

# Chemical Science

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

**Structural study of a small molecule receptor bound to dimethyllysine in lysozyme**

Róise E. McGovern,<sup>a</sup> Brendan D. Snarr,<sup>b</sup> Joseph A. Lyons,<sup>c</sup> James McFarlane,<sup>b</sup> Amanda L. Whiting,<sup>b</sup> Irina Paci,<sup>b</sup> Fraser Hof<sup>b</sup> and Peter B. Crowley<sup>\*,a</sup>

<sup>a</sup>School of Chemistry, National University of Ireland Galway, University Road, Galway, Ireland.

<sup>b</sup>Department of Chemistry, University of Victoria, British Columbia, V8W 3V6, Canada.

<sup>c</sup>School of Biochemistry and Immunology, Trinity College Dublin, Dublin, Ireland.

\*Correspondence to: peter.crowley@nuigalway.ie, +353 91 49 24 80

**Keywords:** Dimethyl lysine, Methylation, Protein interactions, Synthetic receptor

**Abstract**

Lysine is a ubiquitous residue on protein surfaces. Post translational modifications of lysine, including methylation to the mono-, di- or trimethylated amine result in chemical and structural alterations that have major consequences for protein interactions and signalling pathways. Small molecules that bind to methylated lysines are potential tools to modify such pathways. To make progress in this direction, detailed structural data of ligands in complex with methylated lysine is required. Here, we report a crystal structure of *p*-sulfonatocalix[4]arene (sclx<sub>4</sub>) bound to methylated lysozyme in which the lysine residues were chemically modified from Lys-NH<sub>3</sub><sup>+</sup> to Lys-NH(Me<sub>2</sub>)<sup>+</sup>. Of the six possible dimethyllysine sites, sclx<sub>4</sub> selected Lys116-Me<sub>2</sub> and the dimethylamino substituent was deeply buried in the calixarene cavity. This complex confirms the tendency for Lys-Me<sub>2</sub> residues to form cation- $\pi$  interactions, which have been shown to be important in protein recognition of histone tails bearing methylated lysines. Supporting data from NMR spectroscopy and MD simulations confirm the selectivity for Lys116-Me<sub>2</sub> in solution. The structure presented here may serve as a stepping stone to the development of new biochemical reagents that target methylated lysines.

## Introduction

Lysine is one of the most abundant residues on protein surfaces. With four methylenes and an epsilon amino group it is a cation of substantial conformational flexibility. Although native lysine is generally excluded from protein-protein interfaces<sup>1</sup> numerous post-translational modifications<sup>2,3</sup> produce a variety of functional groups with altered interaction properties. In particular, lysine methylation to the mono-, di- or trimethylated amine yields hotspots for protein interactions. Prominent examples include the methyllysines of histone tails that insert into aromatic cage motifs of chromatin remodelling enzymes.<sup>2,4-6</sup> In recent years many other (non-histone) proteins have been shown to contain methylated lysines,<sup>7,8</sup> though the roles of these modifications remain largely uncharacterized.

There is growing interest in the development of small molecule receptors that bind to lysine<sup>9-19</sup> and its methylated derivatives and analogues.<sup>20-30</sup> Synthetic ligands for methylated lysines have potential applications as inhibitors of protein-protein interactions and can be used as reagents in biochemical assays<sup>26,31</sup> and cell biology.<sup>32</sup> In certain cases it has been shown that lysine receptors bind to the methylated side chain with an affinity that is greater by several orders of magnitude and equilibrium dissociation constants ( $K_d$ ) of  $\sim 10$   $\mu$ M have been reported for peptides containing trimethyllysine.<sup>15,24,29</sup> The anionic *p*-sulfonatocalix[4]arene<sup>33</sup> (sclx<sub>4</sub>) and its analogues<sup>34</sup> have proven to be particularly useful for protein surface recognition<sup>17,35</sup> and/or complexation with lysine and methylated lysine.<sup>10,15</sup> A recent study of sclx<sub>4</sub> interactions with cytochrome *c* provided some of the first structural evidence of lysine recognition by a small molecule receptor.<sup>17</sup> And the structure of a phosphate-tweezers bound to a lysine in the 14-3-3 protein<sup>18</sup> further corroborated the use of supramolecular receptors for protein surface recognition.<sup>36,37</sup> Despite the growing literature on ligand binding to methylated lysines,<sup>27-32</sup> the structural characterization of a synthetic receptor bound to methylated lysine in a protein is completely lacking.

To gain structural knowledge of the interaction between a small molecule ligand and a protein bearing post-translationally modified lysines we solved the crystal structure of sclx<sub>4</sub> in complex with dimethylated lysozyme (lysozyme-KMe<sub>2</sub>). The complex was further characterized by NMR spectroscopy and molecular dynamics simulations. We identified a surprisingly selective binding of the calixarene at one of six possible dimethyllysine residues. This selectivity was rationalized in terms of the local chemical environments of the dimethyllysines. A second binding site at Arg14 was also found in the crystal structure.

## Results and Discussion

**Choice of model system.** Lysozyme is a highly-characterized model protein that is frequently used for ligand binding studies.<sup>38-41</sup> Moreover it is a workhorse for structural studies of lysine methylation<sup>42-46</sup> with well-established protocols for dimethylation by reductive alkylation which modifies lysines and the *N*-terminus. With a high proportion of lysine/arginine side chains and an overall positive electrostatic potential (*pI* ~10, Figure 1) lysozyme is suited to binding the anionic  $\text{sclx}_4$ .<sup>35</sup>

**Calixarene binding in solution.** The presence of  $\text{sclx}_4$  resulted in the immediate precipitation of lysozyme- $\text{KMe}_2$ . Precipitation occurred at  $\mu\text{M}$  -  $\text{mM}$  protein concentrations and crystals grew at protein and ligand concentrations as low as 20 and 1  $\mu\text{M}$ , respectively (Figure S1) and in the presence of different sulfate- and chloride-containing salts (Figure S2). Notably, crystal growth occurred in the absence of precipitants such as PEG or ammonium sulfate. These data suggested a relatively high affinity interaction ( $K_d \sim \mu\text{M}$ ).

