictive power for investigations of new protein–cage associations. In this manuscript, we challenge the computational description of the association of **sclx**$_8$ with *Cyt c*, a choice motivated by available X-ray and ITC data that characterize finely association equilibria of the protein with one or two **sclx**$_8$[17]: P + L $\rightarrow$ PL and PL + L $\rightarrow$ PL$_2$. This protein presents many positively-charges residues (16 lysines and 3 arginines, mostly on its surface) and is therefore a legitimate target to probe the competitive binding of **sclx**$_8$ along our MD simulations. Our goal is to hone a robust methodology with two main objectives:

- estimate the binding free energy with one and two calixarene(s) interacting with *Cyt c*
- assess the individual contribution for amino acids binding the calixarene, which would provide a cartography of hotspots for calixarene binding

For organic, small-size ligands binding to proteins, approaches of molecular docking have been long developed and are now globally successful[20], with daily use of routine[21] or more advanced approaches[22], eventually bridged to MD simulations[23,24]. Successful docking approaches have been reported for small size calixarenes, such as calix-[4]-arenes[25–27], in the context of design of drug carriers[28], transport phenomena[12] or protein inhibitors[25,26]. Yet larger calixarenes prone such as **sclx**$_8$ to a higher flexibility may need the use of MD simulations, as recognized since the seminal paper by Gutsche and coworkers[29]. Such supramolecular auxiliaries for protein crystallization can present more versatile and flexible interactions patterns[30] and have been studied much more rarely in the literature[31,32]. For instance, the calixarene **sclx**$_8$, can approach from its upper or lower rims, and its conformation can change depending on the protein environment, from a double cone to a extended pleated conformation[33]. In this paper, we show that the association of the sulfato-calix-[8]-arene with *Cyt c*, which recognizes specific residues in the whole surface can be captured by all-atom MD simulations. We report binding free energies estimated with the Molecular Mechanics/Poisson-Boltzmann Surface Area[34] approach (MM-PBSA), for both the first and second associations that are found to lie in very good agreement with ITC measurements for the *Cyt c*-**sclx**$_8$ system[17], hence validating a computational approach for determining binding sites and corresponding free energies. We also provide an insightful study of the per-residue contribution to the binding by means of MM-PBSA calculations and a supervised machine learning algorithm. Our approach offers possibilities towards an efficient computational screening of supra-biomolecular interactions.

## 2 Methods

### 2.1 Derivation of force field parameters for sclx8 and setting up of the MD simulations

Classical all-atom molecular dynamics (MD) simulations have been performed using the Amber package[35]and the Ambertools suite of programs. A single *Cyt c* protein chain was used (108 residues), whose initial structure was taken from a cristallographic structure from the Protein Databank (*Saccharomyces cerevisae* cytochrome c C102T, PDB ID code 6GD6[11,17]). The presence of the heme group[36] was stabilized within the protein (bound with H18 and M80 residues) through the use of "Metal Center Parameter Builder" (MCPB) Amber tool[37].The heme cofactor (type B) was optimized at the DFT-B3LYP/6-31G(d) level of theory by using a Stuttgart-Dresden SDD pseudopotential for the iron center. A set of RESP charges was generated (see ESI). The topology of the calixarene ligand has been validated through different steps. By starting from the initial sulfonatocalix[4]arene molecular structure, bearing a charge -3, quantum calculations using the framework of the density functional theory (DFT) at the DFT-B3LYP level of theory with the double-$\zeta$ Pople's 6-31+G(d,p) basis set and Grimme's dispersion correction scheme D3BJ has been used to obtain a reliable optimized geometry, and then a suitable RESP[38] charges representation. The obtained partial charges were then generated for describing the sulfonato-calix[8]arene (see ESI). Its structure was optimized at the DFT/M06-2X/6-31+G(d,p) level of theory, with a lowest-energy geometry corresponds to an orientation of the hydroxyl rims promoting a cooperative hydrogen bond network (shown in Figure 1). All DFT calculations were performed using the Gaussian 16 revision B.01 series of programs[39].

