Electronic Supplementary Material (ESI) for ChemComm.
This journal is © The Royal Society of Chemistry 2017

Supporting Information

Site-selective tagging of proteins by pnictogen-mediated self-assembly

Christoph Nitsche, Mithun C. Mahawaththa, Walter Becker, Thomas Huber, Gottfried Otting

Computational design of double cysteine mutations for the Zika virus NS2B-NS3 protease

The crystal structure of the NS2B-NS3 protease (PDB ID: 5LC0)[1] was screened computationally for positioning of two cysteine residues with a geometry suitable for binding to trivalent arsenite. Arsenic–sulfur bond lengths were assumed to be 2.24 Å long and the S–As–S bond angle was required to be less than 95 degrees. All pairs of residues in the crystal structure were in turn mutated to cysteine and standard rotamers (χ_1 angles of 62, -177, and 65 degrees) were enumerated. Whenever the sulfur atoms of two modelled cysteine residues were within 3.027 and 3.303 Å, corresponding to a S–As–S bond angle between 85 and 95 degrees, and the site was solvent accessible and within 15 Å distance of the active site, the possibility of linking both sulfurs by an arsenic atom with correct bond length and bond angle was explored. Among the vicinal cysteine sites identified in this way (A132C/G133C, P131C/A132C, P101C/P131C, T27C/V36C, P101C/E104C, D129C/S160C, G82*C/V72C, T27C/G29C, A56*C/A24C, L86*C/V154C), the double mutation P101C/P131C was chosen for its solvent exposure, close proximity to the active site, and location in a loop region, which may allow for minor conformational adjustments to accommodate pnictogen binding.

Preparation of uniformly ^{15}N-labelled Zika virus NS2B-NS3 protease P101C/P131C

The construct of the Zika virus NS2B-NS3 protease was the same as described previously[2] with small modifications. It included 48 hydrophilic core residues of NS2B (residues 48*–95*) followed by a $Gly_4SerGly_4$ linker, the N-terminal residues of the NS3 protease domain (residues 1–170), a C-terminal TEV protease recognition sequence (ENLYFQG), and a His_6 tag. The mutations R95*A, K15N, and R29G were introduced to increase stability and reduce the potential of auto-cleavage.[2] Proline residues 101 and 131 were mutated to cysteines for pnictogen-mediated labelling. The expression construct used the T7 expression vector pETMCSI.[3] In addition to this construct (referred to as ZiPro P101C/P131C), a construct without the wild-type cysteines (ZiProC80S/C143S) was designed. The plasmid was transformed into *E. coli* BL21(DE3) and uniformly ^{15}N-labelled protein was expressed using a high-cell density protocol.[4] Cells were grown in LB medium at 37 °C until an OD_{600} value of 0.5-0.8 was reached, harvested, and resuspended in M9 medium (6 g/l Na_2HPO_4, 3 g/l KH_2PO_4, 0.5 g/l NaCl) supplied with glucose (10 g/l), $MgSO_4$ (5 mM), $CaCl_2$ (0.2 mM), thiamine (1 μg/ml), and $^{15}NH_4Cl$ (1 g/l). Following incubation at 37 °C for 1 h and subsequent induction with IPTG (1 mM), the cells were incubated overnight at room temperature. The cells were pelleted by centrifuging at 5000 *g* for 10 minutes and lysed by passing twice through a French Press (SLM Aminco, USA) at 830 bars. The cell lysate was centrifuged for 1 h at 34,000 *g* and the supernatant was loaded onto a 5 ml Co-NTA column

(GE Healthcare, USA) pre-equilibrated with loading buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% glycerol). The protein was eluted with a gradually increasing concentration of elution buffer (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 300 mM imidazole, 5% glycerol) and fractions were analysed by 12% SDS-PAGE. The buffer was exchanged to 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 1 mM β-mercaptoethanol, and TEV protease (200 μg/ml) was added. After incubation overnight, the buffer was exchanged to 50 mM Tris-HCl, pH 7.5, and 300 mM NaCl, and the protein was passed through a 5 ml Co-NTA column. Finally, the buffer was exchanged to NMR buffer (20 mM MES pH 6.5, 150 mM NaCl, 1 mM TCEP) and 10% D_2O was added for NMR measurements.

Preparation of reagents

4-(Mercaptomethyl)-2,6-pyridinedicarboxylic acid (4-MMDPA) was synthesised as described elsewhere[5] and used as a 50 mM stock in DMSO. $NaAsO_2$ (BDH) was used as a 50 mM stock in water. $SbCl_3$ (BDH) and $BiBr_3$ (Sigma) were used as 50 mM stocks in DMSO. Lanthanide chlorides ($TmCl_3$, $TbCl_3$) and YCl_3 were used as 50 or 10 mM stocks in water.

