

Supplementary Information

Altered conformational sampling along an evolutionary trajectory changes the catalytic activity of an enzyme

Joe A. Kaczmariski^{1,7}, Mithun C. Mahawaththa^{1,7}, Akiva Feintuch^{2,7}, Ben E. Clifton^{1,6}, Luke A. Adams³, Daniella Goldfarb^{2,8,*}, Gottfried Otting^{1,4,8,*} & Colin J. Jackson^{1,4,5,8,*}

¹ Research School of Chemistry, The Australian National University, Canberra, ACT 2601, Australia.

² Department of Chemical and Biological Physics, Weizmann Institute of Science, Rehovot 76100, Israel.

³ Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, VIC 3052, Australia. (Australian Research Council Training Centre for Fragment Based Design and Monash Fragment Platform)

⁴ Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science.

⁵ Australian Research Council Centre of Excellence in Synthetic Biology.

⁶ Present address: Protein Engineering and Evolution Unit, Okinawa Institute of Science and Technology, 1919-1 Tancha, Onna-son, Okinawa 904-0412, Japan.

⁷These authors contributed equally: Joe A. Kaczmariski, Mithun C. Mahawaththa, Akiva Feintuch

⁸These authors jointly supervised this work: Daniella Goldfarb, Gottfried Otting, Colin J. Jackson

*Corresponding authors: daniella.goldfarb@weizmann.ac.il ; gottfried.otting@anu.edu.au ; colin.jackson@anu.edu.au

Supplementary Table 1. Data collection and refinement statistics for the AncCDT-5 X-ray crystal structure.

AncCDT-5 (HEPES)	
PDB	6WUP
Data collection	
Space group	P 6 ₅ 2 2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	70.95, 70.95, 175.39
α , β , γ (°)	90, 90, 120
Resolution (Å)	35.69–1.49 (1.52–1.49)*
<i>R</i> _{merge}	0.083 (3.91)
<i>R</i> _{pim}	0.014 (0.729)
<i>I</i> / σ <i>I</i>	22.4 (1.40)
CC _{1/2}	1.00 (0.501)
Completeness (%)	100.0 (100.0)
Redundancy	35.3 (30.3)
Refinement	
Resolution (Å)	43.85–1.49 (1.52–1.49)
No. reflections	43560 (4265)
<i>R</i> _{work} / <i>R</i> _{free}	0.1975 (0.2972) / 0.2217 (0.3129)
No. atoms	2194
Protein	1882
Ligand/ion	16
Water	296
<i>B</i> -factors	32.5
Protein	31.0
Ligand/ion	28.9
Water	42.1
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	0.89

* Statistics for the highest-resolution shell are shown in parentheses.

Supplementary Table 2. Ca–Ca distances (tag positions) and calculated Gd(III)–Gd(III) distances for crystal structures and representative MD snapshots.^a

Name of Crystal Structure/MD Snapshot	Ca–Ca distance (nm)	Calculated Gd(III)–Gd(III) distance (nm)	Name of Crystal Structure/MD Snapshot	Ca–Ca distance (nm)	Calculated Gd(III)–Gd(III) distance (nm)
L-Arg AncCDT-1 (PDB 5T0W)	2.61	2.94	HEPES-AncCDT-5 (PDB 6WUP)	2.77	3.09
MD_Anc1_closed_01	2.52	2.89	MD_Anc5_closed_01	2.50	2.79
MD_Anc1_closed_02	2.63	2.88	MD_Anc5_closed_02	2.56	3.09
MD_Anc1_closed_03	2.73	3.28	MD_Anc5_closed_03	2.63	3.11
MD_Anc1_closed_04	2.82	3.18	MD_Anc5_closed_04	2.71	3.42
MD_Anc1_closed_05	2.91	3.39	MD_Anc5_closed_05	2.79	3.41
apo-AncCDT-1 (PDB 5TUJ)	3.36	4.06	MD_Anc5_open_01	3.01	3.40
MD_Anc1_open_01	3.19	3.87	MD_Anc5_open_02	3.11	3.67
MD_Anc1_open_02	3.40	4.90	MD_Anc5_open_03	3.23	4.00
MD_Anc1_open_03	3.53	4.19	MD_Anc5_open_04	3.34	4.18
MD_Anc1_open_04*	3.63	4.10	MD_Anc5_open_05	3.44	4.17
MD_Anc1_open_05	3.65	4.30	MD_Anc5_wide_01	3.56	4.54
MD_Anc1_wide_01	3.79	4.53	MD_Anc5_wide_02	3.68	4.52
MD_Anc1_wide_02	3.95	5.23	MD_Anc5_wide_03	3.93	5.12
MD_Anc1_wide_03	3.97	4.68	MD_Anc5_wide_04	4.06	5.16
MD_Anc1_wide_04	4.07	5.20	MD_Anc5_wide_05	4.21	5.44
MD_Anc1_wide_05	4.17	4.99	HEPES-PaCDT (PDB 3KBR)	2.53	3.47
AncCDT-3/L188 (PDB 5JOS)	4.21	5.25	Acetate-PaCDT (PDB 5HPQ)	2.49	3.45
MD_Anc3_L188_wide_01	4.04	5.03	MD_PaCDT_closed_01	2.58	3.49
MD_Anc3_L188_wide_02	4.07	5.42	MD_PaCDT_closed_02	2.67	3.70
MD_Anc3_L188_wide_03	4.26	5.54	MD_PaCDT_closed_03	2.76	3.84
MD_Anc3_L188_wide_04	4.46	5.85	MD_PaCDT_closed_04	2.82	3.54
MD_Anc3_L188_wide_05	4.67	6.24	MD_PaCDT_closed_05	2.96	3.85
MD_Anc3_P188_wide_01	4.10	5.27	MD_PaCDT_open_01	3.15	3.99
MD_Anc3_P188_wide_02	4.27	5.57	MD_PaCDT_open_02	3.33	4.06
MD_Anc3_P188_wide_03	4.56	5.73	MD_PaCDT_open_03	3.59	4.40
MD_Anc3_P188_wide_04	4.68	6.22	MD_PaCDT_open_04	3.61	4.46
MD_Anc3_P188_wide_05	4.78	6.34	MDPaCDT_open_05	3.81	4.40
			apo-TmArgBP (PDB 4PRS)	4.00	

^a Gd(III)–Gd(III) distances were calculated following modelling of the AzF-propargyl-DO3A-Gd(III) residue at each mutation site on structures obtained from crystallography (bold) or MD simulations. The models were built by identifying the most favourable values of the dihedral angles χ_1 and χ_2 of the tag (using the mutation tool of PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LCC.) and setting χ_6 to 180°¹. For the MD snapshots, structures with

varying C α –C α distances (“MD_...._01-05”) were selected as representative structures from each conformational state (i.e. closed, open, wide-open). The data points representing these structures are shown as colored crosses in Figure 4.

*This snapshot was initially grouped as being open (based on the C α –C α distance), but was allocated into the wide-open group based on PCA analysis and radius of gyration measurements.

Supplementary Table 3. Sites mutated for the incorporation of AzF residues.

