Electronic Supplementary Material (ESI) for ChemComm.

Organoarsenic probes to study proteins by NMR spectroscopy

Mithun C. Mahawaththa,[#] Henry W. Orton,[#] Ibidolapo Adekoya, Thomas Huber, Gottfried Otting, Christoph Nitsche

Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

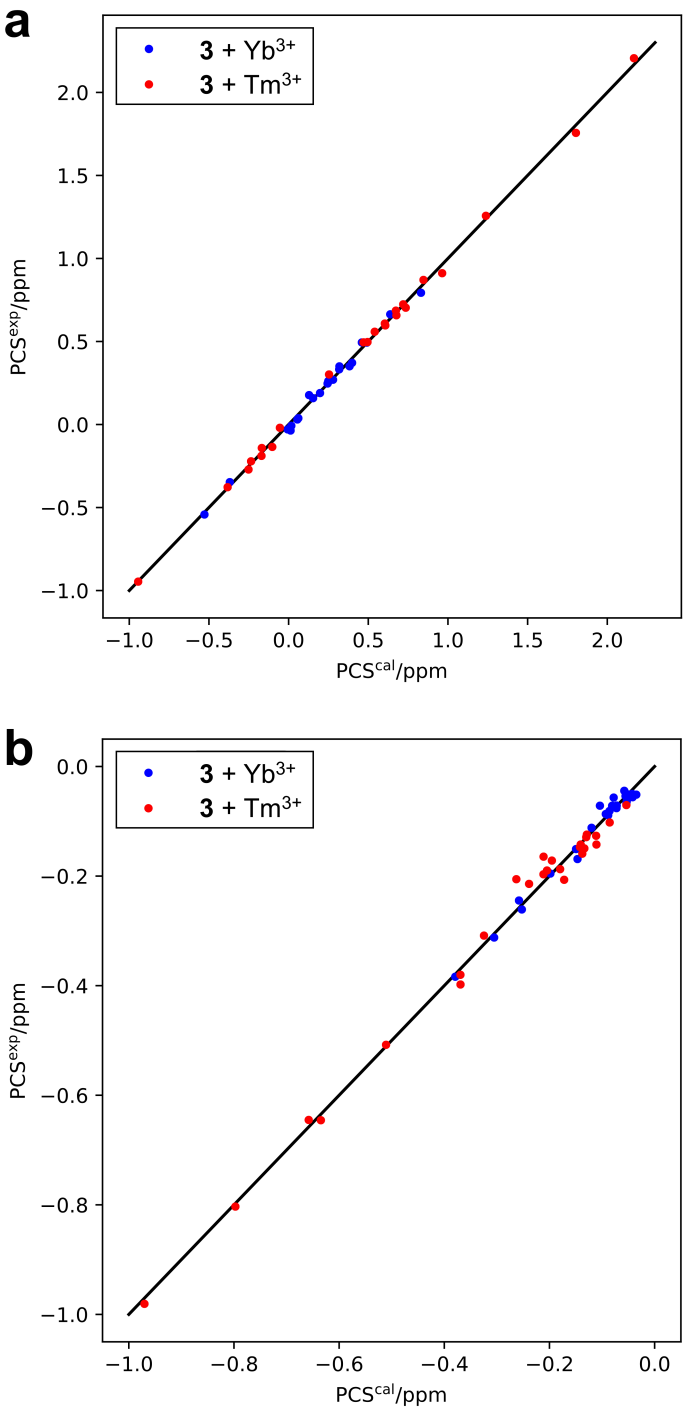

Fig. S1 Correlations between back-calculated (PCS^{cal}) and experimental (PCS^{exp}) PCSs of backbone amide protons obtained from the $\Delta\chi$–tensor fits of Table S2 for (a) GB1 K10C/T11C and (b) ubiquitin E18C/S20C/Q62E, each in the presence of probe **3** and different lanthanide ions as indicated in the Figure.

Table S1. High-resolution mass spectrometry results of double-cysteine mutants of GB1 and ubiquitin tagged with organoarsenic probes **1** and **3**.

