



Figure 1. Sequence and Structure Characterization of GhCPP1-A0A1U8HKT6:

(A) Domain architecture of GhCPP1-A0A1U8HKT6 created using Prosite:MyDomains (Hulo et al. 2008) based on predictions from InterPro (Blum et al. 2021), TM indicates transmembrane, cTP is chloroplast transit peptide. (B) Multi-Sequence Alignment of homologs from *B. angustatum* (BaCPP1), *P. patens* (PpCPP1), *N. benthamiana* (NbCPP1), *A. thaliana* (AtCPP1), and *G. hirsutum* (GhCPP1) created with ClustalOmega (Madeira et al. 2019) and ESPript3 (Robert and Gouet 2014) highlighting locations of the AtCPP1 J-like domain (blue box) and GhCPP1 predicted transmembrane regions (yellow boxes). The green box indicates the corresponding location of the HPD motif present in DnaJ molecular chaperone J domains that is absent in J-like domains. ConSurf (Ashkenazy et al. 2016) identified highly-conserved residues are denoted with an asterisk. (C) AlphaFold model of GhCPP1-A0A1U8HKT6 amino acids 56-254 displayed by ChimeraX (Version 1.3, Pettersen et al. 2021; Jumper et al. 2021) in rainbow format with the N-terminus in blue and C-terminus in red. (D) Hydrophobic surface of the GhCPP1-A0A1U8HKT6 model structure with teal representing low hydrophobicity, and brown representing high hydrophobicity. (E) Proposed J-like domain of GhCPP1-A0A1U8HKT6 in orange (amino acids 63-119) overlaid with the J-domain of the human Tid1 protein (PDB ID: 2DN9) in blue. (F) Overlay of GhCPP1-A0A1U8HKT6 J-like domain (ConSurf coloring based on conservation) with (PDB 2GUZ) yeast mitochondrial import motor protein subunit TIM16 (light blue ribbon), showing potential interaction surface with TIM14 subunit (dark blue ribbon).

Description

Introduction

Cotton plays a significant role in the United States economy as it is estimated to generate over 200,000 jobs and \$25 billion dollars in annual revenue (Raper et al. 2019). By the end of the 2022-2023 growing season, the United States Department of Agriculture (USDA) forecasts that the US will produce 18.2 million bales of cotton, a 4.5% increase of production from the 2021-2022 season (Johnson et al. 2022). Research on cotton is essential as the price of cotton and the financial gain to the cotton grower is dependent on the fiber yield and quality. Recently, the genomes of five allotetraploid cotton species (*Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium mustelinum*, *Gossypium tomentosum*, *Gossypium darwinii*) were sequenced (Chen et al. 2020), and analysis of gene annotations revealed many genes of unknown function in common across the species.

We predict that one of these genes of unknown function, the *Gossypium hirsutum* gene Gohir.A03G073700 (CottonGen: mRNA 'Gohir.A03G073700.1_UTX-TM1_v2.1' https://www.cottongen.org/bio_data/5920329), and associated protein (NCBI: XP_016666667.1; UniProt: A0A1U8HKT6) here referred to as GhCPP1-A0A1U8HKT6, is a chaperone-like protein of POR1 (CPP1). Angiosperms, such as *G. hirsutum*, require light for the biosynthesis of the green pigment chlorophyll. In one of the later steps of the biosynthetic pathway, protochlorophyllide is reduced to chlorophyllide with facilitation and catalysis by protochlorophyllide oxidoreductase (POR), a light-dependent enzyme (Lee et al. 2013). Other oxygenic photosynthetic organisms also contain variations of POR genes, especially those with homology to angiosperms (Masuda and Takamiya 2004). To complete chlorophyll biosynthesis, the POR enzyme must be post-translationally imported into the plastid (Reinbothe et al. 1995). The chaperone-like protein of POR1 (CPP1), previously known as cell growth defect factor-1, (CDF1; At5g23040) has been shown to regulate POR permanence and function during this process (Lee et al. 2013). Based on the function of CPP1 proteins, they are typically predicted to have subcellular localizations within chloroplast membranes with both cytoplasmic and transmembrane domains in their sequence. The GhCPP1-A0A1U8HKT6 protein evaluated here is likely a chaperone-like protein of POR1 found in *G. hirsutum* due to its structural and sequence features depicting similarity to already identified CPP1 proteins in angiosperms and related species.