The simulation boxes with the *Cyt c* protein plus one (*Cyt c*-sclx8) ligand and two (*Cyt c*-2 **sclx**$_8$) calixarene ligands, describing the two equilibria: P + L $\rightarrow$ PL and PL + L $\rightarrow$ PL$_2$. were created with the Amber suite of programs[35]. The AMBER/ff14SB[40] and GAFF2 force fields[41] were used, and potassium counterions, added to neutralize the system total charge. Long-range electrostatic interactions were computed using Particle Mesh Ewald (PME) algorithm[42,43]. A cutoff of 10 Å was applied for the van der Waals, for electrostatic interactions and for the real space of the electrostatic interactions. An exhaustive description of the protocol adopted to build the simulation boxes for the first and second *Cyt c*-sclx8 associations is given as ESI. The last 10 ns of each production run, when the equilibrium and convergence of the system properties have been reached, have been analyzed through the MM-PBSA energy analysis[44–46].

### 2.2 Multilayer Perceptrons

Machine learning (ML) methods have gained enormous amount of attention in recent years. Its power for finding important information out of large amount of data has been embraced by the bio-

chemistry community, many interesting applications have been showcased in the literature[47–52]. Recently, Fleetwood *et al.*[53] have demonstrated ML methods are capable of learning ensemble properties from molecular simulations and providing easily interpretable metrics of important features. In this study, we have performed an analysis of our trajectories with Multilayer Perceptrons (MLP) by utilizing the demystifying package from Fleetwood *et al.*[53]. The MLP is a fully connected artificial neural network (ANN) with one input layer, one output layer and at least one hidden layer. After tests, the architecture of the MLP was chosen to contain a single layer of 200 neurons to provide good accuracy. The rectified linear unit function (ReLU)[54] was used for the activation of neurons, and the Adam algorithm[55] was used for optimization. The inverse of the distances between the geometric centers of the residues were used as the input features for the multilayer perceptrons (MLP) NN, due to better overall performance over Cartesian coordinates, according to Fleetwood *et al*. These internal coordinates were computed for all residue pairs and all frames. Each frame of the trajectories was labelled as either 1 or 0 according to whether the distance between the calixarene and the protein is smaller than 10 Å (bonded) or not (non-bonded). These sets of input features and labels were fed to the MLP classifier for training. Upon completion of the training, layerwise relevance propagation (LRP)[56] was performed to find out the important features for calixarene-protein interaction.

## 3 Results and discussion

Our MD simulations were started from a X-ray structure of isolated cytochrome c *Cyt c* (PDB ID code: 6GD6), where the ligand **sclx**$_8$ was removed and then placed in a water box at a distance of ca. 50 Å far from the protein center of mass (see Figure S1), as we sought to capture the electrostatically-driven association without the bias to start from a X-ray structure co-crystallized with **sclx**$_8$ . Indeed, the **sclx**$_8$ binding hot spots for a protein in solution can differ from the regular, symmetry-driven macromolecular ensemble observed in the crystal. To simulate the binding of the second ligand, another set of starting points was generated, with two solvated ligands or one bound/one solvated ligand (see Figure S2 and Table S1).

The **sclx**$_8$ conformations and protein surface sampled along the MD simulations with one ligand, represented in Figure 2 **A**, clearly covers the key interacting residues identified on the X-ray structure: K4, K5, K11 at Site 1, (in green in Figure 1), K72 and K73 at Site 2 (in red Figure 1)[11]. Arginine interaction is well known for anionic calixarenes[57,58], and R13 (in black in Figure 2), standing between the two sites, can interact with **sclx**$_8$ regardless its binding site. After the addition of a second **sclx**$_8$ (see Figure 2 **B**), this arginine R13 can play a role as a bridge between the two sites, with still a preference for the interaction with the ligand on Site 1. At the end of all simulations, the two **sclx**$_8$ are anchored on the protein surface, on Sites 1-2 or on Site 1-3 (in gray in Figure 2 **B**), observed in only one MD simulation). During MD simulations, the first occupied site can be 1 or 2 alike, without any impact on the *Cyt c*-2**sclx**$_8$ final complex.

