Electronic Supplementary Material (ESI) for ChemComm.

SUPPORTING INFORMATION

Engineered Assembly of a Protein – Cucurbituril Biohybrid

Francesca Guagnini,[a] Sylvain Engilberge,[a] Kiefer O. Ramberg,[a] Javier Pérez[b] and Peter B. Crowley[a]*

[a]School of Chemistry, National University of Ireland Galway, H91 TK33, Ireland

[b]Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP48, 91192 Gif-sur-Yvette Cedex, France

*Correspondence to:

peter.crowley@nuigalway.ie

+353 91 49 24 80

Contents

Experimental

Cucurbit[7]uril. Stock solutions of **Q7** (SIGMA 545201) were prepared in 100 mM NaCl at pH 6.0. The concentration was determined by UV/vis titration with cobaltocenium.[1]

Protein preparation. A modified pET25rsl vector encoding the double mutant N79K/T82Y was produced and sequence-verified by Genscript. Unlabelled and ^{15}N-labelled proteins were produced in *E. coli* BL21. Proteins were purified by mannose-affinity chromatography, using an elution buffer that contained 0.1 M D-fructose.[2,3] Protein samples were dialyzed extensively against water to remove D-fructose. Dimethylation was achieved by using formaldehyde and dimethylaminoborane complex.[4] Mass analysis of the proteins was performed on an Agilent 640 Triple Quadrupole LC/MS and verified complete dimethylation of all 4 lysines and the N-terminus (Table S1).

NMR spectroscopy. ^{1}H-^{15}N HSQC monitored titrations were performed at 30 °C on a 600 MHz Varian spectrometer equipped with a HCN coldprobe, as described.[5] Samples of 0.1 or 1.0 mM ^{15}N-labelled RSLex* were titrated with µL aliquots of a 25 mM **Q7** stock. Spectra were processed in NMRPipe[6] and analysed in CCPN.[7]

Precipitation Tests. Hanging drops were set up in 24 well plates with siliconized glass cover slips at 293 K. The wells contained 30 % PEG 8000 and 100 mM sodium acetate at pH 4.5. Drops were prepared by mixing 1 µl of protein stock solution, 1 µl of **Q7** stock solution and 1 µl of reservoir solution. The stock solutions were prepared to achieve the final concentrations reported in Scheme S1. Each condition was tested in triplicate and was considered positive if precipitate occurred in at least two drops within 1 hour.

Size exclusion chromatography coupled to multi-angle light scattering. SEC-MALS experiments were performed at room temperature on an Äkta Purifier equipped with a Superdex 75, GL 10/300 column (GE Healthcare). Sample elution was monitored by using a multi-angle light scattering detector (miniDAWN TREOS, Wyatt) and a differential refractive index detector (Optilab T-rEX, Wyatt). Filtered, degassed buffer (20 mM sodium acetate, 50 mM NaCl, pH 5.5) was pumped at a constant flow rate of 0.5 mL/min. The samples (300 µL) contained RSLex* (0.1 - 0.3 mM) in the presence of 0 - 10 eq. **Q7** and were centrifuged at 20000 x g for 15 min immediately prior to injection. The mass average molar mass across the elution peak was determined in ASTRA 6 software (Wyatt Technology).

SAXS characterization. Data were collected at the SWING beamline (SOLEIL synchrotron) using the direct injection mode. All samples were exchanged previously into 20 mM sodium acetate, 50 mM NaCl, 5 mM D-fructose, pH 5.6 *via* size exclusion chromatography. Scattering frames were collected with an exposure of 0.99 s (dead time = 0.01 s) on an Eiger 4M detector (Dectris) and integrated in FOXTROT (SOLEIL Synchrotron). Scattering frames from buffer alone and **Q7** in buffer were subtracted from the scattering of protein – **Q7** mixtures. RSL* or RSLex* were characterized at four concentrations (0.125, 0.25, 0.5, 1 mM) using the automated sample changer. The effect of **Q7** complexation was tested at 0.125 mM protein and varying concentrations of ligand. CRYSOL[8] was used to calculate the scattering of the RSL trimer.