	Protein		Protein + 1		Protein + 3	
	calcd	**found**	**calcd**	**found**	**calcd**	**found**
GB1[c]	7492.91[a]	7490.21	7712.19[b]	7712.20	7729.13[b]	7729.13
ubiquitin[d]	8555.56	8555.56	8777.54	8777.54	8794.48	8794.48

[a] Calculated mass for 100% ^{15}N-labelled protein

[b] Mass calculated for 96.57% ^{15}N-labelling

[c] B1 immunoglobulin-binding domain of streptococcal protein G K10C/T11C mutant with N-terminal MASMTG tag and C-terminal TEV cleavage site ENLYFQ.

[d] Human ubiquitin mutant E18C/S20C/Q62E.

Table S2. Δχ-tensor parameters for B1 immunoglobulin-binding domain of streptococcal protein G K10C/T11C (GB1) and ubiquitin E18C/S20C/Q62E (Ubi) in the presence of tag **3** and different lanthanides as 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]
GB1	**Yb**	-12.73	-4.01	-4.46	4.21	14.81	7.20	121.22	81.07	0.07
	Tm	-39.28	-14.07	-4.46	4.21	14.81	32.90	133.16	110.23	0.03
Ubi	**Yb**	4.42	1.28	35.12	13.00	4.03	109.37	141.9	46.9	0.08
	Tm	11.20	3.21	35.12	13.00	4.03	109.69	141.0	47.1	0.06

[a] The Δχ tensors are reported in their unique tensor representation (UTR)[1] as obtained by combined fitting of the PCSs for Yb^{3+} and Tm^{3+} ions to the PDB files 2QMT[2] and 1UBQ[3] for GB1 and ubiquitin, respectively.

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

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

Table S3. PCSs of backbone amide protons of B1 immunoglobulin-binding domain of streptococcal protein G K10C/T11C (GB1) in presence of **3** and different lanthanides.[a]

Residue No.	Yb^{3+}	Tm^{3+}
2	0.157	0.303
3	0.268	0.497
17	0.492	0.726
19	0.349	0.705
20	0.252	0.561
22	0.260	0.598
25	0.331	0.873
26	0.349	0.770
27	0.493	1.259
28	0.662	1.758
29	0.792	1.640
44	-0.348	-0.945
46	-0.031	-0.220
49	-0.009	-0.140
51	0.029	-0.187

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

Table S4. PCSs of backbone amide protons of ubiquitin E18C/S20C/Q62E in the presence of **3** and different lanthanides.[a]

Residue No.	Yb	Tm
4	-0.071	-0.207
6	-0.044	-0.147
7	-0.051	-0.102
11		-0.070
28	-0.195	-0.508
30	-0.169	-0.398
32	-0.072	-0.206
33	-0.076	-0.187
34	-0.068	-0.150
35		-0.124
36	-0.051	-0.129
39	-0.056	-0.142
40	-0.050	-0.126
41	-0.063	-0.159
44	-0.072	-0.190
50	-0.087	-0.214
54	-0.112	-0.309
55	-0.312	-0.803
58	-0.384	-0.981
59	-0.245	-0.645
61	-0.261	-0.646
62	-0.151	-0.380
65	-0.089	-0.165
67	-0.081	-0.197
68	-0.057	-0.172
70	-0.054	-0.142

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

Instrumentation and materials used for chemical synthesis

NMR spectra were recorded on Bruker Avance III 400 MHz and 800 MHz NMR spectrometers. Chemical shifts are quoted in units of parts per million (ppm) and were referenced internally to the solvent resonance. Multiplicities and appearances of NMR resonances are abbreviated as: s, singlet; d, doublet; m, multiplet. Coupling constants (*J*) are given in Hertz (Hz). High-resolution mass measurements were performed on an Orbitrap Elite mass spectrometer in positive mode. Reagents were purchased from Sigma Aldrich (US) and AK Scientific (US) and no further purification steps were performed. All solvents were used as obtained from the commercial sources, except that solvents were dried using a Braun MB-SPS-800 solvent purification system. Analytical thin-layer chromatography (TLC) analysis was performed on pre-coated silica gel aluminum-backed plates (Merck silica gel 60 F_{254}), using visualization under UV light at 254 nm. Flash chromatography was performed using Merck silica gel 60 (40–63 μm).