Attempts to characterize the complex in water / buffer were thwarted by precipitation. Thus, NMR spectroscopy was performed on protein samples in water:DMSO mixtures. Apart from small changes, the  $^1\text{H}$  NMR spectrum of lysozyme was largely unaffected by 20 % DMSO (Figure S3) indicating that the protein was stably-folded under these conditions<sup>47</sup> (>50 % DMSO is required to unfold lysozyme).<sup>48</sup> Titrations were performed by the addition of  $\mu\text{L}$  volumes of a stock solution of  $\text{sclx}_4$  and complex formation was monitored by collecting 1D  $^1\text{H}$  and 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra (Figure 2). A single resonance at  $\sim 2.92$  ppm, assigned to the  $\text{N}^\epsilon\text{Me}$  protons of  $\text{Lys116-Me}_2$ ,<sup>46</sup> demonstrated a large upfield chemical shift. The nature of this chemical shift perturbation was consistent with ring current effects induced by the phenyl rings of the calixarene cavity and suggests that the dimethylamino group was buried inside  $\text{sclx}_4$ .<sup>15</sup> A plot of the chemical shift changes as a function of the ligand concentration resulted in a shallow curve (data not shown) that was unsuited to an accurate  $K_d$  determination. It was not possible to reach saturation as sample precipitation occurred at >10 equivalents of ligand. Similar chemical shift perturbations were observed in 10 % DMSO, although precipitation occurred at lower  $\text{sclx}_4$  concentrations. This indicates that DMSO serves to reduce precipitation without impacting the binding selectivity.

A small downfield shift for the resonance assigned to  $\text{Lys1-Me}_2$  ( $\sim 2.32$  ppm)<sup>46</sup> was observed at >3 equivalents of  $\text{sclx}_4$  suggesting that a weaker interaction occurred at this site. However, the downfield shift indicated that encapsulation of the dimethylamino did not occur

in this case. It is reasonable to assume that the probability of weak interactions at the *N*-terminal Lys1-Me<sub>2</sub> is greater than at the other dimethyllysines due to the relatively higher accessibility of this residue and the presence of two dimethylamines (at the N<sup>ε</sup> and the N<sup>α</sup> atoms).

**Crystal structure of the lysozyme-KMe<sub>2</sub>:sclx<sub>4</sub> complex.** Crystals of the sclx<sub>4</sub> complex with lysozyme-KMe<sub>2</sub> grew under similar conditions and in the same space group as native lysozyme<sup>35</sup> but with a ~50 % smaller unit cell (Table S1). Two almost identical structures (1.9 and 2.2 Å) were refined. The asymmetric units comprised two molecules of lysozyme-KMe<sub>2</sub> and four molecules of sclx<sub>4</sub> (Figure 3A). Similar protein-ligand interactions were observed in each lysozyme-KMe<sub>2</sub> molecule. In agreement with the NMR data (Figure 2) the side chain of Lys116-Me<sub>2</sub> was encapsulated by sclx<sub>4</sub>. A second ligand was observed to bind Arg14. As noted in previous structures, the calixarene appeared to act like “molecular glue” at protein-protein interfaces in the crystal.<sup>17,35</sup>

The sclx<sub>4</sub>-dimethyllysine interaction involves one of the methyl groups of Lys116-Me<sub>2</sub> inserted into the core of the calixarene, which behaves like a four-walled aromatic cage (Figure 4A). The distance between the methyl carbon (C<sup>n</sup>) and the centroids of the calixarene phenyl rings (3.5 - 3.9 Å) is consistent with cation- $\pi$  interactions.<sup>6,24,49</sup> Short range contacts occur between the second methyl group of Lys116-Me<sub>2</sub> and two of the phenyl rings. Interestingly, this methyl is also in van der Waals contact (3.5 and 3.8 Å) with oxygen atoms of two sulfonates, hinting at the possibility of CH $\cdots$ O salt bridges. Water too plays a role in the coordination environment of the dimethyllysine. The tertiary amino nitrogen is 2.7 Å from a water molecule, which is neatly positioned between two sulfonates, indicative of water-mediated salt bridge interactions. The energetic contributions to sclx<sub>4</sub>-dimethyllysine binding are expected to be dominated by cation- $\pi$  interactions (rather than salt bridges)<sup>6</sup> and a contribution from the hydrophobic effect is also to be expected considering the water-bearing capacity of the calixarene cavity.<sup>17,33</sup> These features of the sclx<sub>4</sub>-dimethyllysine interaction are further interesting in terms of their resemblance to how proteins such as chromodomains bind to methylated lysines in histone tails. An aromatic cage, typically comprising three aromatic side chains, provides a pocket for the methylated lysine side chain, which remains partially solvated (Figure 4B).<sup>2,5</sup>

Remarkably the conformation of Lys116-Me<sub>2</sub> bound to sclx<sub>4</sub> was almost identical to the side chain conformation observed in the original structure of dimethylated lysozyme (PDB

132L)<sup>43</sup> even though the crystals were grown from completely different conditions [PEG and low salt at pH 6.0 (Table S1) versus 1.5-2.2 M MgSO<sub>4</sub> at pH 8.0]. The only difference was a rotation about the C<sup>δ</sup>-C<sup>ε</sup> bond, which increases the accessibility of the dimethylamino group in the sclx<sub>4</sub>-bound structure.

**Lysine versus dimethyllysine binding and selectivity.** Substantial differences were observed for the binding of sclx<sub>4</sub> to lysine and to dimethyllysine. In the case of cytochrome *c*, the lysine side chains were fully encapsulated by the calixarene cavity such that all four methylene groups were in van der Waals contact with one or more of the calixarene phenyl rings.<sup>17</sup> To accommodate the entire side chain in this fashion the calixarene adopted a pinched cone conformation and the lysine amino group was positioned off-centre bringing it close to two of the sulfonates. In the case of dimethyllysine, one of the methyl groups is positioned in the centre of the calixarene such that the amino nitrogen is equidistant from all four sulfonates. Only the C<sup>n</sup> and C<sup>ε</sup> atoms make van der Waals contact with the calixarene phenyls while the remainder of the side chain protrudes from the cavity (Figure 4A). Thus it appears that this complex favours a regular cone conformation of sclx<sub>4</sub>, which maximises cation- $\pi$  bonds with the dimethylamino group.<sup>6,49</sup> Similar interactions were found previously in complexes of sclx<sub>4</sub> with tetramethylammonium cations.<sup>22</sup>