Sequence Features

The InterPro webserver (Blum et al. 2021) identified the 254 amino acid GhCPP1-A0A1U8HKT6 protein as a member of the chaperone-like POR1 protein family (CPP1) (IPRO21788, PF11833). Sequence analysis of GhCPP1-A0A1U8HKT6 by the subcellular localization programs Plant-mPLoc (Chou and Shen 2010) and DeepLoc (Thumhuri et al. 2022) predicted the location of this protein to be the plastid. Furthermore, TargetP indicated that GhCPP1-A0A1U8HKT6 contains a chloroplast transit peptide cleavage site at amino acid position 45 (Almagro Armenteros et al. 2019). This is in agreement with experimental localization of the *Arabidopsis* CPP1 protein (At5g23040) to the thylakoid and envelope membranes (Lee et al. 2013). Transmembrane (TM) region predictions were performed using Phobius (Käll et al. 2007), TMHMM (Krogh et al. 2001), and InterPro (Blum et al. 2021). All three bioinformatic programs predicted that four TM regions were found in the GhCPP1-A0A1U8HKT6 protein, between amino acids 152-169, 175-191, 198-220, and 232-253. A domain architecture was created to visualize these sequence features (**Figure 1A**). Interestingly, *Arabidopsis thaliana* and *Nicotiana benthamiana* CPP1 proteins are indicated to contain three, not four, TM regions (Lee et al. 2013). However, a homolog of *Arabidopsis* CPP1,



At2g20920, has been identified to also contain four TM regions (Kawai-Yamada et al. 2014) suggesting a potential structural or functional difference between these family members.

Homology

Homology of GhCPP1-A0A1U8HKT6 to *Arabidopsis* proteins was assessed using the TAIR database (Berardini et al. 2015), and the CPP1 protein (At5g23040, Q9FN50) was identified to be 80.3% identical with an 84% query cover. Additionally, the TAIR database indicates At5g23040 has orthologs in eudicots, monocots, lycophytes, bryophytes, and other photosynthetic organisms such as the single-celled green algae *Ostreococcus tauri* and *Chlamydomonas reinhardtii*. AtCPP1 promoter activity is low in the globular stage of *Arabidopsis* embryogenesis but increases throughout the heart and linear stages and persists through the mature embryo stage (Kawai-Yamada et al. 2014). While homozygous knockouts of AtCPP1 are lethal (Lee et al. 2013), suppression of AtCPP1 by RNA interference completely blocked the *in vitro* import of *Arabidopsis* PORA and PORB isoforms into isolated etioplasts (Reinbothe et al. 2015). Although protein-protein interactions have been demonstrated between CPP1 and POR by fluorescent co-localization and immunoprecipitation (Lee et al. 2013), CPP1 is not involved in the process of protochlorophyllide or NADPH binding to POR (Reinbothe et al. 2015). Instead, due to the structural similarity between CPP1 and the DnaJ/Hsp40 protein family, it has been suggested that CPP1 acts as a “holdase” chaperone to permit preliminary cytoplasmic steps before initiating POR import into the plastid.