Crystallization. Crystals were obtained by using an Oryx 8 robot (Douglas Instruments) and the JCSG++ sparse matrix screen (Jena Bioscience). A solution containing 1 mM RSLex*, 1 mM **Q7** and 12 mM α-methylfucoside was combined with the screen in sitting drop vapour diffusion format (3-drop MRC crystallization plate). Condition D4, containing 30 % PEG 8000, 100 mM sodium acetate pH 4.5 and 200 mM lithium sulfate, yielded hexagonal crystals of ~50 µm dimension. Crystals of RSLex* devoid of **Q7** were obtained from a solution containing 1 mM RSLex* and 1 mM **Q7** in condition H8, containing 25 % PEG 3350, 100 mM BIS-TRIS pH 5.5 and 200 mM sodium chloride. Crystals were transferred to their respective reservoir solutions supplemented with 20 % glycerol and cryo-cooled in liquid N_2. Diffraction data were collected to 1.9 and 1.3 Å resolution for RSLex* – **Q7** and RSLex*, respectively at beamline PROXIMA-2A (SOLEIL synchrotron).

X-ray structure determination. Diffraction frames were processed using the autoPROC pipeline.[9] Data were integrated in XDS[10] and the integrated intensities were scaled and merged in AIMLESS[11] and POINTLESS[12] in CCP4. Crystal pathologies were assessed in POINTLESS and phenix.Xtriage.[13] The structures were solved by molecular replacement in Phaser,[14] using 2BT9-polyalanine as the search model. Iterative model building in COOT[15] and refinement in BUSTER[16] was performed. At all stages of refinement, translation-liberation-screw-rotation (TLS) and automatic water placement were applied. Refinement statistics are reported in Table S2. Structures and associated structure factor amplitudes were deposited in the Protein Data Bank under the codes reported in Table S2 after validation in MolProbity.[17] Crystal packing interfaces were analysed in PDBe-PISA.[18]