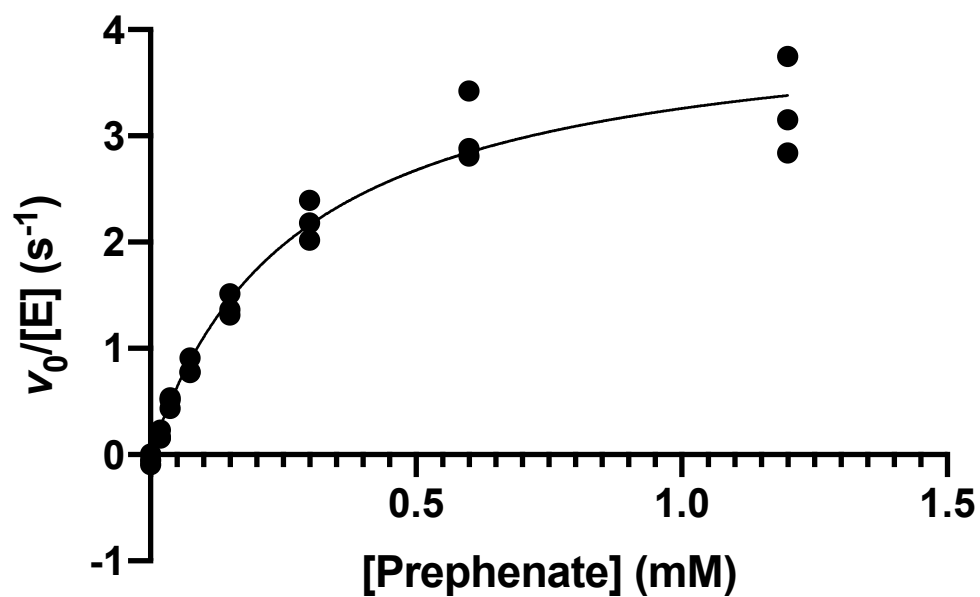
Protein	Mutation sites^a
AncCDT-1	Q68/K138; Q68/Q219; K138/S161
AncCDT-3	Q68/R138
AncCDT-5	Q68/R138
PaCDT^b	A68/A139
SeLAOBP	A61/T151

^a Amber stop codons were introduced to incorporate pairs of AzF residues. The positions in the amino acid sequence are indicated together with the original amino acid types.

^b The mutant A68/R138 was also prepared, but the tagging yields obtained were too low for DEER experiments.

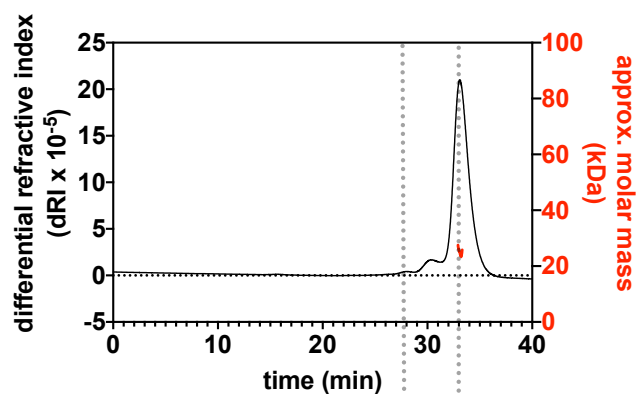
Supplementary Table 4. Primers used in this study.

Primer Name	Sequence
AncCDT-1_Q68F	CCGGCACTGTAGACCGGTAAATTTGATATC
AncCDT-1_Q68R	CAAATTTACCGGTCTACAGTGCCGGAATAATAC
AncCDT-1_K138F	GGCAGCATAGGAATTTCTGCCG
AncCDT-1_K138R	AGAAATTCCTATGCTGCCTGTTGCTG
AncCDT-1_Q219F	TGGCTGAAATAGATGAAAAAGACGGC
AncCDT-1_Q219R	GCCGTCTTTTTTCATCTATTTTCAGCCAATTATTC
AncCDT-1_S161F	GAGGTTGTTTAGGGTCGTGCAGATGCAATG
AncCDT-1_S161R	CACGACCCTAAACAACCTCTTGAAAG
AncCDT-3_Q68F	GCGATCTGTAGGCAGATAAATTTGATATCGC
AncCDT-3_Q68R	TATCTGCCTACAGATCGCTCATCAGGGTCG
AncCDT-3_R138F	GAAAAATTTGCATAGGAACATCTGAAAAAGCCAAAATCACCG
AncCDT-3_R138R	GGCTTTTTTCAGATGTTCTATGCAAATTTTTCATTGGTGCC
AncCDT-5_Q68F	GGATGATTTTTAGGCAGATAAATTTGATATCGCC
AncCDT-5_Q68R	ATCTGCCTAAAAATCATCCATCAGGGTCGG
AncCDT-5_R138F	CGTTTTGCATAGGCCCATCTGAAACAGGCACAG
AncCDT-5_R138R	CAGATGGGCCTATGCAAACGTTTCATTGGTGC
PaCDT_A68F	GCGTGATTTTTAGGATGATCGTTTTGATATTGCCATGAG
PaCDT_A68R	ACGATCATCCTAAAAATCACGCATCAGATTCGGCCAG
PaCDT_A139F	GCACGTTAGAATCTGAAAAAGCACGCATTCTGG
PaCDT_A139R	TCAGATTCTAACGTGCAAATTTTTCATTGGTGCC
SeLAOBP_A61F	CTGAAATAGAAAAAATCGATGCAATTATTAGCAGC
SeLAOBP_A61FR	TTTTTTTCTATTTTCAGGCTCGGAATCAGTGC
SeLAOBP_T151F	GATCTGTAGGCAGGTCGTCTGGATGCAGCACTGCAG
SeLAOBP_T151R	CTGCCTACAGATCGCTATAGATCAGATCCTGATTTGC
AncCDT-1 F	GTTTAACTTTAAGAAGGAGATATACATATG ATCGCAGCAAGCACCCCTG
AncCDT-1 R	GCCCTGAAAATACAGGTTTTTCGAGTTTAAACCATTTTTTCGTACAGTTTGTCATAGG
AncCDT-3 F	ATGGAAAACCTGTATTTTCAGGGCATCGCAGCAAGCCGTCTG
AncCDT-3 R	TACCCGGGAGCTCGAATTCTTATTAGAGTTTCAGCCATTTCTCGTACAGTTTTTC
AncCDT-5 F	ATGGAAAACCTGTATTTTCAGGGCATCGCAGCAAGCCGTCTG
AncCDT-5 R	TACCCGGGAGCTCGAATTCTTAttAGAGTTTCAGCCATTTATCAAAAATACGCTG
PaCDT F	ATGGAAAACCTGTATTTTCAGGGCATCCAAGAAAGCCGTCTGG
PaCDT R	TACCCGGGAGCTCGAATTCTTATTAGAGTTTACCATGTGCGGTCTG
SeLAOBP F	ACCTGTATTTTCAGGGCGCACTGCCGCAGACCGTTCTG
SeLAOBP R	CCCGGGAGCTCGAATTCTTAATCGCCATACACGTTAAAGTCGAAGTATTTTTTC
pETMCSIII F	GAATTCGAGCTCCCGGGTAC
pETMCSIII R	GCCCTGAAAATACAGGTTTTCCATATG

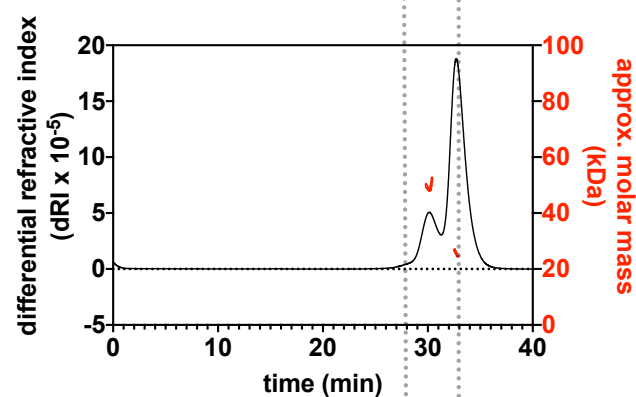


Supplementary Figure 1. Michaelis-Menten plot for AncCDT-5 prephenate dehydratase activity. Data points from three technical replicates are shown ($n = 3$). Source data are provided as a Source Data file.

