

# Molecular dynamics approach for capturing calixarenes–proteins interactions: the case of cytochrome c.

Alessio Bartocci,<sup>†</sup> Natacha Gillet,<sup>†</sup> Tao Jiang,<sup>†</sup> Florence Szczepaniak,<sup>†</sup> and Elise Dumont<sup>\*,†,‡</sup>

<sup>†</sup>*Univ Lyon, ENS de Lyon, CNRS UMR 5182, Université Claude Bernard Lyon 1, Laboratoire de Chimie, F-69342, Lyon, France*

<sup>‡</sup>*Institut Universitaire de France, 5 rue Descartes, 75005 Paris, France*

E-mail: elise.dumont@ens-lyon.fr

## Abstract

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 starting from initially distant protein and ligands. We characterize two main binding sites for the sulfonated calixarene on the cytochrome c surface which are in perfect agreement with the previous experiments with regard to the structure (comparison with the X-ray structure PDB 6GD8) and the binding free energies (comparison between the molecular mechanics Poisson-Boltzmann surface

area (MM-PBSA) analysis and the isothermal titration calorimetry (ITC) measurements). The per-residue decomposition of the interaction energies reveals the detailed picture of this electrostatically-driven association and notably the role of the arginine R13 as a bridging 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 to predict the supramolecular protein-cages association, with the possibility of a computational screening of new promising derivatives for controlled protein assembly and protein surface recognition processes.

## Introduction

The regulation of protein assembly and disassembly mediated by soluble derivatives and based on rational supramolecular approaches has known a growing interest the last decades.<sup>1,2</sup> Indeed, these molecular glues, such as lanthanide coordination complex<sup>3,4</sup>, polyoxometalate clusters<sup>5,6</sup> or organic supramolecular compounds,<sup>7-11</sup> favor protein/additive/protein adduct to create protein material<sup>12</sup> or induce protein crystallization.<sup>13</sup> In that realm, functionalized calixarenes, notably para-sulfonato-calix- $[n]$ -arenes have demonstrated their high potential in biochemistry thanks to their abilities to bind positively charged protein surface to form protein oligomers.<sup>14-17</sup> They offer tremendous perspective in the field of high-resolution X-ray of unfolded protein loops<sup>18</sup> or small cationic proteins<sup>19,20</sup>.

Crowley and coworkers have realized an extended study of the interaction of different sulfonato-calix- $[n]$ -arenes ( $n=4,8$ )<sup>9,10,22,23</sup> with the well-known cytochrome c (*Cytc*) heme protein.<sup>24</sup> The obtained crystallographic structures underline the role of positively charged surface residues (lysines and arginines) in the calixarenes binding.<sup>10</sup> Other residues, prone to more opportunistic non-covalent interactions, may also contribute to the recognition process,

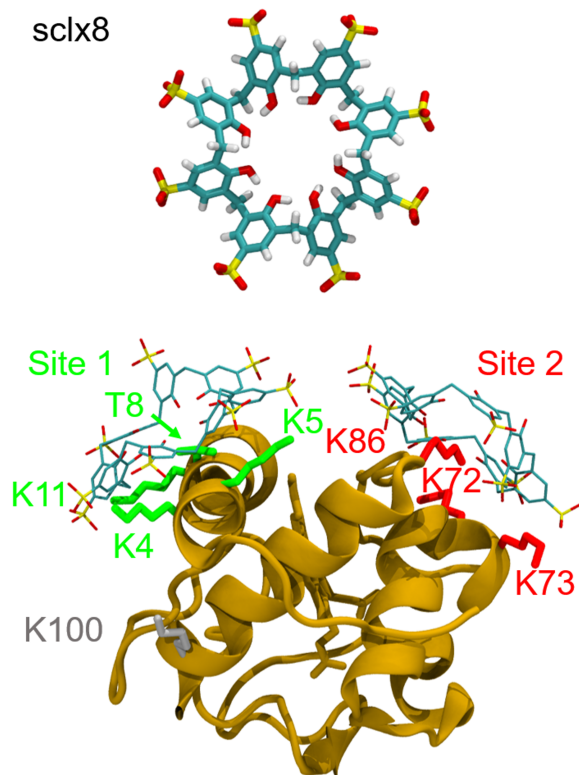


Figure 1: Interaction of the sulfonated calix-[8]-arene (**sclx**<sub>8</sub>, top), with the cytochrome *c* (*Cytc*, down, PDB ID 6GD8)<sup>10</sup>. The important residues are colored according to their corresponding binding site: Site 1 in green, Site 2 in red, site 3 in grey. All picture were rendered using VMD<sup>21</sup>.

which suggests that the driving force of this supramolecular complex is not governed by electrostatic interaction only. For instance, the host-guest complexation between the flexible sulfonato-calix-[8]-arene (**sclx**<sub>8</sub>) (see Figure 1) and *Cytc* leads to the formation of three different crystal packings in space group  $P3_1$ ,  $H3$  and  $P4_32_12$  with various degrees of protein surface coverage by the ligands (see Figure 1 and Figure S1 in ESI).