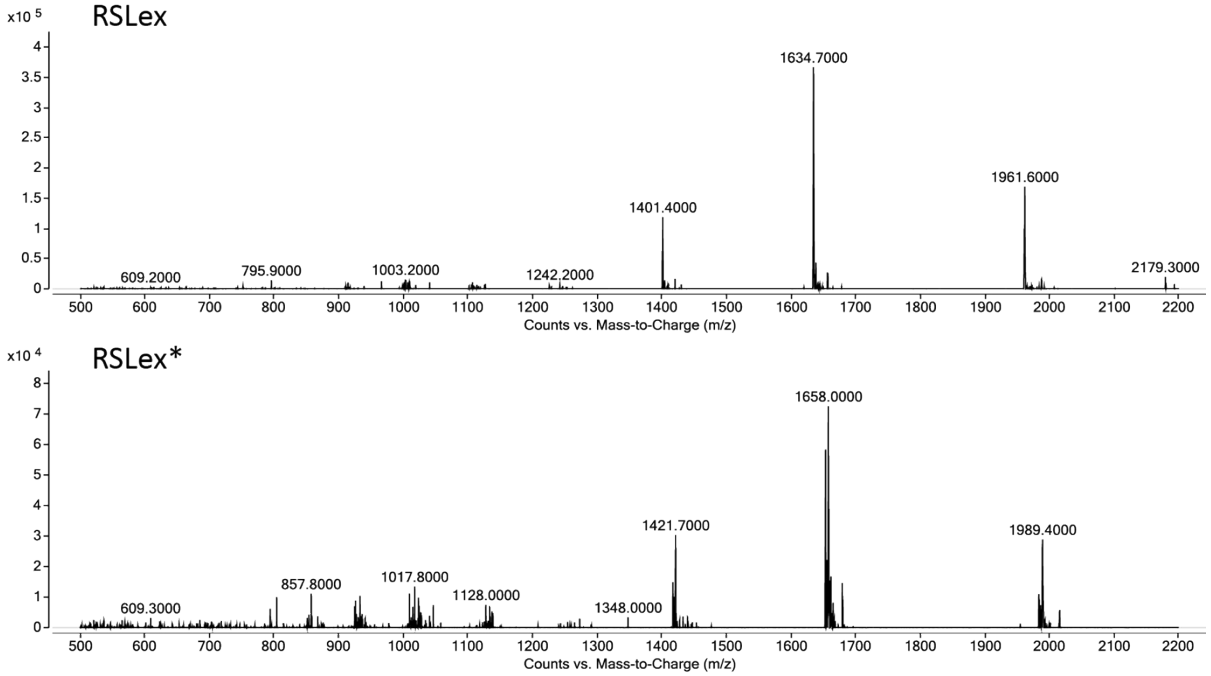


Figure S1. ESI^+ mass spectra for RSLex and RSLex*.

Table S1. Predicted and measured masses from ESI$^+$ mass spectra (Figure S1) using ESI-Prot.[19]

RSLex monomer

m/z	charge	Molecular Weight (Da)	Error (Da)
1401.4	7+	9802.7	0.1
1634.7	6+	9802.2	-0.5
1961.6	5+	9803.0	0.3
Deconvoluted MW (Da)		9802.6	
Predicted MW (Da)		9802.8	

RSLex* monomer

m/z	charge	Molecular Weight (Da)	Error (Da)
1421.7	7+	9943.4	0.5
1658.0	6+	9942.0	-1.0
1989.4	5+	9943.5	0.5
Deconvoluted MW (Da)		9943.0	
Predicted MW (Da)		9942.0	

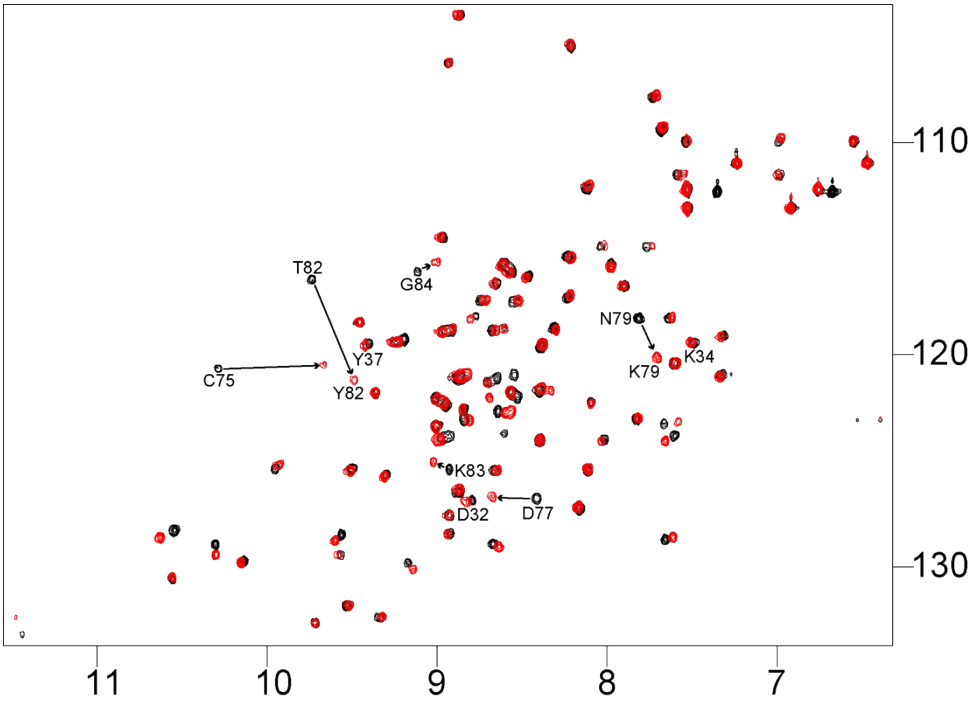


Figure S2. The overlaid ^{1}H-^{15}N HSQC spectra of RSL* (black) and RSLex* (red contours).