More quantitatively, binding free energies can be estimated through the MM-PBSA post-processing of our MD simulations, offering a direct comparison with the isothermal titration calorimetry (ITC) measurements[17]. The first 1:1 equilibrium between *Cyt c* and **sclx**$_8$ was assessed by ITC to -10.8±0.9 kcal.mol$^{-1}$, which is in good agreement with the value we obtained from MD simulations with the MM-PBSA approach: -13.5±6.4 kcal.mol$^{-1}$ (Figure 3). The error bar from the MD simulations is rather important, which is inherent to the MM-PBSA approach. As often for the MM-PBSA approach, the estimated binding free energy is overestimated compared to the experimental value, due to the sampling at short times with no binding–unbinding events. *Cyt c* presents

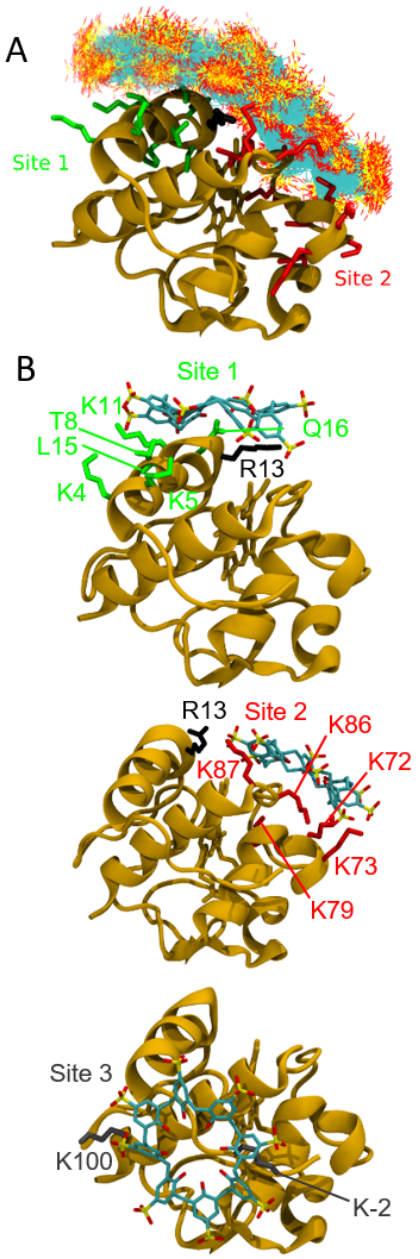


Fig. 2 A. *Representative cartoon of superimposed conformations of different runs (see Table S1 for details) for the Cyt c-***sclx**$_8$ *system. The covered conformations and the sampled protein surface, along four trajectories of 200 ns, are reported with colored lines.* B. *Sites visited along the 12 different trajectories for the Cyt c-2***sclx**$_8$ *system.*

many positively-charged residues and hence is prone to anchor a second calixarene. The second equilibrium PL + L $\rightarrow$ $PL_2$ was also studied (the starting points for these MD simulations are reported in detail in ESI). The overall free energy for the second association is also well reproduced: -7.8±0.9 vs. -8.2±7.0 kcal.mol$^{-1}$ (Figure 3). Free binding energies $\Delta G$ for the two first association equilibria between *Cyt c* and **sclx**$_8$ are shown in Figure 3, alongside with contributions summed over positively charged residues, namely, lysines and arginines, or glutamates and aspartates for the negatively-charged residues. This value is lowered by 3.0 kcal.mol$^{-1}$ in the experiments and 5.3 kcal.mol$^{-1}$ in MD simulations compared to the first association, due to a smaller accessible protein surface area (see ITC results). This confirms the performance of MM-PBSA to evaluate the binding free energies [59–62], and validates in turn our computational approach.

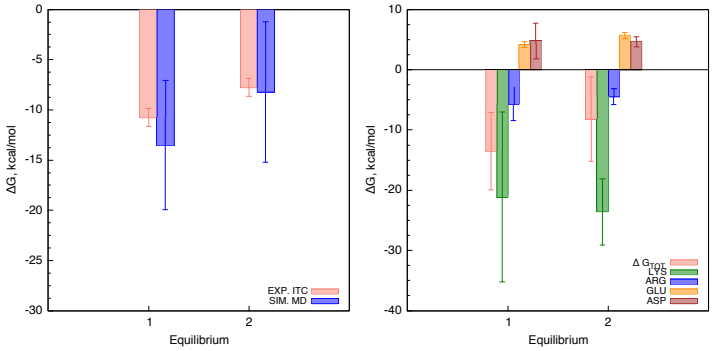


Fig. 3 *Left panel: Experimental ITC vs. computed binding free energies $\Delta G$ in $kcal.mol^{-1}$ (in red and blue respectively) for the first and second associations. Right panel: binding free energy $\Delta G$ in $kcal.mol^{-1}$ for the first and second associations and their decomposition over charged residues (lysine, arginine, glutamate and aspartate), reported as histogram plot.*

