Supplementary Data

Title: The *Paulinella* chromatophore transit peptide part2 adopts a structural fold similar to the γ-glutamyl-cyclotransferase fold

Authors: Victoria Klimenko[1], Jens Reiners[2], Violetta Applegate[2], Katharina Reimann[1], Grzegorz Popowicz[3], Astrid Hoeppner[2], Athanasios Papadopoulos[2], Sander H.J. Smits[2,4], Eva C.M. Nowack[1]

Affiliations:

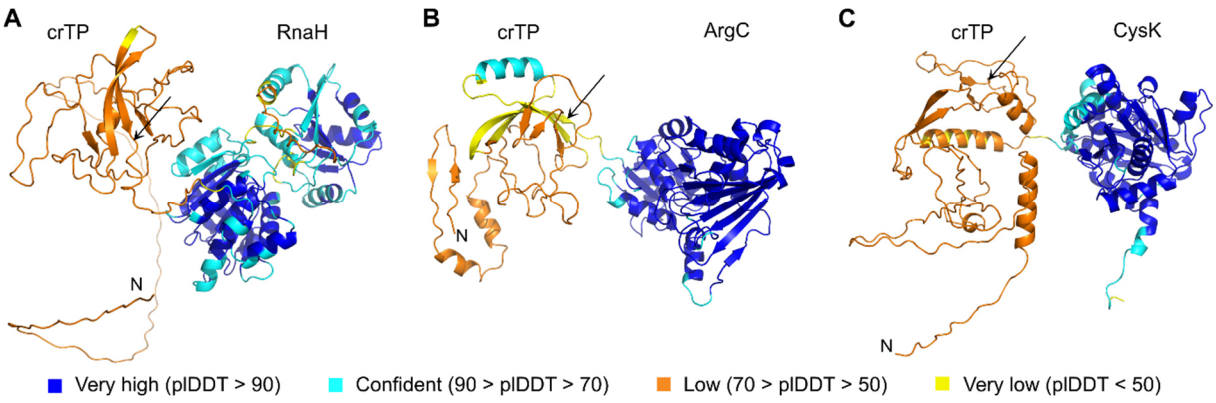
[1]Institute of Microbial Cell Biology, Department of Biology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[2]Center for Structural Studies (CSS), Heinrich Heine University Düsseldorf, Düsseldorf, Germany

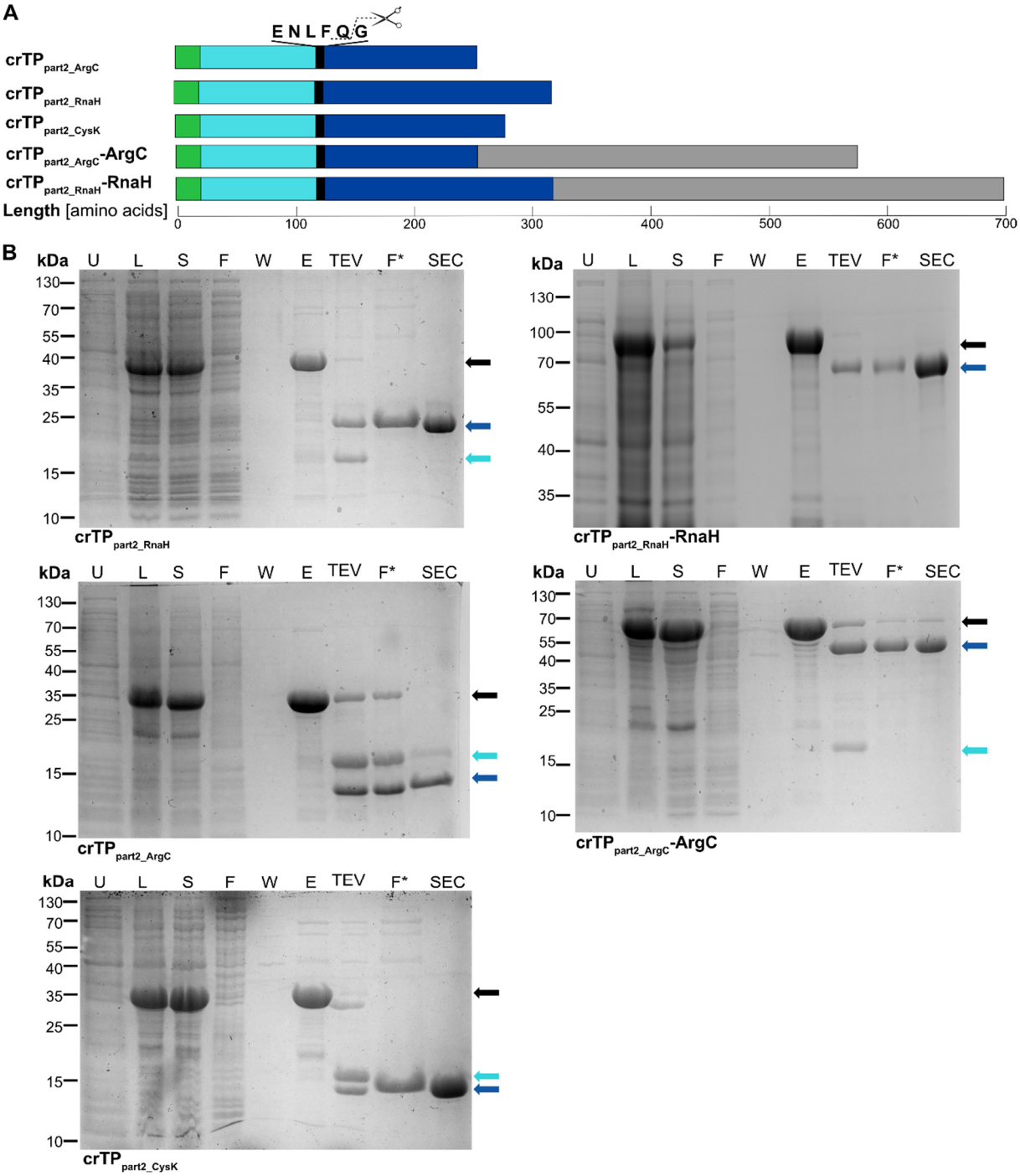
[3]Institute of Structural Biology, Helmholtz Zentrum München, Neuherberg, Germany

[4]Institute of Biochemistry, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