Sample preparation and NMR measurements

All NMR spectra were recorded at 25 °C, using an 800 MHz Bruker Avance NMR spectrometer equipped with a cryoprobe. [^{15}N,^{1}H]-HSQC spectra were recorded at a protein concentration of 0.3 mM in 3 mm NMR tubes (sample volume 180 μl). Samples were prepared by successive addition (titration) of pnictogen, 4-MMDPA, and lanthanide/yttrium stocks to the protein solution in NMR buffer.

Native PAGE

Samples (20 μl) consisting of ZiProC80S/C143S/P101C/P131C (150 μM), 4-MMDPA (150 μM), $TbCl_3$ (150 μM), $NaAsO_2$ (150 μM), $SbCl_3$ (150 μM), or $BiBr_3$ (150 μM) in 50% loading buffer (25 mM Tris-glycine, pH 7.8, 1 mM TCEP, 0.1% bromophenol blue, 50% glycerol) were loaded on a stacking gel (4.5% polyacrylamide, 100 mM Tris-HCl, pH 6.8). Samples were run at 100 V and 4° C for 3 h through a resolving gel (12% polyacrylamide, 400 mM Tris-HCl, pH 8.8) before analysis for terbium luminescence in a UV gel chamber (excitation wave length 254 nm). Subsequently, the gel was stained with Coomassie blue.

Mass spectrometry

Samples consisted of protein, 3-fold excess of trivalent pnictogen, and 10-fold excess of 4-MMDPA in NMR buffer. Mass spectrometry was conducted using an Orbitrap Elite Hybrid Ion Trap-Orbitrap mass spectrometer (Thermo Scientific, USA) coupled with an UltiMate

3000 UHPLC (Thermo Scientific, USA). 7.5 pmol of sample was injected to the mass analyser via an Agilent ZORBAX SB-C3 Rapid Resolution HT Threaded Column (Agilent, USA).

Isothermal titration calorimetry

Calorimetric titration was conducted using a Nano ITC instrument (TA instruments, USA) at 25 °C with an initial cell volume of 164 µl, using 2 µl titration steps of titrant. Between each titration step a sufficiently long time interval was applied for re-equilibration. Protein and titrant were used in the same buffer (20 mM MES, pH 6.5, 150 mM NaCl, 1 mM TCEP).

Molecular modelling

The range of tag conformations compatible with pnictogen coordination by Cys 101 and Cys 131 was modelled by starting from the model built of the As(III) complex with Cys101 and Cys 131 using the crystal structure 5LC0.[1] The local geometry around the trivalent arsenite was modelled based on the crystallographic structure of 5-(2-phenyl-1,3,2-dithiarsinan-4yl)pentanoic acid (CDS: NIDKAM). The dihedral angle about the As–S bond of the protein–As(III)–4-MMDPA–Ln complex was allowed to vary freely, while the following dihedral angles around the S–C and C–C_{Ar} bond were varied within ±10 degrees of the preferred rotamer states (-60/60/180 for S-C and -90/90 for C-C_{Ar}). Only conformations were accepted that did not produce van der Waals clashes with the protein and a total of about 5000 models were computed. The model of Fig. 1 showing all sterically allowed metal positions was generated with the software UCSF Chimera.[6]

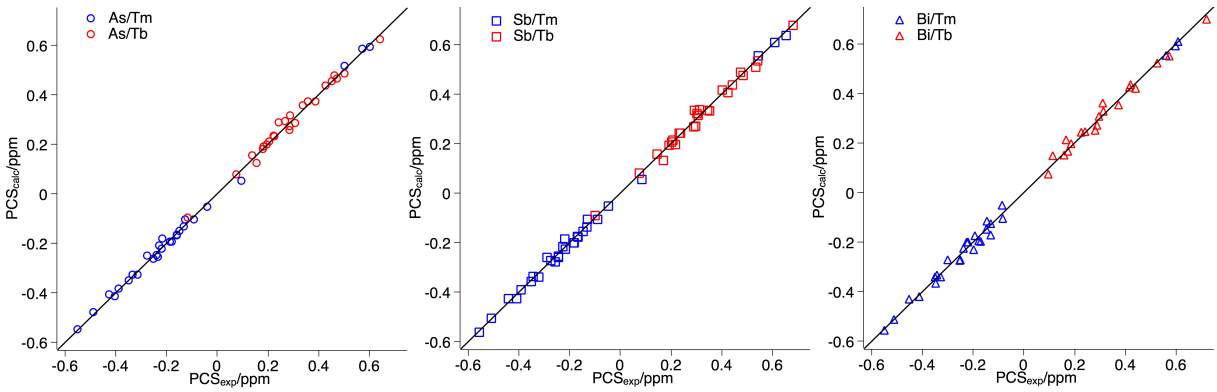

Fig. S1. Correlations between back-calculated and experimental PCSs of backbone amide protons obtained from the Δχ tensor fits of Table S2 for ZiPro P101C/P131C in the presence of 4-MMDPA and different pnictogens and lanthanides as indicated in the Figure.

