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

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]

[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

Corresponding author: Eva C. M. Nowack, email: e.nowack@hhu.de

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is Eva C. M. Nowack.

Short title: Structure of the chromatophore transit peptide

Abstract

The chromatophores of the cercozoan amoeba *Paulinella* are photosynthetic organelles that evolved from a cyanobacterial endosymbiont. Many nucleus-encoded chromatophore-targeted proteins carry unusual N-terminal targeting signals termed crTPs, which are bipartite. $crTP_{part1}$ likely mediates trafficking through the secretory pathway and is cleaved off during import, but $crTP_{part2}$ remains attached to its cargo protein and its function is unknown. To unravel the functional role of $crTP_{part2}$, here we elucidated the structures of $crTP_{part2}$ from two different chromatophore-targeted proteins by X-ray crystallography at ~2.3 Å resolution. Interestingly, the $crTP_{part2}$ of both proteins adopts a structural fold. Both structures share a conserved structured core and a flexible N-terminal arm. The structured core resembles proteins of the γ-glutamyl cyclotransferase superfamily within which $crTP_{part2}$ structures form a protein (sub)-family. The proposed catalytic center typical for proteins with cyclotransferase activity is not conserved in $crTP_{part2}$. A Cys pair that is conserved in $crTP_{part2}$ of many chromatophore-targeted proteins has been captured as a disulfide bridge. Together, our data suggest that chromatophore-targeted

proteins are imported in their folded state and that the fold adopted by $crTP_{part2}$ plays a functional role during import. The characterization of its structure and flexibility provides important steps towards elucidating this protein translocation mechanism.

Introduction

The transformation of bacterial endosymbionts into eukaryotic organelles has been a key process in eukaryote evolution. The only organelles identified so far that evolved by primary endosymbiosis events that were independent of the events that gave rise to mitochondria and plastids, are the photosynthetic "chromatophores" of the cercozoan amoeba *Paulinella* and the nitrogen-fixing "nitroplasts" of the haptophyte *Braarudosphaera.* In both cases, following the establishment of a cyanobacterial endosymbiont, the endosymbiont lost many functions by reductive genome evolution that were compensated by the import of nucleus-encoded proteins (Nowack and Grossman, 2012; Singer et al., 2017; Coale et al., 2024). Many of these organelle-targeted proteins carry conserved sequence extensions that apparently function as unique types of targeting signals (Singer et al., 2017; Coale et al., 2024). Their way of functioning is little understood.

In *Paulinella chromatophora*, the subject of this study, long chromatophore-targeted proteins [lCTPs; typically >250 amino acids (aa)] carry such N-terminal targeting signals that are referred to as 'chromatophore transit peptides' (crTPs) (Singer et al., 2017). CrTPs are ~200 aa long, contain conserved sequence elements, and are bipartite. Upon import, $crTP_{part1}$ is cleaved off, whereas $crTP_{part2}$ remains attached to the N-terminus of most lCTPs (Oberleitner et al., 2022) (**Fig. 1A**). It has been proposed that the conserved hydrophobic helix in $crTP_{part1}$ anchors crTP-carrying proteins co-translationally in the ER membrane in an N-terminus out, C-terminus in conformation and that the N-terminal adaptor protein 1 complex binding site (AP-1 BS) is responsible for packaging lCTPs into clathrin-coated vesicles (Oberleitner et al., 2022). Although the exact timepoint at which $crTP_{part1}$ is cleaved off is unknown, it is reasonable to assume that cleavage happens after this sorting step, possibly following fusion of the vesicles with the outer (host-derived) chromatophore membrane (Sørensen et al., 2025). This would result in a release of the cargo proteins, still attached to $crTP_{part2}$, into the intermembrane space. The function of $crTP_{part2}$ is unclear, but it is likely involved in mediating protein translocation across the two remaining layers (i.e., peptidoglycan (PG) and inner membrane).

Interestingly, $crTP_{part2}$ contains conserved predicted secondary structure elements across proteins (Oberleitner et al., 2022), suggesting that, different from the N-terminal transit peptides of mitochondrion and plastid-targeted proteins, which are generally unstructured, $crTP_{part2}$ adopts a structural fold. This hypothesis guided the experimentation in this study.

Results

$CrTP_{part2}$ adopts a structured fold

To contribute to the understanding of the function of $crTP_{part2}$, we aimed to elucidate its 3D structure. For this purpose, we focused on $crTP_{part2}$ from three chromatophore-targeted proteins. These were derived from the transcripts scaffold2581, scaffold7023, and scaffold4337 (GenBank accessions: GEZN01002575.1, GEZN01007010.1, and GEZN01004327.1, (Nowack et al., 2016)) encoding a predicted RNA helicase (RnaH), N-acetyl-gamma-glutamyl-phosphate reductase (ArgC), and cysteine synthase A (CysK); from here on $crTP_{part2\_RnaH}$, $crTP_{part2\_ArgC}$, and $crTP_{part2\_CysK}$, respectively (**Fig. 1A**). Initial analyses of these proteins by AlphaFold3 (Abramson et al., 2024) predicted the structures of the cargo proteins, RnaH, ArgC, and CysK, with high confidence; however, the crTP structures could not be modeled in high quality and resulted in largely unstructured domains (**Fig. S1**). Hence, we aimed for structure characterization by X-ray crystallography. To this end, we purified recombinant $crTP_{part2\_RnaH}$, $crTP_{part2\_ArgC}$, and $crTP_{part2\_CysK}$-containing constructs following their overexpression in *Escherichia coli* (**Fig. S2-S3**). Two of these proteins, $crTP_{part2\_RnaH}$ and $crTP_{part2\_ArgC}$, readily formed crystals of sufficient quality to determine their structure via X-ray crystallography, at 2.4 Å and 2.2 Å resolution, respectively (for details see **Suppl. Text** and **Table S1**). The quality of the electron density map is visualized in **Fig. S4**.