Synthesis of trimethyl(4-nitrophenyl)silane.[4]

To a solution of 1,4-bis(trimethylsilyl)benzene (3.0 g, 13.5 mmol) in acetic acid (25 ml) was added nitric acid (10 ml, 70%) at a rate which maintained the reaction temperature between 80 and 100 °C of the reaction mixture. After complete addition, the mixture was stirred overnight at 80 °C, before water was added and the pH was adjusted to 3-4 using a concentrated solution of sodium hydroxide. The aqueous solution was extracted with ethyl acetate and the combined organic extracts were dried over sodium sulfate and evaporated to dryness. The residue was purified by flash chromatography (hexane/ethyl acetate) to obtain trimethyl(4-nitrophenyl)silane as a yellow oil (1.5 g, 56%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ (ppm) = 0.32 (s, 9H), 7.68 (d, *J* = 8.5 Hz, 2H), 8.17 (d, *J* = 8.5 Hz, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) = –1.3, 122.4, 134.3, 148.6, 150.0.

Synthesis of 4-(trimethylsilyl)aniline.[5]

Trimethyl(4-nitrophenyl)silane (1.45 g, 7.4 mmol) was dissolved in ethanol (15 ml) and hydrogenated (1 atm) with palladium on carbon (10 %, 125 mg, 1.2 mmol) overnight. The mixture was filtered through celite and washed with ethanol. The solvent was evaporated to obtain 4-(trimethylsilyl)aniline as a pale yellow solid (1.1 g, 90%), which was used without further purification. ^{1}H-NMR (400 MHz, $CDCl_3$): δ (ppm) = 0.23 (s, 9H), 6.70 (d, *J* = 7.3 Hz, 2H), 7.33

(d, J = 7.3 Hz, 2H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) = –0.7, 114.7, 128.6, 134.7, 147.2; MS (ESI): *m/z* = $[M+H]^+$ calcd for $C_9H_{16}NSi$: 166.10, found: 166.10.

Synthesis of (4-(trimethylsilyl)phenyl)arsonic acid.

Sodium nitrite (490 mg, 7.1 mmol) was added to a stirring mixture of 4-(trimethylsilyl)aniline (1.05 g, 6.4 mmol) in concentrated hydrochloric acid (1 ml), water (5 ml) and ethanol (5 ml) at 0 °C. The resulting solution was kept on ice and added dropwise to a cold and vigorously stirring solution of sodium arsenite (1.0 g, 7.7 mmol), sodium carbonate (1.2 g, 11.6 mol) and copper(II) sulfate pentahydrate (125 mg, 0.5 mmol) in water (10 ml) at a rate which maintained the reaction temperature below 15 °C. The resulting slurry was allowed to warm up to room temperature and stirred overnight. The mixture was acidified with hydrochloric acid and stirred at 0 °C for 3 h until full precipitation occurred. The precipitate was collected by filtration, washed with water and dried to obtain (4-(trimethylsilyl)phenyl)arsonic acid (930 mg, 53%) as a colorless solid. ^{1}H-NMR (400 MHz, DMSO-d_6): δ (ppm) = 0.27 (s, 9H), 7.73 (m, 4H); ^{13}C NMR (100 MHz, DMSO-d_6): δ (ppm) = –1.4, 129.0, 134.0, 145.8; MS (ESI): *m/z* = $[M+H]^+$ calcd for $C_9H_{16}AsO_3Si$: 275.01, found: 275.01.

Synthesis of (4-(trimethylsilyl)phenyl)arsonous acid (1).

To a cold solution of (4-(trimethylsilyl)phenyl)arsonic acid (850 mg, 3.1 mmol) and potassium iodide (83 mg, 0.5 mmol) in methanol (30 ml) and concentrated hydrochloric acid (15 ml) was added a saturated solution of sodium metabisulfite (30 ml) at a rate which maintained the reaction temperature below 10 °C. The resulting mixture was diluted with water (20 ml) and methanol (20 ml) before concentrated ammonium hydroxide (30%) was added dropwise at 0 °C until precipitation of the product was complete after 2 h. The precipitate was collected and washed with water to obtain (4-(trimethylsilyl)phenyl)arsonous acid (400 mg, 50%) as a pale yellow solid. ^{1}H-NMR (400 MHz, DMSO-d_6): δ (ppm) = 0.25 (s, 9H), 7.56 (d, J = 7.6 Hz, 2H), 7.62 (d, J = 7.6 Hz, 2H). ^{13}C NMR (100 MHz, DMSO-d_6): δ (ppm) = –1.1, 128.7, 132.9, 141.4, 150.4.