These observations help to explain the selectivity for Lys116-Me<sub>2</sub>, which projects out from the protein surface and has Asn106, Arg112 and Gly117 as neighbours. A hydrogen bond (Lys116 to Asn106) in the native protein is absent in the dimethylated protein where the Lys116-Me<sub>2</sub> side chain flips into the calixarene cavity and Asn106 hydrogen bonds to two of the sulfonates (Figure 4A). The steric accessibility of most of the other Lys-Me<sub>2</sub> side chains is significantly lower, which may preclude sclx<sub>4</sub> binding. Lys1-Me<sub>2</sub> forms a cation- $\pi$  interaction with Phe3, and salt bridges with Glu7 and the sclx<sub>4</sub> bound to Arg14; Lys13-Me<sub>2</sub> is screened by the carboxylates of Asp18 and C-terminal Leu129; Lys33-Me<sub>2</sub> is flanked by Asn37 and the bulky aromatics Phe34, Phe38 and Trp123, only the amine is accessible and it forms a salt bridge with a sclx<sub>4</sub> sulfonate; Lys96-Me<sub>2</sub> is buried and forms weak cation- $\pi$  interactions with both Tyr20 and the sclx<sub>4</sub> bound to Arg14; while Lys97-Me<sub>2</sub> forms a salt bridge with Asp101. Similar conformations of the lysine side chains are present in the native lysozyme structure. To substantiate these observations the solvent accessible surface area (ASA) was calculated for each lysine in a dataset of 15 high resolution structures of lysozyme.<sup>35</sup> On average, Lys116 was the most accessible lysine (Figure 5). While Lys97 has a similar ASA, it may be

the differences in the local charge that tips the scales in favour of  $\text{sclx}_4$  binding to Lys116. Considering charged groups within an 8 Å radius, Lys97-Me<sub>2</sub> forms a salt bridge with Asp101 while Lys116-Me<sub>2</sub> is neighboured by Arg112. The higher positive charge of the latter region will afford a greater attraction for the anionic  $\text{sclx}_4$ . To investigate this hypothesis we performed MD simulations of  $\text{sclx}_4$  binding to Lys-Me<sub>2</sub> side chains.

**Molecular dynamics of  $\text{sclx}_4$  binding to Lys-Me<sub>2</sub>.** Duplicate simulations (10 ns duration) were performed on  $\text{sclx}_4$  binding to each of the six Lys-Me<sub>2</sub> side chains. The goal was to identify the structural features that distinguish Lys116-Me<sub>2</sub> from the other five potential binding sites. Two main features were considered; (1) the accessibility of the side chain, and (2) the local interactions, including those between the ligand substituents and peripheral residues. Site-specific information was determined from the molecular dynamics trajectories, which were examined primarily in terms of the “binding distance” as a means to quantify the degree of encapsulation of the Lys-Me<sub>2</sub> in the  $\text{sclx}_4$  cavity. This was defined as the distance from the Lys-Me<sub>2</sub> N<sup>ε</sup> atom to the best-fit plane through the methylene bridge carbons of the calixarene. Figure 6A shows the evolution of the binding distance at the six sites, during two simulation trials. The trends explicitly illustrate the large differences in the potential for complex formation at the different sites. Lys13-Me<sub>2</sub>, Lys33-Me<sub>2</sub> and Lys96-Me<sub>2</sub> did not form temporally stable complexes with  $\text{sclx}_4$ . In contrast, Lys97-Me<sub>2</sub> and Lys116-Me<sub>2</sub> were stably bound for almost the entire trajectory of each simulation consistent with the greater accessibility of these residues (Figure 5). At a binding distance of 4-5 Å the dimethylamino is positioned deep within the calixarene cavity where it forms cation- $\pi$  interactions. The complex at Lys1-Me<sub>2</sub> was similar though it was unbound over ~3 ns of trial 2, suggesting a lower stability interaction. Attempts to discriminate the sites at Lys1-Me<sub>2</sub>, Lys97-Me<sub>2</sub> and Lys116-Me<sub>2</sub> using calculated binding energies were unsuccessful, likely due to the well-documented incompatibility of the MM-PBSA end-state free energy method with highly charged systems.<sup>50,51</sup>

Further analysis was focused on the complexes at Lys97-Me<sub>2</sub> and Lys116-Me<sub>2</sub>. Representative snapshots of the complexes at Lys116-Me<sub>2</sub> (Figure 6B, see also the MD movie, SI) show the potential role for both Arg112, which can form salt bridge interactions with one of the calixarene sulfonates, and Asn105, which can hydrogen bond to the sulfonates. The two residues interact with the calixarene outer rim alternatively, with the Arg112 in binding position 25-40 % of the time. These observations were corroborated by the crystal structure.



In protein chain B there is a salt bridge between Arg112 and one of the calixarene sulfonates, while in chain A an alternative rotamer is present and Arg112 forms an intramolecular hydrogen bond with the protein backbone. The amide of Asn106 is positioned equidistantly from two sulfonates in the site at chain A, while in chain B only the amide N can form a hydrogen bond with sclx<sub>4</sub>. Similarly, calixarene binding at Lys97-Me<sub>2</sub> can involve salt bridge interactions with Arg21 (80-90% of the simulation time) and hydrogen bonds with Asn93 with a lower incidence. However, as noted above Lys97-Me<sub>2</sub> can also salt bridge with Asp101. The MD data suggests that, once complexed, Lys116-Me<sub>2</sub> and Lys97-Me<sub>2</sub> can interact with sclx<sub>4</sub> in a similar fashion. Selectivity of the calixarene to Lys116-Me<sub>2</sub> is due to the greater steric accessibility at that site. This is supported by the relative values of ASA averages collected during the solution-phase simulations (data not shown), which were in agreement with the crystallographic data (Figure 5), although ASA values in solution tend to be larger than in solid state due to the enhanced conformational freedom.