Both structures contain a structured core (colored gold in **Fig. 1B, C**) and a mostly unstructured N-terminal arm (blue in **Fig. 1B, C**). The structured cores consist of a five-stranded antiparallel β-barrel (β1-β3-β2-β4-β5) flanked by an α-helix (α2), decorated by connecting loops. In $crTP_{part2\_RnaH}$, additional very short β-strands (β1', β2') and α-helical elements (α1'- α3') are embedded into the connecting loops. The N-terminal arm clearly adopts a different conformation in both structures, indicating that this part might be flexible, whereas the structured core is almost identical between $crTP_{part2\_RnaH}$ and $crTP_{part2\_ArgC}$ over a large part of the structure (rmsd 0.7 Å over 63 aligned Cα atoms, **Fig. 1D**). In $crTP_{part2\_RnaH}$, the β-barrel is stabilized by a disulfide bridge formed between Cys113 and Cys169 (**Fig. 1E**). This Cys pair is conserved in many, but not in all crTP sequences, however, if Cys occurs in these positions, it occurs as a pair (see asterisks in **Fig. 1A**). The loop connecting helix α2 with β4 is much larger in $crTP_{part2\_RnaH}$ when compared to $crTP_{part2\_ArgC}$. The loops connecting β2 and β3 as well as α2 with β3 are similar in length but adopt different conformations in the two structures, hinting towards flexibility at these positions whereas the core is rigid.

In $crTP_{part2\_ArgC}$, the N-terminal arm interacts via hydrophobic interactions between Phe64 and Trp110 as well as Ile70 and Tyr112 with the structured core and hence, shows a "closed conformation" (**Fig. 1F**). These hydrophobic residues are highly conserved between different crTP sequences (black arrowheads in **Fig. 1A**). In $crTP_{part2\_RnaH}$, for which the "open conformation" of the N-terminal arm was captured, the hydrophobic residues are exposed (**Fig. 1E**). This open conformation appears to be stabilized by the formation of a crystallographic dimer (**Fig. S5**). The N-terminal arm contains several proline residues, which give the arm a specific conformation. To verify the flexibility of the N-terminal arm, we applied molecular simulations using the program CABS Flex with the all-atom reconstruction implemented

(Wróblewski et al., 2025). Here, $crTP_{part2\_RnaH}$ shows high flexibility at the N-terminus in comparison with the N-terminus of $crTP_{part2\_ArgC}$ (**Fig. S6**). Interestingly, in both structures, flexibility peaks in the proline-rich region, which might function as a hinge for flexibility. Together these results suggest that the N-terminal arm can adopt several conformations, however, in the closed conformation is stabilized by the interactions described above.

Notably, N-terminal crTPs are also found in lCTPs of *Paulinella micropora*, a chromatophore-containing sister of *P. chromatophora* (Lhee et al., 2021). To assess whether the structural features identified in crTPs of *P. chromatophora* are conserved across species, we aligned the sequences of the three crTPs studied here with those of crTPs from *P. micropora* (**Fig. S7A**). Intriguingly, all important structural features identified in *P. chromatophora* are conserved in *P. micropora* including, in $crTP_{part1}$, the AP-1 BS and hydrophobic helix; and, in $crTP_{part2}$, the Cys-pair, the hydrophobic residues supporting interactions between flexible arm and structured core, and the proline-rich region (**Fig. S7A**). Furthermore, homology models obtained using the solved crystal structures of $crTP_{part2\_RnaH}$ and $crTP_{part2\_ArgC}$ as templates, suggest that $crTP_{part2}$ in *P. micropora* can adopt similar folds (**Fig. S7B-C**). However, the predictive value of these models is low, since -as observed before for AlphaFold3 (**Fig. S1**)- the quality estimates for all homology models obtained were very low (see **Fig. S7B-C** and corresponding figure legend). To study $crTP_{part2}$ linked to their natural cargo proteins, we purified recombinant $crTP_{part2\_RnaH}$-RnaH and $crTP_{part2\_ArgC}$-ArgC following their production in *E. coli* (**Fig. S2-S3**). However, despite several attempts, these proteins did not form crystals.

SAXS measurements indicate flexibility of the N-terminal arm and the linker to the cargo protein in solution

To investigate the flexibility of $crTP_{part2}$ in solution as well as the spatial arrangement between the $crTP_{part2}$ regions and their cargo proteins, we performed small angle X-ray scattering (SAXS) measurements. These measurements demonstrated that $crTP_{part2\_ArgC}$, $crTP_{part2\_RnaH}$, and $crTP_{part2\_CysK}$ alone are monomeric in solution (**Tables S2-S3** and **Suppl. Text**). For $crTP_{part2\_RnaH}$, the "open conformation" of the flexible arm, found in the crystal dimer packing, is also present in solution. Here the flexible arm adopts, as expected, multiple conformations (see **Suppl. Text**). Flexibility analyses of the termini with the Ensemble Optimization Method (EOM) revealed four different conformations that all represent different "open conformations", but no indication for a "closed conformation" for $crTP_{part2\_RnaH}$ (**Fig. 2A, B** and **Fig. S8**). In comparison, neither $crTP_{part2\_ArgC}$, nor $crTP_{part2\_CysK}$ showed flexible termini and show a more compact conformation (**Figs. S9-S10** and **Suppl. Text**).

$CrTP_{part2\_RnaH}$-RnaH, too, is a monomer in solution (**Table S2**) and we recovered three distinct conformations, indicating flexibility between the $crTP_{part2\_RnaH}$ domain and the attached cargo protein, RnaH (**Fig. 2C, Fig. S11** and **Suppl. Text**). In contrast, $crTP_{part2\_ArgC}$-ArgC appears dimeric over the whole concentration range (**Table S2**) with the dimer interface predicted within the cargo protein ArgC. The models obtained showed that $crTP_{part2\_ArgC}$-ArgC forms an overall compact molecule, but, again, reveals flexibility between the $crTP_{part2}$ domain

and the attached cargo protein (**Fig. 2D** and **Fig. S12**). Furthermore, the N-terminal arm of $crTP_{part2\_ArgC}$ which appeared "closed" in the monomer crystal and in-solution model of $crTP_{part2\_ArgC}$ alone, now remains flexible, more in line with an "open conformation" (for more details see **Suppl. Text**).