In addition, recent isothermal titration calorimetry (ITC) measurements by Crowley and coworkers<sup>25</sup> provided free energy data for the binding of a first **sclx**<sub>8</sub> and a second ligand on the *Cytc* surface. They explain the formation of a **sclx**<sub>8</sub>-*Cytc* oligomer in solution using a relatively simple model involving four protein states. However, such experimental

measurements for protein–ligand association remain challenging<sup>26</sup> and must be associated with structural characterization to provide an exhaustive insight of the binding sites. Computational approaches can offer a fast and suitable alternative to experiments in order to characterize the thermodynamic of the regulation of the protein assembly *via* supramolecular compounds, as well as the key residues involved in the ligand fixation. They present a potential predictive power for investigations of new protein–cage associations if they fulfill these requirements:

- a good estimation of the binding free energy between the ligands and the protein
- an evaluation of the individual contribution for amino acids binding the ligand, which would provide a cartography of hotspots for protein/molecular glues interaction.

Molecular docking approaches are now available to investigate the binding of small-size organic ligands to proteins<sup>27</sup>. Some are adapted to a daily routine use<sup>28</sup> while other rely on more advanced methods<sup>29</sup>, potentially bridged to MD simulations<sup>30,31</sup>. Successful docking approaches have been reported for small size calixarenes, such as calix-[4]-arenes<sup>32–34</sup>, in the context of design of drug carriers<sup>35</sup>, transport phenomena<sup>36</sup> or protein inhibitors<sup>32,33</sup>. Yet larger calixarenes prone to a higher flexibility may need the use of MD simulations, as recognized since the seminal paper by Gutsche and coworkers<sup>37</sup>. Such supramolecular auxiliaries for protein crystallization can present more versatile and flexible interactions patterns<sup>38</sup>. For instance, **sclx**<sub>8</sub> 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<sup>39</sup>. For these complex molecules, 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 networks.

In this manuscript, we challenge the computational description of the association of **sclx**<sub>8</sub>

with *Cytc*, a choice motivated by available X-ray<sup>10</sup> and ITC data<sup>25</sup> which finely characterize the association equilibria of the protein with one or two **sclx**<sub>8</sub>  $P + L \rightarrow PL$  and  $PL + L \rightarrow PL_2$ . This protein presents many positively-charged residues (16 lysines and 3 arginines), mostly on its surface, and is therefore a legitimate target to probe the competition between the different binding sites of **sclx**<sub>8</sub> along our MD simulations. Our goal is to hone a robust methodology with a descriptive and predictive potential for protein assembly triggered by molecular glues. We first show that the association of the sulfato-calix-[8]-arene with *Cytc* can be captured by all-atom MD simulations. We then report binding free energies estimated with the Molecular Mechanics/Poisson-Boltzmann Surface Area approach (MM-PBSA)<sup>40</sup>, for both the first and second associations. These results are found to lie in very good agreement with ITC measurements for the *Cytc*-**sclx**<sub>8</sub> system<sup>25</sup>, which constitutes a validation of our computational approach for the determination of binding sites and corresponding free energies. Finally, we provide a per-residue analysis of the protein–cage association by means of MM-PBSA calculations and a supervised machine learning algorithm.

## Methods

### Force field parameters and MD simulations

Classical all-atom molecular dynamics (MD) simulations have been performed using the Amber package<sup>41</sup> and the Ambertools suite. A single *Cytc* protein structure was used (108 residues), taken from the H3 form of the **sclx**<sub>8</sub>-*Cytc* complex (*Saccharomyces cerevisiae* cytochrome c C102T, PDB ID 6GD6<sup>10,25</sup>). The heme group<sup>42</sup> was connected to the H18 and M80 residues through the use of "Metal Center Parameter Builder" (MCPB) Ambertool<sup>43</sup>. The atomic charges for the heme were obtained using the RESP method<sup>44</sup> on a geometry optimized at the DFT-B3LYP/6-31G(d) level of theory. A Stuttgart-Dresden SDD pseudopotential was used for the iron center. All DFT calculations were performed using the

Gaussian 16 revision B.01 series of programs<sup>45</sup>. Force field parameters for the **sclx<sub>8</sub>** ligand were taken from GAFF2<sup>46</sup> while the protein was described with the AMBER/ff14SB<sup>47</sup> force field. The atomic charges for **sclx<sub>8</sub>** were assessed from the RESP procedure<sup>44</sup> at the DFT-B3LYP/6-31+G(d,p) level with D3BJ dispersion correction<sup>48</sup> for the sulfonato-calix[4]arene optimized structure in its extended conformation.