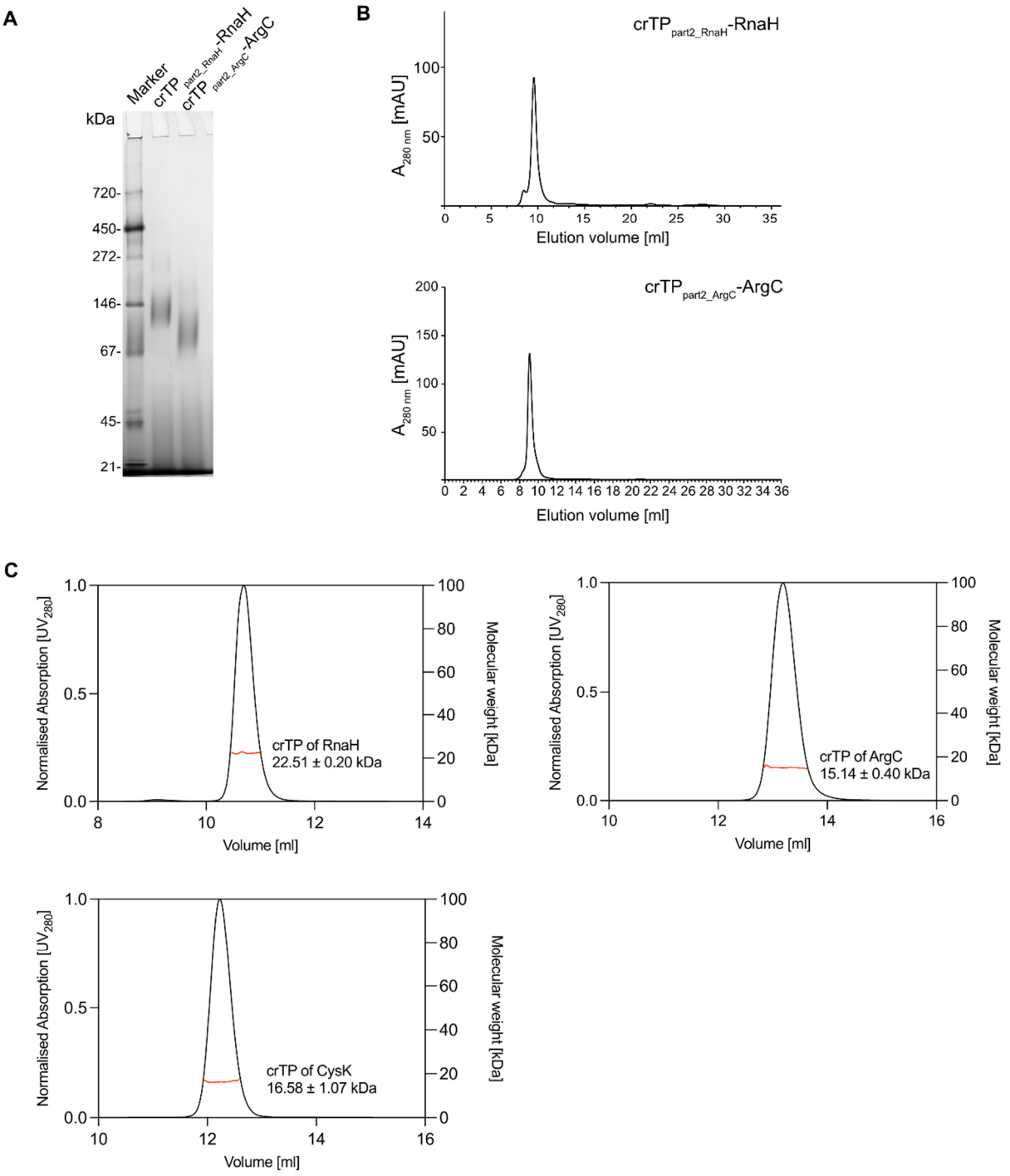
Supplementary Figures



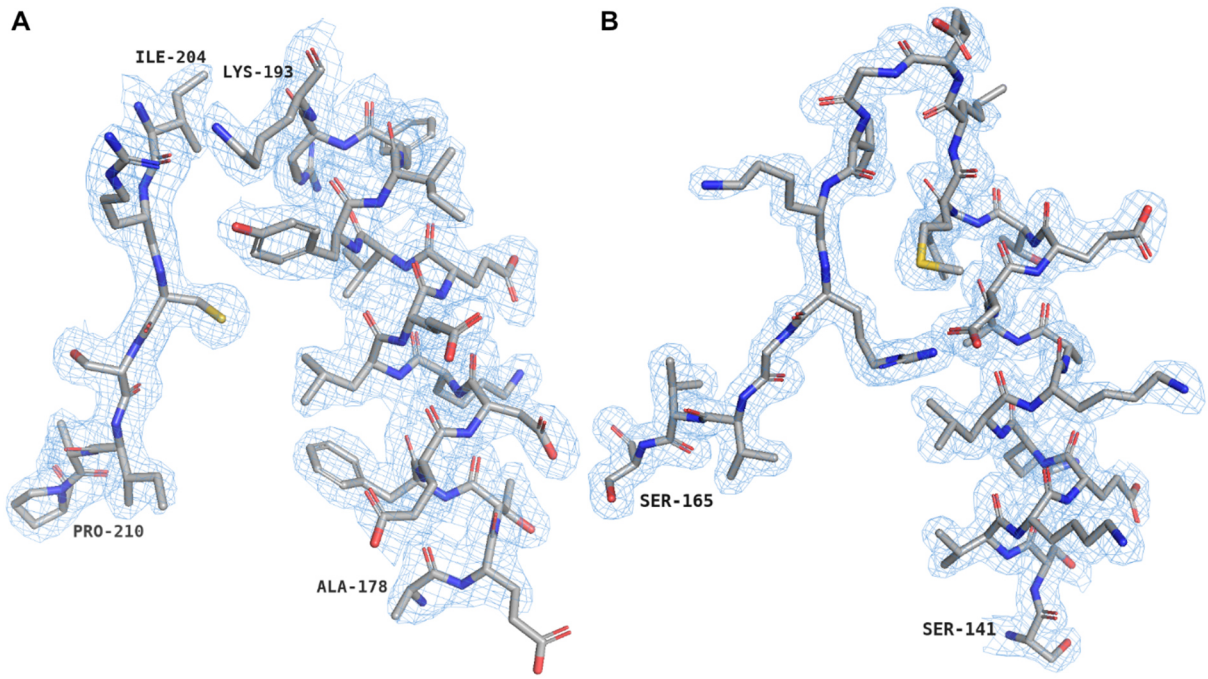
Supplementary Figure S1: AlphaFold3-predicted protein models. Proteins encoded by transcripts scaffold2581 **(A),** scaffold7023 **(B)**, and scaffold7023 **(C)** were analyzed with AlphaFold3. N-termini (N) are indicated and the start of the linker between the crTP and the cargo protein is highlighted by a black arrow. The confidence estimate (pLDDT) is shown as color code in the images of the structures.



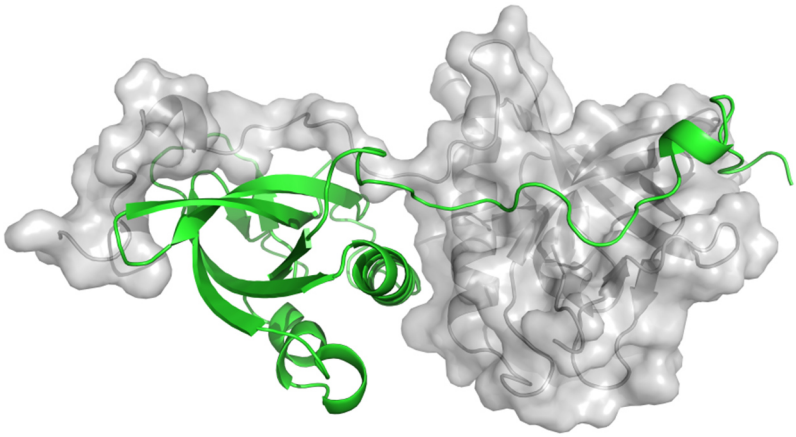
Supplementary Figure S2: Purification of crTP$_{part2}$-containing constructs. (A) Schematic representation of crTP$_{part2}$ constructs. The His$_6$-tag (green) and SUMO solubility-tag (cyan) are cleaved off at the TEV protease recognition site (black) to obtain crTP$_{part2}$ (blue) alone or attached to its corresponding cargo protein (grey). **(B)** All crTP$_{part2}$ constructs were expressed in *E. coli* following induction with IPTG. For un-induced samples (U), 600 µl of expression culture was withdrawn before induction, spun down, and the pellet resuspended in 60 µl PBS. Following expression, lysate (L) was generated spun at 120,000 x g, for 1 h. The supernatant (S) was loaded onto a Ni-NTA column. Column was washed with buffer A. Samples from the flow-through (F) and mid-wash (W) were collected. Proteins of interest were eluted (E). Eluate was diluted 1:2 with buffer A and digested with TEV-protease (TEV). Proteins of interest were isolated by reverse IMAC by collecting the flow-through (F*). Then, proteins of interest were further purified by SEC. Proteins in all samples were solubilized in Laemmli buffer and 5 µl loaded onto a 12.5 % polyacrylamid gel or 10 % for crTP$_{part2_RnaH}$-RnaH. For F* and after SEC, ~5 µg total protein was loaded on the gel. The gel was stained with Coomassie brilliant blue. Full-length constructs are indicated by black arrows, proteins of interest by blue arrows, and His-SUMO (14.2 kDa) by cyan arrows.



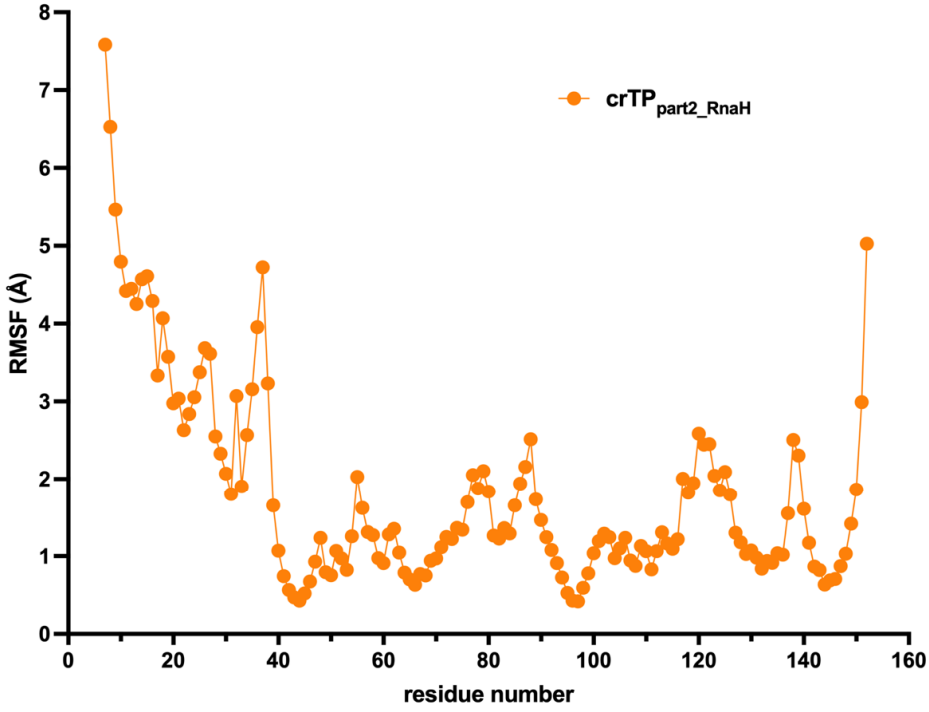
Supplementary Figure S3: Validation of purity and oligomeric states of crTP$_{part2}$-containing constructs. (A) BN PAGE gel of the purified proteins crTP$_{part2_RnaH}$-RnaH and crTP$_{part2_ArgC}$-ArgC. Marker is SERVANativ Marker Liquid Mix for BN/CN (SERVA, Cat. No. 39219). **(B)** SEC profiles for crTP$_{part2_RnaH}$-RnaH and crTP$_{part2_ArgC}$-ArgC. **(C)** SEC-MALS profiles for the smaller constructs (crTP$_{part2_RnaH}$, crTP$_{part2_ArgC}$, and crTP$_{part2_CysK}$). Red lines represent molecular mass distribution (in kDa) over the peak as determined by MALS.