For both associations, the additive decomposition per type of residues reveals that the *Cyt c*-**sclx**$_8$ binding cannot be viewed as a lysine–only problem. Indeed the contribution from all lysines for the first and the second associations is about 8 to 16 kcal.mol$^{-1}$ that the respective overall binding free energy (see Figure 3). Where arginines attractive contribution is smaller ($\simeq$5 kcal.mol$^{-1}$), consistently with the ratio lysine/arginine of 16/3, the overall binding energy is in turn moderated by negatively-charged residues, namely glutamate and aspartate, whose role is nearly equivalent in disfavoring the binding process (ca. 5 kcal.mol$^{-1}$ each). For the second association, the repulsion between the two **sclx**$_8$ can also counterbalance the attractive contribution from positively-charged residues.

In order to further dissect the main interactions between **sclx**$_8$ and *Cyt c*, it is insightful to decompose the overall interaction energy into contributions deriving from single specific residues of the protein (see Table 1 and Figure 4). This systematic per-residue inventory confirms and quantifies the key role of lysines, notably K4, K5, K11, K72, K73, K79, K86, K87, K89 and K100, six of which being present as "tweezers": for instance, K86 contributes to -3.7 ± 2.8 kcal.mol$^{-1}$ to $\Delta G_1$, while K87 role is nearly halved, -2.1 ± 1.4 kcal.mol$^{-1}$; their contributions to $\Delta G_2$ are nearly equal. One can propose a linear dependence between the

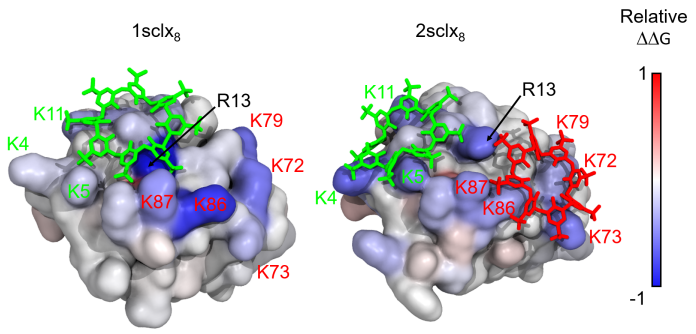


Fig. 4 *Color map of per-residue decomposition $\Delta\Delta G_1$ (left panel) or $\Delta\Delta G_2$ (right panel). The color scale is defined by dividing each contribution by the maximal absolute one ($\Delta\Delta G_1$(R13)) and using the color code: blue for attractive interactions, and red for repulsive interactions. The two calixarenes are colored per main interacting sites.*

strength of the interaction and the distance between the center of mass of the lysine $NH_3^+$ group and the most proximal **sclx**$_8$ $SO_3^-$ sulfate group (see Figure S4). When one **sclx**$_8$ is present in the simulation, four lysines from site 2 (namely K72, K79, K86, K87) define the hotspots for anchoring the ligand, in addition to R13. This residue, reported in Table 1 as part of Site 1, can also take part to the binding on Site 2. This central position explain its large contribution of 4.4±2.6 kcal.mol$^{-1}$ to $\Delta G_1$ (Table 1). The interaction landscape is modified when two **sclx**$_8$ are bounded to the protein, with a more diluted attractive interaction, distributed on more residues from both sites (K4, K5, K11, R13, K73 and K87 for a contribution between 2 and 3 kcal.mol$^{-1}$). Figure 4 highlights this difference with a concentration of bright blue residues around Site 2 for the 1**sclx**$_8$ system whereas the blue surface area for the 2**sclx**$_8$ system is wider but lighter.