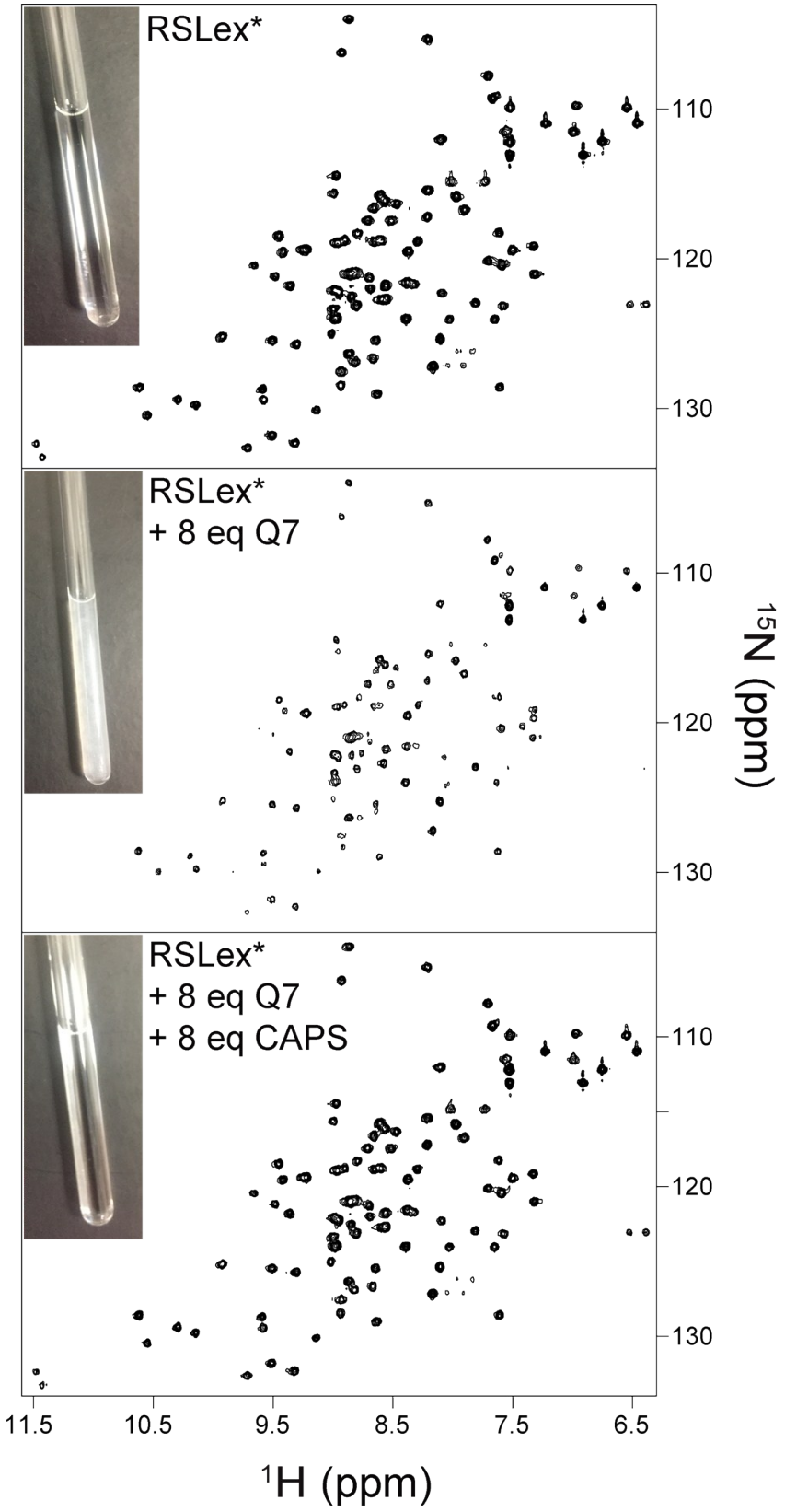