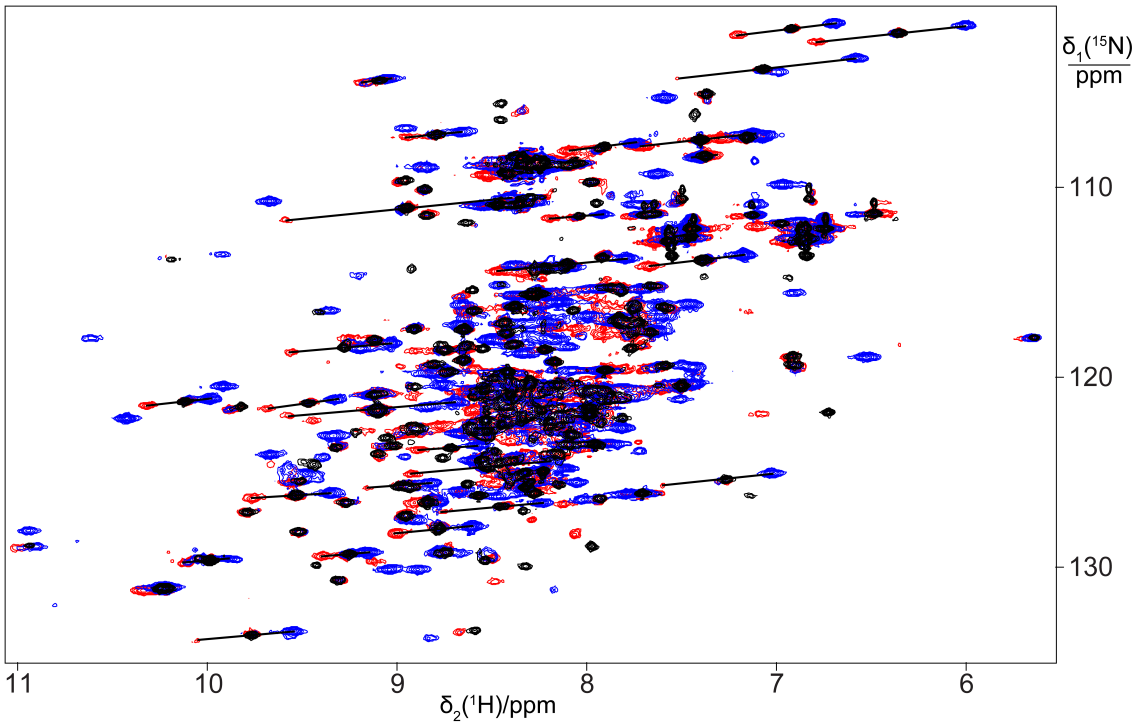

Fig. S2. Superimposition of [^{15}N,^{1}H]-HSQC spectra of 0.3 mM solutions of ZiPro P101C/P131C in the presence of 0.3 mM $NaAsO_2$, 4-MMDPA, and paramagnetic Tm^{3+} (blue), Tb^{3+} (red), or diamagnetic Y^{3+} (black).