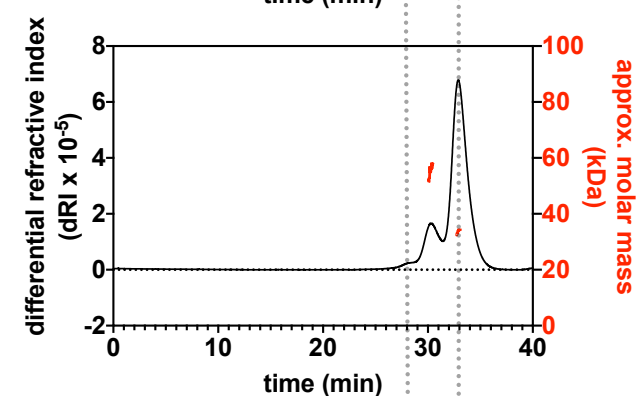
(A) AncCDT-1



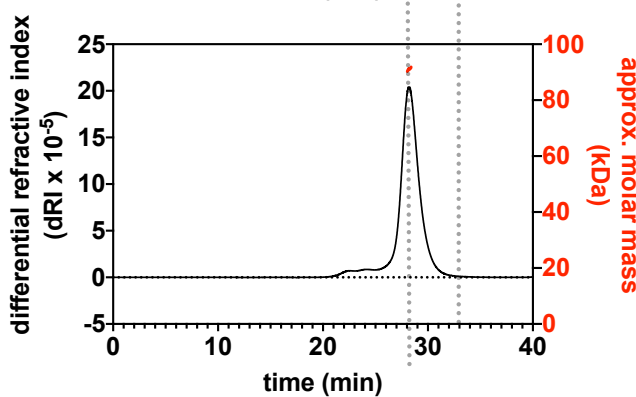
(B) AncCDT-3/P188



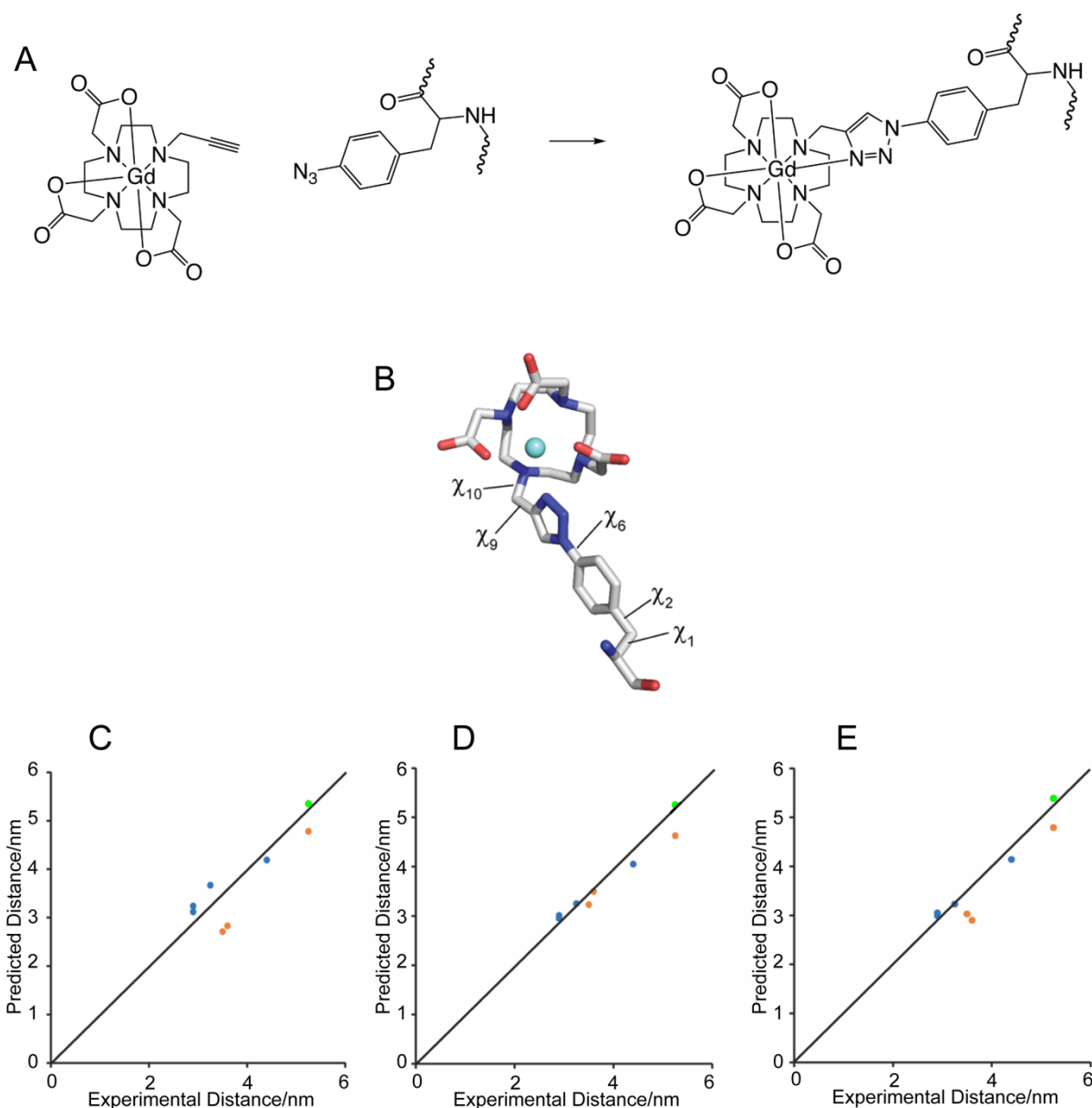
(C) AncCDT-5



(D) PaCDT

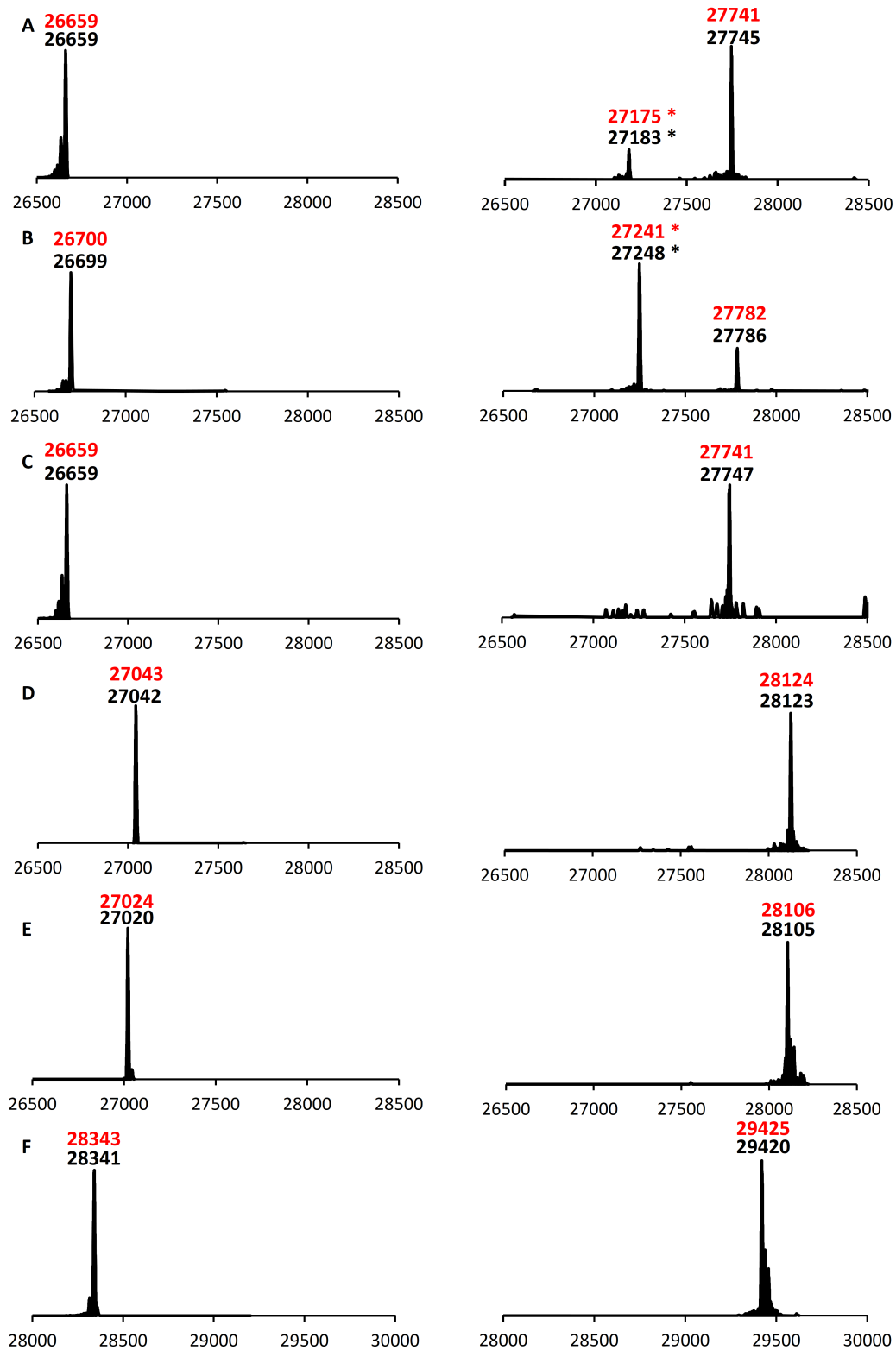


Supplementary Figure 2. SEC-MALS of proteins. Results from analytical size-exclusion chromatography coupled with multiangle light scattering (SEC-MALS) for (A) AncCDT-1, (B) AncCDT-3/P188, (C) AncCDT-5 and (D) PaCDT.



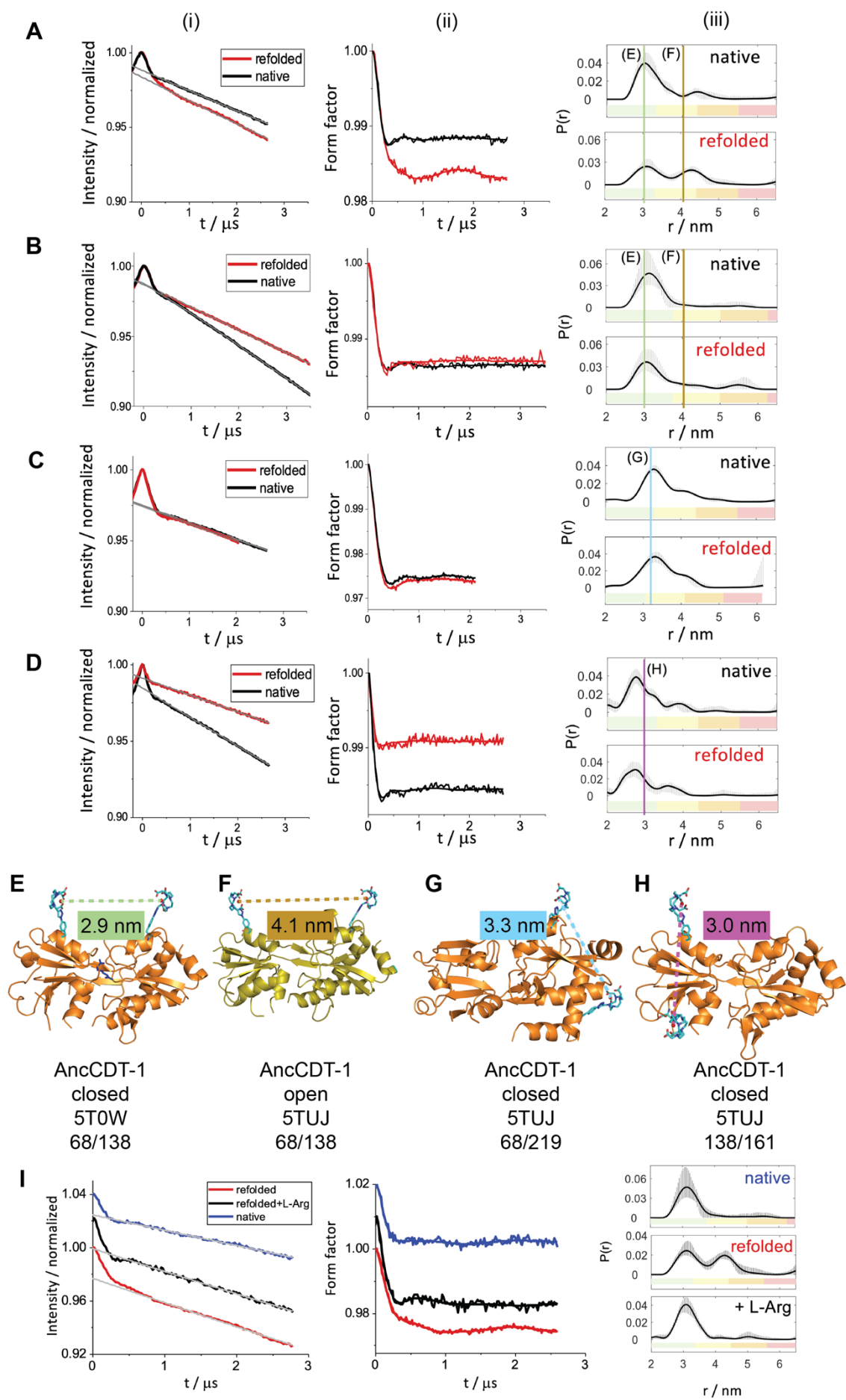
Supplementary Figure 3. The AzF-propargyl-DO3A-Gd(III) residue and correlations between modelled and experimental Gd(III)–Gd(III) distances. (A) Scheme of the Cu(I)-catalyzed cycloaddition reaction used to attach the propargyl-DO3A-Gd(III) tag to the AzF residue⁴. (B) Dihedral angles in the AzF-propargyl-DO3A-Gd(III) residue. $\chi_6 = 0^\circ$ in the conformation shown. (C) Correlation between modelled and experimental Gd(III)–Gd(III) distances, using the crystal structures of AncCDT-1 (blue; PDB ID 5T0W: mutants 68/138, 68/219, 138/161; PDB ID 5TUI: mutant 68/138), AncCDT-3/P188 (green; PDB ID 5JOS: mutant 68/138) and PaCDT (orange; PDB ID 3KBR: single mutants 68 and 139, and double mutant 68/139). The distances correspond to the maxima of the respective experimental or modelled DEER distance distributions. Distances were predicted using $\chi_6 = 0^\circ$. The angles χ_9 and χ_{10} were

fixed to -140° and 70° , respectively, to allow coordination of the metal ion (cyan sphere in B) by the nearest nitrogen of the triazole ring. χ_1 and χ_2 angles were selected by using those conformations, which were identified by the mutation tool of the program PyMOL to generate minimal steric clashes. (D) Same as (C), but using $\chi_6 = 180^\circ$. This conformation was used to represent the modelled distances in all other figures of the present work. (E) Same as (D), but χ_9 was allowed to be random while the χ_{10} angle was randomly set to -60° , 60° , or 180° with equal likelihood. Although the variation of these dihedral angles broadened the modelled distance distributions, the maxima of the Gd(III)–Gd(III) distance distributions did not change by more than 0.7 nm.