Our simulations corroborate that in the protein-**sclx**$_8$ binding, a significant contribution is covered by electrostatics interactions. The downhill as a second **sclx**$_8$ comes into play is effect is notable by using this method, that carries the split of the total complexation $\Delta G$ into electrostatic, van der Waals (vdW) and solvation components. Crowley and coworkers have also defined [11] interacting residues in terms of their importance. Focusing on the $A_{buried}$ (see eq. 4 of ESI), there are master residues, that contribute with an interface area of $\simeq$ 100 Å$^2$ to the total *Cyt c*-**sclx**$_8$ interface, and key residues that contribute with 50-100 Å$^2$ to the total interface (see Table 1). The values of $\Delta^i_{SASA}$ (see ESI for details), where *i* indicates the *i*-th residue, are reported in Table 1. They give an idea of the residues participation during the interaction, and they are a suitable comparison with the interface values between the protein and the calixarene [11]. For some residues, our results are in agreement with the classification proposed by Crowley and coworkers: most of the lysines pointed out for their important interface area belong to the binding hotspots described in the simulations, except K54 and K100. However, T8 and Q16 were considered as key and master residues respectively but their contribution appear minimal compared to the lysines from the same category. Our simulations also reveal the important role of K79, K87 and R13, which were not stressed by the interaction

| | RES | $\Delta\Delta G_1$ (kcal.mol$^{-1}$) | $\Delta\Delta G_2$ (kcal.mol$^{-1}$) | Int. Area (Å$^2$) | $\Delta_{SASA}$ (Å$^2$) |
|---|---|---|---|---|---|
| Site 1 | **K4** | -0.9 ± 0.7 | -2.8 ± 1.0 | > 100 | 62.5±1.7 |
| | **K5** | -1.1 ± 0.6 | -2.3 ± 0.9 | 50-100 | 52.2±2.0 |
| | **T8** | -0.4±0.7 | -0.9 ± 0.2 | 50-100 | 39.1±7.0 |
| | **K11** | -1.7 ± 1.2 | -2.2 ± 0.3 | 50-100 | 49.3±2.6 |
| | R13 | -4.4 ± 2.6 | -2.6 ± 0.7 | - | 32.6±11.4 |
| | **L15** | -0.4 ± 0.5 | -0.6 ± 0.4 | 50-100 | 15.8±4.0 |
| | **Q16** | -1.4 ± 0.8 | -0.8 ± 0.6 | > 100 | 30.5±5.2 |
| Site 2 | **K72** | -2.5 ± 2.2 | -1.7 ± 0.4 | 50-100 | 50.4 ± 7.6 |
| | **K73** | -1.4 ± 1.4 | -2.4 ± 0.8 | 50-100 | 48.5±14.7 |
| | K79 | -2.6 ± 2.4 | -0.8 ± 0.2 | - | – |
| | **K86** | -3.7 ± 2.8 | -1.8 ± 0.2 | > 100 | 51.9±9.6 |
| | K87 | -2.1 ± 1.4 | -2.0 ± 0.4 | - | – |
| Site 3 | K-2 | -0.5 ± 0.1 | -1.4 ± 0.4 | - | 52.3 ± 10.0 |
| | **K100** | -0.5 ± 0.1 | -1.3 ± 0.1 | > 100 | 36.7±24.3 |
| Other | **K54** | -0.6 ± 0.2 | -0.7 ± 0.2 | | |
| | K89 | -1.0 ± 0.5 | -1.4 ± 0.5 | - | |
| **Sum. lysines** | -7.6 | -11.4 | - | - | |
| **Sites 1+2** | -17.9 | -23.6 | - | - | |
| **Total** | All | -13.5 ± 6.4 | -8.2 ± 7.0 | 515 | 407 ± 57 |

Table 1 *Interacting residues defining three different binding sites of* **sclx**$_8$ *on Cyt c Site 1 (in green), Site 2 (in red) and Site 3 (in gray), and their contribution to the association free energy of the first* **sclx**$_8$ *($\Delta\Delta G_1$), the second* **sclx**$_8$ *($\Delta\Delta G_2$), to the experimental interface area (Int. Area), and the theoretical difference in solvent accessible surface area between free or complex Cyt c in the presence of two* **sclx**$_8$ *($\Delta_{SASA}$). Residues identified in the experimental crystal structure are boldfaced*[11]*. R13, here reported as part of in Site 1, can interact simultaneously with two* **sclx**$_8$ *molecules present in Sites 1 and 2, hence acting as a bridge residue.*

area analysis.

