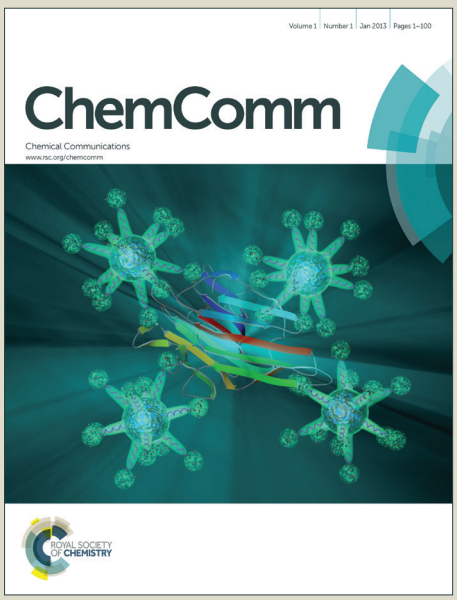
# ChemComm

Accepted Manuscript





ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

# Protein Assembly Mediated by Sulfonatocalix[4]arene

Cite this: DOI: 10.1039/x0xx00000x

Róise E. McGovern,[a] Andrew A. McCarthy,[b] and Peter B. Crowley*[,a]

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

**A crystal structure of lysozyme in complex with *p*-sulfonato-calix[4]arene ($sclx_4$) reveals a linear assembly of protein tetramers glued together by protein-calixarene interactions. One interaction involves encapsulation of the highly exposed *C*-terminal Arg128. The other involves an intricate protein-bound complex of $sclx_4$, $Mg^{2+}$ and a fragment of polyethylene glycol.**

Controlled protein assembly remains a challenging hurdle on the path to nanoscale devices. Current approaches focus on engineering architectures via protein interfaces,[1] binding tags,[2] disulfide bridges,[3] metal co-ordination sites[4] and virus capsid proteins.[5] Valuable progress has been achieved also with small molecule ligands that drive protein-protein interactions.[6] Recently, supramolecular ligands, noted for their ease of synthesis and low cost, have found application in the area of protein assembly.[7,8]

Considering the scale of the protein assembly challenge it is advantageous to have a toolkit of ligands that mediate assembly. Broad-spectrum ligands that recognize common surface features could be used as generic mediators of assembly. The charged residues, especially the cationic side chains of lysine and arginine, stand out as potential targets for generic surface recognition.[8-10] And numerous small molecule receptors have been developed to bind lysine and arginine.[9-15] The highly soluble, symmetric, bowl-shaped and anionic *p*-sulfonatocalix[4]arene[16] ($sclx_4$) has proven to be a particularly versatile ligand for lysine and arginine recognition in water.[8,12,14] We have shown that $sclx_4$ can mediate protein self-assembly *via* lysine binding.[9] In a crystal structure of the cytochrome *c*:$sclx_4$ complex the calixarene was found at interfaces that involved two or more protein chains, suggesting that it functions as "molecular glue". The protein-calixarene contacts were dominated by lysine side chains, bound either inside the cavity or on the outer surface of the calixarene. Having previously characterized the complex of $sclx_4$ and a lysine rich protein (cytochrome *c*; 16 × Lys, 3 × Arg) we sought to investigate how an arginine-rich protein would behave. For this reason we determined the crystal structure of lysozyme (11 × Arg, 6 × Lys) in complex with $sclx_4$. Lysozyme is a well-established model system for protein surface recognition[17] and structural studies of protein-ligand interactions.[18,19]