The simulation TIP3P water<sup>49</sup> boxes with the *Cytc* protein from PDB 6GD6 plus one (*Cytc*-**sclx<sub>8</sub>**) or two (*Cytc*-2 **sclx<sub>8</sub>**) solvated **sclx<sub>8</sub>** ligands were created with the Amber suite of programs<sup>41</sup>. Potassium counterions were added to neutralize the system total charge. Long-range electrostatic interactions were computed using Particle Mesh Ewald (PME) algorithm<sup>50,51</sup>. An exhaustive description of the protocol adopted to build the simulation boxes for the first and second *Cytc*-**sclx<sub>8</sub>** associations and to perform equilibration and production runs is given in ESI. The last 10 ns of each production run in the NPT ensemble at 300K and 1 atm. have been analyzed through the MM-PBSA energy analysis<sup>52-54</sup>.

## Multilayer Perceptrons

Machine learning (ML) methods have gained enormous amount of attention in recent years. Their power for finding important information out of large amount of data has been embraced by the biochemistry community, many interesting applications have been showcased in the literature<sup>55-60</sup>. Recently, ML methods have been applied to learn ensemble properties from molecular simulations and to provide 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.*<sup>61</sup>. 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)<sup>62</sup> was used for the activation of neurons, and the Adam algorithm<sup>63</sup> 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)<sup>64</sup> was performed to find out the important features for calixarene-protein interaction.

## Results and Discussion

At the beginning of our simulation, the **sclx**<sub>8</sub> ligand was placed at a distance of ca. 50 Å far from the protein center of mass (see Table S2 in ESI), as we sought to capture the association **sclx**<sub>8</sub>-*Cytc* avoiding the bias of starting from a X-ray structure of the complex. Indeed, the **sclx**<sub>8</sub> binding hot spots for a protein in solution can differ from the regular, symmetry-driven macromolecular ensemble observed in the crystal.<sup>10</sup> 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 S3).

The **sclx**<sub>8</sub> positions sampled along the MD simulations with one ligand on the protein surface, represented in Figure 2 **A**, clearly covers the key interacting lysines identified in the X-ray structure: K4, K5, K11 at Site 1, (in green in Figure 1), K72 and K73 at Site 2 (in red in Figure 1)<sup>10</sup>. In addition to lysine, arginine is a well-known candidate for electrostatic interactions with anionic calixarenes<sup>22,65</sup>. R13 (in black in Figure 2), stands between the two sites and can interact with **sclx**<sub>8</sub> regardless its binding site. After the addition of a second **sclx**<sub>8</sub> (see Figure 2 **B**), R13 can bridge the two sites, even it shows a preference for

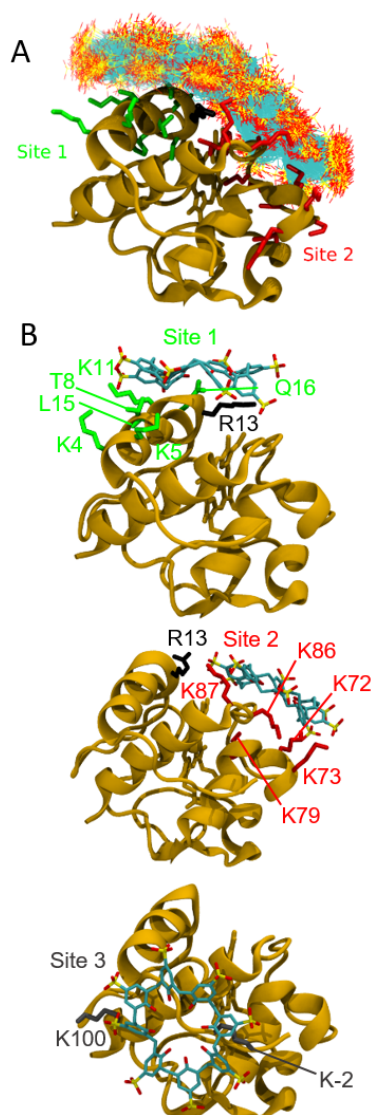


Figure 2: **A.** Representative cartoon of superimposed conformations of different runs (see Table S2 for details) for the *Cytc-1sclx<sub>8</sub>* system. The covered conformations and the sampled protein surface, along four trajectories of 200 ns, are reported with colored lines. **B.** Sites visited along the 12 different trajectories for the *Cytc-2sclx<sub>8</sub>* system.

the interaction with the ligand on Site 1. The first occupied site can be 1 or 2 alike, without any impact on the *Cytc-2sclx<sub>8</sub>* final complex. At the end of all simulations, the two *sclx<sub>8</sub>* are anchored on the protein surface, on Sites 1-2 or on Site 1-3 (in gray in Figure 2 **B**; site 3 is observed in only one MD simulation but also described in 6GD9 structure<sup>10</sup> - see Figure



S1).