The “holdase” activity of the molecular co-chaperone DnaJ has been attributed to a J domain which is an approximately 70 amino-acid sequence containing four helices with a loop region between the second and third helix (Qiu et al. 2006). The J domain is found across all kingdoms and plays an important role in organizing interactions with Hsp70 chaperone partners (Kelley 1998). The presence of a functionally conserved His-Pro-Asp (HPD) motif within J domains of molecular chaperones, such as the human Tid1 protein, facilitates the ATPase activity of DnaK-like co-chaperones with which they interact (Lu et al. 2006). The J-like domain is also found across kingdoms and, although very similar in structure to the J domain, the HPD motif is absent in J-like domain proteins such as CPP1, indicating these domains may perform a similar function but have diverged interaction partners. Amino acid sequences of CPP1 homologs from the eudicots *N. benthamiana* (NbCPP1; AEP68099) and *A. thaliana* (AtCPP1; At5g23040), the bryophyte *Physcomitrella patens* (PpCPP1; XP_024395830.1), and a cyanobacterium *Brasilonema angustatum* (BaCPP1; MBW4597090.1) were aligned with GhCPP1-A0A1U8HKT6 (**Figure 1B**). The blue box indicates the location of the J-like domain in *Arabidopsis*, *N. benthamiana*, and *P. patens* (Lee et al. 2013) and the green box indicates where the HPD motif would be located in a J domain (Qiu et al. 2006) although this motif is lacking in the J-like domain containing CPP1 homologs as well as GhCPP1-A0A1U8HKT6 (Lee et al. 2013). Significant sequence conservation is observed from photosynthetic bacteria to eudicots in both the J-like domain and transmembrane domains (yellow boxes indicate *G. hirsutum* TM domains), with highly-conserved residues in GhCPP1-A0A1U8HKT6 identified by ConSurf (Ashkenazy et al. 2016) indicated with asterisks. The full ConSurf results are available as Extended Data.

Structural Features

AlphaFold and ChimeraX (Version 1.3, Pettersen et al. 2021; Jumper et al. 2021) were used to predict and visualize the model structure of GhCPP1-A0A1U8HKT6. **Figure 1C** depicts two high-confidence modeled regions between amino acids 56-119 and 150-254 and a low-confidence modeled region between them from amino acids 119-150. The structure is oriented vertically to show the cytoplasmic domain at the lower part of the image and the four transmembrane helices at the upper part of the image. The transmembrane domains exhibit expected high hydrophobicity as visualized by the ChimeraX surface calculation tool (**Figure 1D**, same orientation as **1C**). A structural overlay of the J-like domain (amino acids 63-119) of GhCPP1-A0A1U8HKT6 with the J domain of the human DnaJ family member Tid1 (PDB ID: 2DN9) displays a similar anti-parallel hairpin between the horizontally-displayed alpha helices II and III, although GhCPP1-A0A1U8HKT6 lacks the fourth helix of human Tid1 and has a shorter loop (**Figure 1E**). The structural similarity with key sequence differences between a cotton protein and human protein demonstrates this is a highly conserved structural domain that likely has developed a unique function.

When the DALI database (Holm 2020) was searched for proteins having a similar tertiary structure to the proposed J-like domain of GhCPP1-A0A1U8HKT6, the second highest similarity hit (Z-score of 7.9 and RMSD score of 1.5) was with the J-like domain of yeast TIM16 that is part of the translocase of the inner membrane protein complex, TIM16/TIM14 (PDB ID: 2GUZ). The TIM16/TIM14 heterodimer complex serves as a co-chaperone duo, a critical part of the translocase import motor responsible for piloting translocation proteins into the mitochondria (Mokranjac et al. 2006). As previously mentioned, the biosynthesis of chlorophyll in plants requires the light-dependent enzyme POR1 to be post-translationally imported into the plastid (Aronsson et al. 2003). It is possible that the J-like domain found in GhCPP1-A0A1U8HKT6 may allow it to partner with an as yet unknown co-chaperone to function similarly to the TIM16/TIM14 complex, chaperoning the translocation of POR1 into the plastid. This is supported by an overlay (**Figure 1F**) of the GhCPP1-A0A1U8HKT6 J-like domain (transparent



ConSurf colored surface) with the J-like domain of TIM16 (light blue ribbon, PDB 2GUZ) that shows a highly conserved surface in GhCPP1-A0A1U8HKT6 at the modeled protein-protein interaction site with the TIM14 protein (dark blue ribbon).