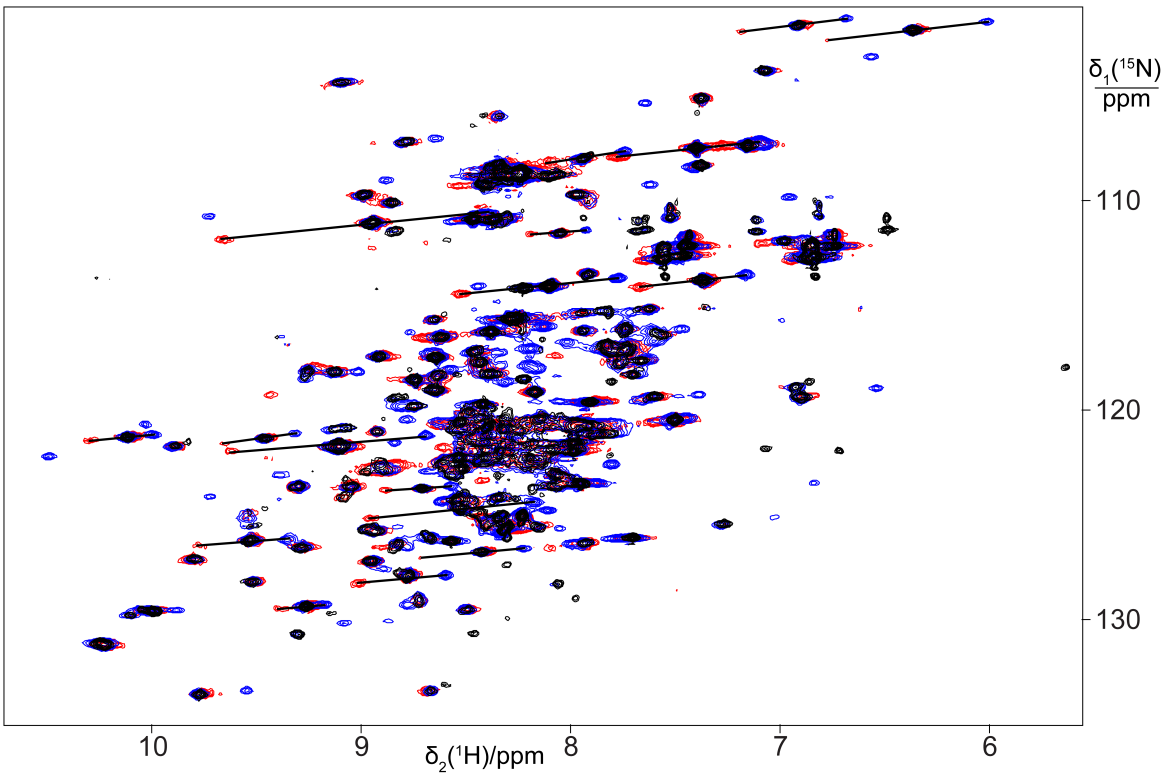

Fig. S3. Same as Fig. S2, except using 0.3 mM $BiBr_3$ instead of $NaAsO_2$.

