Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

SUPPORTING INFORMATION

Protein Recognition by Cucurbit[6]uril: High Affinity N-terminal Complexation

Kiefer O. Ramberg, Sylvain Engilberge, Francesca Guagnini, and Peter B. Crowley*

School of Chemistry, National University of Ireland Galway, University Road, Galway, H91 TK33, Ireland.

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

Contents

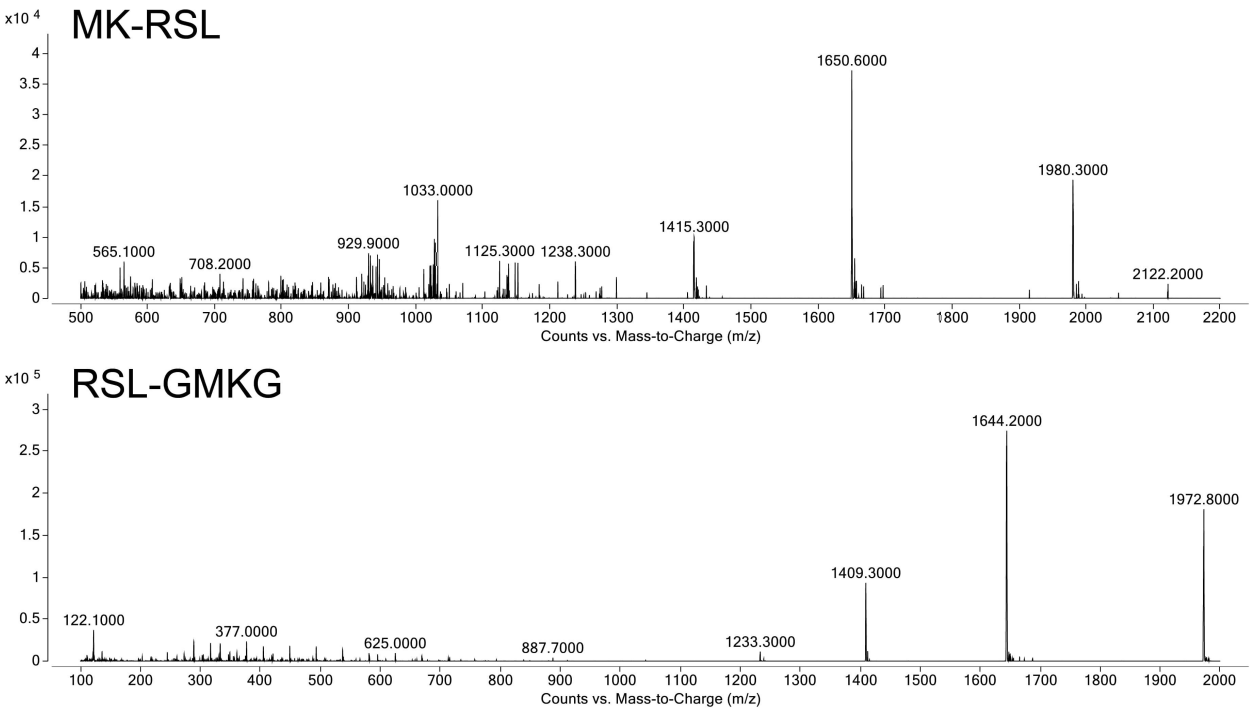

Figure S1. ESI$^+$ mass spectra for MK-RSL and RSL-GMKG.

Table S1. Predicted and measured mass from ESI$^+$ mass spectra using ESI-Prot.[1]

MK-RSL monomer			
m/z	**charge**	**Molecular Weight (Da)**	**Error (Da)**
1415.3	9+	9900.0	2.0
1650.6	8+	9897.6	-0.5
1980.3	7+	9896.5	-1.6
Predicted MW (Da)		9898.9	
Deconvoluted MW (Da)		9898.0	
Standard deviation (Da)		1.8	

RSL-GMKG monomer			
m/z	**charge**	**Molecular Weight (Da)**	**Error (Da)**
1409.3	7+	9858.0	-0.7
1644.2	6+	9859.2	0.4
1972.8	5+	9859.0	0.2
Predicted MW (Da)		9857.8	
Deconvoluted MW (Da)		9858.7	
Standard deviation (Da)		0.6	

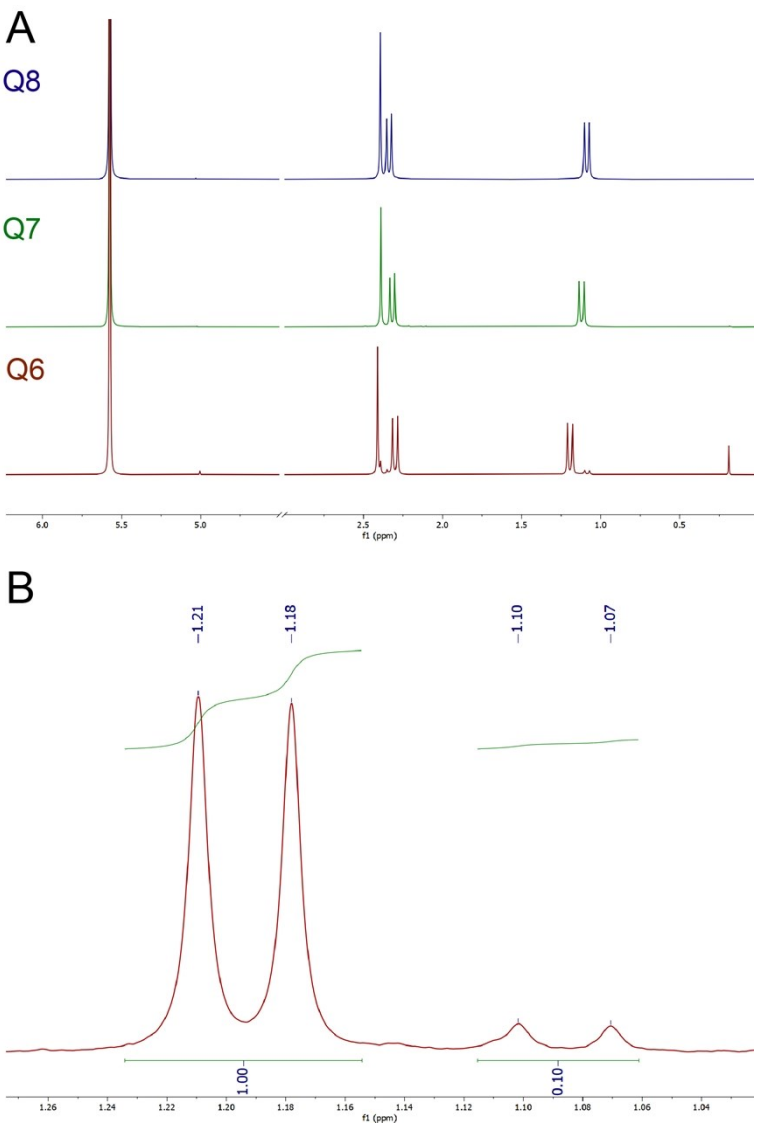

Figure S2. (A) ^{1}H NMR spectra of **Q6**, **Q7**, and **Q8** in 0.1 M HCl. **(B)** Integration of the doublets at 1.1-1.2 ppm suggests that the **Q6** sample contains ~10 % **Q8**.