The structured core of $crTP_{part2}$ shows similarity to γ-glutamyl cyclotransferase fold proteins

Searching the coordinates of the solved $crTP_{part2}$ structures against the PDB database using the DALI server (http://ekhidna2.biocenter.helsinki.fi/dali/) revealed similarity of both structures to members of the γ-glutamyl cyclotransferase-like superfamily (InterPro entry IPR036568) (**Table S4**). This superfamily contains five protein families, the γ-glutamyl cyclotransferase (GGCT), the γ-glutamylamine cyclotransferase (GGACT), the glutathione-specific GGCT (GCG or ChaC), the BrtG-like, and the plant-specific GGCT-like family [**Fig. 3** and (Kumar et al., 2015)].

For 10 members of the GGCT-like superfamily structures have been experimentally solved (**Fig. 3A** and **Table S5**). Highest similarity scores were obtained for the *S. cerevisiae* glutathione-specific GGCT ChaC (pdb id: 5hwi, z-scores = 10.3 and 9.3, sequence identities = 13 and 17%), the human GGCT (pdb id: 2i5t, z-scores = 9.1 and 8.6, sequence identities = 15 and 23%), and the *B. subtilis* protein YkqA (pdb id: 2qik, z-scores = 9.2 and 8.0, sequence identities18 and 23%) (values for $crTP_{part2\_RnaH}$ and $crTP_{part2\_ArgC}$, respectively). Thus, structure comparison did not reveal affiliation of the $crTP_{part2}$ structures to any particular family within the superfamily. In line with this result, a maximum likelihood (ML) phylogenetic analysis of a structure-guided alignment resolves $crTP_{part2}$ sequences as a unique family within IPR036568 that forms a short common branch with the plant-specific GGCT-like family (**Fig. 3A**).

For several proteins of the superfamily, enzymatic activities have been characterized and a common function of many is the cleavage of diverse γ-glutamyl derivatives by cyclotransferase activity (**Suppl. Text**). Despite their low sequence identities, catalysis is based on a similar structural fold, which includes a cavity between the β-barrel and adjacent helix that contains a conserved YGSL motif and a Glu residue that likely represents the active site (Oakley et al., 2010; Chi et al., 2014). The $crTP_{part2}$ structures form a similar cavity build by strand β1 and β5, helix α2, and the loop between β1 and β2 that aligns with the proposed substrate-binding cavity in the human GGACT (3jub, 3juc) (Oakley et al., 2010) (**Fig. 3B**). The YGSL motif is partly conserved in $crTP_{part2}$. Tyr99 of $crTP_{part2\_ArgC}$ at the N-terminal end of the loop connecting β1 and β2 corresponds to Tyr7 in the YGSL motif of the human GGACT. This Tyr residue is conserved throughout different $crTP_{part2}$ sequences (**Fig. S13**) and its side chain is orientated towards the inside of the cavity (**Fig. 3C**). The following Gly is found only in around half of the $crTP_{part2}$ sequences and is replaced in the remaining sequences mostly by other small amino acids (A, S, T, P). Ser and Leu of the YGSL motif are conserved in $crTP_{part2\_RnaH}$ but non-conservatively replaced in $crTP_{part2\_ArgC}$ by Glu and Asp. Finally, Glu82 of the human GGACT that sits at the C-terminal end of cavity-delimiting helix with its side chain oriented towards the inside of the cavity has been proposed to form the catalytic center (Oakley et al.,

2010). Glu in this position is highly conserved in many proteins of the superfamily (**Fig. S13**). Interestingly, in $crTP_{part2\_ArgC}$ and $crTP_{part2\_RnaH}$ this catalytic Glu is replaced by a Tyr and an Arg, respectively (**Fig. 3D** and **Fig. S13**). In other $crTP_{part2}$ sequences, this site harbors several other aa residues (see alignment position 192 in **Fig. 1A**). Hence, it appears unlikely that $crTP_{part2}$ has cylotransferase activity.

Discussion

Here we showed that $crTP_{part2}$ domains of lCTPs in *P. chromatophora* adopt a structured fold that consists of a structured core with similarity to GGCT-like proteins and an N-terminal flexible arm that can interact via conserved hydrophobic residues with the structured core (see arrow heads in **Fig. 1A**). The proline-rich region in the N-terminal arm may represent an anchor for an interaction partner, since proline-rich regions within other proteins are known to form interaction surfaces that are responsible for interactions with e.g. elements of the cytoskeleton, peptidoglycan or biological membranes (Williamson, 1994). Interaction with a potential partner would be facilitated by the inferred flexibility of the N-terminal arm as well as the linker between $crTP_{part2}$ and its cargo protein (**Fig. 2**). The disulfide bridge in $crTP_{part2\_RnaH}$ that has been captured by crystallography (**Fig. 1E**) is formed by a Cys pair that is conserved across $crTP_{part2}$ domains of diverse lCTPs, which suggests that the oxidized state is biologically relevant. This assumption is in line with the oxidizing conditions in the ER lumen (Margittai et al., 2015) that has been suggested as intermediate station in the lCTP import pathway (Oberleitner et al., 2022).

Interestingly, whereas lCTPs apparently require a crTP for import, short chromatophore-targeted proteins (sCTPs; typically <90 aa) that also apparently traffic into the chromatophore via the Golgi (Nowack and Grossman, 2012) lack similar targeting signals (Singer et al., 2017). Since lCTPs and sCTPs comprise overlapping functions [e.g., cytosolic metabolic enzymes, diverse DNA-binding proteins (Singer et al., 2017; Oberleitner et al., 2020; Macorano et al., 2023)], requirement of a crTP does not seem to be tied to a specific function or final localization of the cargo protein but rather its size. This size cutoff could be set by the - so far unknown - import gate in the inner chromatophore membrane or the mesh size of the PG sacculus.