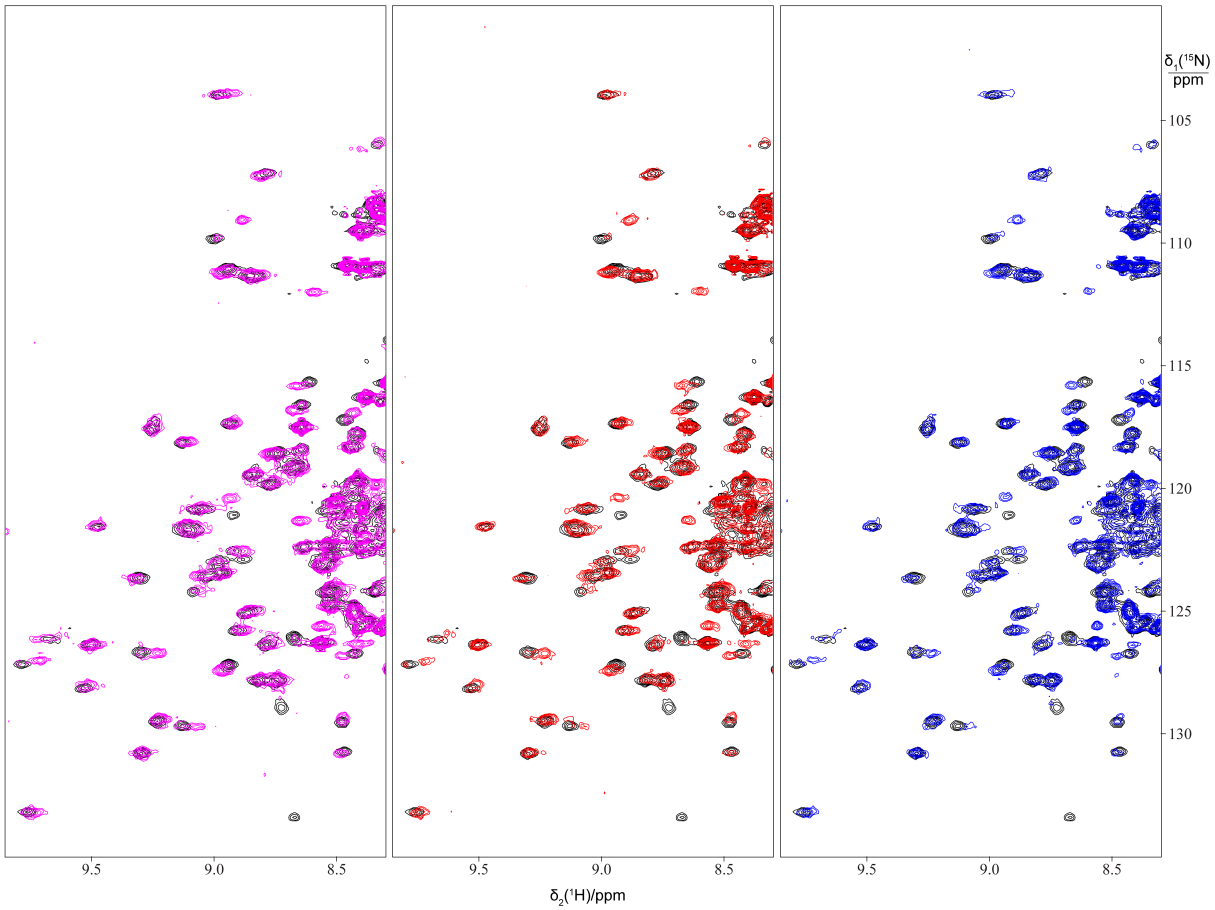

Fig. S4. Selected region of [^{15}N,^{1}H]-HSQC spectra of 0.3 mM solutions of ZiPro P101C/P131C/C80S/C143S (black) superimposed with the corresponding spectrum in the presence of sub-equimolar ratios of $NaAsO_2$ (magenta), $SbCl_3$ (red) or $BiBr_3$ (blue). The co-existence of cross-peaks for the free protein and the complexes with the pnictogens indicates that the exchange of pnictogens on and off the protein is slow on the millisecond time scale.

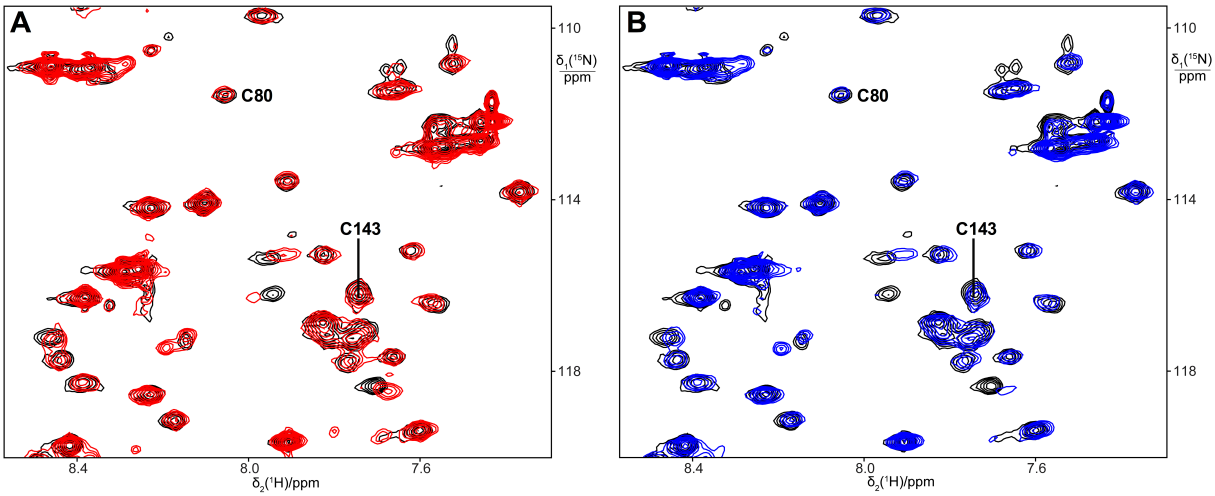

Fig. S5. Selected region of [^{15}N,^{1}H]-HSQC spectra of 0.3 mM solutions of ZiPro P101C/P131C (black) superimposed with the corresponding spectrum in the presence of (A) 0.3 mM $SbCl_3$ (red) or (B) 0.3 mM $SbCl_3$ and 0.3 mM reduced glutathione (blue). The cross-peaks of the backbone amides of the natural ZiPro cysteine residues C80 and C143 are identified.