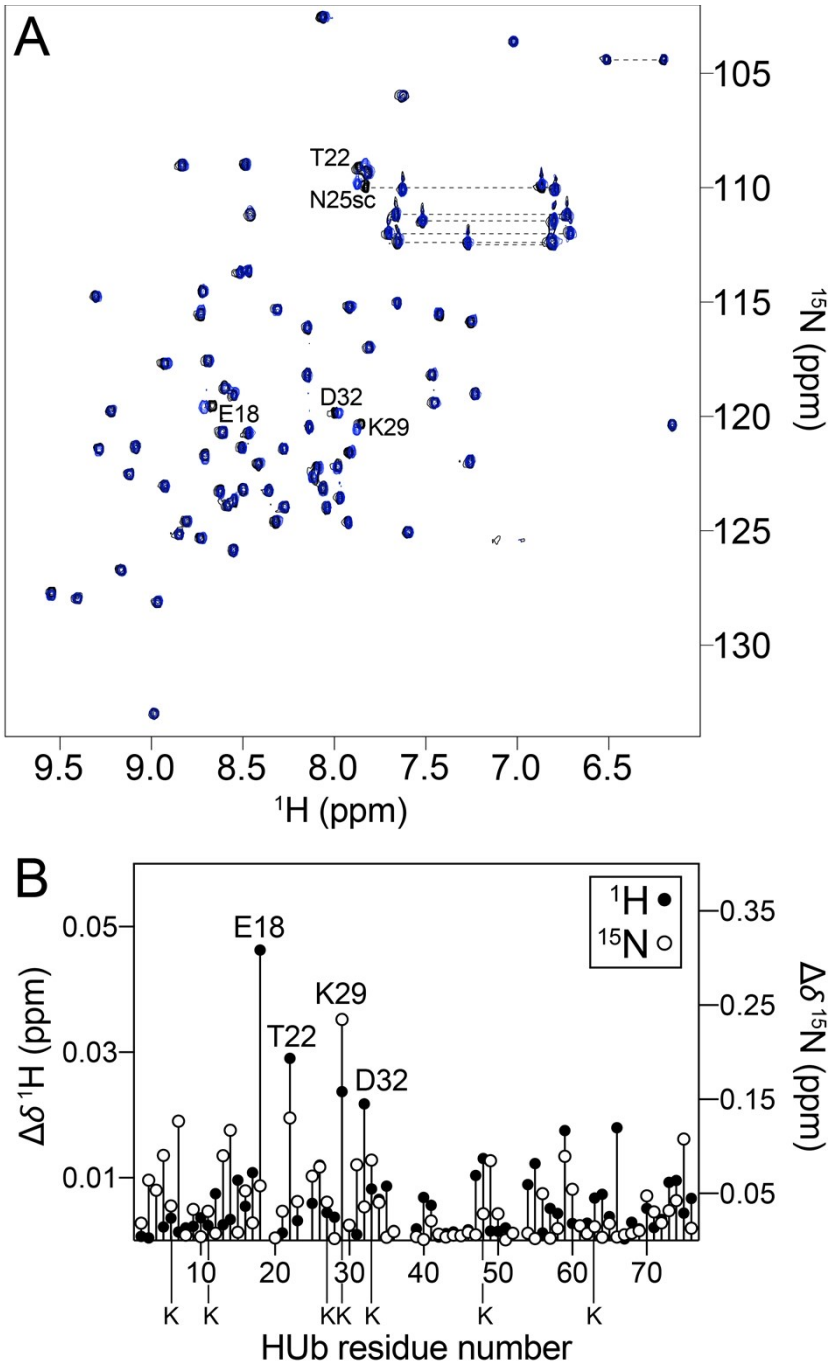

Figure S3. (A) Overlaid ^{1}H-^{15}N HSQC spectra of HUb in the absence (black contours) and presence of **Q6** (blue contours) in 20 mM potassium phosphate pH 6.0. **(B) Q6**-induced chemical shift perturbation plot of HUb backbone amides.

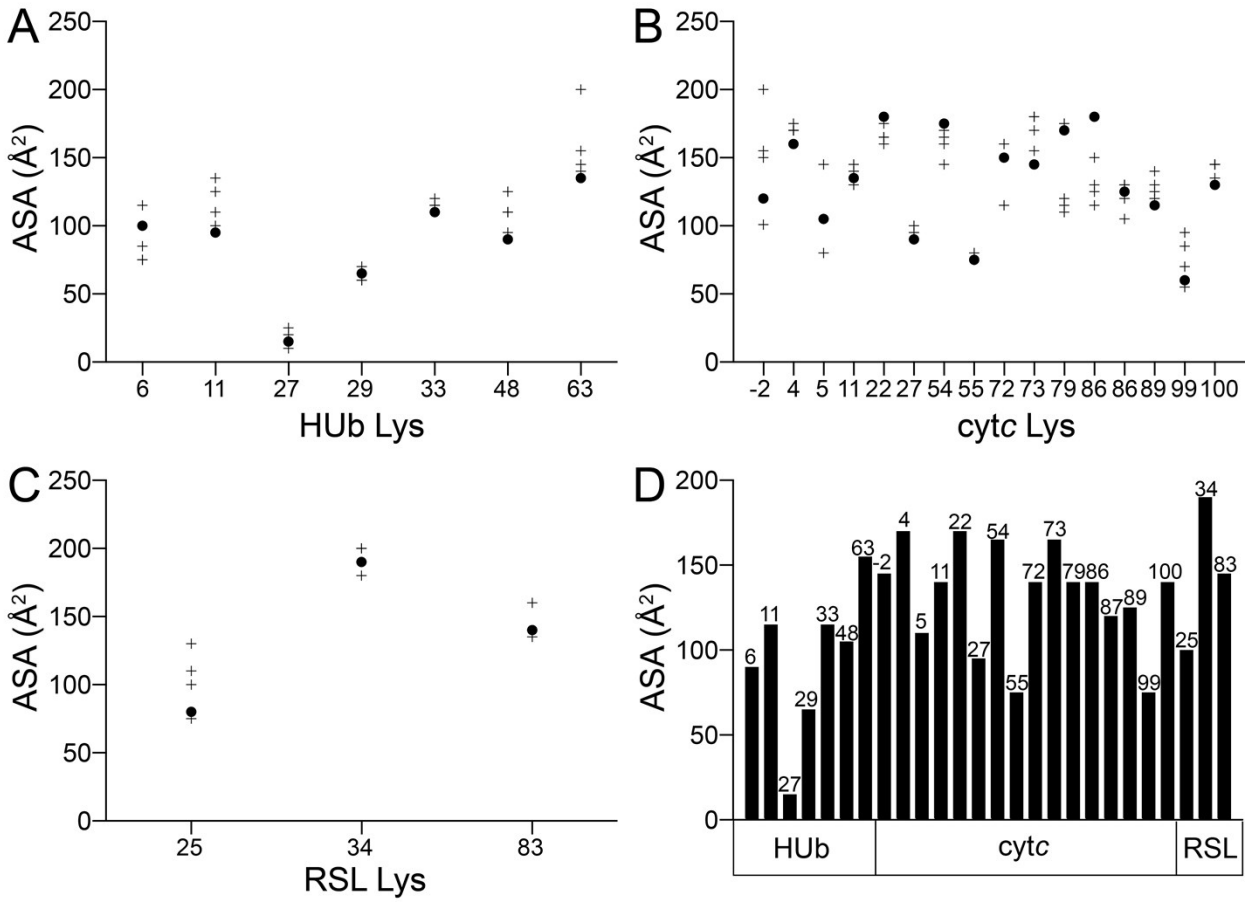

Figure S4. (A-C) Accessible surface areas (ASA) for lysines of HUb, cyt*c*, and RSL obtained from high-resolution crystal structures. The PDB entries analysed included 1ubq, 1ogw, 3nhe, 3ns8, 4xof (HUb), 1ycc, 3cx5, 4n0k, 5cic, 5t8w (cyt*c*), and 2bs5, 2bs6, 2bt9, 3zi8, 4i6s (RSL). Data from entries 1ubq, 1ycc, and 2bs5 are shown as black circles. **(D)** Average ASAs of lysines in the three proteins derived from data in **A**-**C**. Lys residue number indicated atop data bars. SAMP2 is not included as the available data are limited by the flexibility / disorder of this protein (PDBs 5lda and 1sf0).