The presence of $sclx_4$ resulted in the immediate precipitation of lysozyme, thus precluding solution state characterization in water. Precipitation occurred at μM-mM protein concentrations suggesting a relatively high affinity interaction ($K_d$ ~μM). The calixarene-induced precipitation was decreased by the presence of 0.1 M sulfate containing salts, suggesting that sulfate and the sulfonated ligand compete for protein binding. Diffraction-quality crystals of the lysozyme:$sclx_4$ complex were grown from conditions almost identical to those reported for cytochrome *c*.[8] The crystallization drops yielded heavy precipitates (within minutes) from which ~10 μm cubic crystals grew within several days. X-ray diffraction data was collected to 1.7 Å (at Soleil, PROXIMA 1) and the structure was solved by molecular replacement (ESI, Methods). The asymmetric unit was refined with four molecules of lysozyme, five molecules of $sclx_4$, five fragments of polyethylene glycol (PEG) and three $Mg^{2+}$ cations (PDB 4prq, ESI Table S1). The lysozyme molecules form a tetramer in which each monomer buries ~900 $Å^2$ of surface area (Fig. 1A). The core of the tetramer involves residues 79-86 from each monomer positioned around a water-filled channel of ~10 Å diameter. Remarkably, this channel is plugged at either end by a pair of close-packed calixarenes, with their hydroxyl-bearing rims pointing into the channel (Fig. 1B). This close-packing, previously observed in small molecule complexes,[12] brings two sulfonates from each calixarene into van der Waals contact. The resulting accumulation of negative charge is offset partially by salt bridge interactions with the *N*-terminal Lys1 (Fig. 2A). The pairs of calixarenes at the tetramer interfaces (Fig. 1B) are reminiscent of the proposed structure for a calixarene bound to the p53 tetramerization domain.[20]

The structure is further noteworthy in terms of the packing of the lysozyme tetramers and the types of calixarene-complexes that mediate the tetramer-tetramer interactions. Analysis of the crystal packing reveals linear chains of lysozyme tetramers, related by a translation operation along the *a* axis (Fig. 1A). Each calixarene from the close-packed pair is engaged in distinct interactions. One of the calixarenes binds the side chain of Arg128 (Fig. 2A) from a neighbouring lysozyme tetramer. The second calixarene is bound to a $Mg^{2+}$ cation and a PEG fragment, which form a crown ether-like complex[21] (Fig. 1B and 2B). The tetramers are positioned such that