In *E. coli*, a size cutoff of ~50 kDa has been estimated for globular proteins to be able to diffuse through the stretched PG sacculus (Demchick and Koch, 1996). Cyanobacteria generally feature a thicker PG with a much higher degree of crosslinking (Hoiczyk and Hansel, 2000). Hence, although PG composition and crosslinking has not been analyzed for chromatophores yet, it might represent an important hurdle for the transport of folded lCTPs which reach sizes >100 kDa (Singer et al., 2017). Also the plastids of Glaucophytes retained a pronounced PG layer; however, they use a TIC/TOC translocon-based mechanism for importing plastid-targeted pre-proteins in an unfolded state and only the mature stromal proteins fold into their functional conformation (Steiner and Löffelhardt, 2002). Hence, the PG does not represent a relevant size cutoff here.

Since many GGCT-family proteins cleave γ-glutamyl-containing peptides, the γ-glutamyl-containing muropeptides that cross link the PG appeared as possible ligands of $crTP_{part2}$. Although cleavage of these peptides by $crTP_{part2}$ appears unlikely due to the lack of conservation of the proposed catalytic center (**Fig. 3D**), we hypothesized that the binding to muropeptide derivatives could be involved in recognition of non-crosslinked areas in the PG and/or result in a conformational change enabling interaction with interaction partners at the inner membrane. However, we could not experimentally confirm binding of $crTP_{part2}$ to the *E. coli* PG penta or tetrapeptide. Hence, if muropeptides are the natural ligands of $crTP_{part2}$, we could not identify the exact ligand and/or correct conditions yet under which binding occurs.

In sum, our data suggests that lCTPs are imported in their folded state and that the fold adopted by $crTP_{part2}$ plays an as of yet unknown functional role in the import process that appears to be conserved cross chromatophore-containing *Paulinella* species. The characterization of its structure and flexibility provides important steps towards unraveling this protein translocation mechanism.

Materials and Methods

Cultivation of *P. chromatophora* and synthesis of complementary DNA (cDNA). *P. chromatophora* CCAC0185 was grown as described before (Nowack et al., 2016). Total RNA was extracted and cDNA prepared as described in (Macorano et al., 2023).

Construction of expression plasmids. The $crTP_{part2}$ domains alone or $crTP_{part2}$ domains plus their cargo proteins were cloned into the expression vector GPN131. This vector is a derivative of the plasmid pET-22b(+) (Novagene; 69744), in which the pelB sequence and C-terminal His_6-tag were replaced by an N-terminal His_6-tag, thrombin cleavage site, and SUMO-tag. For details see the **Suppl. Text, Fig. S14, and Table S6**.

Heterologous expression of recombinant proteins. For overexpression of the constructs His_6-SUMO-TEV-$crTP_{part2\_RnaH}$, His_6-SUMO-TEV-$crTP_{part2\_ArgC}$, His_6-SUMO-TEV-$crTP_{part2\_CysK}$, His_6-SUMO-TEV-$crTP_{part2\_ArgC}$-ArgC, and His_6-SUMO-TEV-$crTP_{part2\_RnaH}$-RnaH (**see Fig. S2A**), plasmids GPN142, GPN167, GPN168, GPN195, and GPN194, respectively, were individually transformed into *E. coli* strain LOBSTR-BL21(DE3)-RIL (Kerafast, Boston, MA) (Andersen et al., 2013) and proteins were expressed under conditions detailed in the **Suppl. Text**. Finally, cells were harvested, pellets flash frozen and stored at -80°C until use.

Protein purification. Frozen cells from expression cultures were lysed and the His_6-SUMO-tagged proteins of interest isolated by immobilized metal ion chromatography (IMAC). The His_6-SUMO tag was cleaved of by TEV protease and the proteins of interest purified by reverse IMAC followed by size exclusion chromatography (SEC). For details see the **Suppl. Text**. Protein amounts were determined by a nanophotometer (NP80, Implen). Obtained fractions were analyzed by SDS-PAGE under denaturing conditions on 12.5% polyacrylamide (ROTIPHORESE® 30; 29:1; Roth) Tris-glycine gels (Schägger, 2006) (**Fig. S2**) and BN PAGE

on 4-16% gels (SERVA, SERVAGel™ N 4 - 16 Cat. No. 43204) according to the manufacturer's recommendations (**Fig. S3**), both stained with Coomassie Brilliant Blue R250.

Protein crystallization and 3D structure determination by X-ray crystallography. $CrTP_{part2\_RnaH}$ was crystallized at 12°C with 1.5 µl of 12 mg/ml protein in buffer A (see **Suppl. Text**), mixed with 1.5 µl 23% PEG 3350 in 0.1 M HEPES pH 8.5. $CrTP_{part2\_ArgC}$ was crystallized at 12°C with 0.1 µl of 12 mg/ml protein in buffer A mixed with 0.1 µl 0.1 M HEPES pH 6.5, 2.4 M $AmSO_4$ (final pH 7). Diffraction data from obtained crystals of both proteins were collected at the P13 beamline (PETRA III, DESY Hamburg) (Cianci et al., 2017). More details on experimentation, data collection, and refinement statistics are reported in **Table S1** and the **Suppl. Text**. Figures were generated using PyMOL (Schrodinger LLC; www.pymol.org).

Small-angle X-ray scattering. SAXS data of $crTP_{part2\_RnaH}$, $crTP_{part2\_ArgC}$, and $crTP_{part2\_CysK}$ were collected on the P12 beamline at PETRA III, DESY, Hamburg) (Blanchet et al., 2015), and of $crTP_{part2\_ArgC}$-ArgC and $crTP_{part2\_RnaH}$-RnaH on our Xeuss 2.0 Q-Xoom system from Xenocs. Primary data reduction was performed with the program PRIMUS (Konarev et al., 2003). With the Guinier approximation (Guinier, 1939) implemented in PRIMUS, we determine the forward scattering *I(0)* and the radius of gyration (R_g) and used the program GNOM (Svergun, 1992) to estimate the maximum particle dimension (D_{max}) with the pair-distribution function *p(r)*. Comparison of the theoretical scattering intensity of the solved crystal structures against the experimental scattering data was done with CRYSOL (Svergun et al., 1995). Flexible parts of the proteins were analyzed using EOM (Bernadó et al., 2007; Tria et al., 2015) and rigid body modeling with CORAL (Petoukhov et al., 2012). Details are provided in the **Suppl. Text** and **Table S2**.