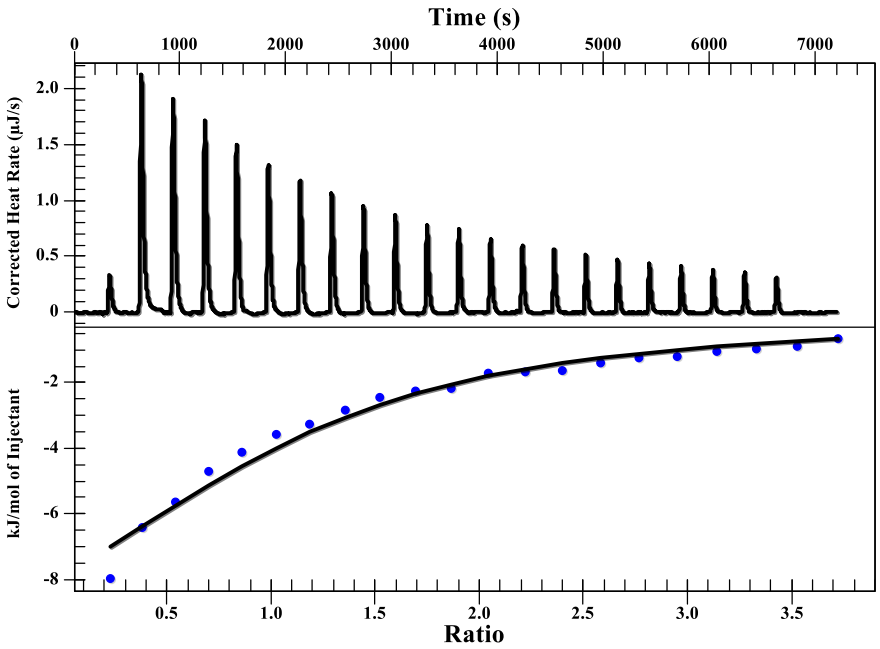

Fig. S6. Isothermal titration calorimetry results for ZiPro C80S/C143S/P101C/P131C (0.41 mM) titrated with $NaAsO_2$ (5.0 mM). A titration of buffer with $NaAsO_2$ (5.0 mM) was subtracted as blank. Data were fitted using an independent binding model assuming a 1:1 binding ratio (n = 1.0). The fitted dissociation constant K_d was 260 µM (200-360 µM), where the uncertainty range corresponds to the 95% confidence interval of the fit. ITC measurements with $SbCl_3$ and $BiBr_3$ were unsuccessful due to formation of the poorly soluble oxides, when the halogenides were titrated into buffer at the high stock concentrations needed for ITC (5.0 mM).

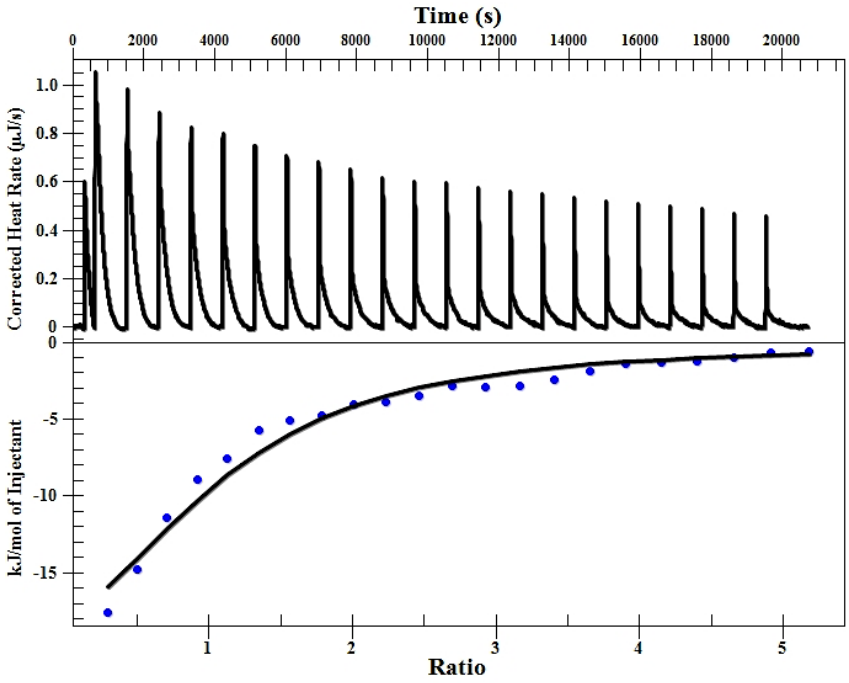

Fig. S7. Isothermal titration calorimetry results for ZiPro C80S/C143S/P101C/P131C (0.33 mM) and $NaAsO_2$ (0.31 mM) titrated with 4-MMDPA (5.0 mM). A titration of buffer with 4-MMDPA (5.0 mM) was subtracted as blank. Data were fitted using an independent binding model assuming a 1:1:1 binding ratio (n = 1.0). The fitted dissociation constant K_d was 170 µM (130-245 µM), where the uncertainty range corresponds to the 95% confidence interval of the fit.