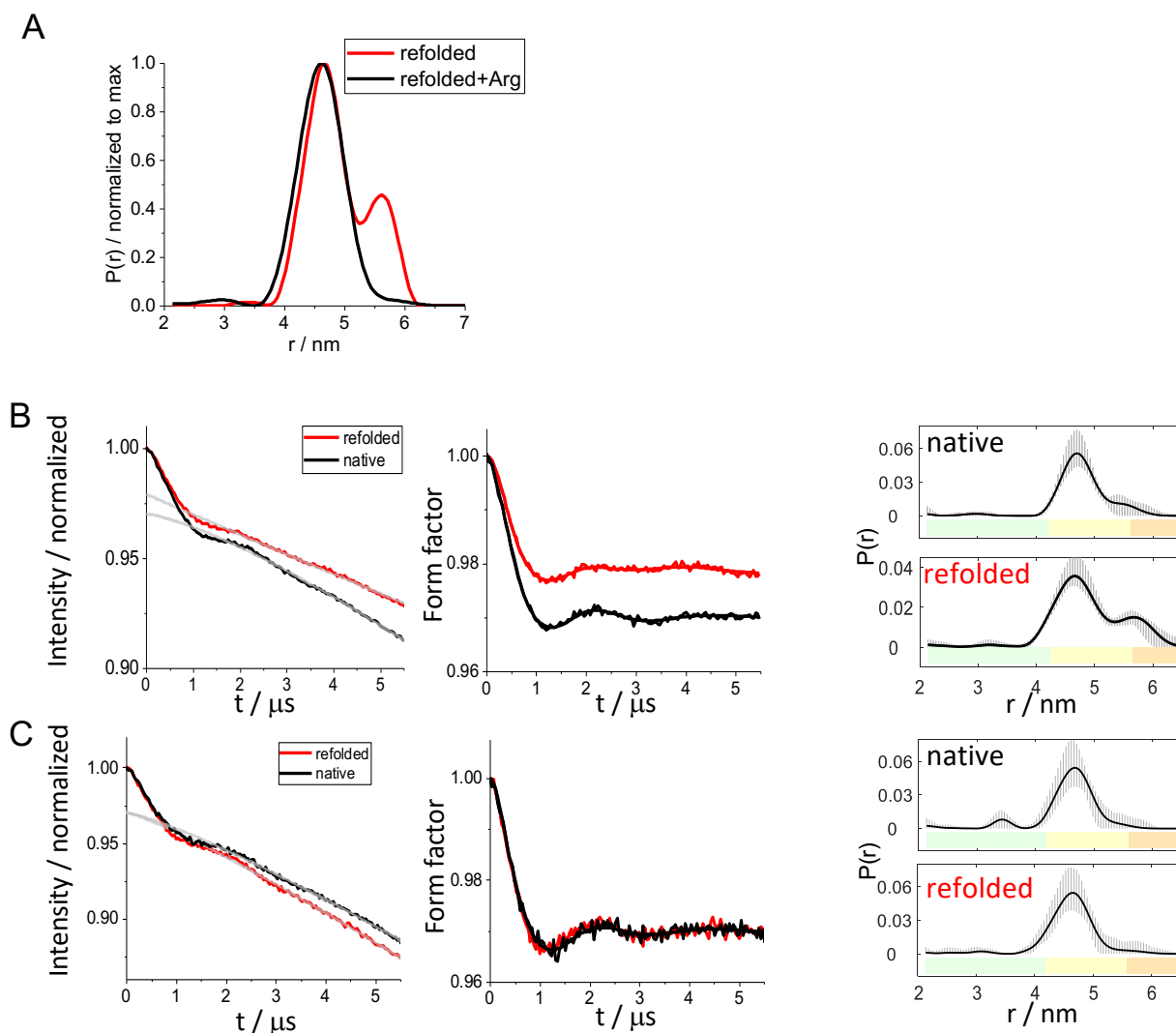


Supplementary Figure 4. Mass spectra of protein samples before and after tagging. The left and right panels show the mass spectra before and after tagging, respectively, together with the expected masses (red). Stars identify mass peaks attributed to singly tagged protein. (A) Refolded AncCDT-1, AzF in positions 68 and 138. (B) Refolded AncCDT-1, AzF in positions 138 and 161.

(C) AncCDT-1, AzF in positions 68 and 219. (D) AncCDT-3, AzF in positions 68 and 138. (E) AncCDT-5, AzF in positions 68 and 138. (F) *Pa*CDT, AzF in positions 68 and 139.