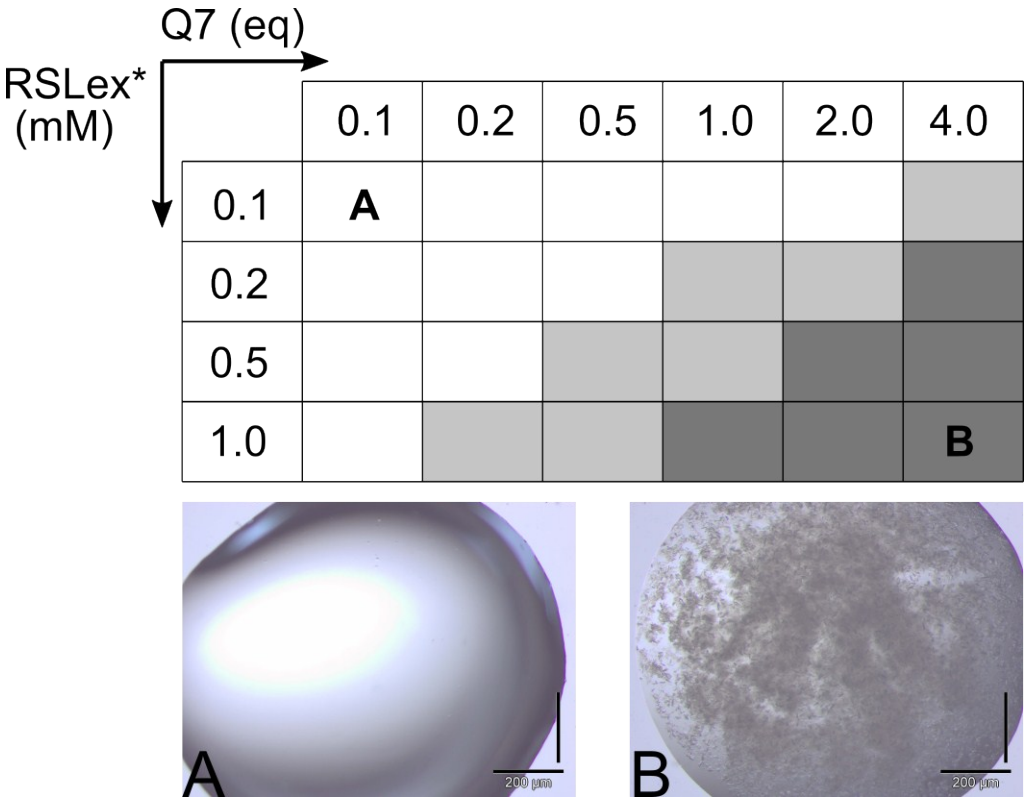