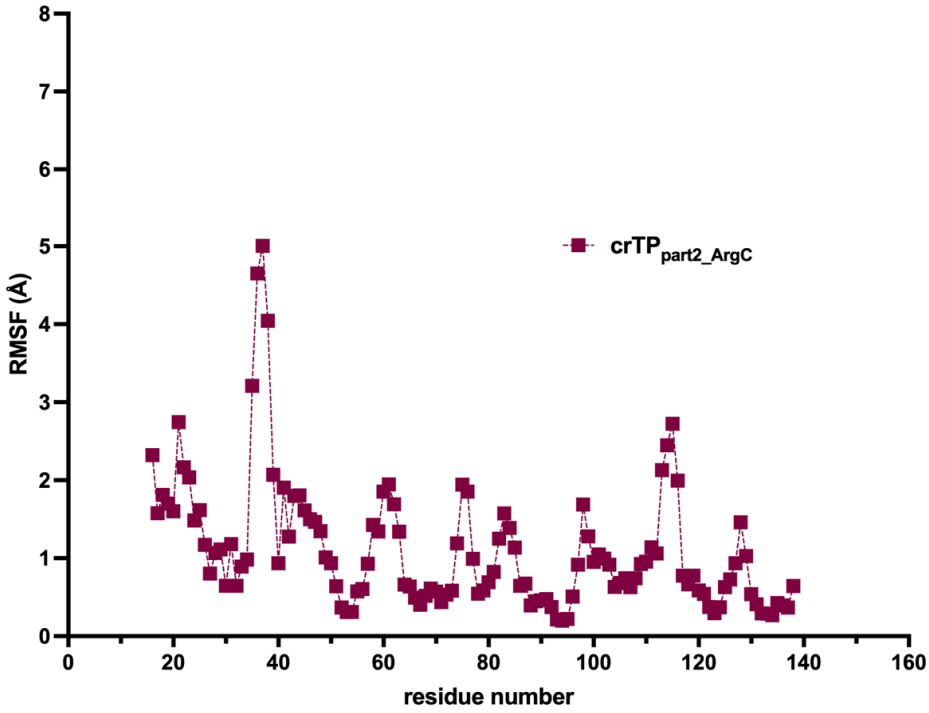


Supplementary Figure S4: Electron density maps of resolved structures. Structure of $crTP_{part2\_RnaH}$ is shown in (A) and $crTP_{part2\_ArgC}$ in (B). For clarity, we selected helix α2 and strand β4 to illustrate the high quality of the electron density of both crTP structures. Both density maps are contoured at 1 σ. Nitrogen atoms, blue; oxygen atoms, red; sulfur atoms, yellow.

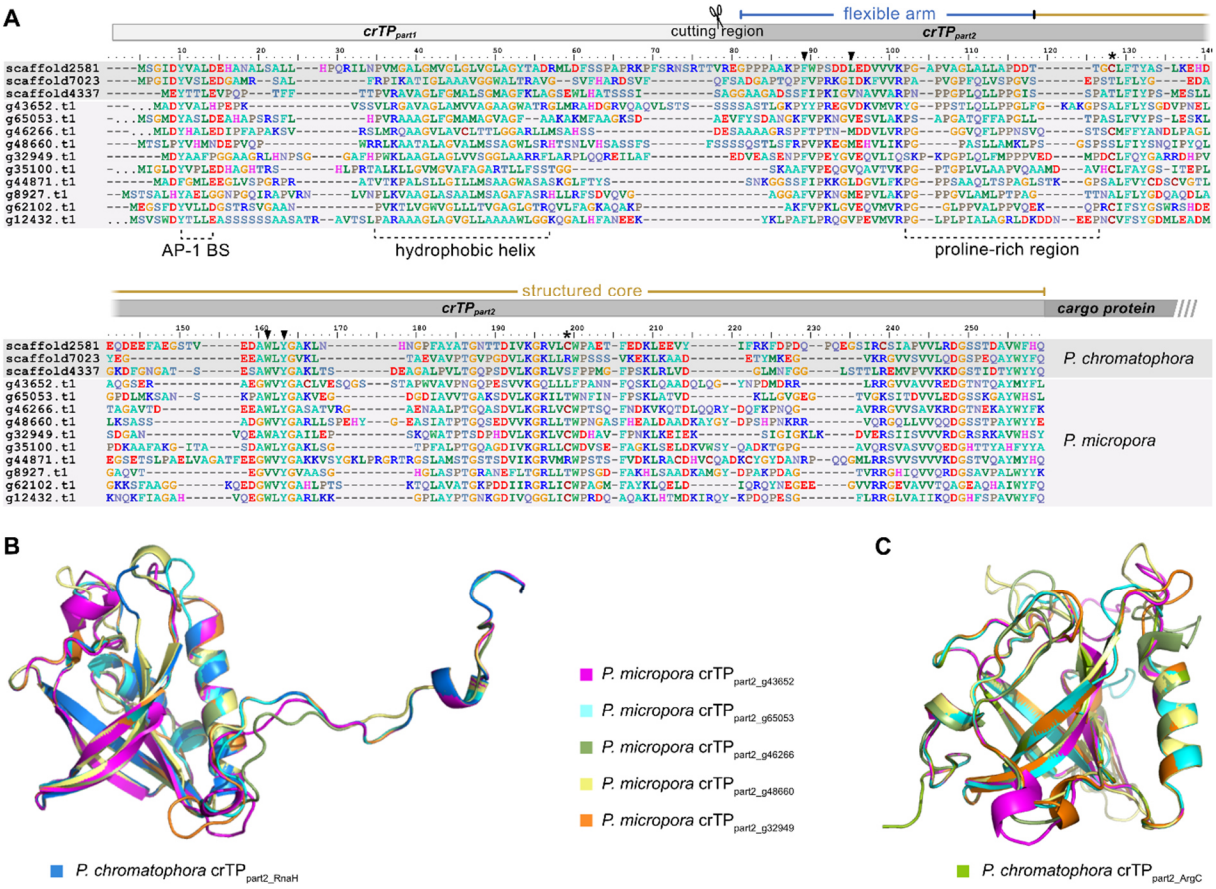


Supplementary Figure S5: Crystallographic dimer of $crTP_{part2\_RnaH}$. One monomer is shown as green cartoon model, the other as grey surface structure.

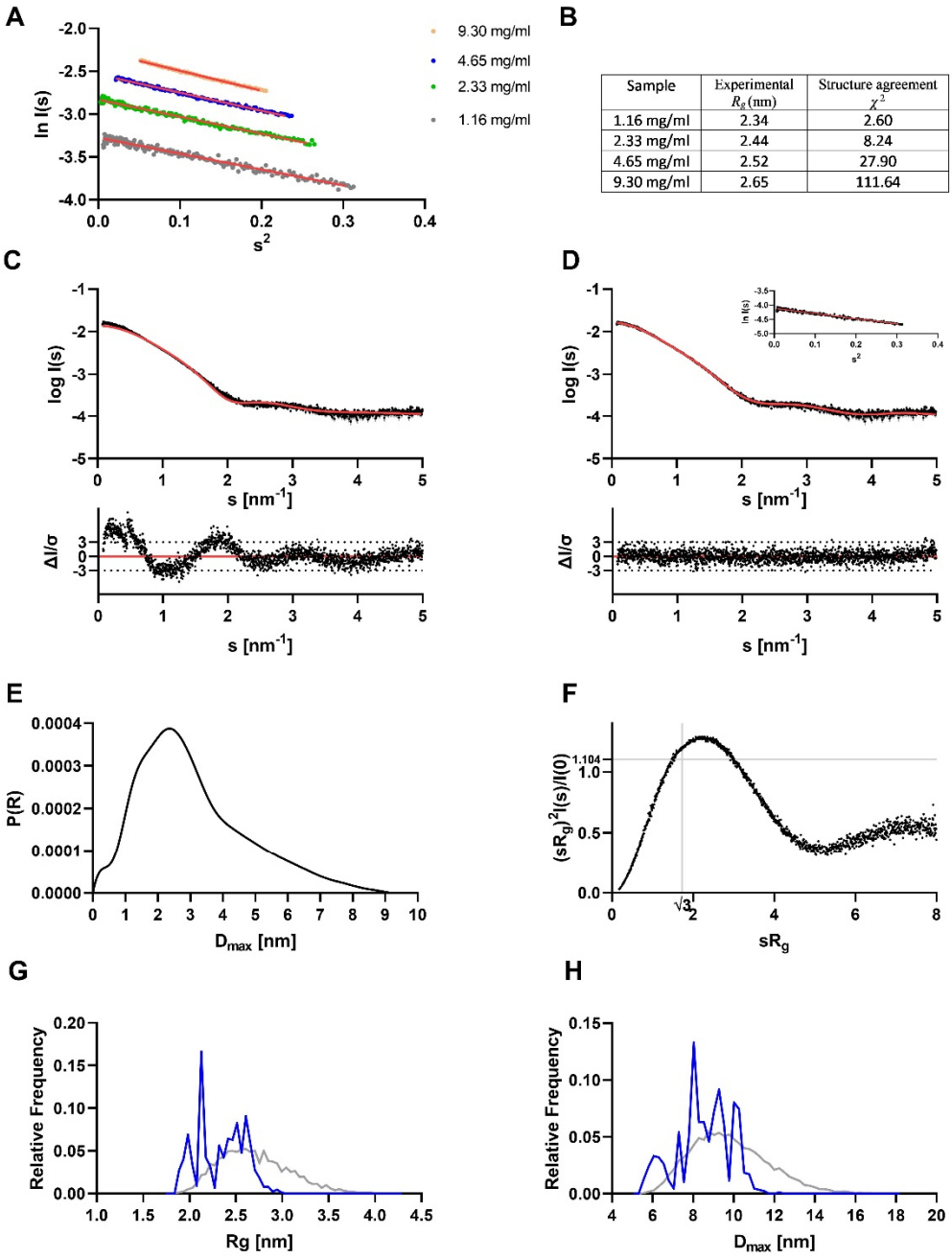




Supplementary Figure S6: Flexibility analysis using the CABS Flex 3.0 program (https://lcbio.pl/cabsflex3/). The N-terminus of crTP$_{part2_RnaH}$ shows high flexibility compared to the N-terminus of crTP$_{part2_ArgC}$ or the structured cores of both crTPs. The flexibility reached an RMSF (root mean square fluctuation) of 6-8 Å indicating large movement in the N-terminal arm of crTP$_{part2_RnaH}$. The N-terminal arm of crTP$_{part2_ArgC}$ appears to be stably bound in the closed conformation. Interestingly, at the end of both arms a peak in flexibility is predicted which is localized in the sequence at the proline-rich motif indicating that this region is important for the final conformation of the N-terminal arm.