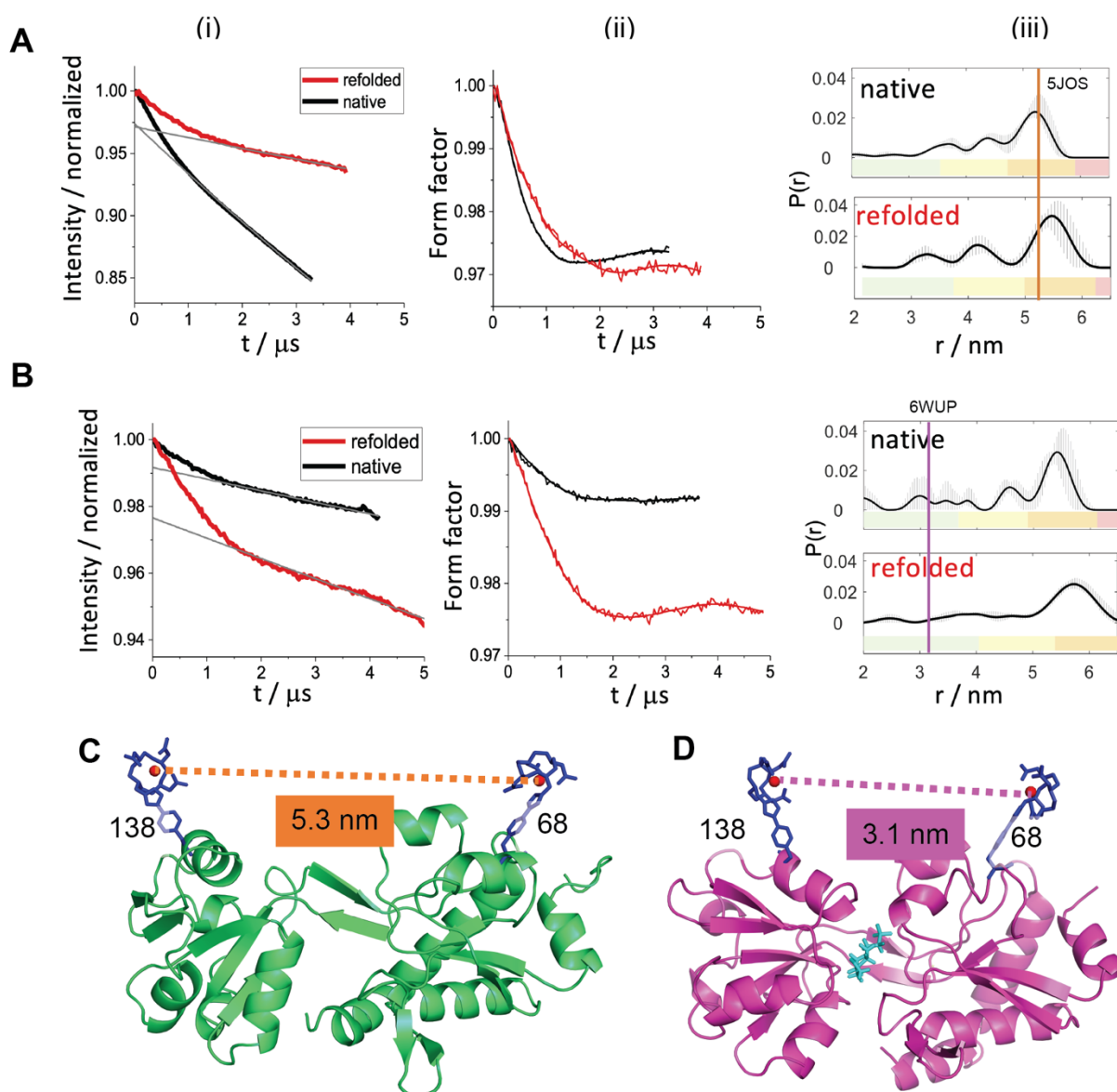


Supplementary Figure 5. Additional and primary DEER data for AncCDT-1. Data from DEER measurements on natively purified (black) and refolded (red) AncCDT-1 samples (100 μ M protein in D₂O and 20 % w/v deuterated glycerol) tagged at positions (A) 68 and 138 in the absence of exogenous L-Arg, (B) 68 and 138 in the presence of 1.5-fold molar excess L-Arg, (C) 68 and 219, or (D) 138 and 161. The left panel (i) shows the primary DEER data for the sample and background (gray) decay function. The central panel (ii) shows the DEER form factors after background correction, with fits corresponding to the distance distributions shown in (iii). The right panel (iii) shows the corresponding distance distributions with color coding of the reliability regions as defined in DeerAnalysis, corresponding to the DEER evolution time used (Pale green: the shape of the distance distribution is reliable. Pale yellow: the mean distance and distribution width are reliable. Pale orange: the mean distance is reliable. Pale red: long-range distance contributions may be detectable, but cannot be quantified.). The solid lines represent the distributions with the best r.m.s.d. from the experimental data and the striped regions represent the variation 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 package⁵. The parameter ranges used for the validations were: white noise 0–1.5, background start $0.2 \cdot t_{\text{max}} - 0.6 \cdot t_{\text{max}}$, and background dimension 3–3.6. A limited variation in background dimension was allowed, as a non-homogeneous background improved the fit for some of the traces. Vertical lines represent the predicted Gd(III)–Gd(III) distances when AzF-propargyl-DO3A-Gd(III) residues are modelled on crystal structures of (E) closed AncCDT-1 (PDB 5T0W) at positions 68 and 138, (F) open AncCDT-1 (PDB 5T0J) at positions 68 and 138, (G) closed AncCDT-1 at positions 68 and 219, or (H) closed AncCDT-1 at positions 138 and 161. (I) Data from a replicate experiment, in which DEER measurements were made on native (blue), refolded (red) and refolded + L-Arg (black) samples of AncCDT-1 (tagged at positions 68 and 138) – as described for (A) and (B).



Supplementary Figure 6. DEER data for *SeLAOBP* samples. (A) Comparison of Gd(III)–Gd(III) distance distributions of refolded *SeLAOBP* tagged at positions 61 and 151, in the absence (red) and presence (black) of a 1.5 molar excess of L-arginine. (B–C) Raw and processed data from DEER measurements of natively purified (black) and refolded (red) *SeLAOBP* samples in the (B) absence and (C) presence of a 1.5 molar excess of L-arginine. Samples were measured using 100 μM protein in D_2O and 20 % w/v deuterated glycerol. The left panel shows the primary DEER data for the sample and background (gray) decay function. The central panel shows the DEER form factors after background correction, with fits corresponding to the distance distributions shown in the right-hand side panel. The right panel shows the corresponding distance distributions: the solid lines represent the distributions with the best r.m.s.d. from the experimental data and the striped regions represent the variation 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 package⁵. The parameter ranges used for the

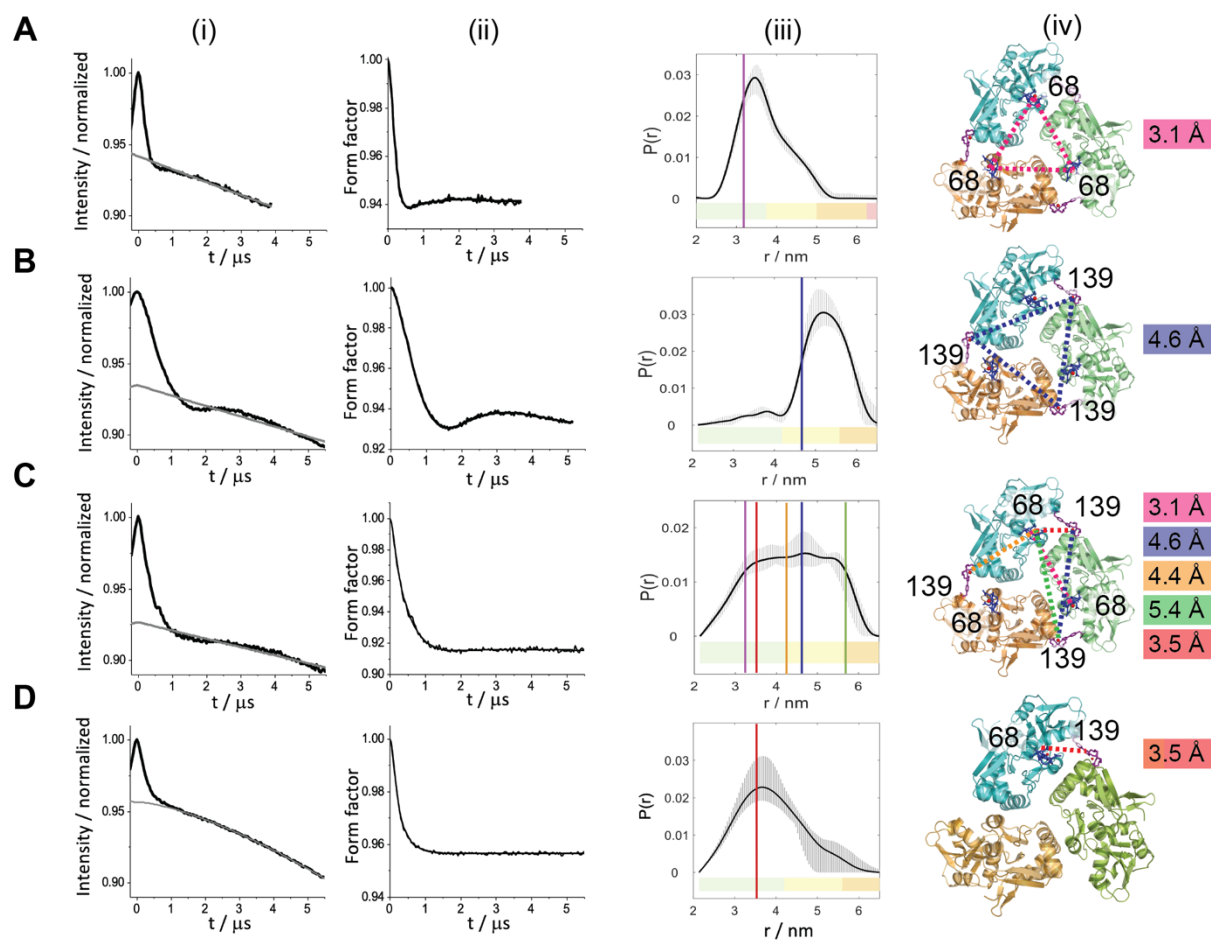
validations were: white noise 0–1.5, background start $0.2 \cdot t_{\max}$ – $0.6 \cdot t_{\max}$, and background dimension 3–3.6. The colored bars at the bottom of each panel reflect the reliability of each region of the distribution, according to DEER evolution time used, as defined in DeerAnalysis (pale green: the shape of the distance distribution is reliable; pale yellow: the mean distance and distribution width are reliable; pale orange: the mean distance is reliable; pale red: long-range distance contributions may be detectable, but cannot be quantified).