More quantitatively, we estimated the binding free energies using the MM-PBSA post-processing, which offers a direct comparison with the isothermal titration calorimetry (ITC) measurements<sup>25</sup>. The first 1:1 equilibrium ( $\Delta G_1$ ) between *Cytc* and **sclx**<sub>8</sub> was assessed by ITC to  $-10.8 \pm 0.9$  kcal.mol<sup>-1</sup>, which is in good agreement with the value we obtained from MD simulations with the MM-PBSA approach:  $-13.5 \pm 6.4$  kcal.mol<sup>-1</sup> (Figure 3). The error bar from the MD simulations is rather important, which is inherent to the MM-PBSA approach. Likewise, the estimated binding free energy is overestimated compared to the experimental value, due to the sampling at short times with no binding–unbinding events. The overall free energy for the equilibrium PL + L  $\rightarrow$  PL<sub>2</sub> ( $\Delta G_2$ ) is also well reproduced:  $-7.8 \pm 0.9$  vs.  $-8.2 \pm 7.0$  kcal.mol<sup>-1</sup> (Figure 3). This value is lowered by 3.0 kcal.mol<sup>-1</sup> in the experiments and 5.3 kcal.mol<sup>-1</sup> in MD simulations compared to the first association, probably because of a smaller accessible protein surface area<sup>25</sup>. Our results confirm the performance of MM-PBSA to evaluate the binding free energies<sup>66–69</sup>, and validate in turn our computational approach for the characterization of the protein-molecular glue association.

The contributions summed over positively (lysines and arginines) or negatively (glutamates and aspartates) charged residues are also represented in Figure 3, alongside with the total binding free energies. For both associations, the additive decomposition per type of residues reveals that the *Cytc*-**sclx**<sub>8</sub> binding cannot be viewed as a lysine–only problem. Indeed the contribution from all lysines for the first and the second associations is 8 and 16 kcal.mol<sup>-1</sup> higher than the respective overall binding free energy (see Figure 3). Then, whereas arginines attractive contribution is smaller than the lysine one ( $\simeq 5$  kcal.mol<sup>-1</sup>), 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, which contributions in disfavoring the binding process are nearly equivalent (ca. 5 kcal.mol<sup>-1</sup> each). During the second association, the repulsion between the two anionic **sclx**<sub>8</sub> can also coun-

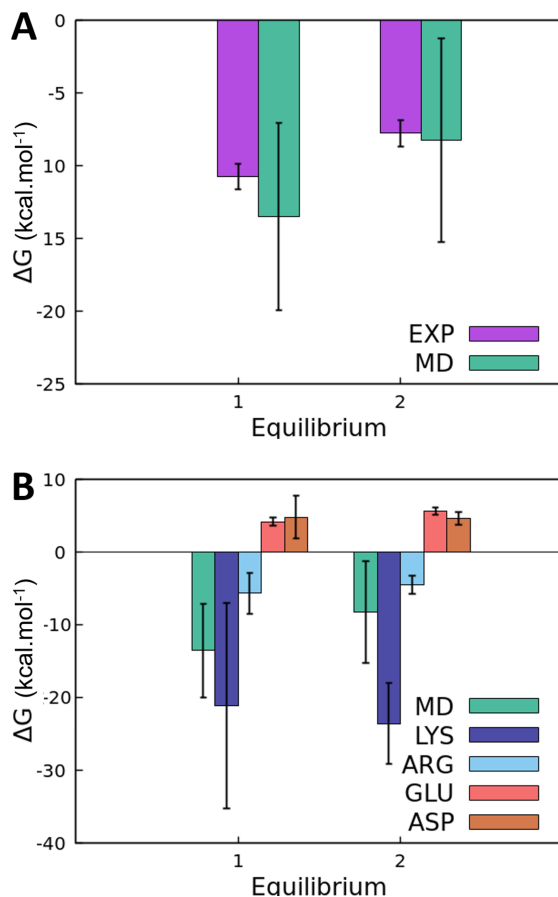


Figure 3: **A.** Experimental ITC vs. computed binding free energies  $\Delta G$  in kcal.mol<sup>-1</sup> for the first (Equilibrium 1) and second associations (Equilibrium 2). **B.** Binding free energy  $\Delta G$  in kcal.mol<sup>-1</sup> for the first and second associations and their decomposition over charged residues (lysine, arginine, glutamate and aspartate).

terbalance the attractive contribution from positively-charged residues and explain the large difference between this latter and the total binding free energy.

In order to further dissect the main interactions between **sclx**<sub>8</sub> and *Cytc*, we 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 and K100, six of which being present as "tweezers": for instance, K86 contributes to  $-3.7 \pm 2.8$