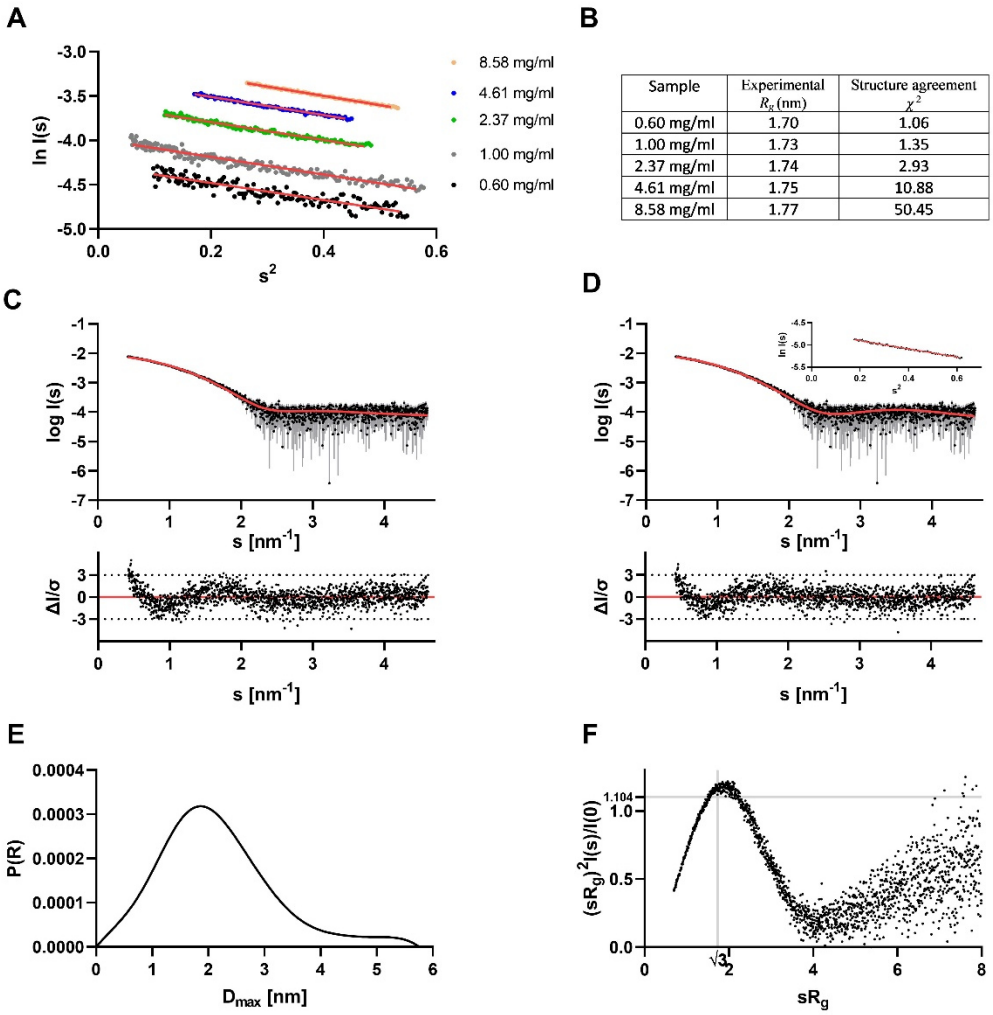


Supplementary Figure S7: Comparison of structural features of crTPs in *P. chromatophora* and *P. micropora*. (A) Multiple sequence alignment of the three *P. chromatophora*-derived crTPs investigated in this study with ten *P. micropora*-derived crTP sequences (ClustalX2 (Larkin et al., 2007), manually refined). As in Fig. 1, interacting Cys residues are indicated by asterisks. Conserved hydrophobic aa involved in the interaction between arm and core in the solved crystal structure of $crTP_{part2\_ArgC}$ (see **Fig. 1F**) are marked by arrow heads. *P. micropora* protein sequences were downloaded from http://cyanophora.rutgers.edu/P_micropora/ and protein ids provided as sequence headers. Sequences starting with '…' were N-terminally truncated by up to 20 aa, as their N-termini were likely mis-predicted to the next AUG codon upstream. **(B and C)** Alignment of the solved crystal structures of $crTP_{part2\_RnaH}$ (B) and $crTP_{part2\_ArgC}$ (C) with corresponding homology models obtained for the upper five *P. micropora* $crTP_{part2}$ sequences shown in panel **A**. For all five proteins, homology models were obtained, in which the arrangement of the β-barrel and helix α2 can be aligned. Conformation of the flexible N-terminal arm and connecting loops is more variable. However, note that the predictive value of these models is low as quality estimates provided by QMEANDisCo Global scores for all models obtained range between 0.29 and 0.45. The QMEANDisCo Global score is the average per-residue QMEANDisCo score that ranges between 0 and 1 with higher values indicating higher quality and residues showing a score <0.6 being expected to be of low quality ((Studer et al., 2020) and SWISS-MODEL documentation, https://swissmodel.expasy.org).



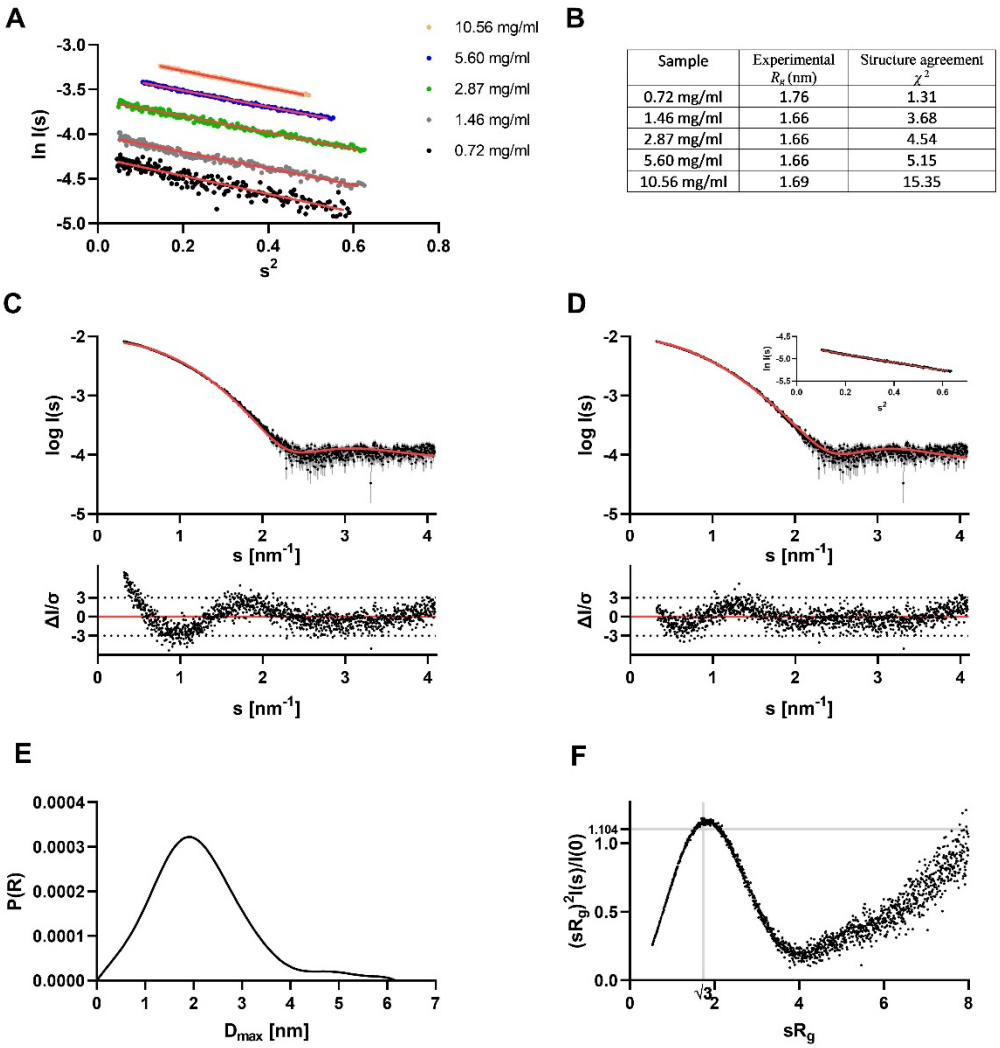
Sample	Experimental R_g (nm)	Structure agreement χ^2
1.16 mg/ml	2.34	2.60
2.33 mg/ml	2.44	8.24
4.65 mg/ml	2.52	27.90
9.30 mg/ml	2.65	111.64

Supplementary Figure S8: Small-angle X-ray scattering data from crTP$_{part2_RnaH}$. A: Guinier plots of the different concentrations used, separated for clarity. **B:** Experimental R_g and χ^2 comparison with the crystal structure at the different concentrations. **C:** Final scattering data of crTP$_{part2_RnaH}$. Experimental data are shown in black dots, with grey error bars representing standard error of the mean (SEM), calculated from 40 replicate measurements at each s-value. The theoretical model fit (χ^2 value 5.833) of the crystal structure is shown as red line and below is the residual plot of the data. **D:** Scattering data of crTP$_{part2_RnaH}$. Experimental data are shown in black dots, with grey error bars. The EOM ensemble model fit (χ^2 value 1.062) is shown as red line and below is the residual plot of the data. The Guinier plot of crTP$_{part2_RnaH}$ is added in the right corner and showed a stable Guinier region with a R_g of 2.34 nm. **E:** The *p(r)* function of crTP$_{part2_RnaH}$ showed an elongated particle with a D_{max} value of 9.17 nm. **F:** The dimensionless Kratky plot of crTP$_{part2_RnaH}$ showed an elongated particle with a degree of flexibility of the termini. **G, H:** R_g and D_{max} distribution of crTP$_{part2_RnaH}$. Ensemble pool is shown in grey, selected EOM models are shown in blue.