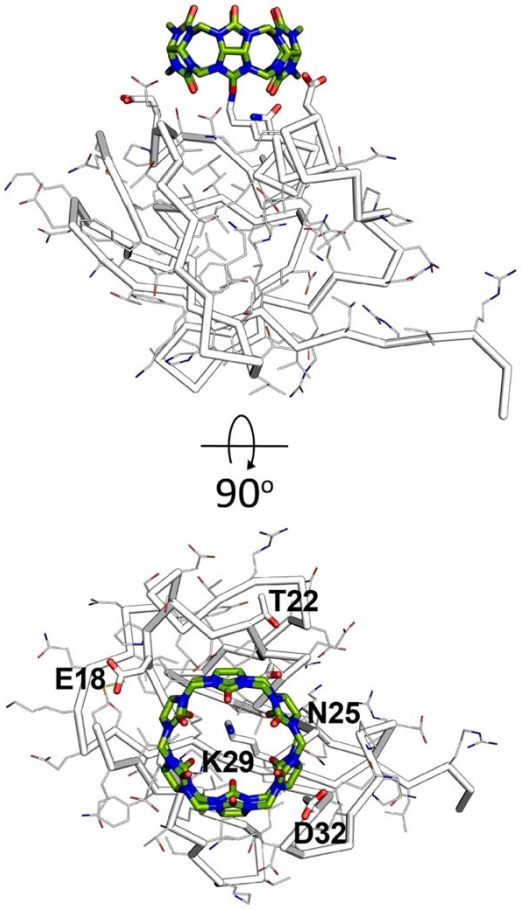

Figure S5. (A) HUb-**Q6** model from top-scoring cluster in HADDOCK.[2]

Table S2. Statistics for HUb-**Q6** clusters determined in HADDOCK.[2]

	Cluster 1	**Cluster 2**
HADDOCK score	-36.0 ± 0.3	-33.1 ± 0.7
Cluster size[a]	66	134
RMSD (Å)[b]	0.1 ± 0.1	0.5 ± 0.0
vdW energy (kcal/mol)	-24.3 ± 0.3	-22.5 ± 0.2
Electrostatic energy (kcal/mol)	-194.4 ± 2.0	-182.1 ± 2.7
Desolvation energy (kcal/mol)	7.7 ± 0.2	7.6 ± 0.3
Restraints violation energy (kcal/mol)	0.0 ± 0.0	0.0 ± 0.0
Buried surface area (Å²)	572.7 ± 7.9	565.5 ± 7.7

[a]HADDOCK clustered 200 structures in 2 clusters only.

[b]From the overall lowest-energy structure.

Table S3. X-K-X motifs in HUb, RSL, cyt*c*, and SAMP2.*

Protein	Lys	X_1–K–X_2
HUb	6	V–K–T
	11	G–K–T
	27	V–K–A
	29	**A–K–I**
	33	D–K–E
	48	G–K–Q
	63	Q–K–E
RSL	25	G–K–I
	34	**G–K–G**
	83	T–K–G
cyt*c*	-2	F–K–A
	4	A–K–K
	5	K–K–G
	11	F–K–T
	22	E–K–G
	27	H–K–V
	54	I–K–K
	55	K–K–E
	72	P–K–K
	73	K–K–Y
	79	T–K–M
	86	L–K–K
	87	K–K–E
	89	E–K–D
	99	L–K–K
	100	K–K–A
SAMP2[a]	**2**	**M–K–M**
	5	I–K–V
	7	V–K–V
	15	E–K–E
	24	M–K–V
	42	A–K–V
	46	G–K–V
	55	V–K–D

***Q6** binding sites are in bold. Buried hydrophobic residues are highlighted grey.

[a]The N-terminus of SAMP2 (residues 1-5) is disordered. **Q6** complexation may occur at Met1 or Lys2.

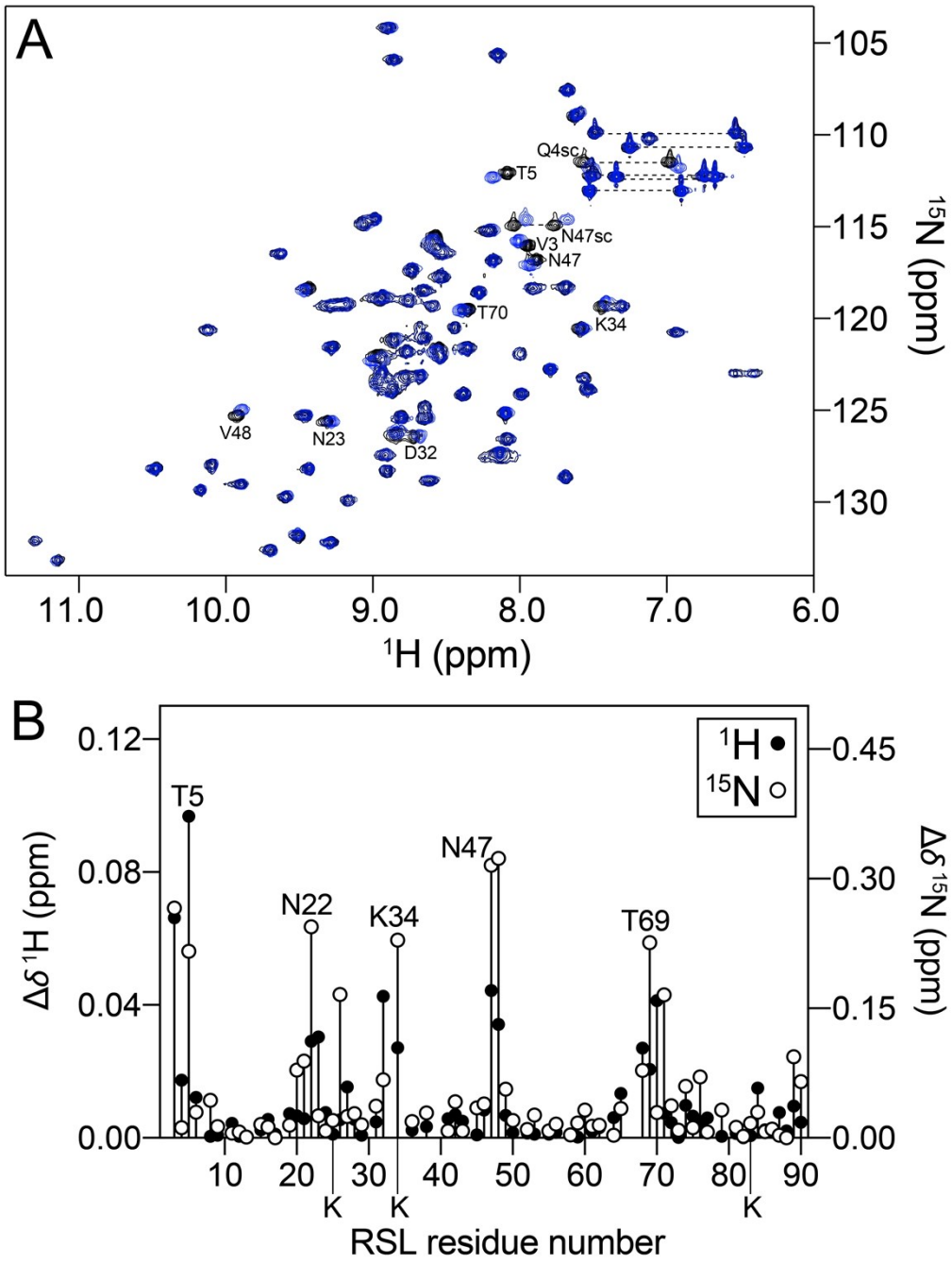

Figure S6. (A) Overlaid ^{1}H-^{15}N HSQC spectra of RSL in the absence (black contours) and presence of **Q6** (blue contours) in 20 mM potassium phosphate pH 6.0. **(B) Q6**-induced chemical shift perturbation plot of RSL backbone amides.