Synthesis of dimethyl 4-chloropyridine-2,6-dicarboxylate.[6]

To a solution of chelidamic acid (1.0 g, 5.5 mmol) in thionyl chloride (5.6 ml, 77 mmol) were added a few drops of DMF at 0 °C to start a strongly exothermic reaction that was allowed to stir at 100 °C overnight. The solvents were evaporated and the residue was resolved in dry dichloromethane (3 ml), before methanol (4 ml) was added dropwise at 0 °C. All solvents were evaporated and a saturated sodium bicarbonate solution (40 ml) was added and extracted with dichloromethane. The combined organic layers were extracted with water and brine, dried over magnesium sulfate and evaporated. The residue was purified by flash chromatography (dichloromethane/methanol) to obtain 4-chloropyridine-2,6-dicarboxylate as a colorless solid (925 mg, 74%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ (ppm) = 4.03 (s, 6H), 8.30 (s, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) = 53.6, 128.5, 147.0, 149.6, 164.3; MS (ESI): *m/z* = $[M+H]^+$ calcd for $C_9H_9ClNO_4$: 230.02, found: 230.02.

Synthesis of dimethyl 4-iodopyridine-2,6-dicarboxylate.[6-7]

A solution of 4-chloropyridine-2,6-dicarboxylate (400 mg, 1.7 mmol) and sodium iodide (2.6 g, 17.4 mmol) in dry acetonitrile (25 ml) was sonicated for 30 min, before acetyl chloride (0.4 ml, 5.6 mmol) was added and sonication was continued for 45 min at 45 °C. Saturated sodium carbonate solution (30 ml) was added, the mixture was extracted with dichloromethane and the combined organic layers were washed with water, dried over magnesium sulfate and evaporated. The residue was purified by flash chromatography (dichloromethane/methanol) to obtain 4-iodopyridine-2,6-dicarboxylate as a pale yellow solid (190 mg, 34%). ^{1}H-NMR (400 MHz, $CDCl_3$): δ (ppm) = 4.02 (s, 6H), 8.66 (s, 2H). ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) = 53.6, 107.1, 137.3, 148.4, 164.0; MS (ESI): *m/z* = $[M+H]^+$ calcd for $C_9H_9INO_4$: 321.96, found: 321.96.

Synthesis of (2,6-bis(methoxycarbonyl)pyridin-4-yl)arsonous acid (2).

To a solution of 4-iodopyridine-2,6-dicarboxylate (640 mg, 2.0 mmol) in dry THF (50 ml) was added a solution of isopropylmagnesium bromide (3 ml, 0.75 M in THF) dropwise at –78 °C. After complete addition, the reaction mixture was slowly warmed up to –20 °C and arsenic(III) chloride (0.5 ml, 6.0 mmol) was added. The reaction mixture was stirred at –20 °C for 2 h and subsequently allowed to slowly warm up to room temperature overnight. The reaction was quenched with water (50 ml) and extracted with dichloromethane. The combined organic layers were dried over

magnesium sulfate and evaporated. The residue was triturated with diethyl ether and purified by flash chromatography (dichloromethane/methanol) to obtain (2,6-bis(methoxycarbonyl)pyridin-4-yl)arsonous acid as a colorless solid (225 mg, 37%). ^{1}H-NMR (400 MHz, DMSO-d_6): δ (ppm) = 3.93 (s, 6H), 8.47 (s, 2H); ^{13}C NMR (100 MHz, DMSO-d_6): δ (ppm) = 52.8, 128.6, 147.0, 163.0, 164.8; MS (ESI): *m/z* = $[M+H]^+$ calcd for $C_9H_{11}AsNO_6$: 303.98, found: 303.98.