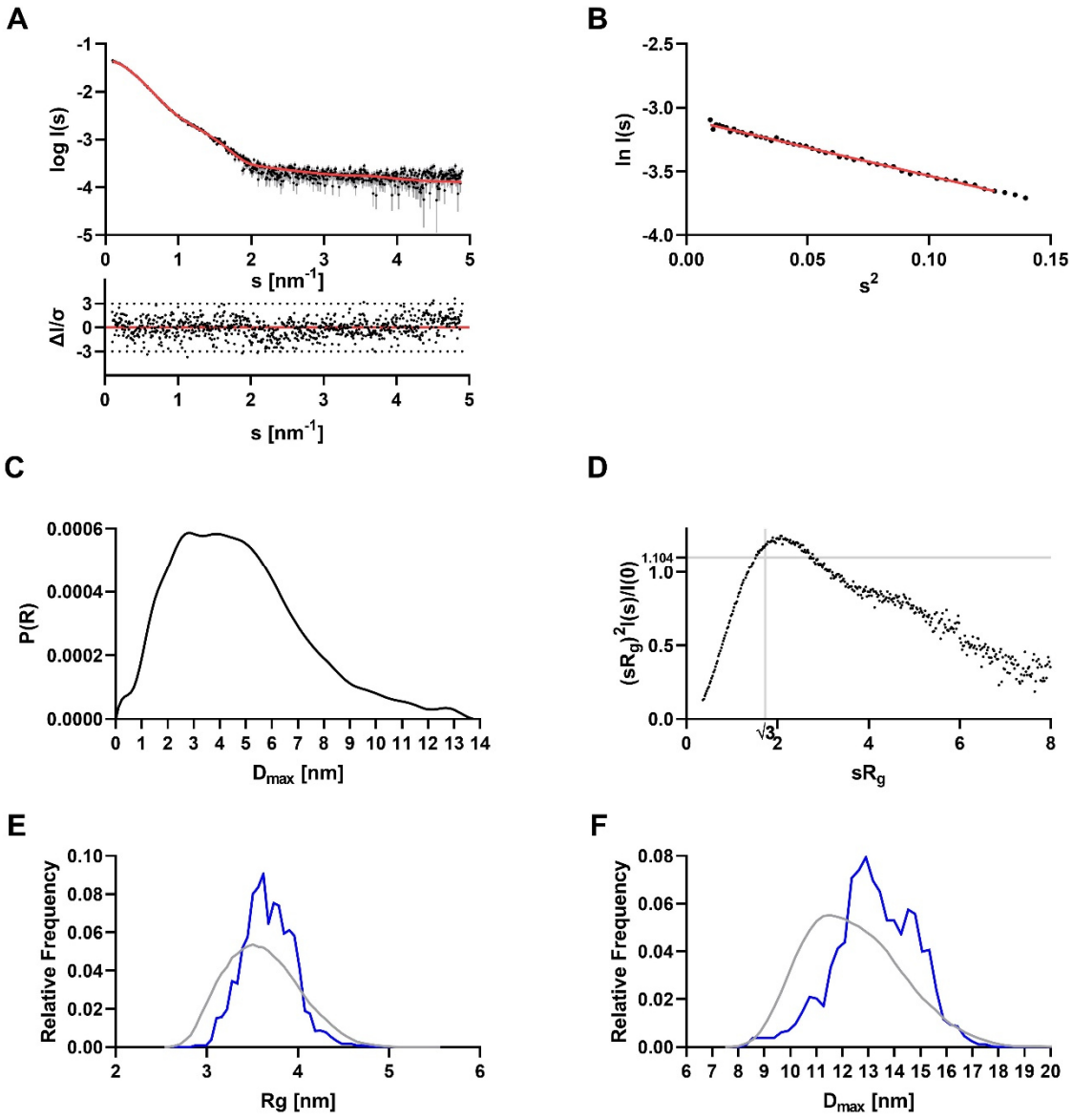
Sample	Experimental R_g (nm)	Structure agreement χ^2
0.60 mg/ml	1.70	1.06
1.00 mg/ml	1.73	1.35
2.37 mg/ml	1.74	2.93
4.61 mg/ml	1.75	10.88
8.58 mg/ml	1.77	50.45

Supplementary Figure S9: Small-angle X-ray scattering data from crTP$_{part2_ArgC}$. A: Guinier plots of the different used concentrations, separated for clarity. **B:** Experimental R_g and χ^2 comparison with the crystal structure at the different concentrations. **C:** Final scattering data of crTP$_{part2_ArgC}$. Experimental data are shown in black dots, with grey error bars (SEM of 40 replicate measurements). The theoretical model fit (χ^2 value 1.420) of the crystal structure is shown as red line and below is the residual plot of the data. **D:** Scattering data of crTP$_{part2_ArgC}$. Experimental data are shown in black dots, with grey error bars. The CORAL model fit (χ^2 value 1.293) is shown as red line and below is the residual plot of the data. The Guinier plot of crTP$_{part2_ArgC}$ is added in the right corner and showed a stable Guinier region with a R_g of 1.66 nm. **E:** The *p(r)* function of crTP$_{part2_ArgC}$ showed a globular molecule with an elongated part and a D_{max} value of 5.76 nm. **F:** The dimensionless Kratky plot of crTP$_{part2_ArgC}$ showed a little elongated, but compact molecule.



Sample	Experimental R_g (nm)	Structure agreement χ^2
0.72 mg/ml	1.76	1.31
1.46 mg/ml	1.66	3.68
2.87 mg/ml	1.66	4.54
5.60 mg/ml	1.66	5.15
10.56 mg/ml	1.69	15.35

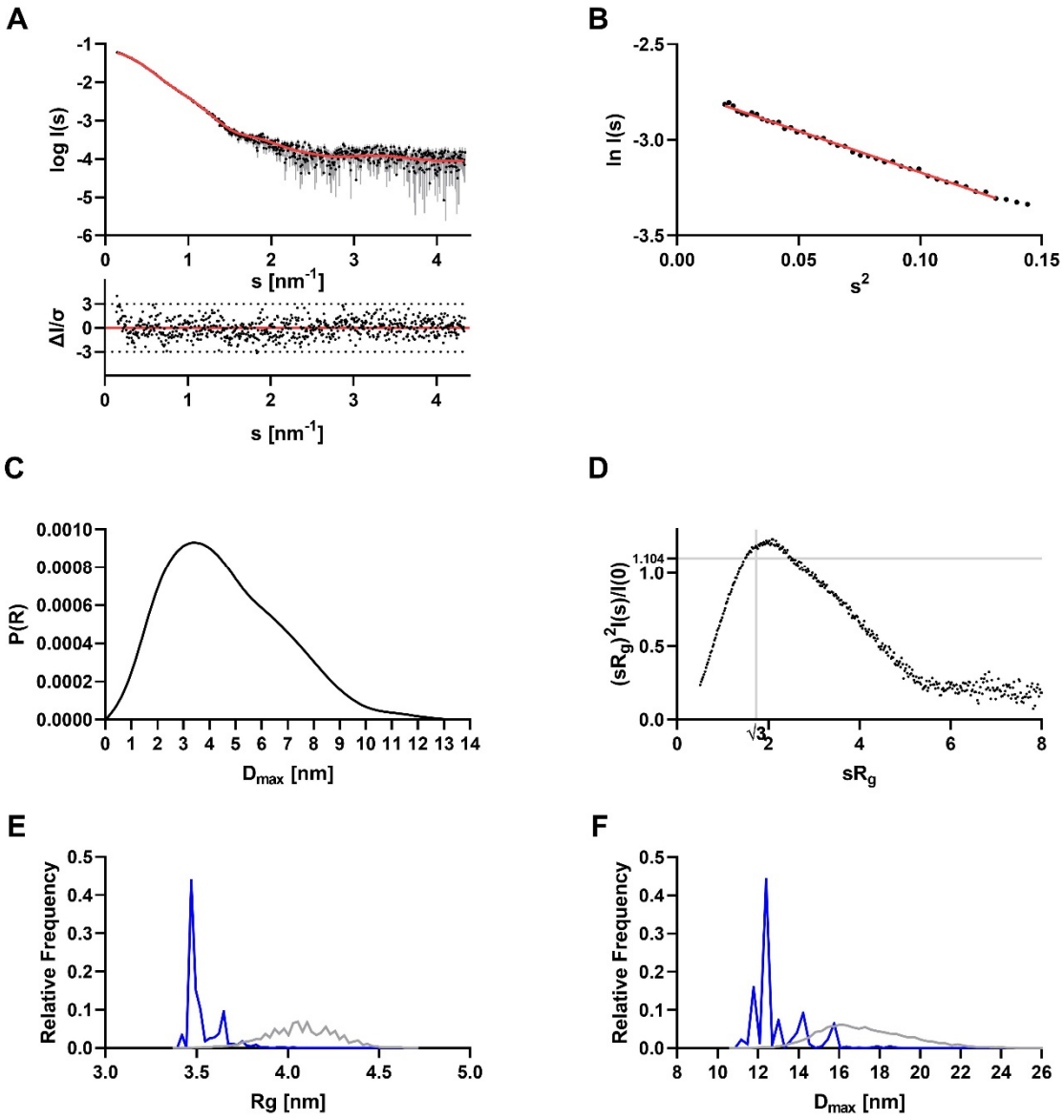
Supplementary Figure S10: Small-angle X-ray scattering data from crTP$_{part2_CysK}$. A: Guinier plots of the different used concentrations, separated for clarity. **B:** Experimental R_g and χ^2 comparison with the homology model at the different concentrations. **C:** Scattering data of crTP$_{part2_CysK}$. Experimental data are shown in black dots, with grey error bars (SEM of 40 replicate measurements). The theoretical model fit (χ^2 value 3.432) of the created homology model is shown as red line and below is the residual plot of the data. **D:** Final scattering data of crTP$_{part2_CysK}$. Experimental data are shown in black dots, with grey error bars. The CORAL model fit (χ^2 value 1.768) is shown as red line and below is the residual plot of the data. The Guinier plot of crTP$_{part2_CysK}$ is added in the right corner and showed a stable Guinier region with a R_g of 1.65 nm. **E:** The *p(r)* function of crTP$_{part2_CysK}$ showed a globular molecule with an elongated part and a D_{max} value of 6.17 nm. **F:** The dimensionless Kratky plot of crTP$_{part2_CysK}$ showed a little elongated, but compact molecule.