**Arginine binding by sclx<sub>4</sub>.** While the selectivity of sclx<sub>4</sub> for Lys116-Me<sub>2</sub> can be rationalised in terms of steric accessibility, there remains the question of why binding also occurred at Arg14 in the crystal structure (Figure 3). This question is interesting for two reasons. (1) NMR-derived binding curves for the interaction of sclx<sub>4</sub> with free amino acids have revealed a ~50-fold greater affinity for Lys-Me<sub>2</sub> over Arg.<sup>15</sup> Thus, it might be expected that sclx<sub>4</sub> would bind only Lys-Me<sub>2</sub> side chains in lysozyme-KMe<sub>2</sub>. However, complex formation at Arg14 (instead of, for example, at Lys13-Me<sub>2</sub>) reinforces the fact that protein surfaces, with their complex topologies and chemistries, can greatly alter the affinity of ligand binding. (2) In a crystal structure of native lysozyme and sclx<sub>4</sub> the ligand was bound at Arg128, the most sterically accessible arginine residue.<sup>35</sup> In lysozyme-KMe<sub>2</sub>, Arg14, the second most accessible Arg residue was selected by sclx<sub>4</sub>. Arg128 provides additional longer range (~ 8 Å) interactions to the sulfonates. It can be concluded that the affinities of sclx<sub>4</sub> for Arg14 and Arg128 are closely matched and the particular complex that prevails in the crystal structure depends on the crystal packing environment where the calixarene mediates protein-protein contacts.<sup>17,35</sup>

## Conclusions

The data presented here illustrate how a protein containing dimethyllysine can be non-covalently modified by a small molecule receptor. Using a combination of X-ray crystallography, NMR spectroscopy and MD simulations we have shown how the symmetric and anionic  $\text{sclx}_4$  selectively binds to a single dimethylated lysine on the surface of a globular, folded protein. We note structural and chemical similarities between the complexes of dimethyllysine bound to the *simple* calixarene or to the aromatic cage motif of a chromodomain. This data will likely benefit the design of synthetic receptors for proteins (including histones) that contain methylated lysines.

## Experimental

**Materials.** Hen egg white lysozyme (62971 Fluka) was dimethylated by using dimethylamine borane complex and formaldehyde according to published methods.<sup>43-46</sup> Electrospray ionization mass spectrometry data (Waters LCT Premier XE) for lysozyme (14302.2 Da) and dimethylated lysozyme (14498.0 Da) indicated complete dimethylation of all six lysines and the *N*-terminus. <sup>13</sup>C-formaldehyde was used to prepare dimethylated samples for <sup>13</sup>C NMR spectroscopy. Chemically-modified protein was purified by carboxymethyl (GE Healthcare) ion exchange chromatography prior to the crystallization experiments.

**NMR Spectroscopy.** 1D <sup>1</sup>H and 2D <sup>1</sup>H-<sup>13</sup>C HSQC spectra were acquired on a Bruker AV500 operating at 500 MHz and 25° C. Protein samples of 0.3 - 10 mM lysozyme-KMe<sub>2</sub> in 40 mM sodium phosphate, 10 % D<sub>2</sub>O, and 20 % DMSO-*d*<sub>6</sub> at pD = 7.0 (pH 7.4) were titrated with μL volumes of a 0.65 M stock of sclx<sub>4</sub> in the same solution.

**Crystallization and X-ray Structure Determination.** The hanging drop vapour diffusion method was used for crystallization at 20° C. Co-crystals of lysozyme-KMe<sub>2</sub> and sclx<sub>4</sub> were grown from similar conditions to those reported for lysozyme.<sup>35</sup> The drops were prepared by combining 1 μL volumes of protein, sclx<sub>4</sub> and the reservoir solution (Table S1). Diffraction data for the lysozyme-KMe<sub>2</sub>:sclx<sub>4</sub> single crystals were collected at the ESRF (BM14, MarCCD detector,  $\phi$  scans of 1° over 180° to a resolution of 1.9 Å) and at the Swiss Light Source (X10SA, 10 μm minibeam, Pilatus 6M detector,  $\phi$  scans of 0.5° over 180°, to a resolution of 2.2 Å). Data processing and scaling were performed in MOSFLM<sup>52</sup> and SCALA,<sup>53</sup> respectively or in xia2<sup>54</sup> using XDS<sup>55</sup>, XSCALE and SCALA. See Table S1 for the data collection and refinement statistics. The structures were solved by molecular replacement in PHASER. Refinement and manual rebuilding were performed in REFMAC5 as implemented in CCP4<sup>56</sup> and COOT,<sup>57</sup> respectively. Solvent molecules were placed automatically using ARP/wARP<sup>58</sup> and refinement was continued until no features remained in the F<sub>o</sub> - F<sub>c</sub> difference maps. Molprobit<sup>59</sup> was used to check the structure quality. Coordinates and structure factors were deposited in the Protein Data Bank with the accession codes 4PRU (2.2 Å) and 4N0J (1.9 Å). The protein-ligand and protein-protein interfaces were analysed in COOT.

**Molecular dynamics of protein-calixarene interactions.** Binding dynamics were followed using classical molecular dynamics over 10 ns intervals, at a temperature of 300 K. To reduce the introduction of conformation bias a structure of native lysozyme (PDB 3RZ4) was used for the initial coordinates. This structure was modified with newly parameterized Lys-Me<sub>2</sub> residues replacing all six of the lysines. Dimethyllysine was not available in the standard AMBER residue library so the parameters were retrieved from the literature.<sup>60</sup> Partial charges were derived from gas phase optimized HF/6-31\*G calculations in Gaussian09, and fit in the preparatory program Antechamber using the RESP charge fitting method. The remaining parameters for the nonstandard amino acids were obtained from Antechamber and fit to the AMBER ff10 force field. Parameters for the calixarene were fit to the general AMBER force field for small organic molecules. The ligand structure was minimized in explicit TIP3P water prior to being placed with the protein for simulation. The calculations used explicit TIP3P water and Cl<sup>-</sup> counter-ions added to charge neutrality.<sup>61,62</sup> Duplicate protein-calixarene complexes were generated for each candidate Lys-Me<sub>2</sub> site by combining the equilibrated protein and the minimized sclx<sub>4</sub> structure. The ligand was placed approximately 7-10 Å above the Lys-Me<sub>2</sub> side chain and the complex was allowed to equilibrate. The binding energy at each site was calculated using the MM-PBSA method (implicit solvent).<sup>50,51</sup> Docking was also monitored through calculations of the average binding distance between the dimethylamino group and the plane of the calixarene methylene bridge carbons.