Figure S3. NMR signal loss due to precipitation of 0.3 mM RSLex* at 8 eq **Q7** and signal recovery upon addition of CAPS, a high-affinity competitor. Loss of signal was accompanied by visible aggregation. Addition of 8 eq CAPS resulted in immediate dissolution of the precipitate. All three HSQC spectra are contoured identically.



Scheme S1. Precipitation test performed at 0.1 - 1 mM RSLex* and 0.1 - 4 eq **Q7**. White, light grey and dark grey boxes represent clear drops, light and heavy precipitate, respectively. Representative images of **(A)** clear drops and **(B)** heavy precipitate.

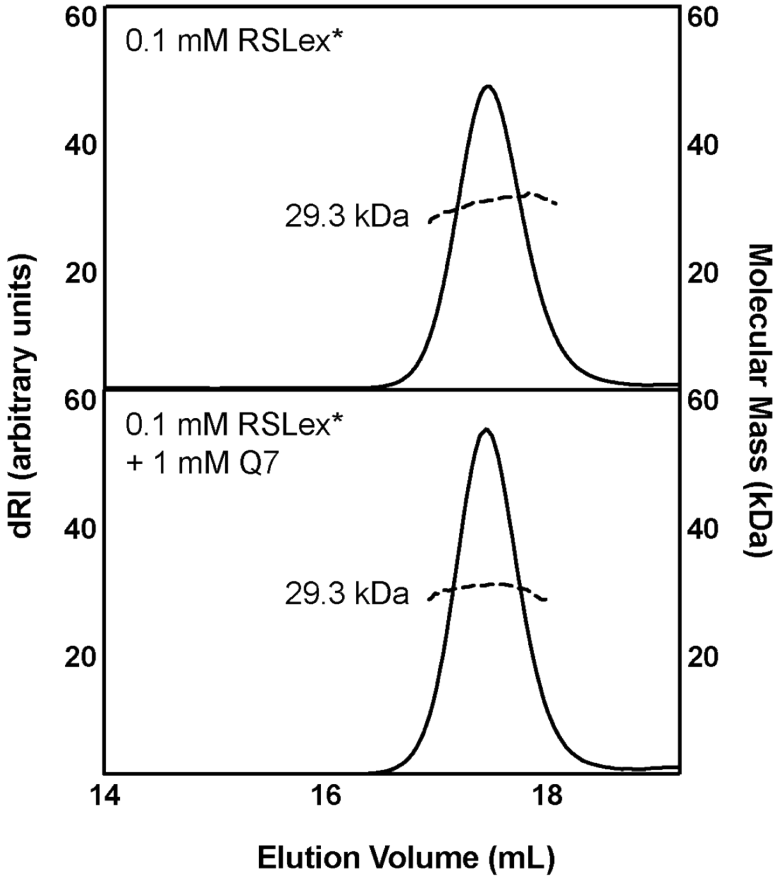


Figure S4. SEC-MALS investigation into **Q7**-induced assembly of RSLex*. Dashed and solid lines correspond to mass average molar masses and differential refractive index (dRI) elution profiles, respectively.

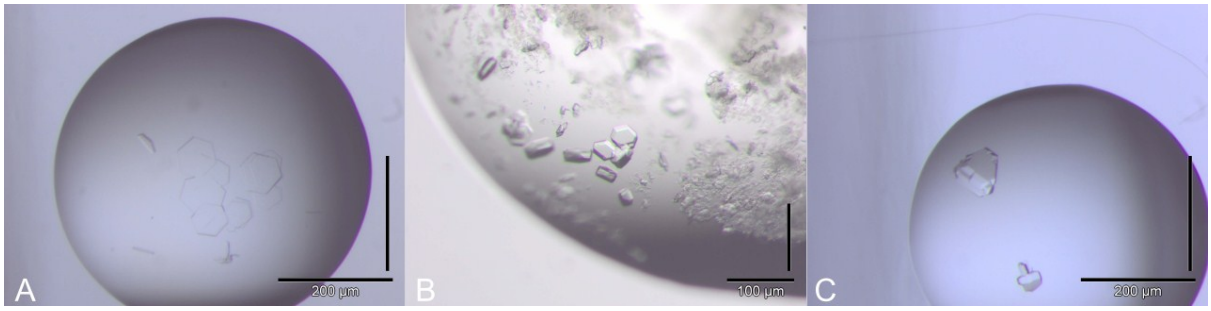


Figure S5. (A) Crystals of the RSLex* – **Q7** complex obtained from condition D4 in JSCG++ Jena Bioscience. **(B)** Representative crystals obtained in hanging drop experiments of RSLex* – **Q7** in 35 % PEG 8000, 100 mM sodium acetate pH 4.5, 200 mM lithium sulfate. **(C)** Crystals of RSLex* in 25 % PEG 3350, 100 mM BIS-TRIS pH 5.5, 200 mM NaCl.