Generation of a stock solution of 4-(dihydroxyarsanyl)pyridine-2,6-dicarboxylic acid (3).

A 50 mM solution of **2** (7.5 mg, 0.025 mmol) and lithium hydroxide hydrate (6 mg, 0.125 mmol) in water (500 µl) was stirred overnight at room temperature. Subsequently, this solution was neutralized and diluted to a 10 mM final stock solution used for protein tagging by addition of 10 mM HEPES pH 7.0, 1 mM TCEP buffer (2 ml). Complete ester cleavage was confirmed by NMR spectroscopy. ^{1}H-NMR (400 MHz, D_2O): δ (ppm) = 8.20 (s, 2H).

Protein expression and purification

The construct for the B1 immunoglobulin-binding domain of streptococcal protein G (GB1) contained an N-terminal MASMTG tag and a C-terminal TEV protease cleavage site followed by a His_6 tag. The ubiquitin construct contained a C-terminal His_6 tag and a Q62E mutation to maintain the pI value after introduction of a cysteine mutation at E18. Both constructs were cloned into a pETMCSI vector.[8] Double-cysteine mutations (K10/T11 in GB1 and E18/S20 in ubiquitin) were introduced by a modified QuikChange protocol using mutant T4 DNA polymerase.[9] All proteins were expressed in *E. coli* BL21(DE3) cells. In order to minimise usage of ^{15}N-labelled ammonium chloride, a top-down expression method was used. Initially, 1000 ml of cell culture was grown in LB medium at 37 °C until the OD_{600} reached 0.6-0.8. Subsequently, the cells were pelleted and resuspended in 300 ml M9 medium (6 g/l Na_2HPO_4, 3 g/l KH_2PO_4, 0.5 g/l NaCl) and supplied with 1 g/l $^{15}NH_4Cl$. Subsequently, the cells were incubated for 30 min at 37 °C and overexpression was induced by addition of 1mM IPTG. Protein expression was conducted at 25 °C overnight. Cells were harvested by centrifugation at 5,000 *g* for 15 min and lysed by passing twice through a French Press (SLM Aminco, USA) at 830 bars. The lysate was centrifuged at 13,000 *g* for 30 min and the filtered supernatant was loaded onto a 5 ml Ni-NTA column (GE Healthcare, USA) equilibrated with binding buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% glycerol). The protein was eluted with elution buffer (binding buffer containing, in addition, 300

mM imidazole) and the fractions were analysed by 12% SDS-PAGE. The His_6 tag of GB1 was removed by digestion with TEV protease in a 1:100 ratio overnight at 4 °C in buffer containing 50 mM Tris-HCl, pH 8.0, 300 mM NaCl and 1 mM β-mercaptoethanol. The His_6 tag of ubiquitin was removed by ubiquitinase treatment in a 1:100 ratio, initially at 37 °C for 3 hours, followed by overnight treatment at 4 °C in buffer containing 50 mM Tris-HCl, pH 8.0, 300 mM NaCl and 2 mM β-mercaptoethanol. After His_6 tag cleavage, protein samples were passed through a 5 ml Ni-NTA column (GE Healthcare, USA) equilibrated with binding buffer. Finally, the buffer was exchanged to NMR buffer (50 mM HEPES pH 7.0) using centrifugal filter units (Amicon Ultra with a MWCO of 3 kDa; Millipore, Billerica, USA). TCEP was added at 1 mM final concentration to the protein sample prior to performing the tagging reactions.

Mass spectrometry of protein samples.

Samples were prepared by mixing 20 μM of protein with 40 μM of compounds **1** or **3** dissolved in water. Mass-spectrometric analysis was conducted using an Elite Hybrid Ion Trap-Orbitrap mass spectrometer (Thermo Scientific, USA) coupled with an UltiMate S4 3000 UHPLC (Thermo Scientific, USA). 7.5 pmol of sample were injected to the mass analyser via an Agilent ZORBAX SB-C3 Rapid Resolution HT Threaded Column (Agilent, USA).