### Supporting Information

Figure S1: Crystals of the lysozyme-KMe<sub>2</sub>:sclx<sub>4</sub> complex grown at different sclx<sub>4</sub> concentrations. Figure S2. Crystals of the complex grown in the presence of chloride- and sulfate-containing salts. Figure S3. 1D <sup>1</sup>H NMR spectra of lysozyme-KMe<sub>2</sub> in buffer and DMSO mixtures. Table S1: Summary of crystallization conditions, data collection and refinement statistics. Movie S1: MD simulation snapshots of sclx<sub>4</sub> binding to Lys116-Me<sub>2</sub>.

**Acknowledgements**

This research was supported by NUI Galway (college scholarship to REM, Millennium Fund to PBC), NSERC grants to IP and FH (Canada Research Chair), and Science Foundation Ireland (10/RFP/BIC2807 to PBC). JAL was funded by grants 07/IN.1/B1836 (SFI) and GM75915 (NIH) to M. Caffrey. ALW was funded by a Michael Smith Foundation for Health Research fellowship. We acknowledge the Swiss light source and the European synchrotron radiation facility for beam time allocation, and the staff of beam lines X10SA (Villigen) and BM14 (Grenoble) for assistance with data collection. Computational work was performed using WestGrid, funded in part by the Canada Foundation for Innovation, Alberta Innovation and Science, BC Advanced Education, and the participating research institutions. For their assistance we thank V. Pye (TCD), A. R. Khan (TCD), C. Greenwood (UVic), C. Barr (UVic) and O. Granot (UVic).

## References

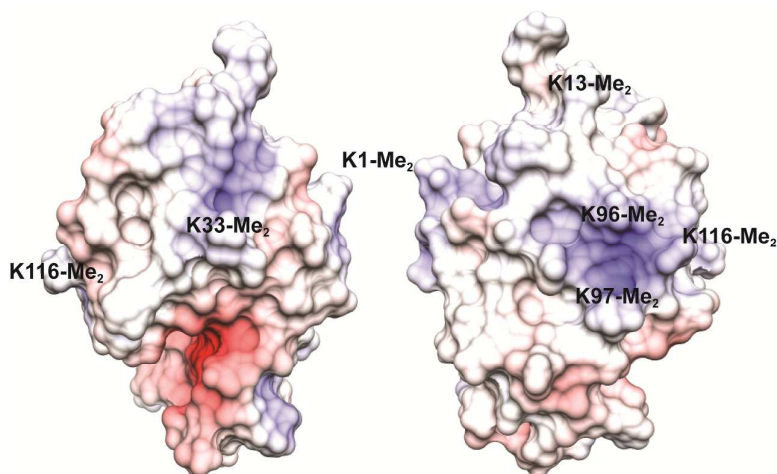
1. J. Janin, R. P. Bahadur and P. Chakrabarti, *Q. Rev. Biophys.* 2008, **41**, 133.
2. S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis and D. J. Patel, *Nat. Struct. Mol. Biol.* 2007, **14**, 1025.
3. Z. Zhang, M. Tan, Z. Xie, L. Dai, Y. Chen and Y. Zhao, *Nat. Chem. Biol.* 2011, **7**, 58.
4. S. Rea, F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, C. D. Allis and T. Jenuwein, *Nature* 2000, **406**, 593.
5. S. A. Jacobs and S. Khorasanizadeh, *Science* 2002, **295**, 2080.
6. R. M. Hughes, K. R. Wiggins, S. Khorasanizadeh and M. L. Waters, *Proc. Natl. Acad. Sci. USA* 2007, **104**, 11184.
7. X. J. Cao, A. M. Arnaudo and B. A. Garcia, *Epigenetics* 2013, **8**, 477.
8. K. E. Moore, S. M. Carlson, N. D. Camp, P. Cheung, R. G. James, K. F. Chua, A. Wolf-Yadlin and O. Gozani, *Mol. Cell.* 2013, **50**, 444.
9. Y. Hamuro, M. C. Calama, H. S. Park and A. D. Hamilton, *Angew. Chem. Int. Ed. Engl.* 1997, **36**, 2680.
10. M. Selkti, A. W. Coleman, I. Nicolis, N. Douteau-Guevel, F. Villain, A. Tomas and de Rango, *C. Chem. Commun.* 2000, **2**, 161.
11. M. Fokkens, T. Schrader and F.-G. Klärner, *J. Am. Chem. Soc.* 2005, **127**, 14415.
12. C. Renner, J. Piehler and T. Schrader, *J. Am. Chem. Soc.* 2006, **128**, 620.
13. O. Hayashida, N. Ogawa and M. Uchiyama, *J. Am. Chem. Soc.* 2007, **129**, 13698.
14. K. Kano and Y. Ishida, *Angew. Chem. Int. Ed. Engl.* 2007, **46**, 727.
15. C. S. Beshara, C. E. Jones, K. D. Daze, B. J. Lilgert and F. Hof, *ChemBioChem* 2010, **11**, 63.
16. M. Florea, S. Kudithipudi, A. Rei, M. J. González-Álvarez, A. Jeltsch and W. M. Nau, *Chem. Eur. J.* 2012, **18**, 3521.
17. R. E. McGovern, H. Fernandes, A. R. Khan, N. P. Power and P. B. Crowley, *Nat. Chem.* 2012, **4**, 527.
18. D. Bier, R. Rose, K. Bravo-Rodriguez, M. Bartel, J. M. Ramirez-Anguila, S. Dutt, C. Wilch, F. G. Klärner, E. Sanchez-Garcia, T. Schrader and C. Ottmann, *Nat. Chem.* 2013, **5**, 234.
19. C. J. Li, J. W. Ma, L. Zhao, Y. Y. Zhang, Y. H. Yu, X. Y. Shu, J. Li and X. S. Jia, *Chem. Commun.* 2013, **49**, 1924.
20. J. L. Atwood, L. J. Barbour, P. C. Junk and G. W. Orr, *Supramol. Chem.* 1995, **5**, 105.