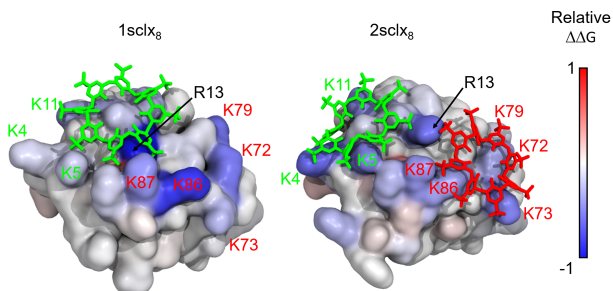


Figure 4: Color map of per-residue decomposition  $\Delta\Delta G_1$  (left panel) or  $\Delta\Delta G_2$  (right panel). The color scale is defined by dividing each contribution by the maximal absolute one ( $\Delta\Delta G_1(R13)$ ) and using the color code: blue for attractive interactions, and red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

kcal.mol<sup>-1</sup> to  $\Delta G_1$ , while K87 role is nearly halved,  $-2.1 \pm 1.4$  kcal.mol<sup>-1</sup>; on the contrary, their contributions to  $\Delta G_2$  are nearly equal. One can propose a linear dependence between the strength of the interaction and the distance between the center of mass of the lysine NH<sub>3</sub><sup>+</sup> group and the most proximal **sclx**<sub>8</sub> SO<sub>3</sub><sup>-</sup> sulfate group (see Figure S4). When one **sclx**<sub>8</sub> is present in the simulation, four lysines from Site 2 (namely K72, K79, K86, K87) define the hot spots 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 explains its large contribution of  $4.4 \pm 2.6$  kcal.mol<sup>-1</sup> to  $\Delta G_1$  (Table 1). The interaction landscape is modified when two **sclx**<sub>8</sub> 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 an individual contribution between 2 and 3 kcal.mol<sup>-1</sup>). Figure 4 highlights this difference with a concentration of bright blue residues (*i.e.* with a strongly attractive interaction with **sclx**<sub>8</sub>) around Site 2 for the **1sclx**<sub>8</sub> system whereas the blue surface area for the **2sclx**<sub>8</sub> system is wider but lighter.

Crowley and coworkers have also defined interacting residues in terms of their importance on the basis of the X-Ray structures for the different crystal packings<sup>10</sup>. Focusing on the participation of each residue to the *Cytc-sclx*<sub>8</sub> interface area ( $A_{buried}$ , see eq. 4 of ESI),

Table 1: *Interacting residues defining three different binding sites of **sclx**<sub>8</sub> on Cytc 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**<sub>8</sub> ( $\Delta\Delta G_1$ ), the second **sclx**<sub>8</sub> ( $\Delta\Delta G_2$ ), to the experimental interface area (Int. Area), and the theoretical difference in solvent accessible surface area between free or complex Cytc in the presence of two **sclx**<sub>8</sub> ( $\Delta_{SASA}$ ). Residues identified in the experimental crystal structure are boldfaced<sup>10</sup>. R13, here reported as part of in Site 1, can interact simultaneously with two **sclx**<sub>8</sub> molecules present in Sites 1 and 2, hence acting as a bridge residue.*

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

they describe master residues, that contribute more than  $\simeq 100$  Å<sup>2</sup> to the total *Cytc-sclx*<sub>8</sub> interface, and key residues that contribute 50-100 Å<sup>2</sup> to the total interface (see Table 1). To compare these results with our simulations, we determine the values of  $\Delta_{SASA}^i$  (see ESI for details), where  $i$  indicates the  $i$ -th residue, are reported in Table 1. These computational

values correspond to the surface of a residue which is covered by the calixarene. Although they cannot be quantitatively compared with the participation to the total interface from crystallographic structure, one can expect that they will follow the same tendency and describe similar residues as "key" or "master" ones. 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 hot spots described in the simulations. The others belong to sites that are not encountered during our simulations or little visited, for instance site 3. Likewise, L15 and Q16 were considered as key and master residues respectively but their contribution appear minimal compared to lysines. Indeed, in our simulations, the site involving these two residues is not visited. We observe conformations which are closer to the  $P3_1$  geometry than to the  $P4_32_12$  packing where the L15-Q16 site is occupied (see Figure S1). Otherwise, our simulations reveal the important role of K79, K87 and R13, which were not stressed by the interaction area from X-Ray analysis<sup>10</sup>.

Association of **sclx**<sub>8</sub> to *Cytc* has a local impact on the flexibility of the residues involving in the binding site but, because of the tertiary structure of protein, this effect can be propagated on a large distance. To identify the possible allosteric effects upon **sclx**<sub>8</sub> we took advantage of a supervised machine-learning protocol recently proposed by Fleetwood and coworkers<sup>61</sup> to monitor the structural changes of the cytochrome upon binding of **sclx**<sub>8</sub>. Important residues known to interact with **sclx**<sub>8</sub> correspond to sites 1, 2 and 3 are denoted in Figure 5 with stars and the respective site color (green, red and gray). The maximum importance corresponds to the bridging R13, supporting our previous conclusions about the central role of this arginine. Besides, the ML analysis reveals a reorganization of a series of residues (26, 38 and 44, as pink triangles below) which belong to a flexible loop of the protein, away from the binding sites, with formation of salt bridges and hydrogen bonds subsequent to the association with **sclx**<sub>8</sub>. The rigidification of this loop probably contributes to a surface entropy reduction. In addition, the importance associated to a leucine, which is close to the

heme, suggests that the  $\text{sclx}_8$  binding may also slightly impact the heme pocket.