EPR measurements

To test the use of the organoarsenic probe **3** for distance measurements by EPR spectroscopy, the SARS-CoV2 main protease was produced with a C-terminal extension by the amino acid sequence GSGCCHHHHHH, in the following referred to as M^{pro}-CC. The protein was expressed and purified using the protocol described previously for the wild-type protein.[10] Samples for EPR measurements contained 200 mM M^{Pro}-CC in EPR buffer (20 mM HEPES-KOH in D_2O, pD 7.4 (uncorrected pH meter reading)), 150 mM NaCl, 200 mM **3**, 200 mM $GdCl_3$, 1 mM TCEP and 20% deuterated glycerol.

DEER measurements employed the standard four-pulse DEER sequence, $\pi/2_{vobs} - \tau_1 - \pi_{vobs} - (\tau_1+t) - \pi_{vpump} - (\tau_2-t) - \pi_{vobs} - \tau_2 -$ echo,[11] including a chirp pump pulse with an eight-step phase cycling. Measurements were carried out at W-band (94.9 GHz) on a home-built pulse spectrometer[12] at a temperature of 10 K. The detection used π/2 and π pulses of 15 ns/30 ns duration, respectively, at 94.85 GHz. The pump pulse length was 128 ns

positioned in the frequency range 94.9–95.1 GHz. The peak of the Gd(III) spectrum was positioned at 94.9 GHz. The repetition time was 0.2 ms and τ_1 = 0.6 µs. The DEER data were analysed using the DeerAnalysis software package.[13]

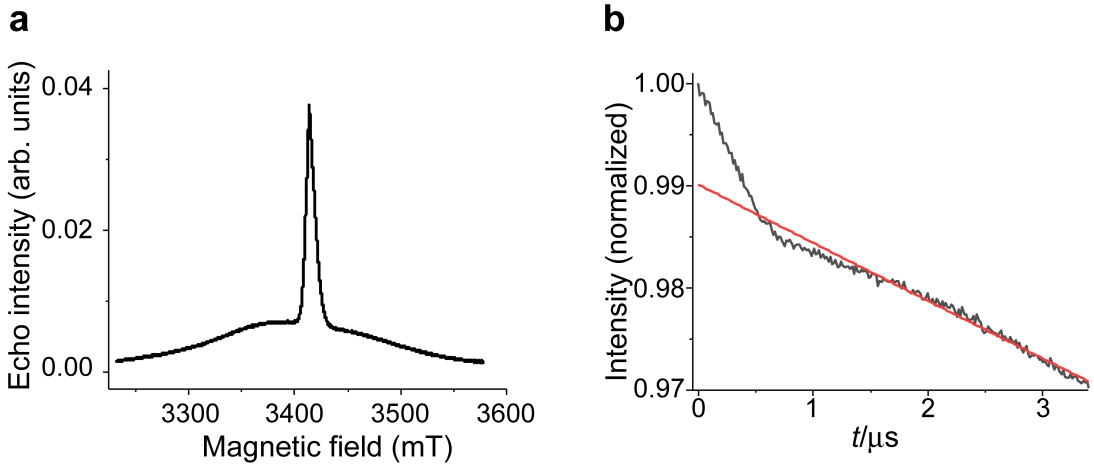

Fig. S2 EPR spectrum and DEER trace of M^{pro}-CC with arsenical tag. (a) Echo-detected field sweep. π/2 and π pulses were 15 ns/30 ns, respectively, and the pulse interval was 500 ns. (b) Primary DEER trace of the data shown in Fig. S3 with the background decay fit shown in red.