Supplementary Figure 7. Primary DEER data for AncCDT-3/P188 and AncCDT-5 samples.

Primary data from the DEER measurements (as shown in Figure 3) on natively purified (black) and refolded (red) of (A) AncCDT-3/P188 tagged at positions 68 and 138, and (B) AncCDT-5 tagged at positions 68 and 138. Samples were measured using 100 μM protein in D_2O and 20 % w/v deuterated glycerol. The left panel (i) shows the primary DEER data for the sample and background (gray) decay function. The central panel (ii) shows the DEER form factors after background correction, with fits corresponding to the distance distributions shown in (iii). The right panel (iii) shows the corresponding distance distributions: the solid lines represent the distributions with the best r.m.s.d. from the experimental data and the striped regions represent the variation 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 package⁵. The parameter ranges used for the validations were: white noise 0–1.5, background start $0.2 \cdot t_{\text{max}} - 0.6 \cdot t_{\text{max}}$, and background dimension 3–3.6. The colored bars at the bottom of each panel reflect the reliability of each region of the distribution, according to DEER evolution time used, as defined in DeerAnalysis (pale green: the shape of the distance distribution is reliable; pale yellow: the mean distance and distribution width are reliable; pale orange: the mean distance is reliable; pale red: long-range distance contributions may be detectable, but cannot be quantified). In (B), the primary DEER traces were shortened in the analysis to avoid an instrumental artefact. Vertical lines represent the predicted Gd(III)–Gd(III) distances when AzF-propargyl-DO3A-Gd(III) residues are modelled onto crystal structures of AncCDT-3/L188 (PDB 5JOS) at positions 68 and 138, or AncCDT-5 (PDB 6WUP) at positions 68 and 138. As the main peak in each DEER distance distribution appears in the orange region, we also calculated means of the distance distributions to assess potential contributions of the smaller peaks nearby. We considered two ranges for the calculation, where the first range (4.0–6.5 nm) included the small peak appearing in the yellow region next to the main peak while the second range included only the most intense peak (5.0–6.5 nm). The mean distance values obtained were: (A) native: 4.96 ± 0.42 and 5.18 ± 0.26 nm (original peak maximum at 5.2 nm); (A) refolded: 5.15 ± 0.57 and 5.44 ± 0.24 nm (original peak maximum at 5.4 nm); (B) native: 5.26 ± 0.41 nm and 5.47 ± 0.20 nm (original peak maximum at 5.4 nm); (B) refolded: 5.62 ± 0.24 (original peak maximum at 5.7 nm).



Supplementary Figure 8. Additional and primary DEER data for *Pa*CDT samples. Data from the DEER measurements on natively purified samples of *Pa*CDT tagged at (A) position 68, (B) position 139, (C) position 68 and 139, or (D) position 68 and 139, but following a 10-fold dilution with unlabelled *Pa*CDT for selective observation of the intramolecular distance between the two domains (as shown in Figure 3). Samples were measured using 100 μ M protein in D₂O and 20 % w/v deuterated glycerol. The left panel (i) shows the primary DEER data for the sample and background (gray) decay function. The central panel (ii) shows the DEER form factors after background correction, with fits corresponding to the distance distributions shown in (iii). The right panel (iii) shows the corresponding distance distributions: the solid lines represent the distributions with the best r.m.s.d. from the experimental data and the striped regions represent the variation 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 package⁵. The parameter ranges used for the validations of (A)-(C) were: white noise 0–1.5, background start $0.2 \cdot t_{\max}$ – $0.6 \cdot t_{\max}$, and background dimension 3–3.6. The parameter ranges used for the validation of (D) were the same, except that the background dimension was varied in the range of 3–4 to take into account the dimensionality of four we used

to get a good fit. The colored bars at the bottom of each panel reflect the reliability of each region of the distribution as defined in DeerAnalysis (pale green: the shape of the distance distribution is reliable; pale yellow: the mean distance and distribution width are reliable; pale orange: the mean distance is reliable; pale red: long-range distance contributions may be detectable, but cannot be quantified.). In (B), the primary DEER traces were shortened in the analysis to avoid an instrumental artefact. Vertical lines represent the predicted Gd(III)–Gd(III) distances when AzF-propargyl-DO3A-Gd(III) residues are modelled onto the crystal of structure *Pa*CDT (PDB 3KBR) at tagged positions, as shown in panel (iv).

AncCDT-1	1	AASRLDEIMKRGTLRVGTTGADYKPFSFKDKNGQYTGFDIDLA	LAKE	LGVKVEFVPTTW
AncCDT-3/P188	1	AASRLDEIMERGTLRVGTTGADYKPFSYRDPDGQYTGFDIDVAKS	LAKS	LGVKVEFVPTTW
AncCDT-3/L188	1	AASRLDEIMERGTLRVGTTGADYKPFSYRDPDGQYTGFDIDVAKS	LAKS	LGVKVEFVPTTW
AncCDT-5	1	AASRLDEIMQRGTLRVGTTGADYKPFSYRDPDGQYTGFDIDMAES	LAKS	LGVKVEFVPTTW
PaCDT	1	QESRLDRILESGVLRVATTGADYKPFSYRTEEGGYAGFDVDMQR	LAES	LGAKLVVPTSW