The amplitude of the buried surface area $A_{buried}$ (see ESI for details) indicates, at this point, a very similar conformation adopted by the two interacting calixarenes (extended conformation). The experimental value, based on the X-ray crystal structures[11] is 515 Å$^2$, describing an interface area larger than the average protein-protein crystal contact[63], and highlighting the role as molecular glue of the **sclx**$_8$. The value obtained from MD simulations (407.0 ± 57.4 Å$^2$) reveals that the calixarene conformation can adapt itself to the protein shape, as its fluctuations reflect the structural flexibility of the calixarene (see also SASA/RMSF part in the ESI).

Association of **sclx**$_8$ to  is guided by a strong electrostatic driving force, and induce larger effect for the structure. To identify the possible allosteric effects upon **sclx**$_8$ we took advantage of a supervised machine-learning protocol recently proposed by Fleetwood and coworkers[53] to monitor the structural changes of the cytochrome upon binding of **sclx**$_8$. Important residues known to interact with **sclx**$_8$ correspond to sites 1 and 2 are denoted in Figure 5 with red stars. But the ML analysis reveals a reorganization of a series of residues (31, 43 and 49, as green triangles below) which belong to a flexible loop of  with formation of salt bridges subsequent to the association with **sclx**$_8$. This rigidification of the flexible loop of  far from the binding site probably contributes to a surface entropy reduction.

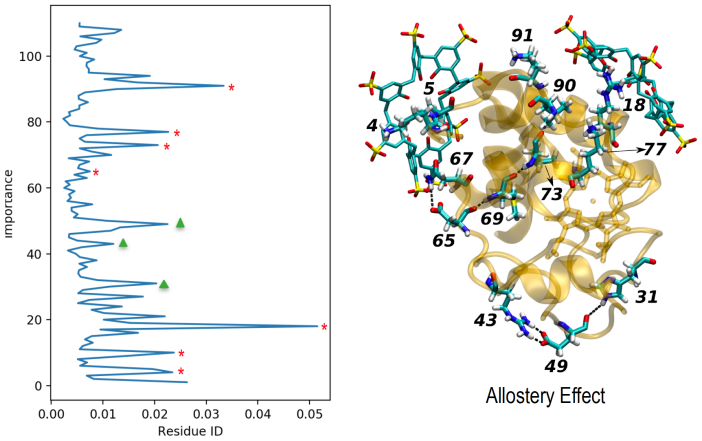


Fig. 5 *Important residues of  upon* **sclx**$_8$ *binding revealed by supervised machine-learning protocol. The formation of salt bridges between residues belonging to a looop at the other side of the protein is revealed (green triangles).*

## 4 Conclusions

In this study, we have relied on molecular dynamics simulations to capture the mainly electrostatically-driven interaction between sulfonato-calix-[8]-arenes ligands and cytochrome c. Our simulations reveal the non-covalent successively taking place over time, beyond the X-ray structure which is obtained ultimately for a crystal after a control self-assembly process. For this electrostatically-driven association phenomenom, the per-residue MM-PBSA decomposition allows discriminate key and master residues coming into play, with a non-negligeable contribution of negatively-charged residues. The computational approach allows to characterize more finely and systematically the surface of the protein implied in the association process, and constitutes a first step to analyze computationally the ligand contribution at protein–protein interfaces. Also, it can be used as a predictive tool to assess the affinity of a given molecular glue onto a protein, which could be used for screening purposes and design of new efficient auxiliaries.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

A.B. is grateful for a post-doctoral fellowship by the Fondation "Maison de la Chimie". The authors thank the SYSPROD project and AXELERA Pôle de Compétitivité for financial support (PSMN Data Center). The authors acknowledge fruitful discussions with Dr. S. Engilberge and Prof. P. Crowley (NUI Galway).

## Notes and references


1 E. Chuang, A. M. Hori, C. D. Hesketh and J. Shorter, *Journal of Cell Science*, 2018, **131**, year.

2 J. W. de Vries, S. Schnichels, J. Hurst, L. Strudel, A. Gruszka, M. Kwak, K.-U. Bartz-Schmidt, M. S. Spitzer and A. Herrmann, *Biomaterials*, 2018, **157**, 98–106.