21. G. Arena, A. Casnati, A. Contino, G. G. Lombardo, D. Sciotto and R. Ungaro, *Chem. Eur. J.* 1999, **5**, 738.
22. E. Da Silva, F. Nouar, M. Nierlich, B. Rather, M. J. Zaworotko, C. Barbey, A. Navaza, A. and W. Coleman, *Cryst. Eng.* 2003, **6**, 123.
23. L. Vial, R. F. Ludlow, J. Leclaire, R. Pérez-Fernández and S. Otto, *J. Am. Chem. Soc.* 2006, **128**, 10253.
24. L. A. Ingeman, M. E. Cuellar and M. L. Waters, *Chem. Commun.* 2010, **46**, 1839.
25. M. Dionisio, G. Oliviero, D. Menozzi, S. Federici, R. M. Yebeutchou, F. P. Schmidtchen, E. Dalcanale and P. Bergese, *J. Am. Chem. Soc.* 2012, **134**, 2392.
26. S. A. Minaker, K. D. Daze, M. C. Ma and F. Hof, *J. Am. Chem. Soc.* 2012, **134**, 11674.
27. K. D. Daze, T. Pinter, C. S. Beshara, A. Ibraheem, S. A. Minaker, M. C. F. Ma, R. J. M. Courtemanche, R. E. Campbell and F. Hof, *Chem. Sci.* 2012, **3**, 2695.
28. M. A. Gamal-Eldin and D.H. Macartney, *Org. Biomol. Chem.* 2013, **11**, 488.
29. L. I. James, J. E. Beaver, N. W. Rice and M. L. Waters, *J. Am. Chem. Soc.* 2013, **135**, 6450.
30. N. K. Pinkin and M. L. Waters. *Org. Biomol. Chem.* 2014, **12**, 7059.
31. S. Tabet, S. F. Douglas, K. D. Daze, G. A. Garnett, K. J. H. Allen, E. M. M. Abrioux, T. H. Quon, J. E. Wulff, and F. Hof, *Biorg. Med. Chem.* 2013, **21**, 7004.
32. H. F. Allen, K. D. Daze, T. Shimbo, A. Lai, C. A. Musselman, J. K. Sims, P. A. Wade, F. Hof and T. G. Kutateladze, *Biochem. J.* 2014, **459**, 505.
33. K. Fucke, K. M. Anderson, M. H. Filby, M. Henry, J. Wright, S. A. Mason, M. J. Gutmann, L. J. Barbour, C. Oliver, A. W. Coleman, J. L. Atwood, J. A. Howard and J. W. Steed, *Chem. Eur. J.* 2011, **17**, 10259.
34. K. D. Daze, M. C. F. Ma, F. Pineux and F. Hof, *Org. Lett.* 2012 **14**, 1512.
35. R. E. McGovern, A. A. McCarthy and P. B. Crowley, *Chem. Commun.* 2014, **50**, 10412.
36. X. Salvatella, M. Martinell, M. Gairí, M. G. Mateu, M. Feliz, A. D. Hamilton, J. De Mendoza and E. Giralt, *Angew. Chem. Int. Ed. Engl.* 2004, **43**, 196.
37. J. M. Chinai, A. B. Taylor, L. M. Ryno, N. D. Hargreaves, C. A. Morris, P. J. Hart and A. R. Urbach, *J. Am. Chem. Soc.* 2011, **133**, 8810.
38. T. M. Hunter, I. W. McNae, X. Liang, J. Bella, S. Parsons, M. D. Walkinshaw and P. J. Sadler, *Proc. Natl. Acad. Sci. USA* 2005, **102**, 2288.
39. J. Muldoon, A. E. Ashcroft and A. J. Wilson, *Chem. Eur. J.*, 2010, **16**, 100.
40. T. Muraoka, K. Adachi, M. Ui, S. Kawasaki, N. Sadhukhan, H. Obara, H. Tochio, M. Shirakawa and K. Kinbara, *Angew. Chem. Int. Ed. Engl.* 2013, **52**, 2430.

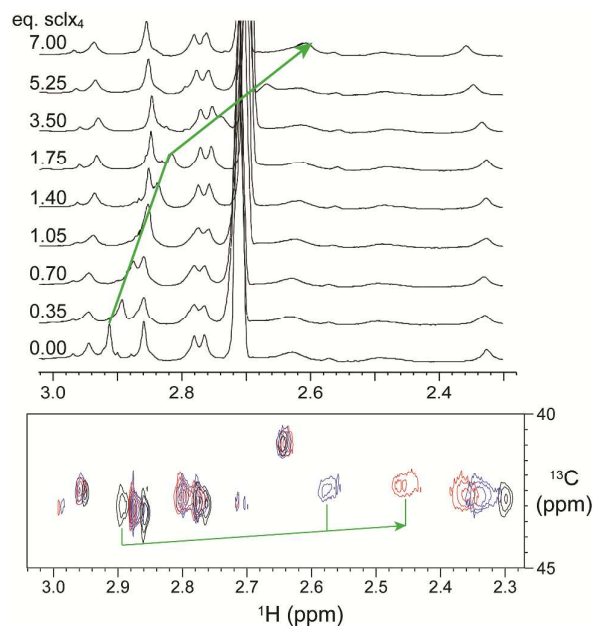
41. M. Calvaresi, F. Arnesano, S. Bonacchi, A. Bottoni, V. Calò, S. Conte, G. Falini, S. Fermani, M. Losacco, M. Montalti, G. Natile, L. Prodi, F. Sparla and F. Zerbetto, *ACS Nano* 2014, **8**, 1871.
42. T. A. Gerken, J. E. Jentoft, N. Jentoft and D. G. Dearborn, *J. Biol. Chem.* 1982, **257**, 2894.
43. W. R. Rypniewski, H. M. Holden and I. Rayment, *Biochemistry* 1993, **32**, 9851.
44. M. A. Macnaughtan, A.M. Kane and J. H. Prestegard, *J. Am. Chem. Soc.* 2005, **127**, 17626.
45. S. J. Abraham, T. Kobayashi, R. J. Solaro and V. Gaponenko, *J. Biomol. NMR* 2009, **43**, 239.
46. S. T. Larda, M. P. Bokoch, F. Evanics and R. S. Prosser, *J. Biomol. NMR* 2012, **54**, 199.
47. M. S. Lehmann and R. F. D. Stansfield, *Biochemistry* 1989, **28**, 7028.
48. S. Bhattacharjya and P. Balaram, *Proteins*, 1997, **29**, 492-507.
49. D. A. Dougherty, *Acc. Chem. Res.* 2013, **46**, 885.
50. B. Kuhn, P. Gerber, T. Schulz-Gasch, and M. Stahl, *J. Med. Chem.* 2005, **48**, 4040.
51. B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, *J. Chem. Theory Comput.* 2012, **8**, 3314.
52. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data; Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography 1992.
53. P. Evans, *Acta Crystallogr. D.* 2006, **62**, 72.
54. G. Winter, *J. Appl. Crystallogr.* 2010, **43**, 186.
55. W. Kabsch, *Acta Crystallogr. D* 2010, **66**, 125.
56. Collaborative Computational Project, N. *The CCP4 suite: programs for protein crystallography*, *Acta Crystallogr. D* 1994, **50**, 760.
57. P. Emsley and K. Cowtan, *Acta Crystallogr. D.* 2004, **60**, 2126.
58. E. J. van Asselt, A. Perrakis, K. H. Kalk, V. S. Lamzin B. W. and Dijkstra, *Acta Crystallogr. D.* 1998, **54**, 58.
59. V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, *Acta Crystallogr. D* 2010, **66**, 12.
60. G. V. Papamokos, G. Tziatzos, D. G. Papageorgiou, S. D. Georgatos, A. S. Politou and E. Kaxiras, *Biophys. J.* 2012, **102**, 1926.
61. I. S. Joung and T. E. Cheatham III, *J. Phys. Chem. B* 2008, **112**, 9020.
62. I. S. Joung and T. E. Cheatham III, *J. Phys. Chem. B* 2009, **113**, 13279.