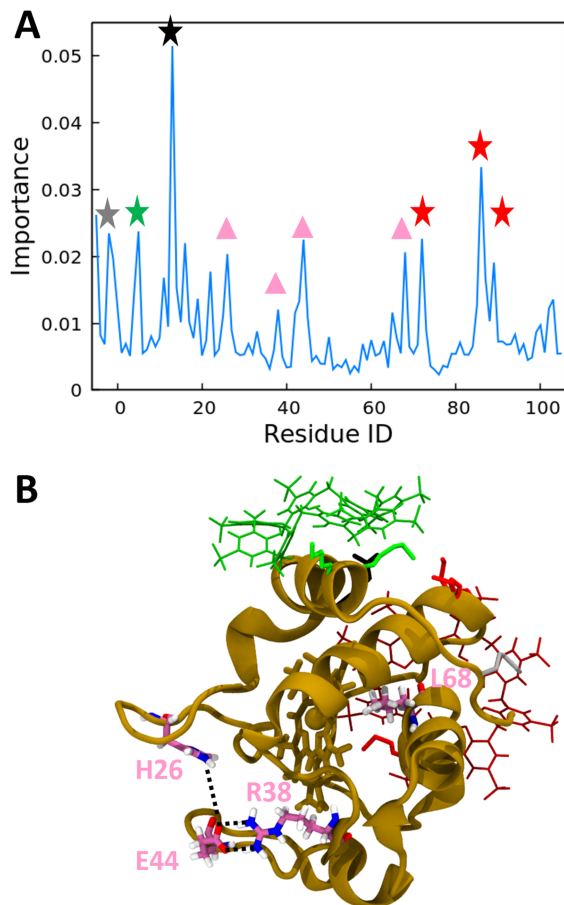


Figure 5: **A.** Important residues of upon  $\text{sclx}_8$  binding revealed by our supervised machine-learning protocol. In addition to the residues previously characterized in the different sites (labelled with colored stars), ML analysis highlights different residues far from the binding sites (pink triangles). **B.** Representation of the different residues denoted by the pink triangles in **A**. Encountered hydrogen bonds are depicted with black dashed lines.

## Conclusions

In this study, we have relied on molecular dynamics simulations to capture the interaction between  $\text{sclx}_8$  ligands and *Cytc*. Our simulations support the predominance in solution of two binding sites which are similar as thus described in the  $P3_1$  crystal structure (PDB

6GD8). The corresponding MM-PBSA binding free energies are in good agreement with the ITC measurement in solution. A further step would be to test the assembly of four *Cytc* using these sites as protein/additive/protein contact region and compare them to the SAXS data obtained for the protein oligomer in solution<sup>10</sup>. Thanks to the computational decomposition of the binding energy in terms of solvation, van der Waals or electrostatic interactions for the whole system or each residue, our approach draws an association mechanism mainly driven by the affinity between cationic lysine and anionic calixarene. It also highlights the role of arginine R13 that bridges the two sites and can counterbalance the electrostatic repulsion between the two **sclx**<sub>8</sub> ligands. The direct attractive contribution from other residues is minimal but a machine learning analysis reveals the formation of a hydrogen bond network concomitant with the **sclx**<sub>8</sub> binding that can help the stiffening of the protein and a better structural resolution. Taken together, our results confirm the ability of our computational protocol to finely investigate the molecular glue/protein surface interaction and its high potential as a predictive tool for screening purposes and design of new efficient auxiliaries.

## Acknowledgement

A.B. is grateful for a post-doctoral fellowship by the Foundation "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).

## References

- (1) van Dun, S.; Ottmann, C.; Milroy, L.-G.; Brunsveld, L. Supramolecular Chemistry Targeting Proteins. *Journal of the American Chemical Society* **2017**, *139*, 13960–13968.

- (2) Gao, C.; Chen, G. Exploring and Controlling the Polymorphism in Supramolecular Assemblies of Carbohydrates and Proteins. *Accounts of Chemical Research* **2020**, *53*, 740–751.
- (3) Pompidor, G.; D’Aléo, A.; Vicat, J.; Toupet, L.; Giraud, N.; Kahn, R.; Maury, O. Protein Crystallography through Supramolecular Interactions between a Lanthanide Complex and Arginine. *Angewandte Chemie International Edition* **2008**, *47*, 3388–3391.
- (4) Talon, R.; Kahn, R.; Durá, M. A.; Maury, O.; Vellieux, F. M. D.; Franzetti, B.; Girard, E. Using lanthanoid complexes to phase large macromolecular assemblies. *Journal of Synchrotron Radiation* **2011**, *18*, 74–78.
- (5) Breibeck, J.; Bijelic, A.; Rompel, A. Transition metal-substituted Keggin polyoxotungstates enabling covalent attachment to proteinase K upon co-crystallization. *Chem. Commun.* **2019**, *55*, 11519–11522.
- (6) Bijelic, A.; Rompel, A. The use of polyoxometalates in protein crystallography – An attempt to widen a well-known bottleneck. *Coordination Chemistry Reviews* **2015**, *299*, 22 – 38.
- (7) 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.
- (8) McGovern, R. E.; Feifel, S. C.; Lisdat, F.; Crowley, P. B. Microscale Crystals of Cytochrome-c and Calixarene on Electrodes: Interprotein Electron Transfer between Defined Sites. *Angewandte Chemie International Edition* **2015**, *54*, 6356–6359.
- (9) Doolan, A. M.; Rennie, M. L.; Crowley, P. B. Protein recognition by functionalized sulfonatocalix [4] arenes. *Chemistry–A European Journal* **2018**, *24*, 984–991.