127

128 **Supplementary Figure S11: Small-angle X-ray scattering data from crTP$_{part2_RnaH}$-RnaH. A:**
129 Scattering data of crTP$_{part2_RnaH}$-RnaH. Experimental data are shown in black dots, with grey error bars
130 (SEM of 18 replicate measurements). The EOM ensemble model fit (χ^2 value 1.335) is shown as red
131 line and below is the residual plot of the data. **B:** The Guinier plot of crTP$_{part2_RnaH}$-RnaH showed a stable
132 Guinier region with a R_g of 3.65 nm. **C:** The *p(r)* function of crTP$_{part2_RnaH}$-RnaH showed an elongated
133 multidomain particle with a D_{max} value of 13.73 nm. **D:** The dimensionless Kratky plot of crTP$_{part2_RnaH}$-
134 RnaH showed an elongated multidomain particle. **E and F:** R_g and D_{max} distribution of crTP$_{part2_RnaH}$-
135 RnaH. Ensemble pool is shown in grey, selected EOM models are shown in blue.

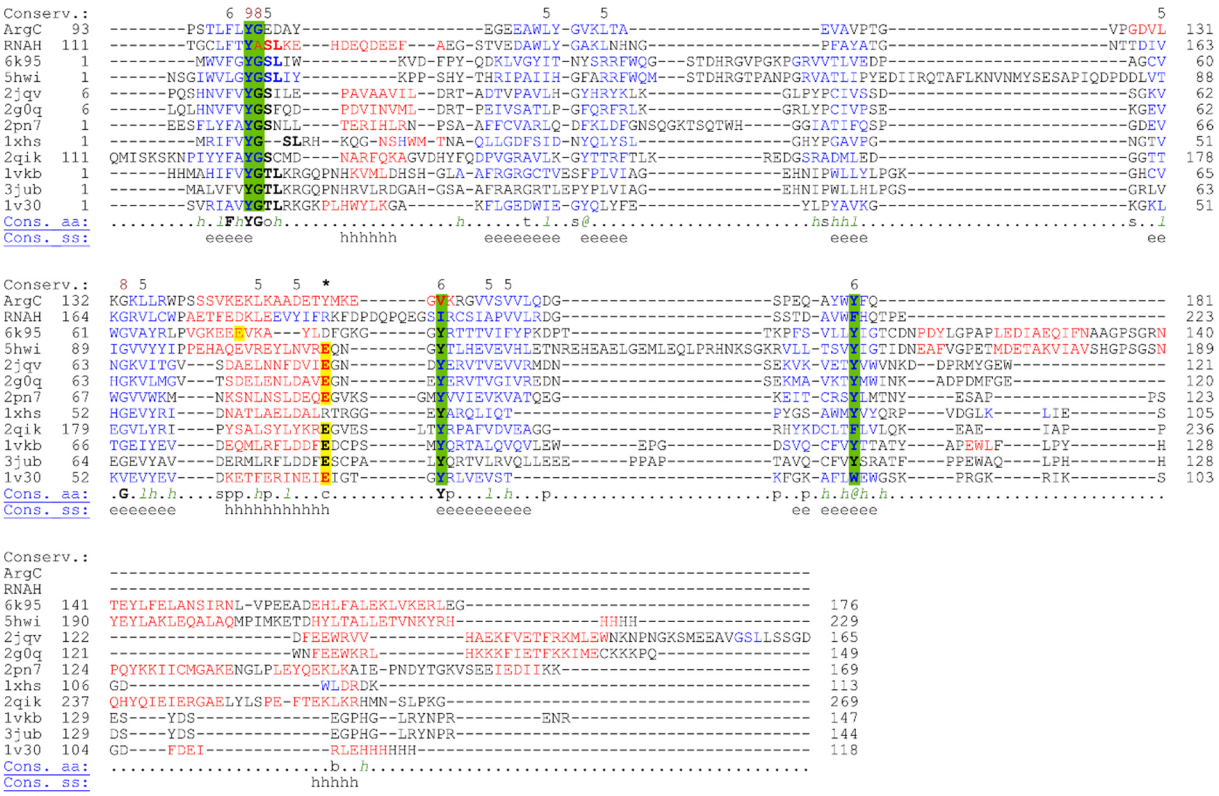
136



137

138 **Supplementary Figure S12: Small-angle X-ray scattering data from crTP$_{part2_ArgC}$-ArgC. A:**
139 Scattering data of crTP$_{part2_ArgC}$-ArgC. Experimental data are shown in black dots, with grey error bars
140 (SEM of 18 replicate measurements). The EOM ensemble model fit (χ^2 value 1.179) is shown as red
141 line and below is the residual plot of the data. **B:** The Guinier plot of crTP$_{part2_ArgC}$-ArgC showed a stable
142 Guinier region with a R_g of 3.60 nm. **C:** The *p(r)* function of crTP$_{part2_ArgC}$-ArgC showed an elongated
143 multidomain particle with a D_{max} value of 13.11 nm. **D:** The dimensionless Kratky plot of crTP$_{part2_ArgC}$-
144 ArgC showed an elongated multidomain particle. **E &F:** R_g and D_{max} distribution of crTP$_{part2_ArgC}$-ArgC.
145 Ensemble pool is shown in grey, selected EOM models are shown in blue.

146



Supplementary Figure S13: The typical GGCT binding pocket is conserved but the active site is not in crTP$_{part2}$ structures. Structure-based alignment of GGCT-like proteins generated with Promals3D (Pei et al., 2008) using default parameters. Secondary structure predictions according to PSIPRED (Jones, 1999): red, α-helix; blue, β-strand. Consensus sequence is provided if the weighted frequency of a certain class of residues in a position is above 0.8. Here, conserved aas are in bold uppercase letters; aliphatic (I, V, L): l; aromatic (Y, H, W, F): @; hydrophobic (W, F, Y, M, L, I, V, A, C, T, H): h; alcohol (S, T): o; polar residues (D, E, H, K, N, Q, R, S, T): p; tiny (A, G, C, S): t; small (A, G, C, S, V, N, D, T, P): s; bulky residues (E, F, I, K, L, M, Q, R, W, Y): b; positively charged (K, R, H): +; negatively charged (D, E): -; charged (D, E, K, R, H): c. Note that in the human ChaC2 (pdb id: 6k95) the catalytic site (Glu74) moved to a long flexible loop. By dimerization of two ChaC2 monomers, the flexible loop of one monomer moves into the catalytic cavity of the other monomer resulting in Glu74 to come close in position to the catalytic centers of other GGCT-like proteins (Nguyen et al., 2020). In the *E. coli* homolog of BtrG (named YftP, of unknown function; pdb id 1xhs) Glu82 of 3jub is replaced by an Arg which suggests that the protein does not have cyclotransferase activity.

```
...cctctagaaataattttgtttaactttaagaaggagatatacatATGGGCAGCAGCCATCATCATCATCATCACGGCAGCGGCCTGGTGCCGCGCG
GCAGCGCTAGCATGTCGGACTCAGAAGTCAATCAAGAAGCTAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCACATCAATTTAAAGGTGTCCG
ATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCTTTAAGAAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCT
TAAGATTCTTGTACGACGGTATTAGAATTCAAGCTGATCAGACCCCTGAAGATTTGGACATGGAGGATAACGATATTATTGAGGCTCACAGAGAACAGA
TTGGTGGTGAAAACCTGTATTTTCAGGGAATGTTGGACTTCAGTAGTCCTGCCCCCAGGAAACCCTTTTCTAGGAATTCTAGGACTACTGTGAGAGAAG
GGCCCCCCCCGGCCGCCAAGCCATTCTGGCCCAGCGACGACCTGGAGGACGTGGTAGTCAAGCCTGGTGCGCCCGTAGCCGGCCTGGCCCTTCTGGCAC
CGGACGACACCACAGGCTGTCTCTTCACCTACGCATCTCTCAAGGAACATGATGAACAGGACGAGGAGTTCGCAGAAGGTTCGACGGTGGAAGATGCAT
GGCTCTATGGCGCCAAACTTAACCATAACGGCCCTTTCGCCTACGCCACTGGAAACACAACGGACATCGTGAAAGGGCGTGTGTTGTGTTGGCCTGCTG
AAACGTTCGAGGACAAACTCGAAGAAGTCTACATATTTAGGAAGTTTGATCCAGACCAGCCACAAGAGGGCAGTATTCGTTGCTCGATTGCTCCTGTCG
TTCTTCGTGACGGTTCTTCTACTGATGCTGTCTGGTTTCATCAGACACCGGAGTCTATAGTCCCCAAGCAGAAGAAAAAAACCAACACGCCCAAGATCG
CCGAGTCATTCTAAtgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataac...
```

Supplementary Figure S14: The GPN142 expression vector is a pET22b(+) derivative, carrying the depicted insert. Into the backbone pET22b(+) backbone (lowercase letters) an insert containing the coding sequences for His$_6$ (blue), a thrombin cleavage site (brown), a SUMO-tag (green), a TEV cleavage site (cyan), and the crTP$_{part2_RnaH}$ domain (violet) was inserted at the indicated position.

Supplementary Tables

Supplementary Table S1: X-ray crystallography data collection and refinement statistics for $crTP_{part2\_RnaH}$ and $crTP_{part2\_ArgC}$. Statistics for the highest resolution shell are shown in parentheses.

