*ELECTRONIC SUPPLEMENTARY INFORMATION*

# Protein Assembly Mediated by Sulfonatocalix[4]arene

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

[a]School of Chemistry, National University of Ireland Galway, Ireland

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

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

## Methods

**Co-crystallization of lysozyme and $sclx_4$.** Hen white egg lysozyme (62971 Fluka) was purchased from Sigma-Aldrich and used without further purification. The hanging drop vapour diffusion method was used for crystallization at 20° C. Co-crystals of lysozyme and $sclx_4$ were grown from identical conditions to those reported for cytochrome *c*[1] except that an oil barrier was not used. Drops were prepared by combining 1 μL volumes of lysozyme (1.4 mM), $sclx_4$ (17 mM) and the reservoir solution (20-30 % polyethylene glycol (PEG) 8000, 50 mM NaCl, 100 mM $MgCl_2$ and 50 mM sodium cacodylate pH 6.3). Premixing the protein and ligand was not possible as this resulted in complete precipitation.

**X-ray Diffraction and Structure Determination.** Crystals were transferred to reservoir solution supplemented with 25 % glycerol and flash-frozen under a stream of nitrogen gas at 100 K (X-stream 2000). Diffraction data were collected from a single crystal of the lysozyme:$sclx_4$ complex at Soleil (PROXIMA 1, Pilatus 6M detector, $\varphi$ scans of 0.1° over 180° to a resolution of 1.72 Å). Data processing and scaling were performed in MOSFLM[2] and SCALA,[3] respectively. The data collection and refinement statistics are given in Table S1. The structure was solved by molecular replacement in PHASER.[4] Refinement and manual rebuilding were performed in REFMAC5 as implemented in CCP4[5] and COOT,[6] respectively. Solvent molecules were placed automatically using ARP/wARP[7] and refinement was continued until no features remained in the $F_o - F_c$ difference maps. Molprobity[8] was used to check the structure quality. Coordinates and structure factors were deposited in the Protein Data Bank with the accession code 4PRQ. The protein-ligand and protein-protein interfaces were analysed in COOT and PISA.[9] Interface areas [the inaccessible surface ($Å^2$) of the protein or ligand in the complex] were calculated as described previously.[1]

**Arginine solvent accessibility in lysozyme.** 15 high resolution (0.9-2.1 Å) crystal structures of hen egg white lysozyme were analyzed using the Accessible Surface Areas calculation in CCP4.[4] Each file (PDB codes: 194L, 1GWD, 1JIS, 1LPI, 1YIK, 2CDS, 2D4K, 2FBB, 2I25, 2ZQ3, 3A67, 3AGH, 3AW7, 3LZT, 4J1A) was manually edited to contain a single copy of lysozyme. Other proteins, water molecules, ions and alternate conformations were removed prior to the calculations.

**Table S1.** Summary of crystallization conditions, data collection and refinement statistics

| | |
|---|---|
| *Crystallization Conditions*[a] | |
| [protein], [$sclx_4$] (mM) | 1.4, 17 |
| PEG 8,000 (%) | 24 |
| Buffer, $(CH_3)_2AsO_2Na$ | 0.05 M, pH 6.3 |
| Salts | 0.05 M NaCl, 0.1 M $MgCl_2$ |
| *Data Collection*[b] | |
| Space group | $P2_1$ |
| Cell constants | $a$ = 44.01 Å<br>$b$ = 81.75 Å<br>$c$ = 72.10 Å<br>$\alpha = \gamma = 90^o$<br>$\beta = 105.37^o$ |
| Resolution (Å) | 81.75-1.72 (1.77-1.72) |
| Wavelength (Å) | 1.00792 |
| Unique reflections | 1386759 (51142) |
| Multiplicity | 3.0 (2.6) |
| $I/\sigma$ | 9.4 (1.8) |
| Completeness (%) | 98.7 (94.9) |
| $R_{merge}$[c] (%) | 7.3 (55.6) |
| Solvent content (%) | 43.86 |
| *Refinement* | |
| $R_{factor}$ (%) | 18.48 |
| $R_{free}$ (%) | 22.42 |
| rmsd[d] bonds (Å) | 0.01 |
| rmsd angles (°) | 1.12 |
| # molecules in asymmetric unit | |
| Protein | 4 |
| $sclx_4$ | 5 |
| PEG | 5 |
| $Mg^{2+}$ | 3 |
| Solvent | 322 |
| Average $B$ factors[e] (Å$^2$) | |
| Protein | 21.51 |
| $sclx_4$ | 19.70 |
| PEG | 45.46 |
| $Mg^{2+}$ | 28.88 |
| Solvent | 28.44 |
| Ramachandran analysis[f] | |
| % residues (favoured regions) | 98.8 |
| % residues (allowed regions) | 100.0 |