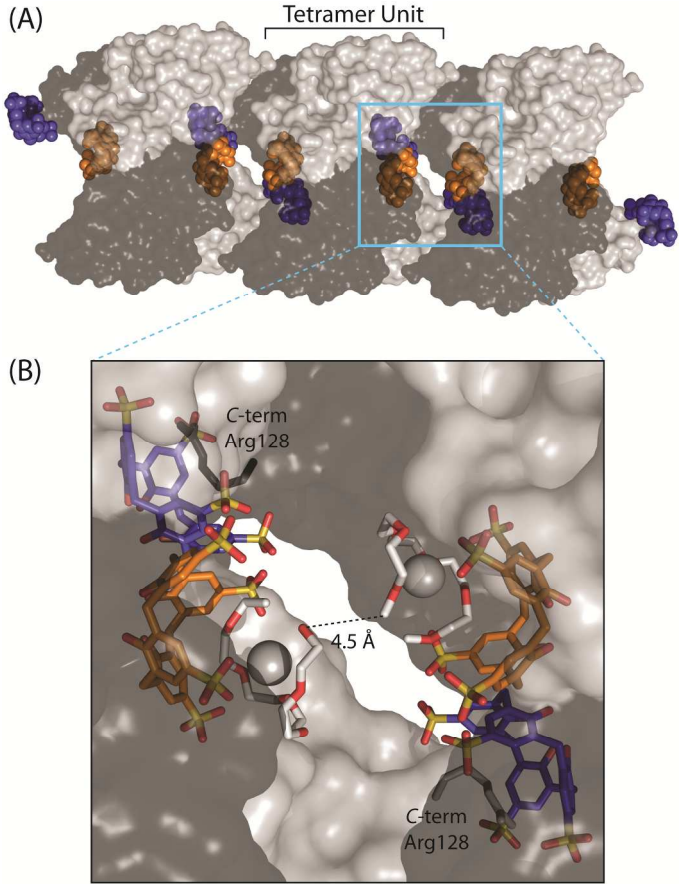



**Fig. 1** The supramolecular architecture in the lysozyme:sclx$_4$ co-crystal. (A) The asymmetric unit comprises a lysozyme tetramer, which assembles into linear chains (three tetramer units shown, obtained by translation along the *a* axis). The proteins, two light grey and two dark grey, are rendered as semi-transparent surfaces. The arginine-binding and the PEG-binding calixarenes are coloured purple and orange, respectively. (B) Detail of the tetramer-tetramer interface highlighting the protein-calixarene contacts and the PEG molecules that thread through sclx$_4$-$Mg^{2+}$ complexes. Two PEG fragments are within van der Waals contact suggesting that the same PEG molecule can interlace adjacent tetramers. The calixarenes, Arg128 and PEG fragments are represented as sticks and the $Mg^{2+}$ cations are grey spheres.

two sclx$_4$-$Mg^{2+}$-PEG complexes oppose each other ($Mg^{2+}$-$Mg^{2+}$ separation of 9.5 Å). Interestingly, the PEG fragments are in van der Waals contact, raising the possibility that the same PEG molecule interlaces two tetramer assemblies (Fig. 1B). Thus, it appears that sclx$_4$ gives rise to chains of lysozyme tetramers, which are held together in part by sclx$_4$-$Mg^{2+}$-PEG complexes.

A fifth sclx$_4$ is bound near the active site in one of the monomers. This calixarene also forms a complex with $Mg^{2+}$ and PEG. Here, the PEG fragment makes van der Waals contacts with the indole rings of the active site residues Trp62 and Trp63 (Fig. S1), confirming the results of an early NMR study of lysozyme-PEG interactions.[22]

Of the 11 possible Arg residues, the *C*-terminal Arg128 was selected for binding by sclx$_4$ (Fig. 2A). It appears that steric accessibility of the side chain is a key determining factor of selectivity. Arg128 was calculated to be the most accessible arginine residue in 15 structures of lysozyme (Fig. 3). Conformational flexibility and the increased accessibility of the *C*-terminus may additionally promote binding at this site. The second most accessible residue, Arg14, is also involved in sclx$_4$ binding (Fig. 2). Arg128 is almost entirely engulfed by the calixarene (Fig. 2A) with ~230 Å$^2$ of protein surface buried upon sclx$_4$ binding. The side chain is planar from $C^{\gamma}$ to the guanidino and sits into the long axis of sclx$_4$, which adopts an elliptical cone conformation. The guanidino points out of the plane of the calixarene's upper rim and forms salt bridge interactions with two of the sulfonates. The side chain conformation is such that the partially cationic $C^{\delta}$ atom is within van der Waals distance (3.8-4.1 Å) of two of the sclx$_4$ phenyl rings, suggesting that cation-π interactions[23] contribute to the binding (The $C^{\gamma}$ also forms short range contacts with three of the phenyl rings). Notably, the upper face of the guanidino remains solvent accessible, though only one water molecule was located here in the crystal structure. Apart from steric effects, the selection of Arg128 is likely to be controlled by charge-charge interactions. Binding at this site involves the insertion of a sclx$_4$ sulfonate into an anion binding pocket on the lysozyme surface, which can accommodate sulfate.[24] This structural detail helps to rationalize our observation that high concentrations of sulfate containing salts reduced the amount of sclx$_4$-induced protein precipitation.

While we currently lack solution state data (due to precipitation of the lysozyme:sclx$_4$ complex in buffer) it is likely that the selection of Arg128 observed in the crystal is representative of what occurs in solution (rather than being a result of crystal packing). Previously, we observed a strong agreement between the crystal structure data and the NMR binding maps for the related complex of cytochrome *c* and sclx$_4$.[8] On the other hand the formation of soluble lysozyme tetramers is less probable as chains of tetramers mediated by sclx$_4$-$Mg^{2+}$-PEG complexes were observed in the crystal.