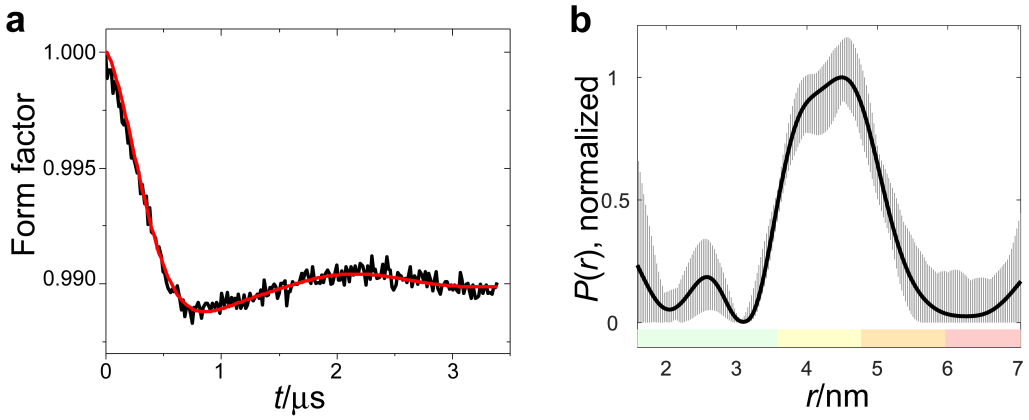

Fig. S3 Results of DEER measurements. (a) DEER trace after background subtraction (black) and fit obtained from the Tikhonov regularization using DeerAnalysis software. (b) Distance distribution resulting from Tikhonov regularization as obtained from the DeerAnalysis validation procedure. The colour bar underneath the distance distribution indicates the reliability regions as defined in DeerAnalysis and determined by the DEER evolution time (green: the shape of the distance distribution is reliable; yellow: the mean distance and distribution width are reliable; orange: the mean distance is reliable; red: unreliable long-range distances). The solid lines

represent the distribution with the smallest r.m.s.d. to the experimental data and the striped regions indicate the range of alternative distributions (±2 times the standard deviation) obtained by varying the parameters of the background correction and noise as calculated by the validation tool in the DeerAnalysis software package with default values. In the crystal structure of the M^{pro} homodimer (PDB ID 7NTS),[10] the distance between the C-termini of the individual monomers is about 3 nm.

References

1 C. Schmitz, M. J. Stanton-Cook, X. C. Su, G. Otting and T. Huber, *J Biomol NMR*, 2008, **41**, 179–189.

2 H. L. Frericks Schmidt, L. J. Sperling, Y. G. Gao, B. J. Wylie, J. M. Boettcher, S. R. Wilson and C. M. Rienstra, *J. Phys. Chem. B*, 2007, **111**, 14362–14369.

3 S. Vijay-Kumar, C. E. Bugg and W. J. Cook, *J. Mol. Biol.*, 1987, **194**, 531–544.

4 F. B. Deans and C. Eaborn, *J. Chem. Soc.*, 1957, 498–499.

5 G. Félix, J. Dunoguès and R. Calas, *Angew. Chem. Int. Ed.*, 1979, **18**, 402–404.

6 M. Qi, M. Hülsmann and A. Godt, *Synthesis*, 2016, **48**, 3773–3784.

7 A. Picot, C. Feuvrie, C. Barsu, F. Malvolti, B. Le Guennic, H. Le Bozec, C. Andraud, L. Toupet and O. Maury, *Tetrahedron*, 2008, **64**, 399–411.

8 C. Neylon, S. E. Brown, A. V. Kralicek, C. S. Miles, C. A. Love and N. E. Dixon, *Biochemistry*, 2000, **39**, 11989–11999.

9 R. Qi and G. Otting, *PLoS One*, 2019, **14**, e0211065.

10 J. Johansen-Leete, S. Ullrich, S. E. Fry, R. Frkic, M. J. Bedding, A. Aggarwal, A. S. Ashhurst, K. B. Ekanayake, M. C. Mahawaththa, V. M. Sasi, T. Passioura, M. Larance, G. Otting, S. Turville, C. J. Jackson, C. Nitsche and R. J. Payne, *bioRxiv*, 2021, DOI: 10.1101/2021.08.23.457419.

11 M. Pannier, S. Veit, A. Godt, G. Jeschke and H. W. Spiess, *J. Magn. Reson.*, 2000, **142**, 331–340.

12 D. Goldfarb, Y. Lipkin, A. Potapov, Y. Gorodetsky, B. Epel, A. M. Raitsimring, M. Radoul and I. Kaminker, *J. Magn. Reson.*, 2008, **194**, 8–15.

13 G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger and H. Jung, *Appl. Magn. Reson.*, 2006, **30**, 473–498.