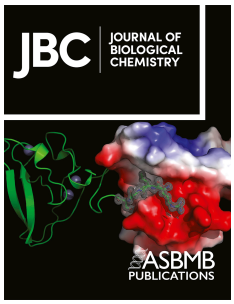
Journal Pre-proof



Functional diversification within the heme-binding split-barrel family

Nicolas Grosjean, Lifang Zhang, Desigan Kumaran, Meng Xie, Audrey Fahey, Kassandra Santiago, Fangle Hu, Michael Regulski, Ian K. Blaby, Doreen Ware, Crysten E. Blaby-Haas

PII: S0021-9258(24)02390-1

DOI: https://doi.org/10.1016/j.jbc.2024.107888

Reference: JBC 107888

To appear in: *Journal of Biological Chemistry*

Received Date: 14 May 2024

Revised Date: 29 September 2024

Accepted Date: 7 October 2024

Please cite this article as: Grosjean N, Zhang L, Kumaran D, Xie M, Fahey A, Santiago K, Hu F, Regulski M, Blaby IK, Ware D, Blaby-Haas CE, Functional diversification within the heme-binding split-barrel family, *Journal of Biological Chemistry* (2024), doi: https://doi.org/10.1016/j.jbc.2024.107888.

Functional diversification within the heme-binding split-barrel family

Nicolas Grosjean[1,#,$], Lifang Zhang[2,$], Desigan Kumaran[1,$], Meng Xie[1], Audrey Fahey[2], Kassandra Santiago[1], Fangle Hu[2], Michael Regulski[2], Ian K. Blaby[3], Doreen Ware[2,4*], Crysten E. Blaby-Haas[3,5*]

[1] Biology Department, Brookhaven National Laboratory, Upton, NY, 11973, USA
[2] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, USA
[3] US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA
[4] USDA ARS NEA Plant, Soil & Nutrition Laboratory Research Unit, Ithaca, NY, 14853, USA
[5] The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[#] present affiliation: US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA
[$]these authors contributed equally
*ware@cshl.edu
*cblaby@lbl.gov

Running title: Heme-binding split-barrel family in plants

Keywords: iron, chloroplast, AT3G03890, HOZ, AT1G51560, HOZ2A, AT3G21140, HOZ2B, GluTRBP

Abstract

Due to neofunctionalization, a single fold can be identified in multiple proteins that have distinct molecular functions. Depending on the time that has passed since gene duplication and the number of mutations, the sequence similarity between functionally divergent proteins can be relatively high, eroding the value of sequence similarity as the sole tool for accurately annotating the function of uncharacterized homologs. Here, we combine bioinformatic approaches with targeted experimentation to reveal a large multi-functional family of putative enzymatic and non-enzymatic proteins involved in heme metabolism. This family (homolog of HugZ (HOZ)) is embedded in the "FMN-binding split barrel" superfamily and contains separate groups of proteins from prokaryotes, plants, and algae, which bind heme and either catalyze its degradation or function as non-enzymatic heme sensors. In prokaryotes these proteins are often involved in iron assimilation, whereas several plant and algal homologs are predicted to degrade heme in the plastid or regulate heme biosynthesis. In the plant *Arabidopsis thaliana*, which contains two HOZ subfamilies that can degrade heme *in vitro* (HOZ1 and HOZ2), disruption of At*HOZ1* (AT3G03890) or At*HOZ2A* (AT1G51560) causes developmental delays, pointing to important biological roles in the plastid. In the tree *Populus trichocarpa*, a recent duplication event of a *HOZ1* ancestor has resulted in localization of a paralog to the cytosol. Structural characterization of this cytosolic paralog and comparison to published homologous structures suggests conservation of heme-binding sites. This study unifies our understanding of the sequence-structure-function relationships within this multi-lineage family of heme-binding proteins and presents new molecular players in plant and bacterial heme metabolism.

Introduction

Iron is an essential micronutrient that functions in a wide range of biological processes. To control the chemical properties of the iron atom and catalyze a specific activity, multiple proteins and complexes bind a prosthetic group where iron is bound to a protoporphyrin ring (*i.e.*, heme). The specific interactions between heme and the polypeptide tailors the function of the iron atom, such as tuning the redox potential or enabling oxygen transport. While the malleable function of heme has resulted in its use by a plethora of proteins, free heme can be highly toxic (1). As an example, the plant plastid has an absolute requirement for heme in many of the electron transfer reactions during photosynthesis, but free heme can react with lipoproteins and oxygen species leading to membrane and protein damage (2). As such, heme biosynthesis, availability, and turnover must be tightly regulated.