Phylogenetic analysis. Sequences of $crTP_{part2}$ from indicated transcripts were aligned with diverse GGCT-like superfamily proteins downloaded from NCBI. A structure-guided alignment was generated using PROMALS3D (Pei et al., 2008). The ML tree was inferred with iqtree2 (Nguyen et al., 2015; Minh et al., 2020) using automatic model selection and 1000 ultrafast bootstrap replicates.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers _GEZN01002575.1, GEZN01007010.1, and GEZN01004327.1.

Data availability

Solved protein structures were deposited in the Worldwide Protein Data Bank (https://www.rcsb.org) with accession codes provided in **Table S1**. SAXS data were uploaded to the Small Angle Scattering Biological Data Bank (SASBDB) (Kikhney et al., 2020) with accession codes provided in **Table S2**.

Acknowledgements

We acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for provision of experimental facilities. Parts of this research were carried out at PETRA III and we thank Tobias Gräwert, Dmytro Soloviov, and Cy M. Jeffries (EMBL Hamburg) for assistance in using beamline P12 and Gleb Bourenkov for P13. We thank the beamline scientist from ID23-eh1 at the ESRF Grenoble for their help in the initial stages of this project where the initial crystals were tested, and Christian Mammen and Alice Pawlowski for their help with the BN PAGE.

Author contributions

V.K., J.R., S.H.J.S., and E.C.M.N. designed the research; all authors performed research and analyzed data; V.A. and A.H. crystalized the proteins; A.P. performed SEC-MALS experiments; E.C.M.N., S.H.J.S, and J.R. wrote the paper with contributions of all co-authors.

Funding

This work was partially funded by SFB 1208 of the DFG, Project ID 267205415 (to ECMN), SFB1535 of the DFG, Project ID 458090666 (to SHJS). The Center for Structural Studies is funded by the DFG (Grant number 417919780 and INST 208/761-1 FUGG, INST 208/740-1 FUGG, and INST 208/868-1 FUGG to SS).

References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Zemgulyte A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Zídek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature **630:** 493-500

Andersen KR, Leksa NC, Schwartz TU (2013) Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins: Structure, Function and Bioinformatics **81:** 1857-1861

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

Chi ZK, Byrne ST, Dolinko A, Harraz MM, Kim MS, Umanah G, Zhong J, Chen R, Zhang JM, Xu JC, Chen L, Pandey A, Dawson TM, Dawson VL (2014) Botch is a gamma-glutamyl cyclotransferase that deglycinates and antagonizes Notch. Cell Reports **7:** 681-688

Cianci M, Bourenkov G, Pompidor G, Karpics I, Kallio J, Bento I, Roessle M, Cipriani F, Fiedler S, Schneider TR (2017) P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. Journal of Synchrotron Radiation **24:** 323-332

Coale TH, Loconte V, Turk-Kubo KA, Vanslembrouck B, Mak WKE, Cheung S, Ekman A, Chen J-H, Hagino K, Takano Y, Nishimura T, Adachi M, Le Gros M, Larabell C, Zehr JP (2024) Nitrogen-fixing organelle in a marine alga. Science **384:** 217–222

Demchick P, Koch AL (1996) The permeability of the wall fabric of *Escherichia coli* and *Bacillus subtilis*. Journal of Bacteriology **178:** 768-773

Guinier A (1939) Small-angle X-ray diffraction: application to the study of ultramicroscopic phenomena. Annales de Physique **11:** 161-237

Hoiczyk E, Hansel A (2000) Cyanobacterial cell walls: News from an unusual prokaryotic envelope. Journal of Bacteriology **182:** 1191-1199

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

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

Kumar S, Kaur A, Chattopadhyay B, Bachhawat AK (2015) Defining the cytosolic pathway of glutathione degradation in *Arabidopsis thaliana*: role of the ChaC/GCG family of gamma-glutamyl cyclotransferases as glutathione-degrading enzymes and AtLAP1 as the Cys-Gly peptidase. Biochemical Journal **468:** 73-85

Lhee D, Lee J, Ettahi K, Cho CH, Ha JS, Chan YF, Zelzion U, Stephens TG, Price DC, Gabr A, Nowack ECM, Bhattacharya D, Yoon HS (2021) Amoeba genome reveals dominant host contribution to plastid endosymbiosis. Molecular Biology and Evolution **38:** 344-357

Macorano L, Binny TM, Spiegl T, Klimenko V, Singer A, Oberleitner L, Applegate V, Seyffert S, Stefanski A, Gremer L, Gertzen CGW, Höppner A, Smits SHJ, Nowack ECM (2023) DNA-binding and protein structure of nuclear factors likely acting in genetic information processing in the *Paulinella* chromatophore. Proc. Natl. Acad. Sci. USA **120:** e2221595120

Margittai É, Enyedi B, Csala M, Geiszt M, Bánhegyi G (2015) Composition of the redox environment of the endoplasmic reticulum and sources of hydrogen peroxide. Free Radical Biology and Medicine **83:** 331-340

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution **37:** 1530-1534

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution **32:** 268-274

Nowack ECM, Grossman AR (2012) Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. Proc. Natl. Acad. Sci. USA **109:** 5340-5345

Nowack ECM, Price DC, Bhattacharya D, Singer A, Melkonian M, Grossman AR (2016) Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. Proc. Natl. Acad. Sci. USA **113:** 12214-12219