3 K. L. Chen and G. D. Bothun, *Environmental Science & Technology*, 2014, **48**, 873–880.

4 S. Salassi, F. Simonelli, A. Bartocci and G. Rossi, *Journal of Physics D: Applied Physics*, 2018, **51**, 384002.

5 G. Rossi, S. Salassi, F. Simonelli, A. Bartocci and L. Monticelli, *Biomembrane Simulations: Computational Studies of Biological Membranes*, 2019, 163.

6 M. R. Arkin, Y. Tang and J. A. Wells, *Chemistry & biology*, 2014, **21**, 1102–1114.

7 M. Giuliani, I. Morbioli, F. Sansone and A. Casnati, *Chem. Commun.*, 2015, **51**, 14140–14159.

8 R. E. McGovern, H. Fernandes, A. R. Khan, N. P. Power and P. B. Crowley, *Nature chemistry*, 2012, **4**, 527.

9 F. Perret, A. N. Lazar and A. W. Coleman, *Chemical Communications*, 2006, 2425–2438.

10 A. M. Doolan, M. L. Rennie and P. B. Crowley, *Chemistry–A European Journal*, 2018, **24**, 984–991.

11 M. L. Rennie, G. C. Fox, J. Pérez and P. B. Crowley, *Angewandte Chemie*, 2018, **130**, 13960–13965.

12 E. S. Español and M. M. Villamil, *Biomolecules*, 2019, **9**, 90.

13 B. J. G. E. Pieters, M. B. van Eldijk, R. J. M. Nolte and J. Mecinović, *Chem. Soc. Rev.*, 2016, **45**, 24–39.

14 A. F. Danil de Namor, R. M. Cleverley and M. L. Zapata-Ormachea, *Chemical reviews*, 1998, **98**, 2495–2526.

15 J. M. Alex, P. McArdle and P. B. Crowley, *CrystEngComm*, 2020, **22**, 14–17.

16 S. B. Nimse and T. Kim, *Chem. Soc. Rev.*, 2013, **42**, 366–386.

17 M. L. Rennie and P. B. Crowley, *ChemPhysChem*, 2019, **20**, 1011–1017.

18 I. Bertini, G. Cavallaro and A. Rosato, *Chemical Reviews*, 2006, **106**, 90–115.

19 S. Denis-Quanquin, F. Riobé, M.-A. Delsuc, O. Maury and N. Giraud, *Chemistry – A European Journal*, 2016, **22**, 18123–18131.

20 L. Duan, X. Liu and J. Z. Zhang, *Journal of the American Chemical Society*, 2016, **138**, 5722–5728.

21 D. S. Goodsell, G. M. Morris and A. J. Olson, *Journal of molecular recognition*, 1996, **9**, 1–5.

22 J. Yang, M. Baek and C. Seok, *Journal of Computational Chemistry*, 2019, **40**, 2739–2748.

23 V. Salmaso and S. Moro, *Frontiers in Pharmacology*, 2018, **9**, 923.

24 P. Kollman, *Chemical reviews*, 1993, **93**, 2395–2417.

25 A. I. Vovk, V. I. Kalchenko, S. A. Cherenok, V. P. Kukhar, O. V. Muzychka and M. O. Lozynsky, *Org. Biomol. Chem.*, 2004, **2**, 3162–3166.

26 M. G. Chini, S. Terracciano, R. Riccio, G. Bifulco, R. Ciao, C. Gaeta, F. Troisi and P. Neri, *Organic Letters*, 2010, **12**, 5382–5385.

27 B. Khairutdinov, E. Ermakova, A. Sitnitsky, I. Stoikov and Y. Zuev, *Journal of Molecular Structure*, 2014, **1074**, 126 – 133.

28 Y. C. K. J. Jang, Y. and S.-U. Kim, *J. Mater. Sci.*, 2018, **53**, 5125–5139.

29 C. D. Gutsche and L. J. Bauer, *Journal of the American Chemical Society*, 1985, **107**, 6052–6059.

30 S. Engilberge, F. Riobé, T. Wagner, S. Di Pietro, C. Breyton, B. Franzetti, S. Shima, E. Girard, E. Dumont and O. Maury, *Chemistry–A European Journal*, 2018, **24**, 9701–9701.