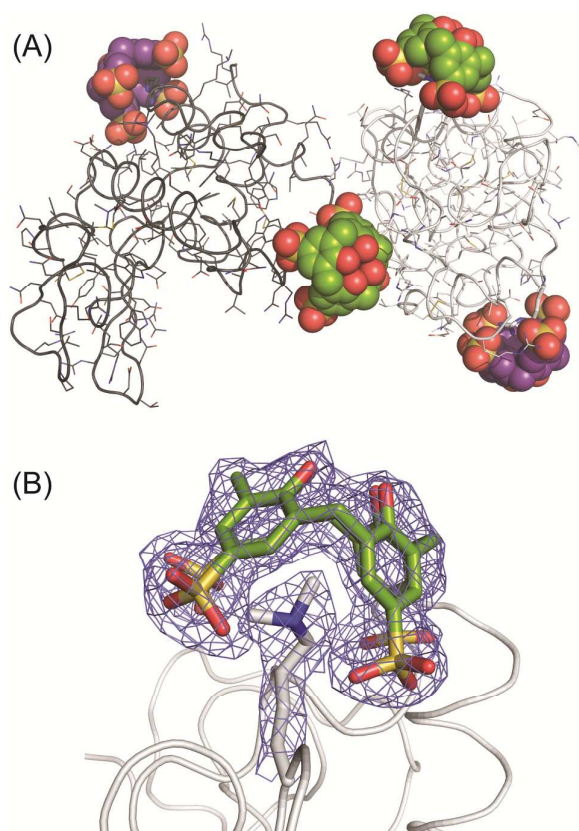
## Figures



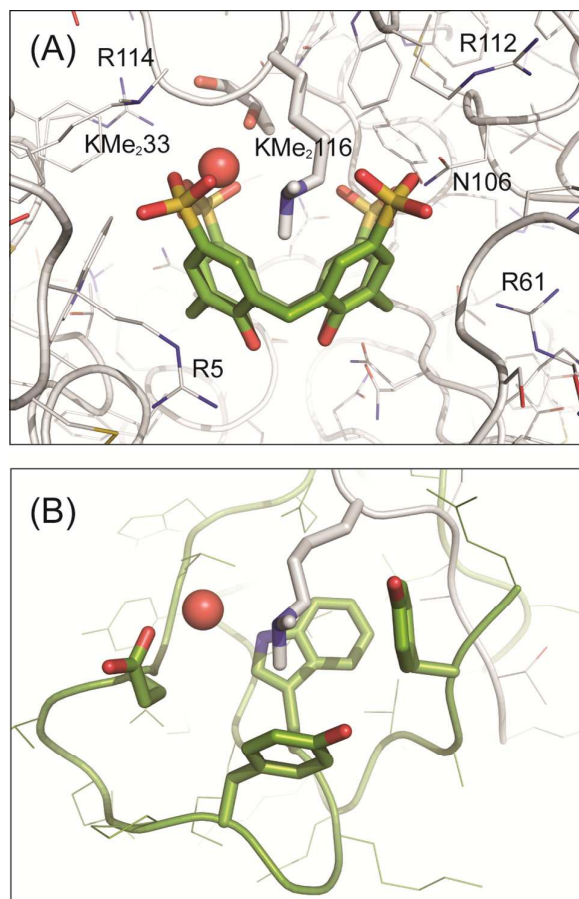
**Figure 1.** The electrostatic potential surface (Adaptive Poisson-Boltzmann Solver) of lysozyme-KMe<sub>2</sub> with positive and negative patches coloured blue and red, respectively (the two views are related by a 180° rotation). Labels indicate the approximate locations of each of the six dimethyllysine residues.