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

Oberleitner L, Perrar A, Macorano L, Huesgen PF, Nowack ECM (2022) A bipartite chromatophore transit peptide and N-terminal protein processing in the *Paulinella* chromatophore. Plant Physiology **189:** 152-164

Oberleitner L, Poschmann G, Macorano L, Schott-Verdugo S, Gohlke H, Stühler K, Nowack ECM (2020) The puzzle of metabolite exchange and identification of putative octotrico peptide repeat expression regulators in the nascent photosynthetic organelles of *Paulinella chromatophora*. Frontiers in Microbiology **11:** 607182

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

Schägger H (2006) Tricine-SDS-PAGE. Nature Protocols **1:** 16-22

Singer A, Poschmann G, Mühlich C, Valadez-Cano C, Hänsch S, Rensing SA, Stühler K, Nowack ECM (2017) Massive protein import into the early evolutionary stage photosynthetic organelle of the amoeba *Paulinella chromatophora*. Current Biology **27:** 2763-2773

Sørensen MES, Stiller ML, Kröninger L, Nowack ECM (2025) Protein import into bacterial endosymbionts and evolving organelles. FEBS Journal **292:** 2992-3013

Steiner JM, Löffelhardt W (2002) Protein import into cyanelles. Trends in Plant Science **7:** 72-77

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

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

Williamson MP (1994) The structure and function of proline-rich regions in proteins. Biochemical Journal **297:** 249-260

Wróblewski K, Zalewski M, Kuriata A, Kmiecik S (2025) CABS-flex 3.0: an online tool for simulating protein structural flexibility and peptide modeling. Nucleic Acids Research **53:** W95-W101

Figure Legends

Figure 1: $CrTP_{part2}$ adopts a structured fold. (A) Multiple sequence alignment of 9 representative crTP sequences (ClustalX2, manually refined). Identifiers of proteins that were experimentally studied are highlighted in color. In $crTP_{part1}$, conserved hydrophobic helix and adapter protein complex 1 binding site (AP-1 BS) are indicated. In $crTP_{part2}$, secondary structure elements resolved by X-ray crystallography are provided underneath the alignment. Interacting Cys residues are indicated by asterisks. Conserved hydrophobic aa involved in the interaction between arm and core are marked by arrow heads. **(B-C)** Cartoon representation of crystal structure of $crTP_{part2\_RnaH}$ (B; pdb id 9I09) and $crTP_{part2\_ArgC}$ (C; pdb id 9I08). Flexible arm, blue; structured core, gold. **(D)** Superposition of the cores of $crTP_{part2\_RnaH}$ (blue; Leu114 to Gln220) and $crTP_{part2\_ArgC}$ (green; Pro93 to Gln181). **(E)** Disulfide bridge between Cys113 and Cys169

(red) stabilizes the β-barrel of $crTP_{part2\_RnaH}$. In the "open conformation", extension of the flexible arm results in exposure of hydrophobic aas (orange). **(F)** In the "closed conformation" of $crTP_{part2\_ArgC}$, the flexible arm interacts with the core via hydrophobic interactions (aas marked in orange).

Figure 2: SAXS refined models of the analyzed proteins. **(A)** The $crTP_{part2\_RnaH}$ core from the crystal structure is shown as orange cartoon and the solved part of the N-terminal arm as blue sphere representation. The core remained rigid for the EOM analysis and only the N and C-terminal parts were completed and used as flexible tails for the modeling. For clarity, only the N-terminal part is shown in sphere representations (volume fractions in green 22%, yellow 56%, cyan 11%, and magenta 11%). **(B)** Volume fractions from the $crTP_{part2\_RnaH}$ EOM analysis in the corresponding color code shown as spheres in (A). **(C)** EOM models of $crTP_{part2\_RnaH}$-RnaH. The RnaH core (orange) was used as rigid body. The solved crystal structure of $crTP_{part2}$ was used as flexible template and the missing linker regions were remodeled with EOM. **(D)** Dimer model of $crTP_{part2\_ArgC}$-ArgC. The ArgC protomer dimer interface (grey and orange) was used as rigid body and the solved $crTP_{part2}$ as flexible template. The flexible linkers and $crTP_{part2}$ are colored in green and cyan for each protomer. The corresponding volume fractions are indicated below of each model in (C) and (D).

Figure 3: ML tree depicting the inferred phylogenetic relationship between the structured core of $crTP_{part2}$ and members of the GGCT-like superfamily. Ultrafast bootstrap values ≥95 are shown at branches. Eukaryotic sequences are in black, bacterial in orange, archaeal in blue. Protein structures available are represented as cartoon models in which the part that is structurally similar to the $crTP_{part2}$ structured core is highlighted in colors. Note that the *B. subtilis* homolog of BtrG (named YkqA, of unknown function; pdb id 2qik) contains two cyclotransferase domains within one polypeptide chain. Only the second is shown here and in Fig. S13. Species abbreviations: At, *A. thaliana*; Bs, *B. subtilis*; Ce, *C. elegans*; Dm, *Drosophila melanogaster*; Dr, *Danio rerio*; Ec, *E. coli*; Gg, *Gallus gallus*; Hs, *H. sapiens*; Mj, *Methanocaldococcus jannaschii* DSM; Mm, *Mus musculus*; Nc, *Niallia circulans*; Os, *Oryza sativa*; Pf, *Pyrococcus furiosus*; Ph, *Pyrococcus horikoshii*; Sc, *S. cerevisiae*; Vc, *Vibrio cholerae*; Xl, *Xenopus laevis*; Yp, *Yersinia pestis*. **(B-D)** Superposition of the substrate-binding pocket of the human GGACT (pdb id: 3juc, violet with single residues highlighted in magenta) in complex with the product 5-oxoproline (shown in blue) and the one of $crTP_{part2\_ArgC}$ (green with single residues highlighted in light green). **(C)** Detail showing the conserved Tyr and Gly of the YGSL motif highlighted as well as other conserved or conservatively replaced hydrophobic side chains facing the substrate-binding pocket. **(D)** Detail showing the replacement of the catalytic Glu82 typical for GGCT-like proteins by a Tyr in $crTP_{part2\_ArgC}$.