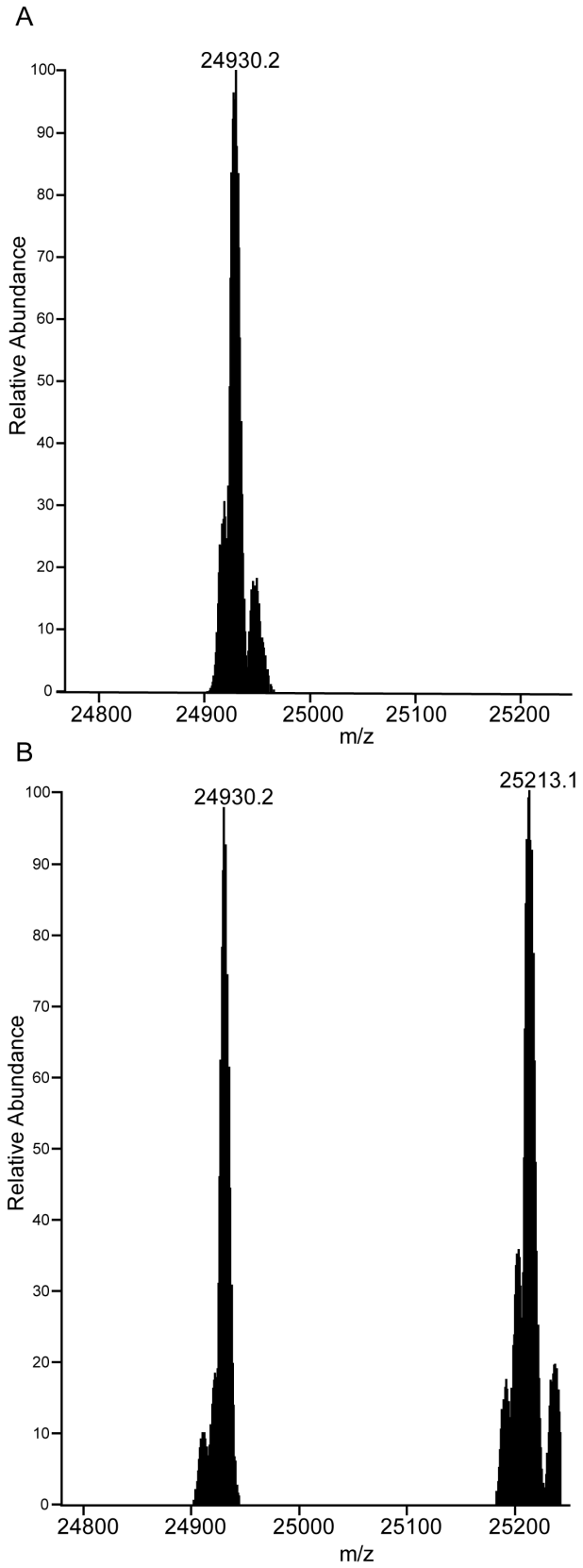

Fig. S8. Mass spectrometry data for ZiPro C80S/C143S/P101C/P131C. A) ZiPro: *m/z* = **24930** [ZiPro (87% ^{15}N-labelled)]$^{+}$ calcd: 24930. B) ZiPro, $NaAsO_2$, 4-MMDPA: *m/z* = **24930** [ZiPro (87% ^{15}N-labelled)]$^{+}$ calcd: 24930; **25213** [ZiPro (87% ^{15}N-labelled) + As(III) + 4-MMDPA –3H]$^{+}$ calcd: 25215; 25213 (deprotonated 4-MMDPA). The difference of 283 atomic mass units between the two peaks corresponds to arsenic (75), deprotonated 4-MMDPA (210) and two deprotonated cysteine side chains (–2).

Table S1. PCSs of backbone amide protons of ZiPro P101C/P131C in the presence of an equimolar amount of 4-MMDPA and the different pnictogens and lanthanides indicated.[a]