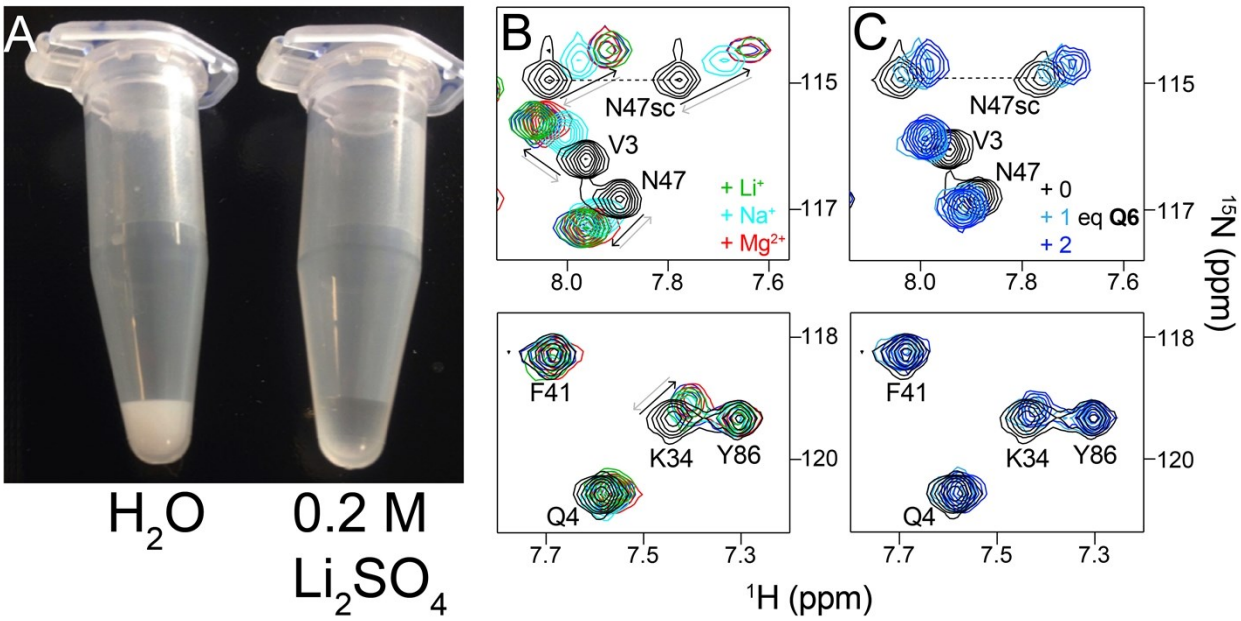

Figure S7. (A) Solubilization of **Q6** to 1 mM in 0.2 M Li_2SO_4. **(B)** Regions of overlaid ^{1}H-^{15}N HSQC spectra of RSL in the absence (black contours) and presence of **Q6** (blue contours) in water. The effect of different cations on RSL-Q6 binding was tested via addition of 50 mM sulfate salts. **(C)** The same spectral regions during titration of **Q6** into RSL in 50 mM Li_2SO_4.

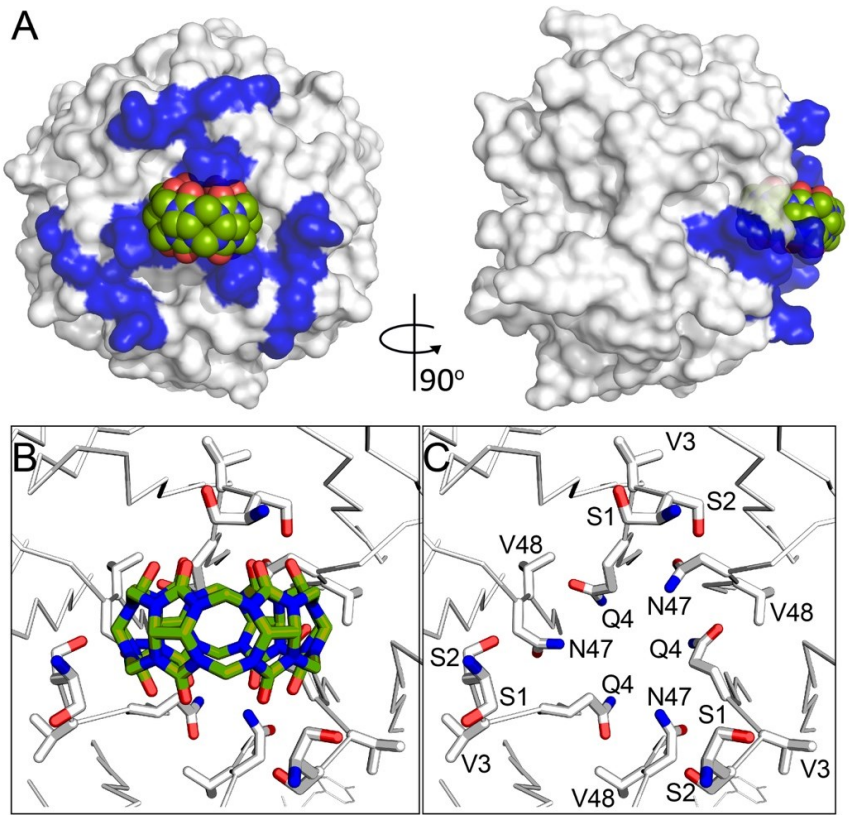

Figure S8. (A) RSL-**Q6** model from the top-scoring cluster in HADDOCK.[2] RSL is a grey surface with NMR-derived binding map in blue. **Q6** is shown as spheres. **(B)** Detail of the **Q6** interaction with the N-terminal region of RSL. **Q6** and interacting side chains shown as sticks. **(C)** Same view with **Q6** omitted.

Table S4. Statistics for the top two RSL-**Q6** clusters determined in HADDOCK.[2]

	Cluster 1	**Cluster 2**
HADDOCK score	-42.8 ± 0.5	-40.7 ± 1.1
Cluster size[a]	21	4
RMSD (Å)[b]	0.1 ± 0.1	0.2 ± 0.0
vdW energy (kcal/mol)	-29.9 ± 0.9	-28.1 ± 0.6
Electrostatic energy (kcal/mol)	-143.9 ± 9.7	-133.6 ± 6.5
Desolvation energy (kcal/mol)	1.3 ± 0.4	0.6 ± 0.7
Restraints violation energy (kcal/mol)	2.0 ± 0.1	2.2 ± 0.2
Buried surface area (Å²)	806.1 ± 7.9	805.1 ± 7.3
Z-score[d]	-2.0	-0.8

[a]HADDOCK generated 186 models in 16 clusters, accounting for 93 % of the water-refined models.
[b]From the overall lowest-energy structure.
[d]Indication of how many standard deviations from the average the cluster is located in terms of HADDOCK score.