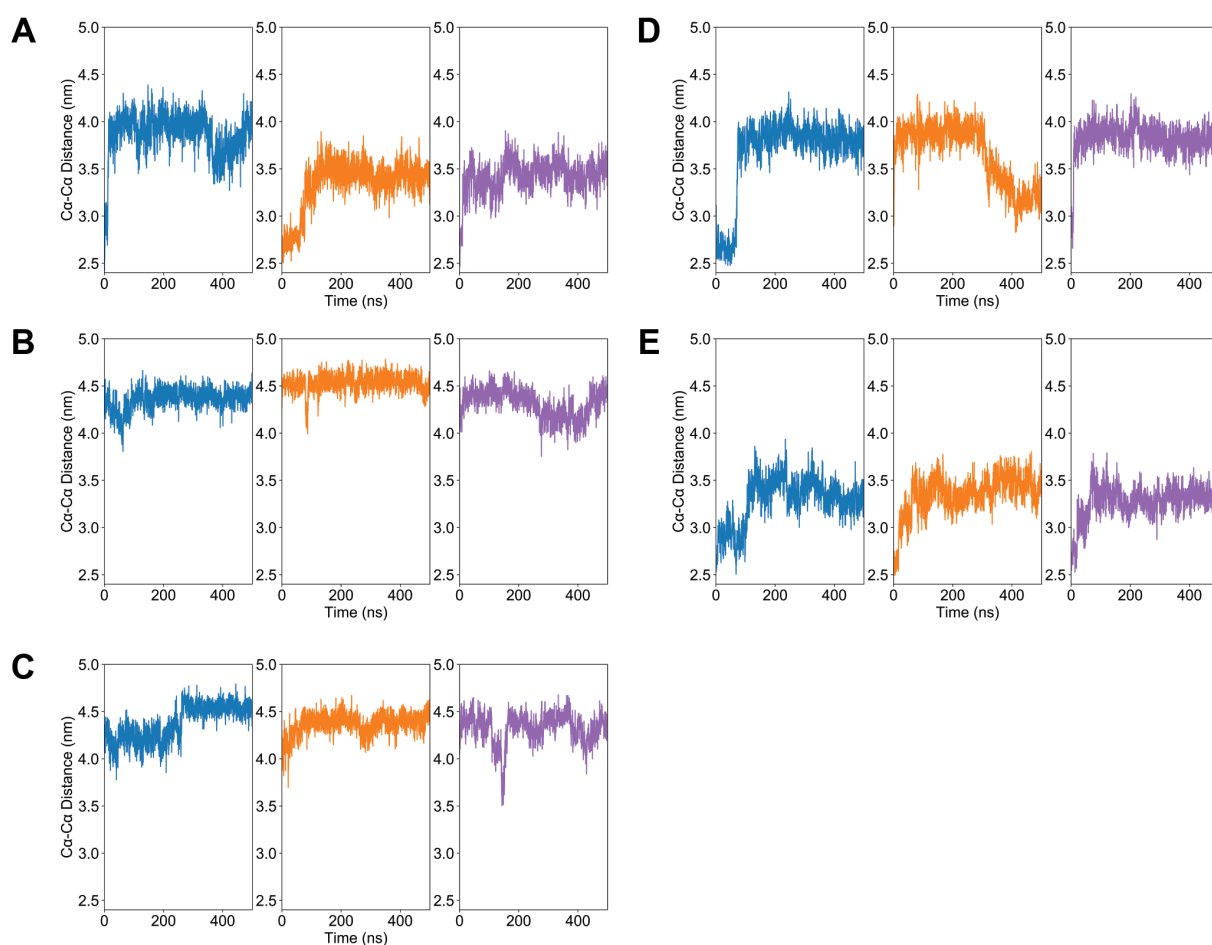
AncCDT-1	61	DGIIPALQTKGFDIVMSGMTITPERKKKVFSDPYMTAGQTILVK	KNADKIKSFE	DLNK		
AncCDT-3/P188	61	PTLMSDLQADKFDIAMGGVTVTPERQKKADFSDPYMTFGKTPLVR	KEDADKFKSLED	INR		
AncCDT-3/L188	61	PTLMSDLQADKFDIAMGGVTVTPERQKKADFSDPYMTFGKTPLVR	KEDADKFKSLED	INR		
AncCDT-5	61	PTLMDDFQADKFDIAMGGVSVTPERQKKADFSEPYMTDGKTPIVR	CEADAKYQTTLE	QIDR		
PaCDT	61	PNLMRDFADDRFDIAMS	GISINLERORQAYFSIPYLRD	GKTPITLCS	FEARFQTTLE	QIDQ

AncCDT-1	121	PDVKVAVQLGTSEQA	AKFELPKAKIRTFENNAEAFQEVVS	GRADAMVTDSPVAAYYAKK
AncCDT-3/P188	121	PDVRVAVNPGGTNEKFAREHLKKAKITVYENNVEIFQEVAS	GRADVMITDTVEALYYAKK	
AncCDT-3/L188	121	PDVRVAVNPGGTNEKFAREHLKKAKITVYENNVEIFQEVAS	GRADVMITDTVEALYYAKK	
AncCDT-5	121	PDVRVVNPGGTNERFARAHKQAQITVYPDNVTIFQEI	VAGRADVMMTDAVETRYQOKL	
PaCDT	121	PGVTAIVNPGGTNEKFA	RANLKKARILVHPDNVTIFQEI	VAGKADLMMTDAIEARLQSR

AncCDT-1	181	NPGLAVVVDKPFTH	EPPLGFAIRKGDPELLN	WVNNWLKQMKKD	GTYDKLYEKWFK			
AncCDT-3/P188	181	HPGLAAVPVDKPFTHS	EKGYYMMPKGDQEFLN	YVNQWLDQMKQO	GTYEKLYEKWLK			
AncCDT-3/L188	181	HPGLAAVLVDKPFTHS	EKGYYMMPKGDQEFLN	YVNQWLDQMKQO	GTYEKLYEKWLK			
AncCDT-5	181	HPGLCAVHVDKPFTHS	EKA	YLLPRGDPAFKA	YVDQWLHQA	MQSGTYQRIFDKWLK		
PaCDT	181	HPELCAVHPQQPFDF	A	EKA	YLLPRDEAFKRY	VVDQWLHIAEQS	GLLRQRMEHWLE	YRWPTA

AncCDT-1	...	
AncCDT-3/P188	...	
AncCDT-3/L188	...	
AncCDT-5	...	
PaCDT	241	HGK

Supplementary Figure 9. Multiple sequence alignment. Multiple sequence alignment of the proteins studied in this work. N-terminal residues that were introduced during cloning, including the His₆-tag, have been omitted. The residue numbering corresponds to that used throughout this work. The sequences were aligned using Clustal Omega². The figure was generated using ESPrpt 3³. Every 10th residue position in the alignment is indicated by a black dot. Shading is based on the percentage of equivalent residues at that position (i.e. black = 100 % conserved).



Supplementary Figure 10. Time evolution of Ca-Ca distances in MD simulations. The time evolution of distances between the alpha carbons of tagging sites in simulations of (A) AncCDT-1, (B) AncCDT-3/P188, (C) AncCDT-3/L188, (D) AncCDT-5 and (E) *PaCDT*. For AncCDT-1, AncCDT-3 and AncCDT-5, individual replicates ($n=3$) are shown as different colored traces. For the *PaCDT* plot, the Ca-Ca distances between the tagged sites of each monomer of the *PaCDT* trimer ($n=1$) are shown as different colored traces.

Supplementary References

1. Mahawaththa, M. C. *et al.* Small neutral Gd(III) tags for distance measurements in proteins by double electron-electron resonance experiments. *Phys. Chem. Chem. Phys.* **20**, 23535–23545 (2018).
2. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
3. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
4. Abdelkader, E. H. *et al.* Protein conformation by EPR spectroscopy using gadolinium tags clicked to genetically encoded *p*-azido-L-phenylalanine. *Chem. Commun.* **51**, 15898–15901 (2015).
5. Jeschke, G. *et al.* DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* **30**, 473–498 (2006).