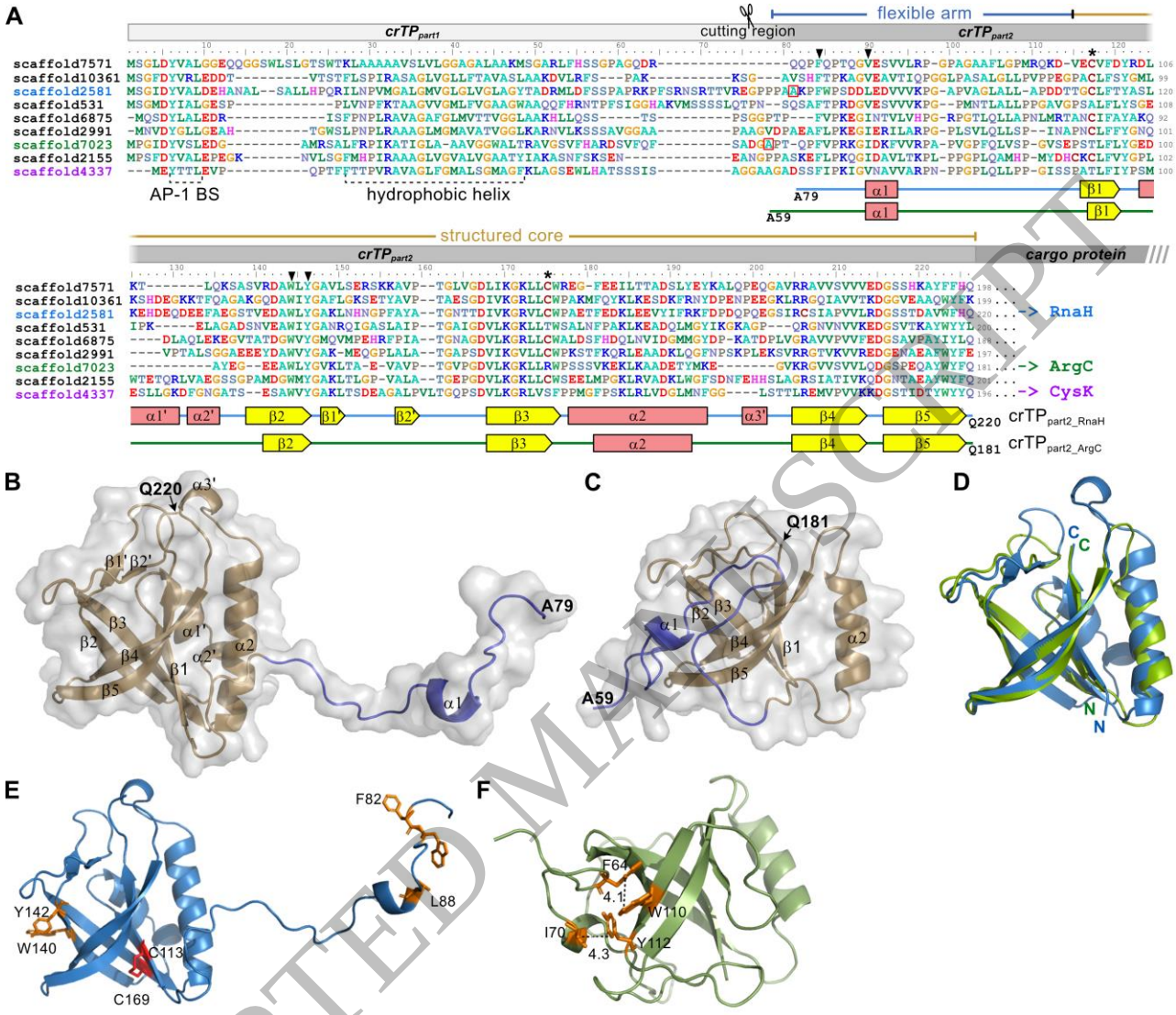

Figure 1
169x146 mm (x DPI)