	$crTP_{part2\_RnaH}$ (PDB ID: 9I09)	**$crTP_{part2\_ArgC}$ (PDB ID: 9I08)**
Wavelength	1.000	0.9763
Resolution range	35.15 - 2.4 (2.49 - 2.4)	43.75 - 2.2 (2.28-2.2)
Space group	P 31 2 1	P 1 21 1
Unit cell	81.164 81.164 63.678 90 90 120	36.069 67.227 43.774 90 92.03 90
Total reflections	19658 (1874)	20934 (1855)
Unique reflections	9829 (937)	10544 (944)
Multiplicity	2.0 (2.0)	2.0 (2.0)
Completeness (%)	100 (99.9)	98.5 (87.6)
Mean I/sigma(I)	34.0 (11.4)	15.6 (6.9)
Wilson B-factor	46.61	22.48
R-merge	0.009 (0.043)	0.03 (0.162
R-meas	0.013 (0.06)	0.042 (0.162)
R-pim	0.009 (0.043)	0.03 (0.115)
CC1/2	1.00 (0.997)	0.998 (0.97)
Reflections used in refinement	18682 (820)	19417 (1710)
Reflections used for R-free	976 (117)	977 (145)
R-work	0.1860 (0.2017)	0.2034 (0.2685)
R-free	0.2203 (0.2676)	0.2260 (0.3033)
Number of non-hydrogen atoms	1223	2235
macromolecules	1157	1910
ligands	4	16
solvent	62	309
Protein residues	146	247
RMS(bonds)	0.007	0.009
RMS(angles)	0.95	0.81
Ramachandran favored (%)	96.53	96.71
Ramachandran allowed (%)	3.47	2.47
Ramachandran outliers (%)	0.00	0.82
Rotamer outliers (%)	0.81	0.98
Clashscore	4.44	7.25
Average B-factor	52.96	26.28
macromolecules	53.00	25.44
ligands	41.42	38.48
solvent	52.80	30.85

Supplementary Table S2: Overall SAXS Data.

Data collection parameters					
SAXS Device	P12, PETRA III, DESY Hamburg (Blanchet et al., 2015)			Xenocs Xeuss 2.0 with Q-Xoom	
Detector	PILATUS 6 M (423.6 x 434.6 mm^2)			PILATUS 3 R 300K windowless	
Detector distance (m)	3.0			0.550	
Beam size	120 µm x 200 µm			0.8 mm x 0.8 mm	
Wavelength (nm)	0.124			0.154	
Sample environment	Quartz glass capillary, 1 mm ø			Low Noise Flow Cell, 1 mm ø	
Absolute scaling method	Comparison with scattering from pure H_2O				
Normalization	To transmitted intensity by beam-stop counter			To transmitted intensity by direct beam	
Scattering intensity scale	Absolute scale scale, cm^{-1}				
s range (nm^{-1})‡	0.03 – 7.0			0.05 – 6.0	
Sample	**$crTP_{part2\_RnaH}$**	**$crTP_{part2\_ArgC}$**	**$crTP_{part2\_CysK}$**	**$crTP_{part2\_ArgC}$ -ArgC**	**$crTP_{part2\_RnaH}$ -RnaH**
Organism	*Paulinella chromatophora*				
GenBank:	GEZN01002575.1	GEZN01007010.1	GEZN01004327.1	GEZN01007010.1	GEZN01002575.1
Mode of measurement	Batch mode				
Temperature (°C)	10				
Exposure time (# frames)	0.095 s (40)			600 s (18)	
Protein buffer	20 mM HEPES, 300 mM NaCl, pH 8.0				
Protein concentration (mg/ml)	Merged from concentrations 1.16 and 9.30 mg/ml	8.6 mg/ml – 0.6 mg/ml (extrapolated to zero concentration)	10.6 mg/ml – 0.7 mg/ml (extrapolated to zero concentration)	10.00	9.00
Structural parameters					
Guinier Analysis (PRIMUS)					
$I(0) \pm \sigma$ (cm^{-1})	0.016 ± 0.00005	0.0089 ± 0.00003	0.0089 ± 0.00001	0.065 ± 0.0002	0.045 ± 0.0001
$R_g \pm \sigma$ (nm)	2.34 ± 0.011	1.66 ± 0.009	1.65 ± 0.004	3.60 ± 0.017	3.65 ± 0.016
s-range (nm^{-1})	0.086 – 0.551	0.419 – 0.779	0.321 – 0.788	0.140 – 0.362	0.099 – 0.356
min < sRg < max limit	0.201 – 1.291	0.697 – 1.296	0.530 – 1.300	0.505 – 1.300	0.362 – 1.300
Data point range	4 - 171	1 - 130	1 - 168	1 - 39	1 - 45
Linear fit assessment (R^2)	0.9747	0.9861	0.9963	0.9961	0.9947
PDDF/P(r) Analysis (GNOM)					
$I(0) \pm \sigma$ (cm^{-1})	0.017 ± 0.00005	0.0089 ± 0.00003	0.0089 ± 0.00001	0.064 ± 0.0002	0.046 ± 0.0002
$R_g \pm \sigma$ (nm)	2.48 ± 0.009	1.68 ± 0.007	1.67 ± 0.005	3.61 ± 0.017	3.81 ± 0.019
D_{max} (nm)	9.17	5.76	6.17	13.11	13.73
Porod volume (nm^3)	45.61	27.20	28.34	130.92	105.55
s-range (nm^{-1})	0.086 – 6.755	0.419 – 4.606	0.321 – 4.09	0.140 – 4.340	0.099 – 5.642
χ2 / CorMap P-value	0.970 / 0.253	0.978 / 0.519	1.106 / 0.280	1.088 / 0.160	1.098 / 0.978
Molecular mass (kDa)					
From *I*(0)	21.63	12.32	12.32	90.01	62.32
From Qp (Porod, 1951)	20.26	9.85	9.05	96.33	71.12
From MoW2 (Fischer et al., 2010)	22.40	13.62	12.63	95.25	60.22
From Vc (Rambo and Tainer, 2013)	21.84	13.57	13.69	87.06	66.01
Bayesian Inference (Hajizadeh et al., 2018)	21.18	13.45	12.03	91.18	63.88
From sequence	21.29 (monomer)	13.99 (monomer)	15.82 (monomer)	96.62 (dimer)	63.49 (monomer)
Atomistic modeling					
CRYSOL (with default parameters)					
Constant subtraction allowed					
Structure template	Crystal structure	Crystal structure	Homology model	-	-
s-range for fit (nm^{-1})	0.077 – 4.988	0.419 – 4.603	0.321 – 4.09-	-	-

χ^2, CorMap *P*-value	5.833 / 2.78e-68	1.420 / 2.01e-14	3.432 / 1.54e-29	-	-
Predicted R_g (nm)	2.00	1.41	1.32	-	-
Predicted *Diameter* (nm)	8.10	5.10	4.47	-	-
CORAL					
Symmetry	-	P1	P1	-	-
s-range for fit (nm^{-1})	-	0.419 – 4.603	0.321 – 4.09	-	-
χ^2, CorMap *P*-value	-	1.293 / 0.00000002	1.768 / 0.000000000000004	-	-
EOM					
Symmetry	P1	-	-	P1	P1
s-range for fit (nm^{-1})	0.077 – 4.988	-	-	0.140 – 4.340	0.099 – 4.895
χ^2, CorMap *P*-value	1.062 / 0.000013	-	-	1.179 / 0.160	1.335 / 0.0122
SASBDB accession codes (Kikhney et al., 2020)	SASDWV3	SASDWT3	SASDWW3	SASDWS3	SASDWU3
Software					
ATSAS Software Version (Manalastas-Cantos et al., 2021)	3.0.5				
Primary data reduction	PRIMUS (Konarev et al., 2003)				
Data processing	GNOM (Svergun, 1992)				
Structure evaluation	CRYSOL (Svergun et al., 1995)				
Rigid body modelling	CORAL (Petoukhov et al., 2012)				
Flexibility ensemble modelling	EOM (Bernadó et al., 2007; Tria et al., 2015)				
Model visualization	PyMOL (PyMOL, 2022)				

‡s = 4πsin(θ)/λ, 2θ – scattering angle, n.d. not determined

Supplementary Table S3: Comparison of the structural dimensions from the solved crystal structures with the experimental SAXS data.

Template	Predicted R_g (nm) from crystal structure	Experimental R_g (nm) from SAXS	Predicted *Diameter* (nm) from crystal structure	Experimental D_{max} (nm) from SAXS	Structure agreement with experimental data from SAXS χ^2, CorMap *P*-value
$crTP_{part2\_RnaH}$ Crystal structure	2.00	2.34	8.10	9.17	5.833 / 2.78e^{-68}
$crTP_{part2\_ArgC}$ Crystal structure	1.41	1.66	5.10	5.76	1.420 / 2.01e^{-14}

Supplementary Table S4: Best matches from DALI searches against PDB25. Z value cutoff is ≥7.

crTP$_{part2, RnaH}$							
No	**Chain**	**Z**	**rmsd**	**lali**	**nres**	**%id**	**PDB Description**
1	5hwi-A	10.3	2.0	101	229	13	GLUTATHIONE-SPECIFIC GAMMA-GLUTAMYLCYCLOTRANSFERA
2	2qik-A	9.2	2.2	94	269	18	UPF0131 PROTEIN
3	2i5t-A	9.1	2.6	100	169	15	PROTEIN C7ORF24
4	2jqv-A	8.4	2.4	97	165	11	AIG2 PROTEIN-LIKE
5	3juc-A	8.1	2.5	97	150	14	AIG2-LIKE DOMAIN-CONTAINING PROTEIN 1
6	1v30-A	7.8	2.3	89	118	15	HYPOTHETICAL UPF0131
7	6ky1-C	7.2	1.9	77	152	12	GLUTATHIONE-SPECIFIC GAMMA-GLUTAMYLCYCLOTRANSFERA
crTP$_{part2, ArgC}$							
No	**Chain**	**Z**	**rmsd**	**lali**	**nres**	**%id**	**PDB Description**
1	5hwi-A	9.3	1.7	86	229	17	GLUTATHIONE-SPECIFIC GAMMA-GLUTAMYLCYCLOTRANSFERA
2	2i5t-A	8.6	1.6	83	169	23	PROTEIN C7ORF24
3	2qik-A	8.0	1.8	79	269	23	UPF0131 PROTEIN YKQA
4	3juc-A	7.4	2.0	81	150	17	AIG2-LIKE DOMAIN-CONTAINING PROTEIN 1
5	2jqv-A	7.4	1.7	80	165	21	AIG2 PROTEIN-LIKE

Supplementary Table S5: Protein structures available for the γ-glutamyl cyclotransferase-like superfamily.