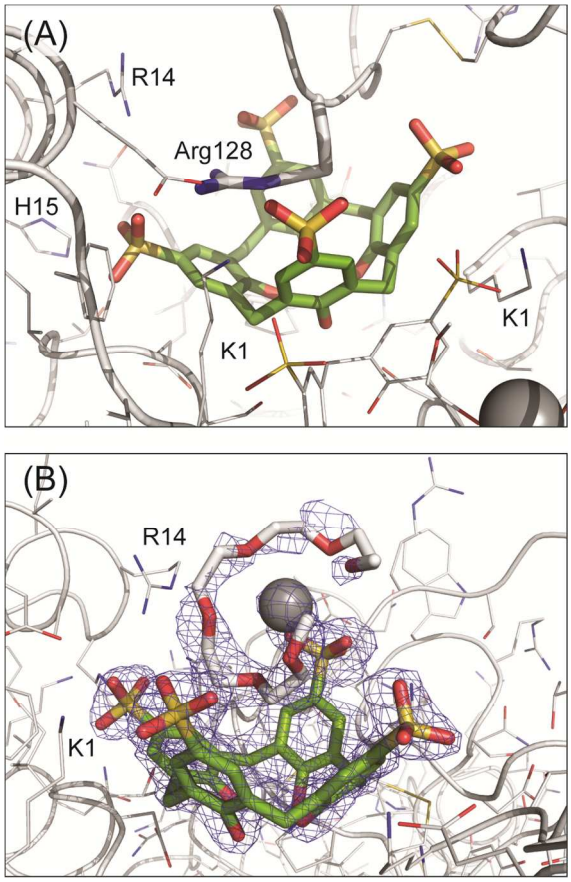



**Fig. 2** Detailed views of (A) the sclx$_4$-Arg128 binding site and (B) the sclx$_4$-$Mg^{2+}$-PEG complex (at the tetramer-tetramer interface, Fig. 1B) showing the $2F_o$-$F_c$ electron density map contoured at 1.0 σ. The sclx$_4$ and the PEG fragment are shown as sticks and the $Mg^{2+}$ is a grey sphere.

In conclusion, small molecule mediated protein assembly is an area of growing interest[7,8] and a repertoire of ligands is necessary to permit protein assembly under different conditions. Here, we have established that $sclx_4$ is an attractive agent to generate protein assemblies via interactions with arginine side chains. Simple considerations of steric accessibility can be used to explain the selectivity of $sclx_4$ for one of eleven Arg residues in lysozyme. The symmetry of $sclx_4$ and the similarity of the *endo*- and *exo*-surfaces facilitate its function as "molecular glue" for protein assembly.[8,20] More generally, crystal structures of protein-$sclx_4$ complexes are useful as they contain information that may benefit our understanding of the interactions between sulfonated-(bio)polymers and cationic proteins.[19] Finally, considering the pivotal roles that arginine plays in protein-protein interfaces,[25] the structure of $sclx_4$ bound to arginine serves as a valuable reference point for the development of interaction inhibitors.

The authors acknowledge NUI Galway (college scholarship to REM, Millennium Fund to PBC) and Science Foundation Ireland (grant 10/RFP/BIC2807 to PBC). We thank the French national synchrotron facility, the staff of beam line PROXIMA 1 (Soleil) and N. P. Power for providing the calixarene.