- (10) 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.
- (11) Guagnini, F.; Antonik, P. M.; Rennie, M. L.; O’Byrne, P.; Khan, A. R.; Pinalli, R.; Dalcanale, E.; Crowley, P. B. Cucurbit[7]uril-Dimethyllysine Recognition in a Model Protein. *Angewandte Chemie International Edition* **2018**, *57*, 7126–7130.
- (12) Lai, Y.-T.; Reading, E.; Hura, G. L.; Tsai, K.-L.; Laganowsky, A.; Asturias, F. J.; Tainer, J. A.; Robinson, C. V.; Yeates, T. O. Structure of a designed protein cage that self-assembles into a highly porous cube. *Nature Chemistry* **2014**, *6*, 1065–1071.
- (13) McPherson, A.; Nguyen, C.; Cudney, R.; Larson, S. B. The Role of Small Molecule Additives and Chemical Modification in Protein Crystallization. *Crystal Growth & Design* **2011**, *11*, 1469–1474.
- (14) Danil de Namor, A. F.; Cleverley, R. M.; Zapata-Ormachea, M. L. Thermodynamics of calixarene chemistry. *Chemical reviews* **1998**, *98*, 2495–2526.
- (15) Nimse, S. B.; Kim, T. Biological applications of functionalized calixarenes. *Chem. Soc. Rev.* **2013**, *42*, 366–386.
- (16) Perret, F.; Lazar, A. N.; Coleman, A. W. Biochemistry of the para-sulfonato-calix [n] arenes. *Chemical Communications* **2006**, 2425–2438.
- (17) Giuliani, M.; Morbioli, I.; Sansone, F.; Casnati, A. Moulding calixarenes for biomacromolecule targeting. *Chem. Commun.* **2015**, *51*, 14140–14159.
- (18) Engilberge, S.; Rennie, M. L.; Crowley, P. B. Calixarene capture of partially unfolded cytochrome c. *FEBS Letters* **2019**, *593*, 2112–2117.

- (19) Alex, J. M.; Rennie, M. L.; Engilberge, S.; Lehoczki, G.; Dorottya, H.; Fizil, Á.; Batta, G.; Crowley, P. B. Calixarene-mediated assembly of a small antifungal protein. *IUCrJ* **2019**, *6*, 238–247.
- (20) Alex, J. M.; Rennie, M. L.; Volpi, S.; Sansone, F.; Casnati, A.; Crowley, P. B. Phosphonated Calixarene as a "Molecular Glue" for Protein Crystallization. *Crystal Growth & Design* **2018**, *18*, 2467–2473.
- (21) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.
- (22) McGovern, R. E.; McCarthy, A. A.; Crowley, P. B. Protein assembly mediated by sulfonatocalix [4] arene. *Chemical Communications* **2014**, *50*, 10412–10415.
- (23) Alex, J. M.; Brancatelli, G.; Volpi, S.; Bonaccorso, C.; Casnati, A.; Geremia, S.; Crowley, P. B. Probing the determinants of porosity in protein frameworks: co-crystals of cytochrome c and an octa-anionic calix[4]arene. *Org. Biomol. Chem.* **2020**, *18*, 211–214.
- (24) Bertini, I.; Cavallaro, G.; Rosato, A. Cytochrome c: Occurrence and Functions. *Chemical Reviews* **2006**, *106*, 90–115.
- (25) Rennie, M. L.; Crowley, P. B. A Thermodynamic Model of Auto-regulated Protein Assembly by a Supramolecular Scaffold. *ChemPhysChem* **2019**, *20*, 1011–1017.
- (26) 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.
- (27) Duan, L.; Liu, X.; Zhang, J. Z. Interaction Entropy: A New Paradigm for Highly

- Efficient and Reliable Computation of Protein-Ligand Binding Free Energy. *Journal of the American Chemical Society* **2016**, *138*, 5722–5728.
- (28) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *Journal of molecular recognition* **1996**, *9*, 1–5.
- (29) Yang, J.; Baek, M.; Seok, C. GalaxyDock3: Protein–ligand docking that considers the full ligand conformational flexibility. *Journal of Computational Chemistry* **2019**, *40*, 2739–2748.
- (30) Salmaso, V.; Moro, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Frontiers in Pharmacology* **2018**, *9*, 923.
- (31) Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chemical reviews* **1993**, *93*, 2395–2417.
- (32) 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. *Org. Biomol. Chem.* **2004**, *2*, 3162–3166.
- (33) 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.
- (34) 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.
- (35) Jang, Y. C. K. J., Y.; Kim, S.-U. Ab initio design of drug carriers for zoledronate guest