Conclusion

Evidence from protein sequence, homology, and structure support that GhCPP1-A0A1U8HKT6 is likely a chaperone-like protein of POR1 in *Gossypium hirsutum*. GhCPP1-A0A1U8HKT6 is predicted to be anchored in the plastid membrane of the chloroplast where it could interact with POR and support its entry into the chloroplast where photochemical reactions of chlorophyll biosynthesis occur, similar to Arabidopsis CPP1 to which it has high sequence similarity. The molecular function of the J-like domain is still unclear. Similarity in structure to the J domain suggests a similar protein-protein interaction domain function as in the “holdase” activity of molecular chaperones, although the lack of a HPD motif indicates different binding partners. This is supported by the J-like domain in mitochondrial TIM16 being part of a heterodimer complex. It would be interesting for further studies to explore potential binding partners of GhCPP1-A0A1U8HKT6.

Proteopedia Molecular Tour

Acknowledgements: This research used resources provided by the SCINet project of the USDA Agricultural Research Service, ARS project number 0500-00093-001-00-D. The authors also thank the Bethel Chemistry Department for supporting the Course-based Undergraduate Research Experience that was the basis for data collection.

Extended Data

Description: ConSurf sequence conservation results. Resource Type: Dataset. File: [A0A1U8HKT6 ConSurf results.pdf](#). DOI: [10.22002/a9fcw-h5n51](#)

References

- Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, Nielsen H. 2019. Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance* 2: . PubMed ID: [31570514](#)
- Aronsson H, Sundqvist C, Dahlin C. 2003. POR - import and membrane association of a key element in chloroplast development. *Physiol Plant* 118: 1-9. PubMed ID: [12702007](#)
- Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 44: W344-50. PubMed ID: [27166375](#)
- Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E. 2015. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* 53: 474-85. PubMed ID: [26201819](#)
- Blum M, Chang HY, Chuguransky S, Grego T, Kandasamy S, Mitchell A, et al., Finn RD. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* 49: D344-D354. PubMed ID: [33156333](#)
- Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, et al., Schmutz J. 2020. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat Genet* 52: 525-533. PubMed ID: [32313247](#)
- Chou KC, Shen HB. 2010. Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One* 5: e11335. PubMed ID: [20596258](#)
- Holm L. 2020. Using Dali for Protein Structure Comparison. *Methods Mol Biol* 2112: 29-42. PubMed ID: [32006276](#)
- Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche BA, de Castro E, et al., Sigrist CJ. 2008. The 20 years of PROSITE. *Nucleic Acids Res* 36: D245-9. PubMed ID: [18003654](#)
- Johnson J, MacDonald S, Meyer L, Soley G. 2022. Cotton Outlook for 2022. *Agricultural Outlook Forum 2022* 321078, United States Department of Agriculture, Agricultural Outlook Forum. DOI: [10:22004/ag.econ.321078](#)
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al., Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583-589. PubMed ID: [34265844](#)
- Käll L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res* 35: W429-32. PubMed ID: [17483518](#)
- Kawai-Yamada M, Nagano M, Kakimoto M, Uchimiya H. 2014. Plastidic protein Cdf1 is essential in Arabidopsis embryogenesis. *Planta* 239: 39-46. PubMed ID: [24097264](#)



Kelley WL. 1998. The J-domain family and the recruitment of chaperone power. *Trends Biochem Sci* 23: 222-7. PubMed ID: [9644977](#)

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305: 567-80. PubMed ID: [11152613](#)

Lee JY, Lee HS, Song JY, Jung YJ, Reinbothe S, Park YI, Lee SY, Pai HS. 2013. Cell growth defect factor1/chaperone-like protein of POR1 plays a role in stabilization of light-dependent protochlorophyllide oxidoreductase in *Nicotiana benthamiana* and *Arabidopsis*. *Plant Cell* 25: 3944-60. PubMed ID: [24151298](#)