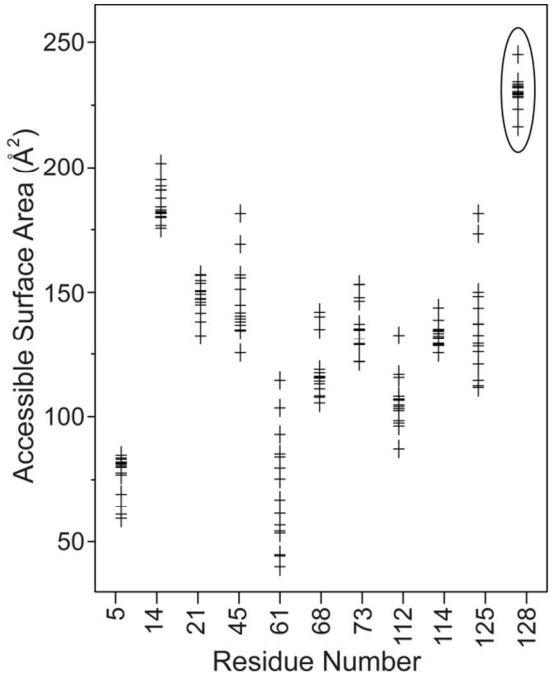



**Fig. 3** The accessible surface area of the arginine residues in 15 high resolution crystal structures of lysozyme. The ellipse highlights Arg128 the most accessible side chain, which was selectively bound by $sclx_4$.

## Notes and references

[a] School of Chemistry, National University of Ireland Galway, University Road, Galway, Ireland. Email: peter.crowley@nuigalway.ie; Tel: +353 91 49 24 80

[b] European Molecular Biology Laboratory and Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, Grenoble Outstation, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France.

† Electronic Supplementary Information (ESI) available: Methods, Table S1 and Fig. S1. See DOI: 10.1039/c000000x/