Protein	**UniProt**	**PDB code of the structure of the natural protein**	**PDB code of additional structures**	**References**
Arabidopsis thaliana At5g39720.1	Q9FIX2	2g0q		(Lytle et al., 2006)
Arabidopsis thaliana At3g28950.1	Q9MBH1	2jqv		(de la Cruz et al., 2008)
Homo sapiens γ-glutamylamine cyclotransferase (GGACT)	Q9BVM4	3jub	3juc (Complex w/ 5-oxoproline) 3jud (E82Q mutant)	(Oakley et al., 2010)
*Mus musculus*aig2-like protein (a2ld1, ggact, mgc7867)	Q923B0	1vkb (crystal) 2kl2 (NMR)		(Klock et al., 2005) (Serrano et al., 2010)
Pyrococcus horikoshii PH0828	O58558	1v30		(Tajika et al., 2004)
Escherichia coli YtfP	P0AE48	1xhs (NMR)		(Aramini et al., 2007)
Bacillus subtilis YkqA	P39759	2qik		-
Homo sapiens γ-glutamyl cyclotransferase (GGCT)	O75223	2pn7 2i5t	2rbh (E98A mutant) 3cry (E98Q mutant)	(Oakley et al., 2008) (Bae et al., 2008)
Saccharomyces cerevisiae ChaC2	P32656	5hwi (seleno-methionine mutant)	5hwk (benzoic acid complex)	(Kaur et al., 2017)
Homo sapiens Glutathione-specific γ-glutamylcyclotransferase 2 (ChaC2)	Q8WUX2	6k95	6ky0 (E74Q mutant) 6ky1 (E83Q mutant)	(Nguyen et al., 2020)

Supplementary Table S6: Nucleotide sequences of primers used in this study. The table provides sequence, internal primer number and indicates forward (fw) or reverse (rv) orientation for each primer.

Primer sequence (5' to 3')	Primer #	Orientation
atgaaaacctgtattttcagggaATGTTGGACTTCAGTAGTCCTGCCCCCA	1515	fw
ggatcctcgagcataTTAGAATGACTCGGCGATCTTGG	1514	rv
acagagaacagattggtggtGAAAACCTGTATTTTCAGGGA	1782	fw
ttgttagcagccggatctcaTTAGAATGACTCGGCGAT	1783	rv
gagatccggctgctaacaaa	1785	fw
accaccaatctgttctctgtg	1784	rv
CAAGATCGCCGAGTCATTCGAAGACTTCCATCTCCTTGGCTCT	2640	fw
cagcttcctttcgtcaATTAGGCGGCTTGGTCGTG	2641	rv
CACGACCAAGCCGCCTAATTGAcgaaaggaagctgagt	2639	fw
CAAGGAGATGGAAGTCTTCGAATGACTCGGCGATCTTG	2638	rv
aaacctgtattttcagggaTTCTCTGCCGACGGAGCGC	2297	fw
ttagcagccggatctcaCTGGAAATACCAATAAGCCTGTTCTG	2298	rv
gcttattggtatttccagTGAGATCCGGCTGCTAACAAAG	2299	fw
tccgtcggcagagaaTCCCTGAAAATACAGGTTTTCAC	2300	rv
CTTATTGGTATTTCCAGTTCCACAAGAGCAAG	2669	fw
gttagcagccggatctcaCTAGCAGAGACCCGCACGTTCG	2645	rv
GAACGTGCGGGTCTCTGCTAGtgagatccggctgctaacaaag	2643	fw
CTTGCTCTTGTGGAACTGGAAATACCAATAAGCCTGTTCTG	2670	rv
tgtattttcagggaGCGACCAGCAGCAGCATCAG	2293	fw
gttagcagccggatctcaTTACTGGTAGTACCAGTAAGTGTCGATGGTG	2294	rv
ggtactaccagTAATGAGATCCGGCTGCTAACAAAGCC	2295	fw
ctgctgctggtcgcTCCCTGAAAATACAGGTTTTCACCACC	2296	rv

References

Aramini JM, Huang YJ, Swapna GVT, Cort JR, Rajan PK, Xiao R, Shastry R, Acton TB, Liu J, Rost B, Kennedy MA, Montelione GT (2007) Solution NMR structure of *Escherichia coli* ytfP expands the structural coverage of the UPF0131 protein domain family. Proteins-Structure Function and Bioinformatics **68:** 789-795

Bae E, Bingman CA, Aceti DJ, Phillips GN (2008) Crystal structure of *Homo sapiens* protein LOC79017. Proteins-Structure Function and Bioinformatics **70:** 588-591

Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. Journal of the American Chemical Society **129:** 5656-5664

Blanchet CE, Spilotros A, Schwemmer F, Graewert MA, Kikhney A, Jeffries CM, Franke D, Mark D, Zengerle R, Cipriani F, Fiedler S, Roessle M, Svergun DI (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). Journal of Applied Crystallography **48:** 431-443

de la Cruz NB, Peterson FC, Volkman BE (2008) Solution structure of at3g28950 from *Arabidopsis thaliana*. Proteins-Structure Function and Bioinformatics **71:** 546-551

Fischer H, Neto MD, Napolitano HB, Polikarpov I, Craievich AF (2010) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. Journal of Applied Crystallography **43:** 101-109

Hajizadeh NR, Franke D, Jeffries CM, Svergun DI (2018) Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. Scientific Reports **8:** 7204

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology **292:** 195-202

Kaur A, Gautam R, Srivastava R, Chandel A, Kumar A, Karthikeyan S, Bachhawat AK (2017) ChaC2, an enzyme for slow turnover of cytosolic glutathione. Journal of Biological Chemistry **292:** 638-651

Kikhney AG, Borges CR, Molodenskiy DS, Jeffries CM, Svergun DI (2020) SASBDB: Towards an automatically curated and validated repository for biological scattering data. Protein Science **29:** 66-75

Klock HE, Schwarzenbacher R, Xu QP, McMullan D, Abdubek P, Ambing E, Axelrod H, Biorac T, Canaves JM, Chiu HJ, Deacon AM, DiDonato M, Elsliger MA, Godzik A, Grittini C, Grzechnik SK, Hale J, Hampton E, Han GW, Haugen J, Hornsby M, Jaroszewski L, Koesema E, Kreusch A, Kuhn P, Miller MD, Moy K, Nigoghossian E, Paulsen J, Quijano K, Reyes R, Rife C, Sims E, Spraggon G, Stevens RC, van den Bedem H, Velasquez J, Vincent J, White A, Wolf G, Hodgson KO, Wooley J, Lesley SA, Wilson IA (2005) Crystal structure of a conserved hypothetical protein (gi : 13879369) from Mouse at 1.90 Å resolution reveals a new fold. Proteins-Structure Function and Bioinformatics **61:** 1132-1136

Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography **36:** 1277-1282

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and clustal X version 2.0. Bioinformatics **23:** 2947-2948

Lytle BL, Peterson FC, Tyler EM, Newman CL, Vinarov DA, Markley JL, Volkman BF (2006) Solution structure of *Arabidopsis thaliana* protein At5g39720.1, a member of the AIG2-like protein family. Acta Crystallographica Section F-Structural Biology Communications **62:** 490-493

Manalastas-Cantos K, Konarev PV, Hajizadeh NR, Kikhney AG, Petoukhov MV, Molodenskiy DS, Panjkovich A, Mertens HDT, Gruzinov A, Borges C, Jeffries CM, Svergun DI, Franke D (2021) *ATSAS 3.0*: expanded functionality and new tools for small-angle scattering data analysis. Journal of Applied Crystallography **54:** 343-355

Nguyen YTK, Park JS, Jang JY, Kim KR, Vo TTL, Kim KW, Han BW (2020) Structural and functional analyses of human ChaC2 in glutathione metabolism. Biomolecules **10:** 31

Oakley AJ, Coggan M, Board PG (2010) Identification and characterization of gamma-glutamylamine cyclotransferase, an enzyme responsible for gamma-glutamyl-epsilon-lysine catabolism. Journal of Biological Chemistry **285:** 9642-9648

Oakley AJ, Yamada T, Liu D, Coggan M, Clark AG, Board PG (2008) The identification and structural characterization of C7orf24 as gamma-glutamyl cyclotransferase - An essential enzyme in the gamma-glutamyl cycle. Journal of Biological Chemistry **283:** 22031-22042

Pei JM, Kim BH, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Research **36:** 2295-2300

Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev PV, Svergun DI (2012) New developments in the *ATSAS* program package for small-angle scattering data analysis. Journal of Applied Crystallography **45:** 342-350

Porod G (1951) Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen .1. Teil. Kolloid-Zeitschrift and Zeitschrift für Polymere **124:** 83-114

PyMOL (2022) The PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC.

Rambo RP, Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. Nature **496:** 477-481

Serrano P, Pedrini B, Geralt M, Jaudzems K, Mohanty B, Horst R, Herrmann T, Elsliger MA, Wilson IA, Wuthrich K (2010) Comparison of NMR and crystal structures highlights conformational isomerism in protein active sites. Acta Crystallographica Section F-Structural Biology Communications **66:** 1393-1405

Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T (2020) QMEANDisCo—distance constraints applied on model quality estimation. Bioinformatics **36:** 1765-1771

Svergun D, Barberato C, Koch MH (1995) CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. Journal of Applied Crystallography **28:** 768-773

Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of Applied Crystallography **25:** 495-503

Tajika Y, Sakai N, Tamura T, Yao M, Watanabe N, Tanaka I (2004) Crystal structure of hypothetical protein PH0828 from *Pyrococcus horikoshii*. Proteins-Structure Function and Bioinformatics **57:** 862-865

Tria G, Mertens HDT, Kachala M, Svergun DI (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ **2:** 207-217