31 S. Gordo, V. Martos, E. Santos, M. Menéndez, C. Bo, E. Giralt and J. de Mendoza, *Proceedings of the National Academy of Sciences*, 2008, **105**, 16426–16431.

32 N. K. Beyeh, Nonappa, V. Liljeström, J. Mikkilä, A. Korpi, D. Bochicchio, G. M. Pavan, O. Ikkala, R. H. A. Ras and M. A. Kostiainen, *ACS Nano*, 2018, **12**, 8029–8036.

33 S. Engilberge, M. L. Rennie, E. Dumont and P. B. Crowley, *ACS nano*, 2019, **13**, 10343–10350.

34 C. Wang, D. Greene, L. Xiao, R. Qi and R. Luo, *Frontiers in Molecular Biosciences*, 2018, **4**, 87.

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

36 C. Rovira, P. Carloni and M. Parrinello, *The Journal of Physical Chemistry B*, 1999, **103**, 7031–7035.

37 P. Li and K. M. Merz Jr, *MCPB. py: A python based metal center parameter builder*, 2016.

38 J. Wang, P. Cieplak and P. A. Kollman, *Journal of Computational Chemistry*, 2000, **21**, 1049–1074.

39 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian~16 Revision C.01*, 2016, Gaussian Inc. Wallingford CT.

40 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *Journal of Chemical Theory and Computation*, 2015, **11**, 3696–3713.

41 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of Computational Chemistry*, 2004, **25**, 1157–1174.

42 T. Darden, D. York and L. Pedersen, *The Journal of Chemical Physics*, 1993, **98**, 10089–10092.

43 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577–8593.

44 N. Homeyer and H. Gohlke, *Molecular Informatics*, 2012, **31**, 114–122.

45 B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *Journal of Chemical Theory and Computation*, 2012, **8**, 3314–3321.

46 E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. H. Zhang and T. Hou, *Chemical Reviews*, 2019, **119**, 9478–9508.

47 J. Wang, H. Cao, J. Z. H. Zhang and Y. Qi, *Scientific Reports*, 2018, **8**, 6349.

48 Y. Ofran and B. Rost, *Bioinformatics*, 2007, **23**, e13–e16.

49 S. Liu, C. Liu and L. Deng, *Molecules*, 2018, **23**, year.

50 D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello and J. J. Collins, *Cell*, 2018, **173**, 1581 – 1592.

51 M. Ahmad, V. Helms, O. V. Kalinina and T. Lengauer, *Journal of Chemical Theory and Computation*, 2019, **15**, 2166–2178.

52 M. T. Degiacomi, *Structure*, 2019, **27**, 1034 – 1040.e3.

53 O. Fleetwood, M. A. Kasimova, A. M. Westerlund and L. Delemotte, *Biophysical Journal*, 2019.

54 X. Glorot, A. Bordes and Y. Bengio, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 2011, pp. 315–323.

55 D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014, `http://arxiv.org/abs/1412.6980`, cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

56 G. Montavon, S. Lapuschkin, A. Binder, W. Samek and K.-R. MÃŒller, *Pattern Recognition*, 2017, **65**, 211 – 222.

57 F. Perret and A. W. Coleman, *Chem. Commun.*, 2011, **47**, 7303–7319.

58 R. E. McGovern, A. A. McCarthy and P. B. Crowley, *Chemical Communications*, 2014, **50**, 10412–10415.

59 D. Huang, Y. Qi, J. Song and J. Z. H. Zhang, *Journal of Computational Chemistry*, 2019, **40**, 1045–1056.

60 M. Retegan, A. Milet and H. Jamet, *Journal of Chemical Information and Modeling*, 2009, **49**, 963–971.

61 E. Bignon, C.-H. Chan, C. Morell, A. Monari, J.-L. Ravanat and E. Dumont, *Chemistry – A European Journal*, 2017, **23**, 12845–12852.

62 E. Jeamet, J. Septavaux, A. Héloin, M. Donnier-Maréchal, M. Dumartin, B. Ourri, P. Mandal, I. Huc, E. Bignon, E. Dumont, C. Morell, J.-P. Francoia, F. Perret, L. Vial and J. Leclaire, *Chem. Sci.*, 2019, **10**, 277–283.

63 J. Janin, R. Bahadur and P. Chakrabarti, *Q. Rev. Biophys*, 2008, **41**, 133–180.