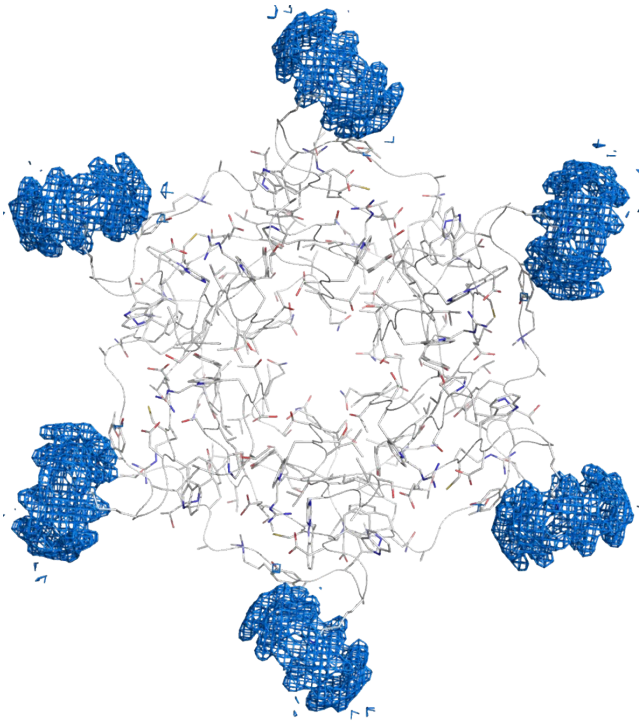


Figure S6. 2Fo - Fc electron density map at 1.9 Å resolution (contoured at 1.5 σ) showing **Q7** bound to the RSLex* trimer.

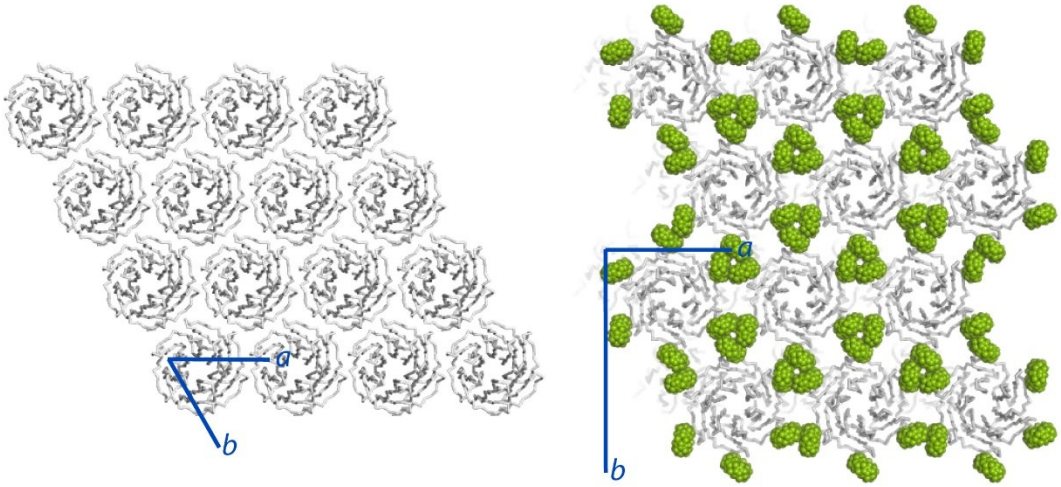


Figure S7. Comparison of RSLex* and RSLex* – **Q7** crystal packing in the *ab* plane. The honeycomb assembly (~30 % solvent content) is expanded to accommodate the **Q7** clusters, resulting in a solvent content of ~55 %. Proteins shown as grey ribbons, **Q7** as green spheres and unit cell axes in blue.

Table S2. X-ray data collection, processing and refinement statistics.