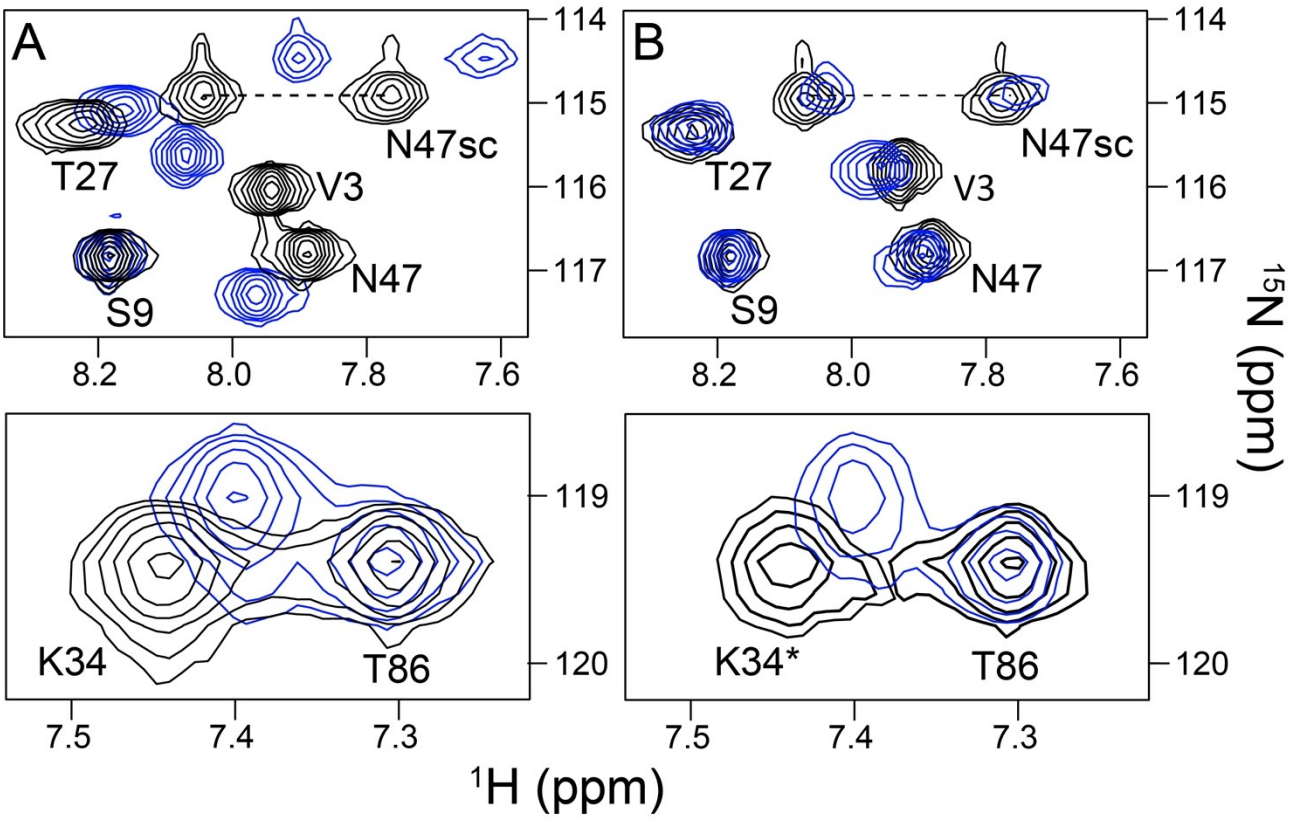

Figure S9. Regions of overlaid ^{1}H-^{15}N HSQC spectra of **(A)** native and **(B)** dimethylated RSL in the absence (black contours) and presence of **Q6** (blue contours) in water.

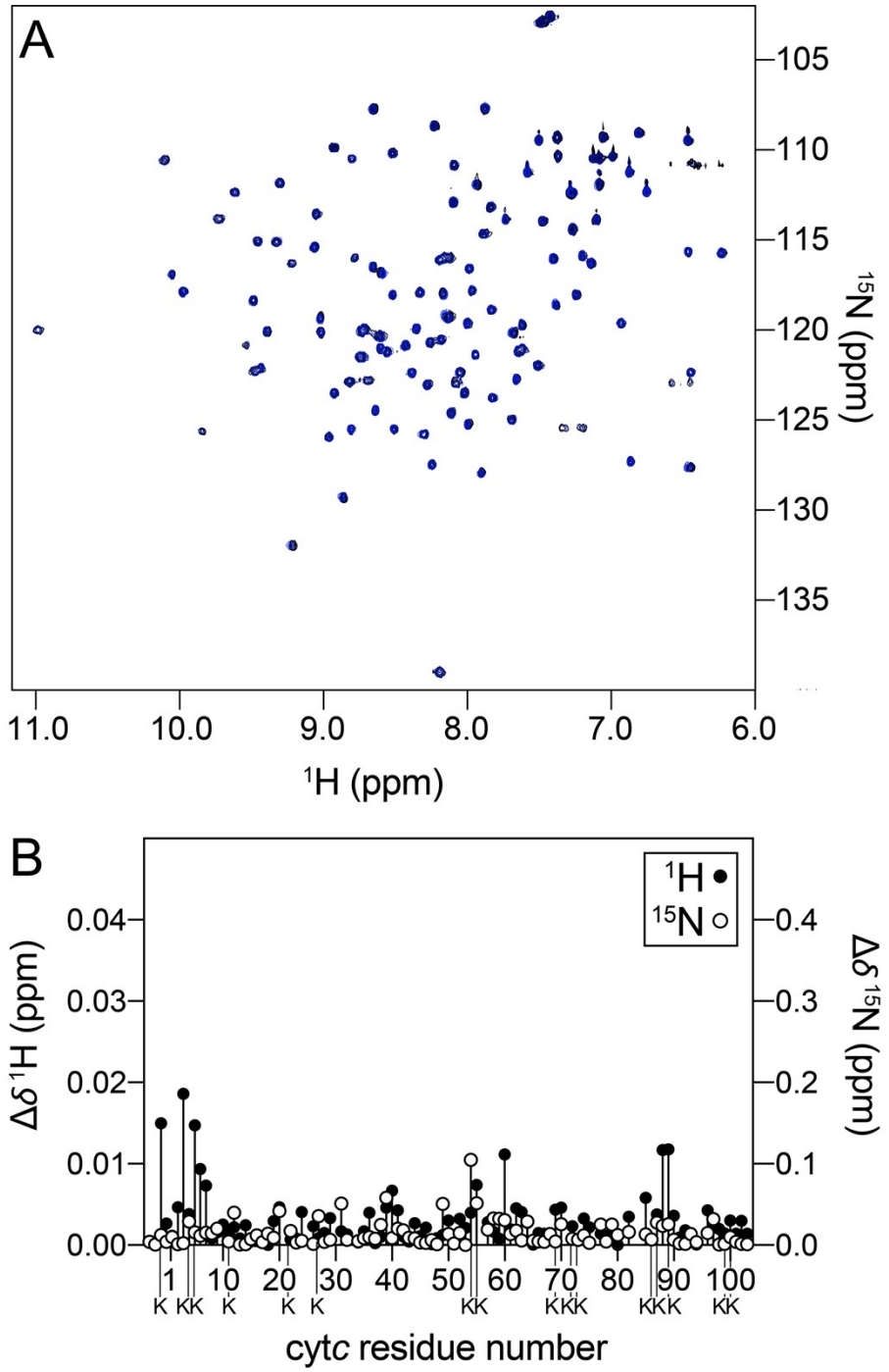

Figure S10. (A) Overlaid ^{1}H-^{15}N HSQC spectra of oxidized cyt*c* in the absence (black contours) and presence of **Q6** (blue contours) in 20 mM potassium phosphate pH 6.0. **(B) Q6**-induced chemical shift perturbation plot of cyt*c* backbone amides.

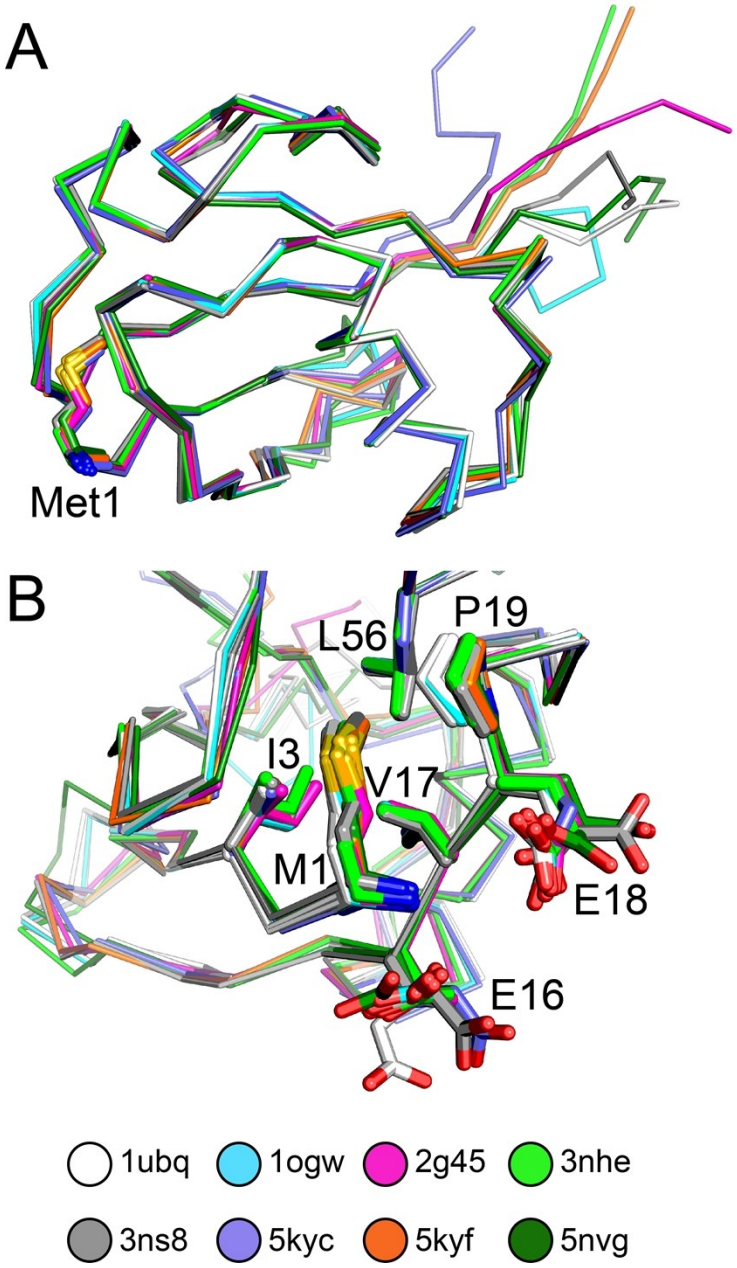

Figure S11. (A) Alignment of HUb models from eight high-resolution crystal structures. Met1 is shown as sticks. **(B)** The Met1 side chain points into the hydrophobic core, and the alpha ammonium is flanked by the carboxylates of Glu16 and Glu18.