Residue	PCS_{exp}/ppm					
	Tm			Tb		
	As	Sb	Bi	As	Sb	Bi
Tyr52*	-0.404	-0.411	-0.412	0.462	0.533	0.525
Gly57*	-0.489	-0.509	-0.512	0.452	0.474	n.m.
Ile59*	0.096	0.084	n.m.	-0.115	-0.101	n.m.
Thr60*	0.573	0.609	0.560	n.m.	n.m.	n.m.
Gly21	-0.127	-0.132	-0.147	0.157	0.168	0.158
Thr27	-0.550	-0.557	-0.551	0.643	0.681	0.720
Arg28	-0.333	-0.345	-0.343	0.389	0.423	0.441
Gly29	-0.312	-0.324	-0.328	0.357	0.401	0.420
Leu31	-0.233	-0.245	-0.253	0.244	0.291	0.310
Gly32	-0.276	-0.290	-0.301	0.308	0.350	0.373
Ser33	-0.425	-0.442	-0.454	0.472	0.540	0.572
Met41	-0.216	-0.221	-0.220	0.286	0.288	0.288
Gln42	-0.092	-0.090	-0.084	0.138	0.144	0.114
Gly44	-0.039	-0.048	-0.086	0.076	0.073	0.097
His47	-0.218	-0.218	-0.199	0.289	0.302	0.295
Trp50	-0.239	-0.246	-0.239	0.337	0.312	n.m.
Val52	-0.389	-0.393	-0.349	0.504	0.484	n.m.
Tyr53	-0.348	-0.352	-0.351	0.427	0.440	0.415
Lys54	n.m.	-0.275	n.m.	n.m.	0.343	n.m.
Gly55	-0.227	-0.229	-0.225	0.287	0.295	0.280
Ala57	-0.179	-0.183	-0.179	0.224	0.232	0.227
Gly61	-0.250	-0.259	-0.250	0.268	0.305	0.313
Glu62	-0.158	-0.170	-0.132	0.180	0.215	n.m.
Gly63	-0.159	-0.168	-0.193	0.184	0.200	0.166
Leu65	-0.185	-0.186	-0.172	0.225	0.236	0.241
Ser78	-0.132	-0.133	-0.130	0.195	0.191	0.174
Tyr79	-0.148	-0.149	-0.149	0.206	0.207	0.187
Asp90	0.503	0.544	0.597	n.m.	n.m.	n.m.
Ser93	0.604	0.652	0.608	n.m.	n.m.	n.m.

[a] The PCSs were measured as the chemical shifts observed in the presence of paramagnetic lanthanide minus the chemical shift observed with diamagnetic Y^{3+}.

n.m.: not measured

Table S2. Δχ tensor parameters of ZiPro P101C/P131C in the presence of 4-MMDPA and the different pnictogens and lanthanides indicated.[a]

		$\Delta\chi_{ax}$ (10^{-32} m^3)	$\Delta\chi_{rh}$ (10^{-32} m^3)	x (Å)	y (Å)	z (Å)	α (°)	β (°)	γ (°)	Q[b]
Tm	**As**	-14.9	-8.4	90.4	40.4	166.0	96.0	86.7	23.1	0.05
	Sb	-16.0	-9.8	90.2	40.1	165.5	94.4	82.5	19.9	0.05
	Bi	-16.4	-7.2	90.7	40.0	164.1	96.8	77.4	13.8	0.06
Tb	**As**	16.9	10.2	90.4	40.4	166.0	96.9	89.4	5.3	0.06
	Sb	18.9	10.5	90.2	40.1	165.5	97.1	84.0	179.7	0.05
	Bi	20.2	7.3	90.7	40.0	164.1	98.9	78.3	1.6	0.06

[a] The Δχ tensors are reported in their unique tensor representation (UTR)[7] as obtained by combined fitting of the PCSs for Tm^{3+} and Tb^{3+} ions (Table S1) for each pnictogen to the PDB file 5LC0.[1]

[b] Quality factors were calculated using the following equation:

$$Q = \sqrt{\frac{\sum(PCS_{exp} - PCS_{calc})^2}{\sum(PCS_{exp})^2}}$$

References

1 J. Lei, G. Hansen, C. Nitsche, C. D. Klein, L. Zhang and R. Hilgenfeld, *Science*, 2016, **353**, 503.

2 M. C. Mahawaththa, B. J. G. Pearce, M. Szabo, B. Graham, C. D. Klein, C. Nitsche and G. Otting, *Antiviral Res.*, 2017, **142**, 141.

3 C. Neylon, S. E. Brown, A. V. Kralicek, C. S. Miles, C. A. Love and N. E. Dixon, *Biochemistry*, 2000, **39**, 11989.

4 A. Sivashanmugam, V. Murray, C. Cui, Y. Zhang, J. Wang and Q. Li, *Protein Sci.*, 2009, **18**, 936.

5 X. C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, T. Huber, B. A. Messerle and G. Otting, *J. Am. Chem. Soc.*, 2008, **130**, 10486.

6 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605.

7 C. Schmitz, M. J. Stanton-Cook, X.-C. Su, G. Otting and T. Huber, *J. Biomol. NMR*, 2008, **41**, 179.