Enzymatic degradation to detoxify free heme and recycle the iron atom usually involves oxidative cleavage by a heme oxygenase. This enzymatic activity is proposed to have evolved at least three independent times. Once with the canonical heme oxygenases (*i.e.*, HO from mammals, HMX1 from yeast, and HmuO from *Corynebacterium diphtheriae*), which compose a family of proteins that share an overall α-helical structural fold (3, 4). At another time, heme oxygenases evolved from within the dimeric alpha-beta barrel superfamily (*i.e.*, IsdG and IsdI from *Staphylococcus aureus* (5, 6) and LFO1 from *Chlamydomonas reinhardtii* (7)) and are related to the antibiotic biosynthesis monooxygenases (ABM) (6). The third example of convergent evolution includes the proposed heme-degradation proteins from the "FMN-binding split barrel" superfamily (*i.e.*, HugZ from *Helicobacter pylori*, HutZ from *Vibrio cholerae*, and HOZ from *Arabidopsis thaliana*) (8–10).

This latter superfamily, "FMN-binding split barrel", comprises a collection of structurally similar proteins, which was named based on an FMN-binding protein of unknown biological function from *Desulfovibrio vulgaris* (11). Although some family members, such as pyridoxine 5′-phosphate oxidase and a ferric reductase from *Archaeoglobus fulgidus* (locus: AF_0830), have been found to bind FMN, many members do not, such as the heme-degradation proteins. Therefore, the name "FMN-binding split barrel" can be a source of confusion and has led to an unfortunate instance of mis-annotation when membership in this family is used to infer FMN binding. The superfamily has also been named flavin (FMN/FAD)/deazaflavin (F_{420}) oxidoreductases (FDOR) based on studies of a large group of related enzymes in mycobacteria and elsewhere (12), but, again, a number of family members do not bind either cofactor, nor do all family members function as oxidoreductases. Additionally, multiple members of the split-barrel superfamily contain an array of domains fused to the common split-barrel fold, such as the bifunctional epimerase/oxidase PDX3 in plants (13). Other family members, such as human CREG (14) and plant GluTRBP/GBP (15, 16) have no known enzymatic activity and, instead, function as regulators. Herein, we refer to this superfamily as the "split-barrel superfamily" to avoid presumption of function due to superfamily membership. Further, we refer to the heme-binding subfamily as the "heme-binding split-barrel family" for inclusion of proteins that function as enzymes or regulators.

The putative heme-degradation enzymes from the heme-binding split-barrel family have been found to have different substrates and products. HugZ from *Helicobacter pylori* can bind hemin and produce biliverdin IXδ in the presence of a reductant (17), while, for HutZ from *Vibrio cholerae*, the cleavage site is in the β- or δ-meso position (18). Diverging from these homologs, Pden_1323 from *Paracoccus denitrificans* cleaves *c*-type hemopeptides producing peptide-linked biliverdin (19), while HupZ from *Streptococcus pyogenes* is proposed to function as a heme chaperone (20). Analogous to HugZ and HutZ, the plant homolog HOZ is also capable of degrading heme to biliverdin *in vitro* (8), but while these bacterial proteins are involved in heme utilization, the biological function of HOZ in plants is unknown.

Here, we combine sequence and structure analyses with *in vitro* and *in vivo* experimentation to better understand the evolution and presence of heme-degradation activity among the split-barrel superfamily,

with a focus on plant homologs. We identify several distinct families of putative heme-binding proteins across plants, algae, fungi, and prokaryotes. Using a yeast-based complementation assay, we identify two *A. thaliana* family members: the recently described plastidial heme-degradation protein HOZ (renamed HOZ1) from *A. thaliana* (8) and a paralog that we have named *HOZ2A*, which rescue loss of the *HMX1* gene in *Saccharomyces cerevisiae.* Using CRISPR-Cas to generate loss-of-function mutants in *HOZ1* and a T-DNA insertion line for *HOZ2A*, we find that *HOZ1* and *HOZ2A* are important for plant development. Using structural and biochemical characterization of *HOZ1* orthologs from taxonomically diverse plants, we find that heme-degradation activity is conserved in grasses (represented by *Sorghum bicolor*) and trees (represented by *Populus trichocarpa*). A gene duplication event in a *Saliceae* ancestor followed by loss of the transit peptide resulted in the presence of a HOZ1 paralog in the cytosol of *P. trichocarpa*, in addition to retention of the ancestral plastidial enzyme. A crystal structure of the cytosolic poplar HOZ1-like protein suggests conservation of the heme-binding sites observed in bacterial proteins HugZ, ChuZ, and AtHOZ1.