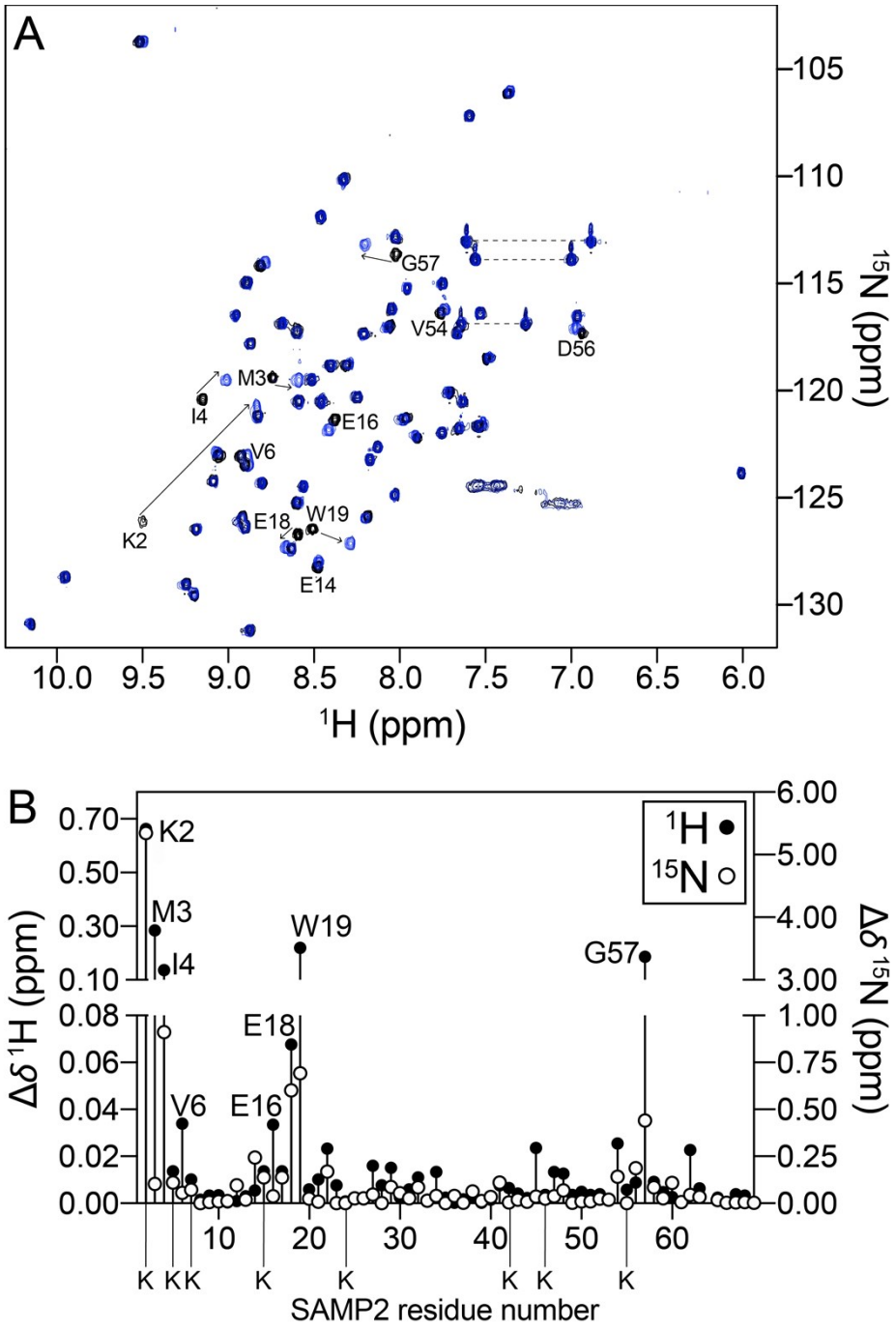

Figure S12. (A) Overlaid ^{1}H-^{15}N HSQC spectra of SAMP2 in the absence (black contours) and presence of **Q6** (blue contours) in 20 mM potassium phosphate, 50 mM NaCl, pH 6.0. **(B) Q6**-induced chemical shift perturbation plot of SAMP2 backbone amides.

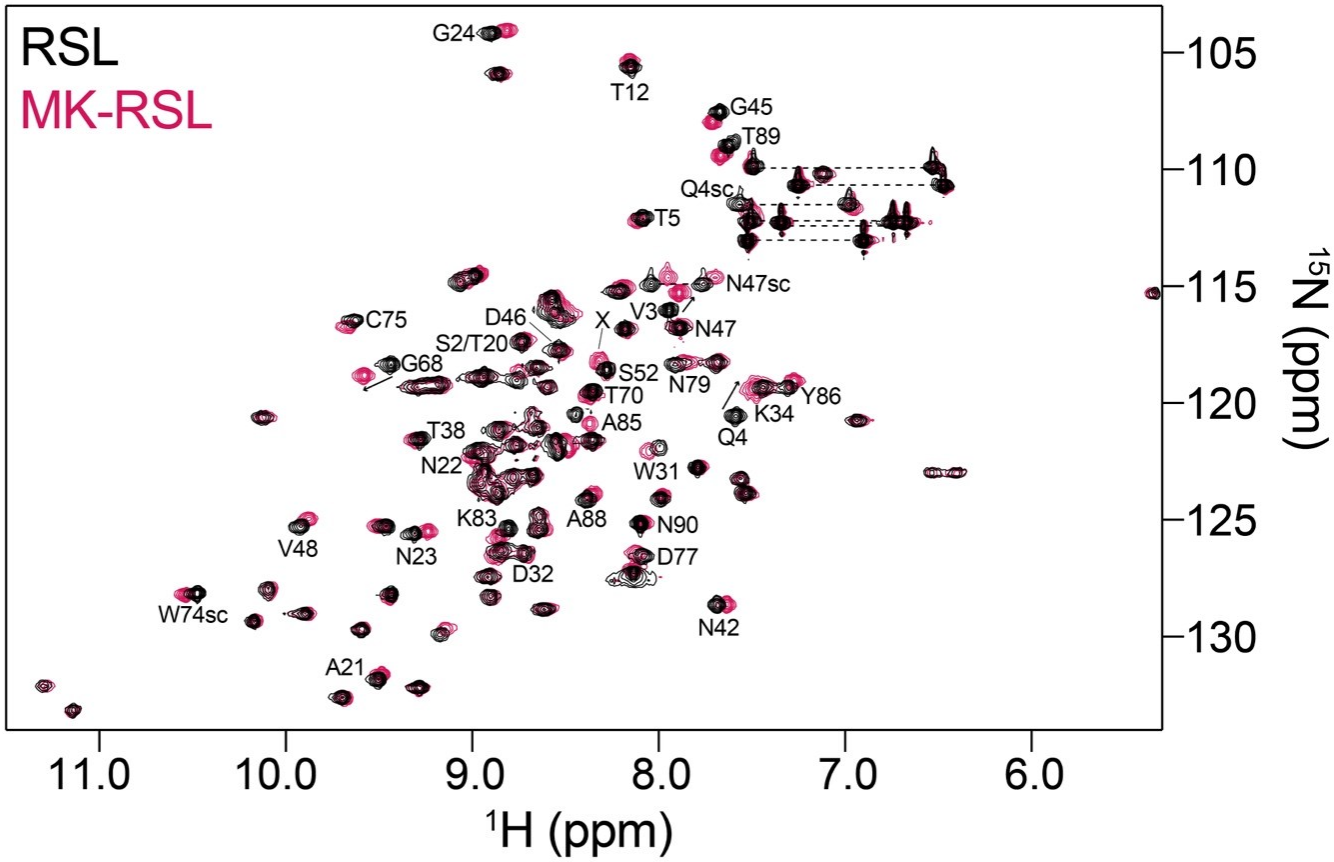

Figure S13. Overlaid ^{1}H-^{15}N HSQC spectra of RSL and MK-RSL in 20 mM KP_i, 50 mM NaCl pH 6.0. The new resonance in MK-RSL is labelled X.