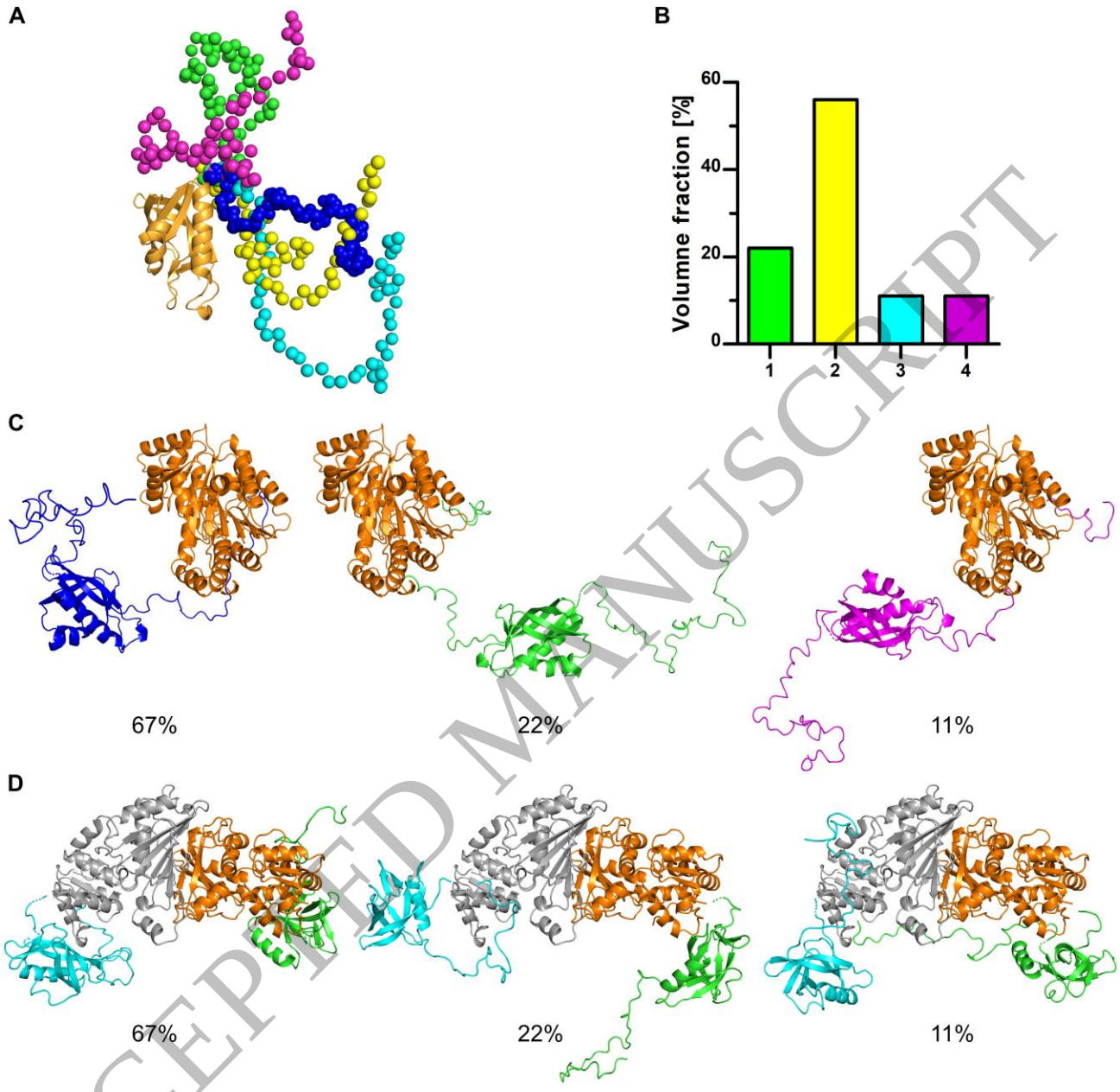

Figure 2
171x165 mm (x DPI)

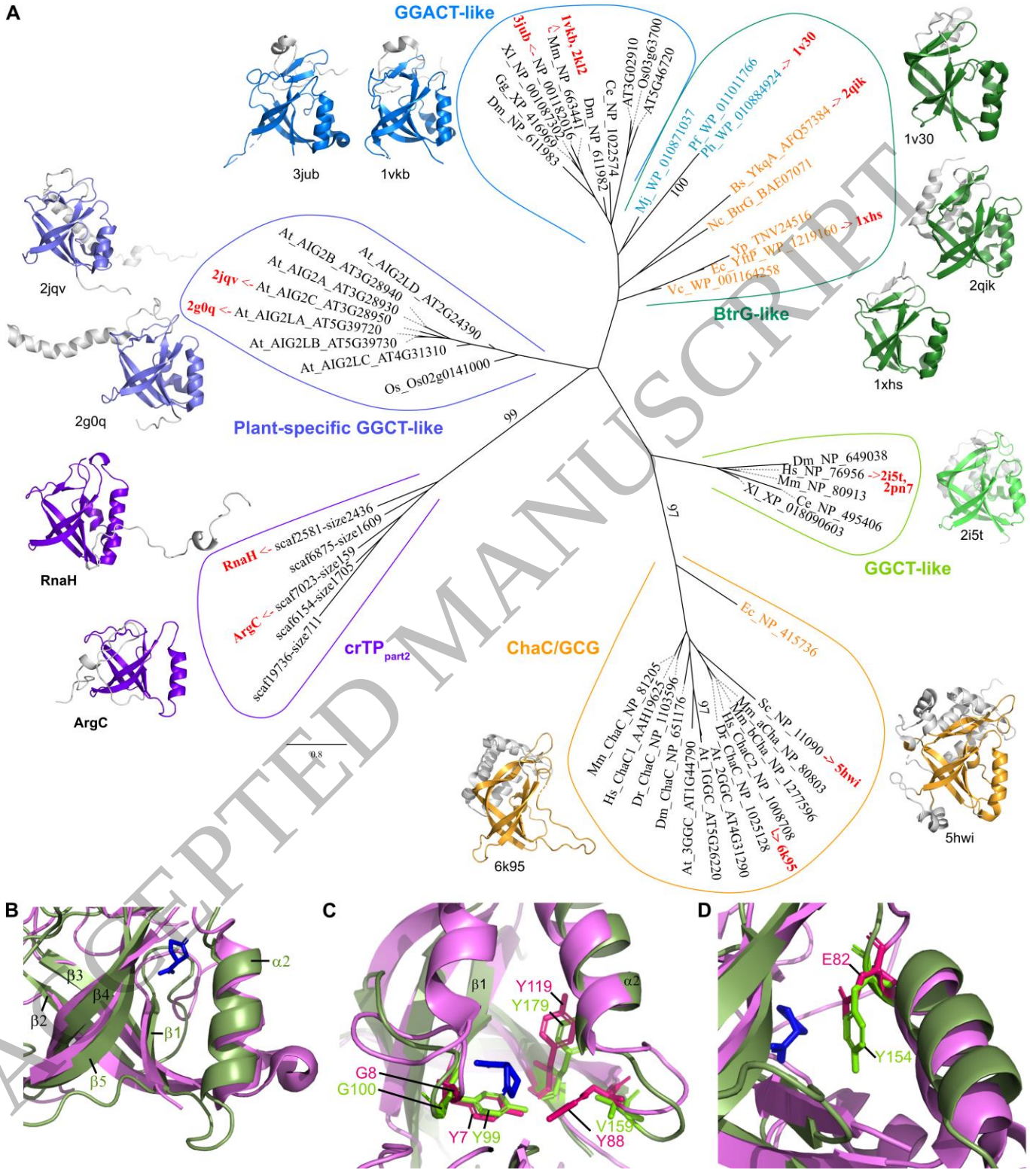

Figure 3
181x205 mm (x DPI)