Data Collection		
Light source	SOLEIL, PROXIMA-2A	
Wavelength (Å)	0.98013	
Structure	RSLex*	RSLex* – **Q7**
Space group	$P3_2$	$C121$
Cell constants	43.94, 43.94, 207.27 Å	54.63, 94.53, 149.64 Å
	90.00, 90.00, 120.00 °	90.00, 90.06, 90.00 °
Resolution (Å)	69.09 - 1.13 (1.15 - 1.13)	149.64 - 1.98 (2.01 - 1.98)
# reflections	1539737 (40729)	338138 (16198)
# unique reflections	163694 (8000)	52443 (2608)
Multiplicity	9.4 (5.1)	6.4 (6.2)
I/σ (I)	8.7 (2.2)	6.0 (2.2)
Completeness (%)	99.8 (96.8)	99.4 (98.4)
R_{meas}[b] (%)	17.9 (85.4)	22.6 (89.8)
R_{pim}[c] (%)	5.7 (36.3)	8.9 (35.8)
$CC_{1/2}$	99.2 (63.7)	98.7 (90.0)
Solvent content (%)	30	55
Refinement		
R_{work}	17.2	24.7
R_{free}	19.3	31.1
rmsd bonds (Å)	0.010	0.010
rmsd angles (°)	1.13	0.95
# molecules in asymmetric unit		
RSL monomer	6	6
Q7	0	12
Water	811	279
Ave. B-factor (Å^2)	13.9	46.0
Clashscore	3.5	9.0
Ramachandran analysis,[d] % residues in		
favoured regions	99.1	94.2
allowed regions	0.9	5.1
PDB code	6STZ	6SU0

[a]Values in parentheses correspond to the highest resolution shell [b]$R_{meas} = \sum_{hkl} \sqrt{(n/n-1)} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$; [c]$R_{pim} = \sum_{hkl} \sqrt{(1/n-1)} \sum^n_{i=1} |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$; [d]Calculated in MolProbity.

References

1. S. Yi and A. E. Kaifer, *J. Org. Chem.*, 2011, **76**, 10275.
2. N. Kostlánová, E. P. Mitchell, H. Lortat-Jacob, S. Oscarson, M. Lahmann, N. Gilboa-Garber, G. Chambat, M. Wimmerová and A. Imberty, *J. Biol. Chem.*, 2005, **280**, 27839.
3. P. M. Antonik, A. N. Volkov, U. N. Broder, D. Lo Re, N. A. J. van Nuland and P. B. Crowley, *Biochemistry*, 2016, **55**, 11953.
4. S. T. Larda, M. P. Bokoch, F. Evanics, R. S. Prosser, *J. Biomol. NMR*, 2012, **54**, 199.
5. F. Guagnini, P. M. Antonik, M. L. Rennie, P. O'Byrne, A. R. Khan, R. Pinalli, E. Dalcanale and P. B. Crowley, *Angew. Chem. Int. Ed.*, 2018, **57**, 7126.
6. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, *J. Biomol. NMR*, 1995, **6**, 277.
7. W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins*, 2005, **59**, 687.
8. D. I. Svergun, C. Barberato and M. H. J. J. Koch, *J. Appl. Cryst.*, 1995, **28**, 768.
9. C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, O. S. Smart, W. Paciorek, T. Womack and G. Bricogne, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2011, **D67**, 293.
10. W. Kabsch, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **D66**, 125.
11. P. R. Evans and G. N. Murshudov, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2013, **D69**, 1204.
12. P. R. Evans, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2011, **D67**, 282.
13. P. D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger and P.H. Zwart, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **D66**, 213.
14. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. App. Cryst.*, 2007, **40**, 658.
15. P. Emsley and K. Cowtan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2004, **D60**, 2126-2132.
16. G. Bricogne, E. Blanc, M. Brandl, C. Flensburg, P. Keller, W. Paciorek, P. Roversi, A. Sharff, O. S. Smart, C. Vonrhein and T. O. Womack, 2011, BUSTER version X.Y.Z. Cambridge, United Kingdom: Global Phasing Ltd.
17. V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy DA, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **D66**, 12.
18. E. Krissinel and K. J. Henrick, *J. Mol. Biol.*, 2007, **372**, 774.
19. R. Winkler, *Rapid Commun. Mass Spectrom.,* 2010, **24**, 285.