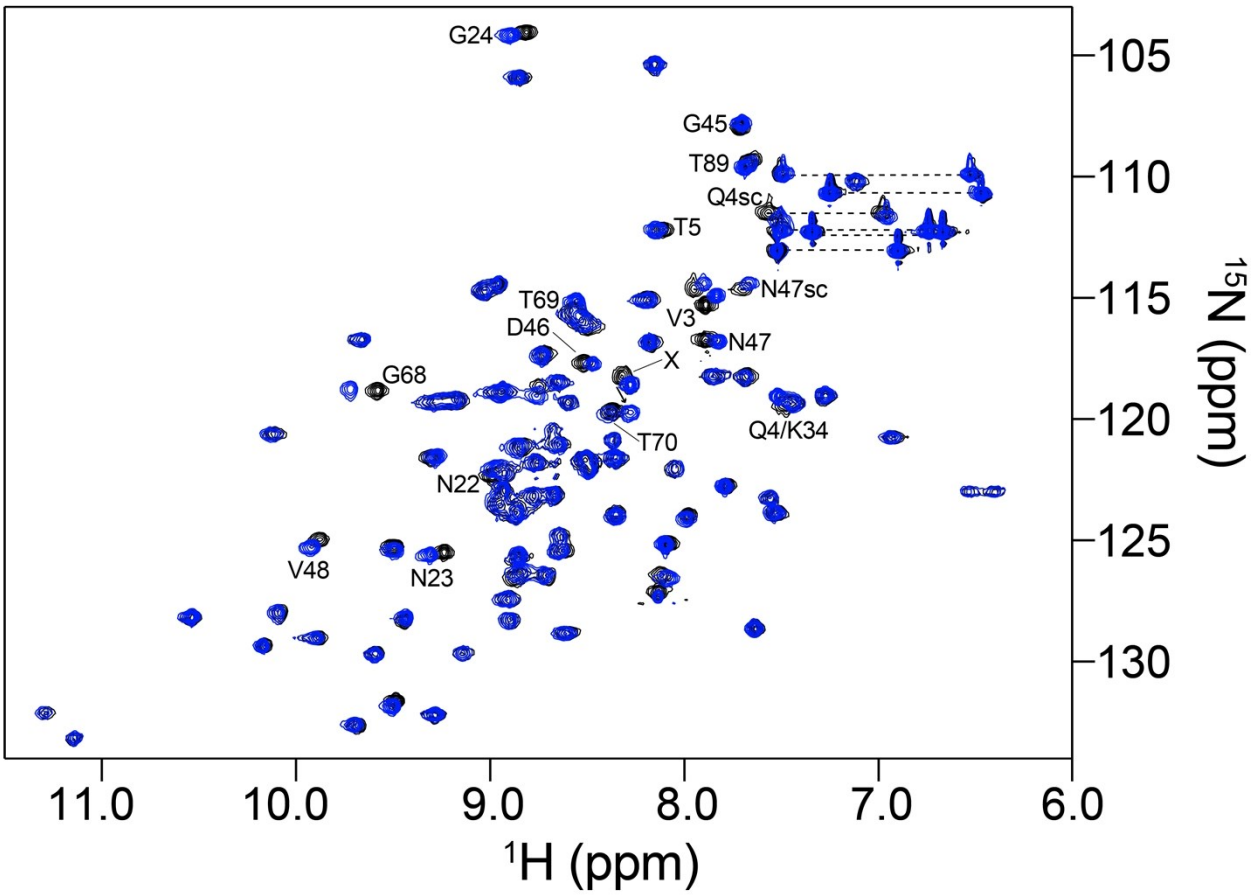

Figure S14. Overlaid ^{1}H-^{15}N HSQC spectra of MK-RSL in the absence (black contours) and presence of **Q6** (blue contours) in 20 mM potassium phosphate, 50 mM NaCl, pH 6.0.

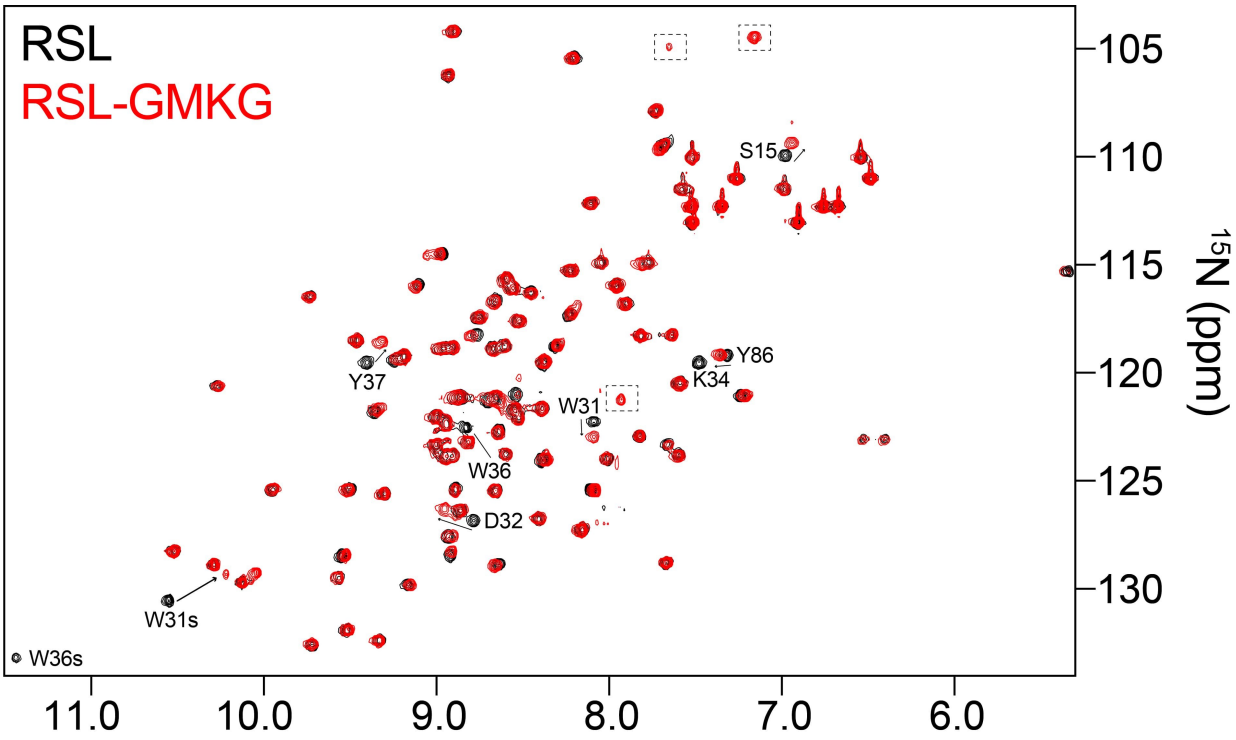

Figure S15. Overlaid ^{1}H-^{15}N HSQC spectra of RSL and RSL-GMKG in 20 mM KP_i, 50 mM NaCl pH 6.0. Resonances belonging to residues of the GMKG loop are highlighted with dashed boxes.