Lu B, Garrido N, Spelbrink JN, Suzuki CK. 2006. Tid1 isoforms are mitochondrial DnaJ-like chaperones with unique carboxyl termini that determine cytosolic fate. *J Biol Chem* 281: 13150-13158. PubMed ID: [16531398](#)

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al., Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47: W636-W641. PubMed ID: [30976793](#)

Masuda T, Takamiya K. 2004. Novel Insights into the Enzymology, Regulation and Physiological Functions of Light-dependent Protochlorophyllide Oxidoreductase in Angiosperms. *Photosynth Res* 81: 1-29. PubMed ID: [16328844](#)

Mokranjac D, Bourenkov G, Hell K, Neupert W, Groll M. 2006. Structure and function of Tim14 and Tim16, the J and J-like components of the mitochondrial protein import motor. *EMBO J* 25: 4675-85. PubMed ID: [16977310](#)

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* 30: 70-82. PubMed ID: [32881101](#)

Qiu XB, Shao YM, Miao S, Wang L. 2006. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell Mol Life Sci* 63: 2560-70. PubMed ID: [16952052](#)

Raper TB, Pilon C, Singh V, Snider J, Stewart S, Byrd S. 2019. Cotton Production in the United States of America. *Cotton Production* : 217-247. DOI: [10.1002/9781119385523.ch11](#)

Reinbothe S, Runge S, Reinbothe C, van Cleve B, Apel K. 1995. Substrate-dependent transport of the NADPH:protochlorophyllide oxidoreductase into isolated plastids. *Plant Cell* 7: 161-72. PubMed ID: [7756827](#)

Reinbothe S, Gray J, Rustgi S, von Wettstein D, Reinbothe C. 2015. Cell growth defect factor 1 is crucial for the plastid import of NADPH:protochlorophyllide oxidoreductase A in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 112: 5838-43. PubMed ID: [25901327](#)

Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42: W320-4. PubMed ID: [24753421](#)

Thumhuri V, Almagro Armenteros JJ, Johansen AR, Nielsen H, Winther O. 2022. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res* 50: W228-W234. PubMed ID: [35489069](#)

Wu Q, Peng Z, Zhang Y, Yang J. 2018. COACH-D: improved protein-ligand binding sites prediction with refined ligand-binding poses through molecular docking. *Nucleic Acids Res* 46: W438-W442. PubMed ID: [29846643](#)

Yu J, Jung S, Cheng CH, Lee T, Zheng P, Buble K, et al., Main D. 2021. CottonGen: The Community Database for Cotton Genomics, Genetics, and Breeding Research. *Plants (Basel)* 10: . PubMed ID: [34961276](#)

Funding: This research was funded in part by the U.S. Department of Agriculture Agricultural Research Service (USDA-ARS) project number 6066-21310-005-00D and Cotton Incorporated project 18-274 to AMH-K. We have no additional external funding to report.

Author Contributions: Alana N. Osborne: formal analysis, investigation, visualization, writing - original draft, writing - review editing. Andrew Osagiede: formal analysis, investigation, visualization, writing - original draft. Amanda R. Storm: conceptualization, data curation, formal analysis, methodology, supervision, validation, project administration, writing - review editing. Amanda M. Hulse-Kemp: conceptualization, data curation, writing - review editing, funding acquisition. Angela K. Stoeckman: formal analysis, investigation, project administration, supervision, validation, visualization, writing - original draft, writing - review editing.

Reviewed By: Paul Craig

History: Received May 22, 2023 **Revision Received** June 23, 2023 **Accepted** July 27, 2023 **Published Online** July 30, 2023

Copyright: © 2023 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any



7/30/2023 - Open Access

medium, provided the original author and source are credited.

Citation: Osborne, AN; Osagiede, A; Storm, AR; Hulse-Kemp, AM; Stoeckman, AK (2023). *Gossypium hirsutum* gene of unknown function Gohir.A03G0737001 encodes a potential Chaperone-like Protein of protochlorophyllide oxidoreductase (CPP1). microPublication Biology. [10.17912/micropub.biology.000867](https://doi.org/10.17912/micropub.biology.000867)