Results

Phylogenomic analysis of the split-barrel family in plants and algae

Although commonly annotated based on their similarity to pyridoxamine 5'-phosphate oxidase (often abbreviated to PNPO or PPO), plant and algal members of the split-barrel family belong to multiple distinct subfamilies (Figure 1). Domain analysis combined with phylogenetic reconstruction encompassing prokaryotic homologs further supports similarity among members of individual subfamilies and divergence between subfamilies (Figure 2). Across the split-barrel family, three groups of plant proteins also contain the recently named Domain Related to Iron (DRI) found in the bacterial heme-utilization proteins, HugZ (10) and ChuZ (21), and in the heme-binding protein Dri1 from cyanobacteria (22) (Figures 1 and 2). Of the DRI-containing *A. thaliana* proteins, GBP (GluTR Binding Protein) orthologs form a sequence similarity cluster with two additional subclusters (Figure 1). Based on phylogeny of the split-barrel domain sequences, these proteins form three distinct clades (Figure 2F). One subfamily, which we have renamed HOZ1, contains the previously described HOZ protein from *A. thaliana* (8) and other uncharacterized plant and algal proteins. The second subfamily, which we have named HOZ2, contains uncharacterized plant and algal proteins. Given the phylogenetic relatedness between these eukaryotic split-barrel domain proteins, we refer to this subfamily of the heme-binding split-barrel family (*i.e.*, HOZ1, HOZ2, and GBP) as HOZ (Figure 2F).

The HOZ family is related to a large number of prokaryotic proteins. Like the *A. thaliana* HOZ family, the majority of prokaryotic proteins contain the split-barrel domain followed by DRI, which is opposite to the domain architecture in HugZ (Figure 2). To provide insight into the function of the plant HOZ family and related proteins, we analyzed the genomic context of the related prokaryotic homologs (Figure 3) (23). We identified three subfamilies of HOZ-like prokaryotic homologs that are commonly encoded by genes next to, and potentially sharing operons with, genes encoding proteins related to iron or heme transport and homeostasis, such as proteins similar to ferritin, transferrin, siderophore-interacting proteins, ChuX, HtaA, TonB-dependent receptors similar to CirA, and putative iron transporters from the NRAMP, BPD, and FecCD families (Figure 3B). These gene neighborhoods may represent heme-uptake and -utilization gene clusters. We also identified predicted protein fusions (in four separate species), represented by PAI11_25360 from *Patulibacter medicamentivorans*, involving a HOZ-like protein and an ABC-type periplasmic binding domain (Figure 3C). Specifically, the periplasmic binding protein region is similar to FepB and FhuD, which are periplasmic components of ABC-type Fe^{3+}-siderophore transport systems. Combined, these bioinformatic analyses suggest that the ancestor of the plant HOZ family was involved in heme utilization and, as such, a role in heme degradation could be conserved among different uncharacterized members of the plant HOZ family.

Yeast complementation suggests conservation and presence of a second HOZ-like heme-degradation subfamily in plants

To determine whether heme-degradation activity may be conserved among these plant homologs, we leveraged a yeast-based complementation assay. The *S. cerevisiae* genome encodes only one canonical heme oxygenase, Hmx1p. Deletion of *HMX1* has been previously shown to result in H_2O_2 sensitivity (24). Presumably, this sensitivity is at least partially due to the loss of enzymatic degradation of heme, a pro-oxidant that subsequently accumulates in the mutant, and to the loss of bilirubin, an antioxidant formed from the heme-degradation product biliverdin. Although, the exact mechanism of Hmx1p and/or its degradation products in abating oxidative stress may be more complex (24). Since recombinantly produced AtHOZ1 can degrade heme and produce biliverdin *in vitro* (8), we reasoned that expression of At*HOZ1* should be able to rescue the H_2O_2-sensitivity of the yeast *hmx1*Δ mutant, and that this strain and phenotype can be used to screen for conservation of heme-degradation activity. Indeed, like expression of the yeast *hmx1* gene *in trans*, At*HOZ1* can rescue the H_2O_2-dependent growth defect (Figure 4B). With the complementation assay established, we screened the other HOZ family *A. thaliana* genes. Of the 7 other *A. thaliana* proteins that contain DRI (Figure 2A), we only observed rescue by At*HOZ2A* (AT1G51560), suggesting that this closely related HOZ1-like subfamily also contains members that can degrade heme (Figure 4B). However, we note that the paralog At*HOZ2B* (AT3G21140) did not rescue the growth defect.