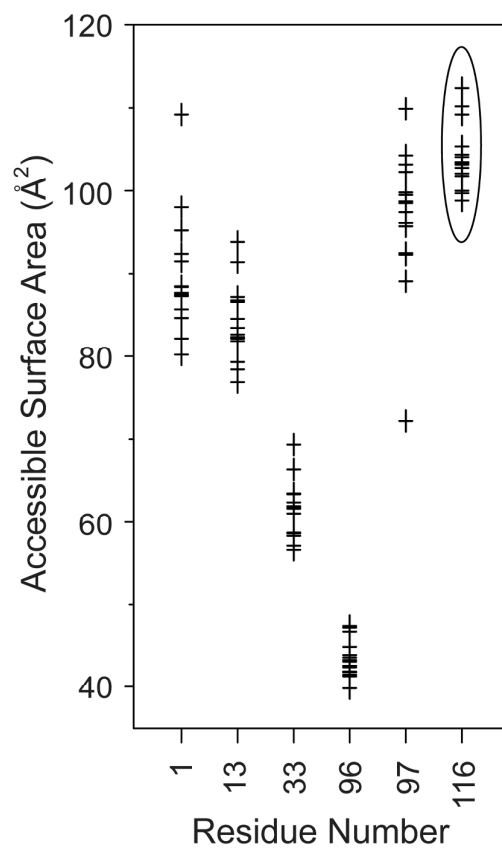
**Figure 2.** NMR spectroscopic characterization of sclx<sub>4</sub> interactions with lysozyme-KMe<sub>2</sub>. Upper panel, 1D <sup>1</sup>H NMR spectra (showing the region corresponding to -NMe<sub>2</sub> resonances) of dimethylated lysozyme in the presence of 0 - 7 equivalents of sclx<sub>4</sub>. Lower panel, <sup>1</sup>H-<sup>13</sup>C HSQC spectra of <sup>13</sup>C-labeled dimethylated lysozyme (black contours) in the presence of 7 (blue) and 10 (red) equivalents of sclx<sub>4</sub>. The green arrows indicate the upfield shifts of the resonance assigned to Lys116-Me<sub>2</sub>. A smaller downfield shift was observed for the Lys1-Me<sub>2</sub> resonance. The signal at ~2.7 ppm corresponds to DMSO. Samples were in 40 mM sodium phosphate, 10 % D<sub>2</sub>O, and 20 % DMSO-*d*<sub>6</sub>, pH 7.4.



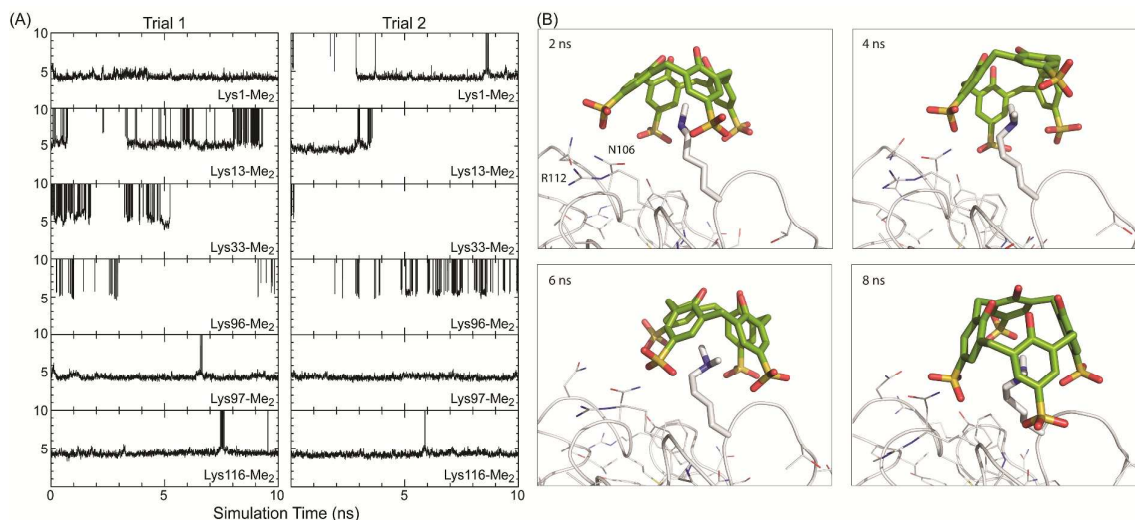
**Figure 3.** The lysozyme-KMe<sub>2</sub>:sclx<sub>4</sub> complex. (A) The asymmetric unit comprises two molecules of lysozyme-KMe<sub>2</sub> (rendered as light and dark grey ribbons) and four molecules of sclx<sub>4</sub>. The dimethyllysine-binding and the arginine-binding calixarenes are coloured green and purple, respectively. (B) The calixarene binding site at Lys116-Me<sub>2</sub> showing the 2F<sub>o</sub>-F<sub>c</sub> electron density map around the Lys116-Me<sub>2</sub> side chain and sclx<sub>4</sub> (contoured at 1.0  $\sigma$ ). See Figure 4A for a detailed view of the sclx<sub>4</sub>:Lys116-Me<sub>2</sub> interaction. The crystals used for this structure were grown from a 1:5 protein:ligand mixture (Table S1).



**Figure 4.** Sclx<sub>4</sub> mimics the aromatic cage motif for binding dimethyllysine. (A) Detailed view of the sclx<sub>4</sub> complex with Lys116-Me<sub>2</sub>. Neighbouring cationic side chains and Asn106 are labelled, see main text for details. (B) The aromatic cage (green) in the chromodomain of HP1 bound to dimethyllysine in a histone H3 tail peptide (PDB 1KNA).<sup>5</sup> The chromodomain side chains Tyr24, Trp45, Tyr48 and Glu52 are shown as sticks. In both structures the dimethyllysine amino is solvated by a water molecule (red spheres), which is hydrogen bonded to one or more acidic substituents on the receptor. The proteins are shown as C<sup>α</sup> traces with side chains as lines or sticks and the C atoms of sclx<sub>4</sub> are green.



**Figure 5.** The accessible surface area of the lysine residues in 15 high resolution crystal structures of lysozyme. Lys116 (highlighted by an ellipse) was on average the most accessible residue. The dimethylated form of Lys116 was selectively bound by sclx<sub>4</sub> (Figures 2 and 4A).



**Figure 6.** (A) Binding distance (see main text for description) between sclx<sub>4</sub> and individual Lys-Me<sub>2</sub> side chains over the course of two MD simulations of 10 ns duration. Time points when the ligand was unbound are off the scale (0-10 Å). (B) Representative snapshots of ligand binding at Lys116-Me<sub>2</sub>. The protein is represented as the C<sup>α</sup> trace with side chains and Lys116-Me<sub>2</sub> shown as lines and sticks, respectively. Refer to the movie (SI) for a more comprehensive view of the binding conformations.