[a]The crystallization drops comprised 1 μL each of the protein, ligand and reservoir solutions; [b]Values in parentheses correspond to the highest resolution shell; [c]$R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$; [d]root mean square deviation; [e]calculated from the B values of all non-hydrogen atoms; [f]calculated with Molprobity.

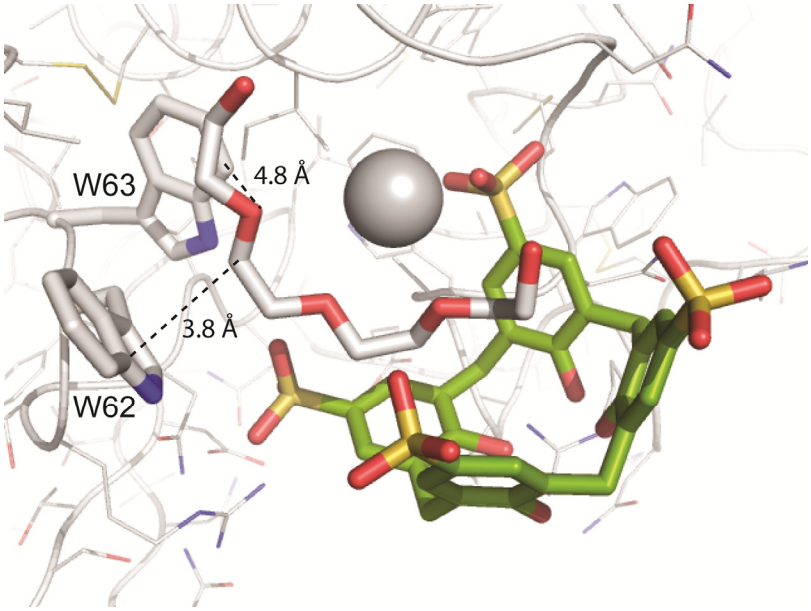


**Fig S1.** A calixarene bound near the active site in chain B. This calixarene is complexed with $Mg^{2+}$ and a fragment of PEG (in a crown ether like conformation). The PEG also makes van der Waals contacts with the indole rings of the active site residues Trp62 and Trp63.

**References**


1. R. E. McGovern, H. Fernandes, A. R. Khan, N. P. Power and P. B. Crowley, *Nat. Chem.* 2012, **4**, 527.
2. A. G. W. Leslie, Recent changes to the MOSFLM package for processing film and image plate data; Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography **1992**.
3. P. Evans, *Acta Crystallogr. D.* 2006, **62**, 72.
4. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read. *J. Appl. Crystallogr.* 2007, **40**, 658.
5. Collaborative Computational Project, N. *The CCP4 suite: programs for protein crystallography.*; *Acta Crystallogr. D* 1994, **50**, 760.
6. P. Emsley, and K. Cowtan, *Acta Crystallogr. D.* 2004, **60**, 2126.
7. E. J. van Asselt, A. Perrakis, K. H. Kalk, V. S. Lamzin and B. W. Dijkstra, *Acta Crystallogr. D.* 1998, **54**, 58.
8. 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.
9. E. Krissinel, and K. Henrick, *J. Mol. Biol.* 2007, **372**, 774.