- molecule using phosphonated and sulfonated calix[4]arene and calix[4]resorcinarene host molecules. *J. Mater. Sci.* **2018**, *53*, 5125â5139.
- (36) 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.
- (37) 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.
- (38) Engilberge, S.; Riobé, F.; Wagner, T.; Di Pietro, S.; Breyton, C.; Franzetti, B.; Shima, S.; Girard, E.; Dumont, E.; Maury, O. Cover Feature: Unveiling the Binding Modes of the Crystallophore, a Terbium-based Nucleating and Phasing Molecular Agent for Protein Crystallography (Chem. Eur. J. 39/2018). *Chemistry–A European Journal* **2018**, *24*, 9701–9701.
- (39) Engilberge, S.; Rennie, M. L.; Dumont, E.; Crowley, P. B. Tuning Protein Frameworks via Auxiliary Supramolecular Interactions. *ACS nano* **2019**, *13*, 10343–10350.
- (40) 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.
- (41) 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*
- (42) Rovira, C.; Carloni, P.; Parrinello, M. The Iron- Sulfur Bond in Cytochrome c. *The Journal of Physical Chemistry B* **1999**, *103*, 7031–7035.

- (43) Li, P.; Merz Jr, K. M. MCPB. py: A python based metal center parameter builder. 2016.
- (44) 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.
- (45) Frisch, M. J. et al. Gaussian 16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.
- (46) 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.
- (47) 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.
- (48) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.
- (49) 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.
- (50) 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.
- (51) 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.

- (52) Homeyer, N.; Gohlke, H. Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. *Molecular Informatics* **2012**, *31*, 114–122.
- (53) 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.
- (54) 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.
- (55) Wang, J.; Cao, H.; Zhang, J. Z. H.; Qi, Y. Computational Protein Design with Deep Learning Neural Networks. *Scientific Reports* **2018**, *8*, 6349.
- (56) Ofran, Y.; Rost, B. ISIS: interaction sites identified from sequence. *Bioinformatics* **2007**, *23*, e13–e16.
- (57) Liu, S.; Liu, C.; Deng, L. Machine Learning Approaches for Protein-Protein Interaction Hot Spot Prediction: Progress and Comparative Assessment. *Molecules* **2018**, *23*.
- (58) Camacho, D. M.; Collins, K. M.; Powers, R. K.; Costello, J. C.; Collins, J. J. Next-Generation Machine Learning for Biological Networks. *Cell* **2018**, *173*, 1581 – 1592.
- (59) Ahmad, M.; Helms, V.; Kalinina, O. V.; Lengauer, T. Relative Principal Components Analysis: Application to Analyzing Biomolecular Conformational Changes. *Journal of Chemical Theory and Computation* **2019**, *15*, 2166–2178.
- (60) Degiacomi, M. T. Coupling Molecular Dynamics and Deep Learning to Mine Protein Conformational Space. *Structure* **2019**, *27*, 1034 – 1040.e3.
- (61) Fleetwood, O.; Kasimova, M. A.; Westerlund, A. M.; Delemotte, L. Molecular insights from conformational ensembles via machine learning. *Biophysical Journal* **2019**,

- (62) Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, FL, USA, 2011; pp 315–323.
- (63) Kingma, D. P.; Ba, J. 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.
- (64) Montavon, G.; Lapuschkin, S.; Binder, A.; Samek, W.; Müller, K.-R. Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognition* **2017**, *65*, 211 – 222.
- (65) Perret, F.; Coleman, A. W. Biochemistry of anionic calix[n]arenes. *Chem. Commun.* **2011**, *47*, 7303–7319.
- (66) Huang, D.; Qi, Y.; Song, J.; Zhang, J. Z. H. Calculation of hot spots for protein-protein interaction in p53/PMI-MDM2/MDMX complexes. *Journal of Computational Chemistry* **2019**, *40*, 1045–1056.
- (67) Retegan, M.; Milet, A.; Jamet, H. Exploring the Binding of Inhibitors Derived from Tetrabromobenzimidazole to the CK2 Protein Using a QM/MM-PB/SA Approach. *Journal of Chemical Information and Modeling* **2009**, *49*, 963–971.
- (68) Bignon, E.; Chan, C.-H.; Morell, C.; Monari, A.; Ravanat, J.-L.; Dumont, E. Molecular Dynamics Insights into Polyamine-DNA Binding Modes: Implications for Cross-Link Selectivity. *Chemistry - A European Journal* **2017**, *23*, 12845–12852.
- (69) 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.