1 N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. André, T. Gonen, T. O. Yeates and D. Baker, *Science* 2012, **336**, 1171.
2 P. Ringler and G. E. Schulz, *Science* 2003, **302**, 106.
3 D. R. Banatao, D. Cascio, C. S. Crowley, M. R. Fleissner, H. L. Tienson and T. O. Yeates, *Proc. Natl. Acad. Sci. USA* 2006, **103**, 16230l.
4 E. N. Salgado, J. Faraone-Mennella and F. A. Tezcan, *J. Am. Chem. Soc.* 2007, **129**, 13374; P. B. Crowley, P. M. Matias, A. R. Khan, M. Roessle and D. I. Svergun, *Chem. Eur. J.* 2009, **15**, 12672.
5 M. B. van Eldijk, J. C. Y. Wang, I. J. Minten, C. L. Li, A. Zlotnick, R. J. M. Nolte, J. J. L. Cornelissen, and J. C. M. van Hest, *J. Am. Chem. Soc.* 2012, **134**, 18506.
6 D. M. Spencer, T. J. Wandless, S. L. Schreiber and G. R. Crabtree, *Science* 1993, **262**, 1019; Q. Li, C. R. So, A. Fegan, V. Cody, M. Sarikaya, D. A. Vallera and C. R. Wagner, *J. Am. Chem. Soc.* 2010, **132**, 17247; A. Rutkowska and C. Schultz, *Angew. Chem. Int. Ed. Engl.* 2012, **51**, 8166.
7 L. M. Heitmann, A. B. Taylor, P. J. Hart and A. R. Urbach, *J. Am. Chem. Soc.* 2006, **128**, 12574; 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; H. D. Nguyen, D.T. Dang, J.L. van Dongen and L. Brunsveld, *Angew. Chem. Int. Ed. Engl.* 2010, **49**, 895.
8 R. E. McGovern, H. Fernandes, A. R. Khan, N. P. Power and P. B. Crowley, *Nat. Chem.* 2012, **4**, 527
9 M. Fokkens, T. Schrader and F. G. Klärner, *J. Am. Chem. Soc.* 2005, **127**, 14415.
10 D. Bier, R. Rose, K. Bravo-Rodriguez, M. Bartel, J. M. Ramirez-Anguita, S. Dutt, C. Wilch, F. G. Klärner, E. Sanchez-Garcia, T. Schrader and C. Ottmann, *Nat. Chem.* 2013, **5**, 234.
11 Y. Hamuro, M. C. Calama, H. S. Park and A. D. Hamilton, *Angew. Chem. Int. Ed. Engl.* 1997, **36**, 2680.
12 M. Selkti, A. W. Coleman, I. Nicolis, N. Douteau-Guevel, F. Villain, A. Tomas and C. de Rango, *Chem. Commun.* 2000, **2**, 161; A. Lazar, E. Da Silva, A. Navaza, C. Barbey and A. W. Coleman, *Chem. Commun.* 2004, **19**, 2162.
13 C. Renner, J. Piehler and T. Schrader, *J. Am. Chem. Soc.* 2006, **128**, 620; O. Hayashida, N. Ogawa, M. Uchiyama, *J. Am. Chem. Soc.* 2007, **129**, 13698; W. M. Nau, G. Ghale, A. Hennig, H. Bakirci and D. M. Bailey, *J. Am. Chem. Soc*. 2009, **131**, 11558; C. T. Öberg, A. L. Noresson, H. Leffler and U. J. Nilsson. *Chem. Eur. J.* 2011, **17**, 8139; M. Dionisio, G. Oliviero, D. Menozzi, S. Federici, R. M. Yebeutchou, F. P. Schmidtchen, E. Dalcanale and P. Bergese, P. *J. Am. Chem. Soc.* 2012, **134**, 2392.
14 C. S. Beshara, C. E. Jones, K. D. Daze, B. J. Lilgert and F. Hof, *ChemBioChem* 2010, **11**, 63; S. A. Minaker, K. D. Daze, M. C. Ma and F. Hof, *J. Am. Chem. Soc.* 2012, **134**, 11674.
15 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.
16 J. L. Atwood, L. J. Barbour, P. C. Junk and G. W. Orr, *Supramol. Chem.* 1995, **5**, 105; 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. K. Howard and J. W. Steed, *Chem. Eur. J.*, 2011, **17**, 10259.
17 A. J. Wilson, J. Hong, S. Fletcher and A. D. Hamilton, *Org. Biomol. Chem.* 2007, **5**, 276; K. Wenck, S. Koch, C. Renner, W. Sun and T. Schrader, *J. Am. Chem. Soc.* 2007, **129**, 16015; J. Muldoon, A. E. Ashcroft and A. J. Wilson, *Chem. Eur. J.* 2010, **16**, 100

18 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; M. J. Panzner, S. M. Bilinovich, W. J. Youngs and T. C. Leeper, *Chem. Commun.* 2011, **47**, 12479.
19 I. Morfin, E. Buhler, F. Cousin, I. Grillo and F. Boué, *Biomacromolecules* 2011, **12**, 859.
20 S. Gordo, V. Martos, E. Santos, M. Menéndez, C. Bo, E. Giralt, J. de Mendoza, *Proc. Natl. Acad. Sci. USA* 2008, **105**, 16426.
21 A. Drljaca, M .J. Hardie, C. L. Raston and L. Spiccia, *Chem. Eur. J.* 1999, **5**, 2295; S.J. Dalgarno, J. Fisher and C. L. Raston, *Chem. Eur. J.* 2006, **12**, 2772.
22 E. L Furness, A. Ross, T. P. Davis and G. C. King, *Biomaterials* **1998**, *19*, 1361-1369.
23 D. A. Dougherty, *Acc. Chem. Res.* 2013, **46**, 885.
24 R. L. Stanfield, H. Dooley, P. Verdino, M. F. Flajnik and I. A. Wilson. *J. Mol. Biol.* 2007, **367**, 358.
25 P. B. Crowley and A. Golovin, *Proteins* 2005, **59**, 231; J. Janin, R. P. Bahadur and P. Chakrabarti, *Q. Rev. Biophys.* 2008, **41**, 133.