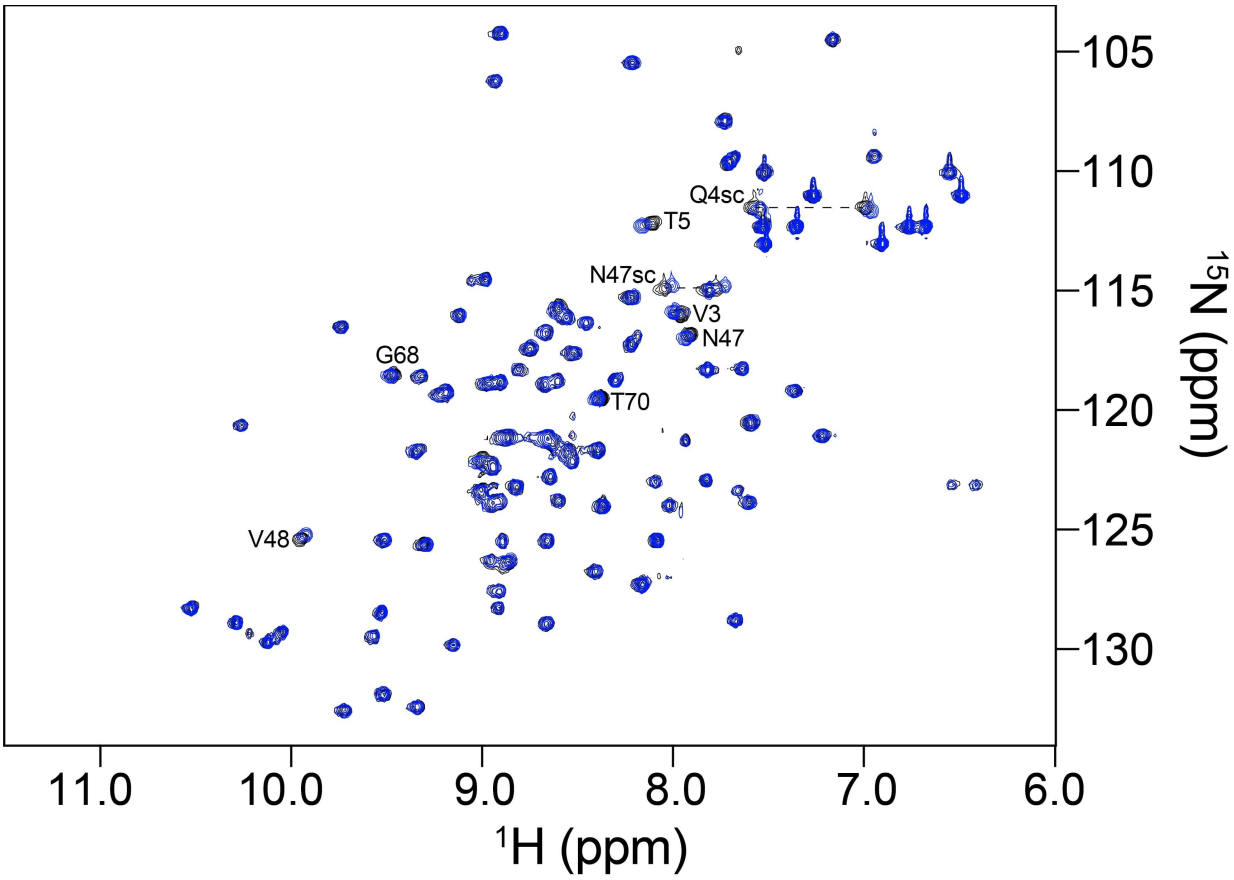

Figure S16. Overlaid ^{1}H-^{15}N HSQC spectra of RSL-GMKG in the absence (black contours) and presence of **Q6** (blue contours) in 20 mM potassium phosphate, 50 mM NaCl, pH 6.0.

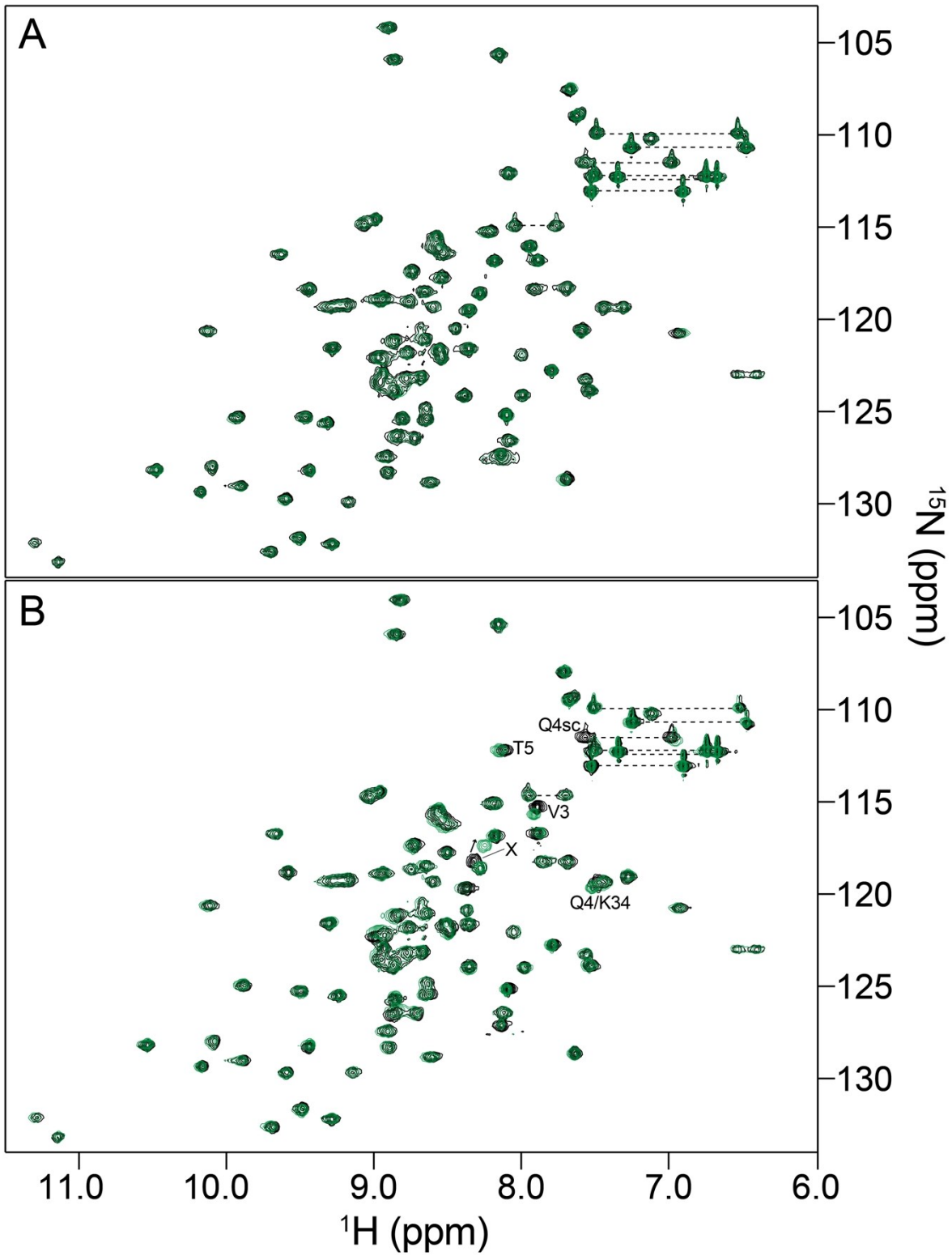

Figure S17. Overlaid ^{1}H-^{15}N HSQC spectra of **(A)** RSL and **(B)** MK-RSL in the absence (black contours) and presence of **Q8** (green contours) in 20 mM potassium phosphate, 50 mM NaCl, pH 6.0. The 'reporter' resonance of MK-RSL is labelled X.

REFERENCES

1. R. Winkler, *Rapid Commun. Mass Spectrom.* 2010, **24**, 285.
2. G. C. P. Van Zundert, J. P. G. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastritis, E. Karaca, A. S. J. Melquiond, M. van Dijk, S. J. De Vries and A. M. J. J. Bonvin, *J. Mol. Biol.* 2016, **428**, 720-725.