We also tested conservation of HOZ1 function across land plants and green algae. The HOZ1 family is typically found as single-copy genes in plant and algal genomes, and predicted plastid-localization transit peptides are generally conserved (Figure 4A). An exception is the tree *P. trichocarpa* (Figure 4A), where there are two paralogs: one with a predicted plastid transit peptide (Potri.019G035200; PtHOZ1A) and one predicted to localize to the cytosol (Potri.013G057700; PtHOZ1B). The presence of the putatively cytosolic paralog suggests that neofunctionalization may have resulted in the evolution of HOZ1-catalyzed heme degradation in the cytosol. To test this hypothesis, we performed localization experiments with the two *P. trichocarpa* paralogs and the single ortholog from *S. bicolor*. As expected, based on the sequence analysis, one *P. trichocarpa* paralog has retained the ancestral localization to the plastid, while the duplicated copy has lost the transit peptide and localizes to the cytosol (Figure 5A-C). HOZ1 from *S. bicolor* localizes to the chloroplast, as observed for PtHOZ1A and previously observed for AtHOZ1 (8), but was also found in the nucleus, as has been observed for the canonical heme oxygenase in yeast and animals (25). Three subcellular localization patterns of *S. bicolor* HOZ1 (Figure 5C) could suggest possible retrograde translocation (26).

Expression of the AtHOZ1 ortholog from *S. bicolor* and the paralogs from *P. trichocarpa* rescued the H_2O_2-sensitive phenotype, while the ortholog from the green alga *C. reinhardtii* failed to rescue, either because of poor expression or functional divergence. However, we did observe *in vitro* that the recombinantly produced *S. bicolor* and *P. trichocarpa* HOZ1 proteins could bind and degrade heme, while the algal protein had relatively poor heme-binding and -degradation activities (Figure 5D-F). *In vitro*, the purified land plant HOZ1 proteins displayed a characteristic Soret peak at 405 nm in the presence of heme, while the characteristic shift from 385 nm to 405 nm in the UV-VIS spectra was far less obvious with the *C. reinhardtii* ortholog (Figure 5D). To investigate whether the proteins are able to degrade heme, ascorbate was added as an electron donor together with purified protein and heme, and spectral changes were collected over time. As previously observed for AtHOZ1, there was noticeable reduction of the Soret peak for the land plant proteins, indicative of their heme degradation ability *in vitro*, but this reduction was less obvious with the *C. reinhardtii* protein (Figure 5F).

Although we could not find evidence of heme binding for the algal protein, the function of related bacterial proteins in Fe- and heme-utilization suggests that these eukaryotic proteins could have similar biological functions. We therefore tested whether the algal gene is required for growth in suboptimal iron availability and when growth is dependent on photosynthesis (photoautotrophic growth). We did not observe a growth defect when acetate (a reduced carbon source; TAP medium) is provided (Figure 5G), but a growth defect was apparent during photoautotrophic growth (in the absence of acetate; TP medium) (Figure 5H and I), which is consistent with a biological role of CrHOZ1 in chloroplast physiology. Under the conditions tested

in this study, the amount of iron in the medium did not appear to have a major impact on the *C. reinhardtii* mutant growth compared to WT (Figure 5H).

At*HOZ1* and At*HOZ2A*, but not At*HOZ2B*, are important for development

We generated mutant lines of At*HOZ1* using CRISPR-Cas9 and acquired T-DNA lines of At*HOZ2A* and At*HOZ2B*. The resulting At*HOZ1* mutants *hoz1*-1 and *hoz1*-3 contain 5 bp or 38 bp deletions, respectively, resulting in premature stop codons in both mutants (Figure S3). Compared to the parent, the At*HOZ1* and At*HOZ2A* mutants, but not the At*HOZ2B* mutant, showed developmental delays when grown in soil (Figures 6 and S7). Given the experimentally determined location of AtHOZ1 (8) and predicted location of AtHOZ2A in the plastid, these proteins appear to play important roles in chloroplast biology. Presumably, the function of At*HOZ1* and At*HOZ2A* are independent to some extent, and their roles cannot be performed by plastid-localized members of the canonical heme oxygenase family. The homozygous *hoz2a hoz2b* mutant is similar to the *hoz2a* mutant. The triple *hoz1*-1 *hoz2a hoz2b* mutant had a similar developmental delay to the *hoz1*-1 and *hoz2a* mutants (Figures 6 and S7).

Structure of the cytosolic poplar paralog

PtHOZ1B, the recently evolved cytosolic version of HOZ1, was recombinantly expressed in *Escherichia coli*, purified using Ni-NTA affinity chromatography, and crystallized. The crystal structure of PtHOZ1B was determined to 1.8 Å resolution by the single wavelength anomalous dispersion (SAD) method. The crystal belonged to the tetragonal system, with space group $P4_22_12$ and contains one molecule per asymmetric unit (Table S1). As with the plastid-localized AtHOZ1 (8), PtHOZ1B forms a dimer through a crystallographic two-fold symmetry via the split-barrel domains, with the DRI regions oriented away from the dimerization interface (Figure 7A and 7B). Unlike AtHOZ1, we were unable to acquire a crystal structure bound to heme. Additionally, the PtHOZ1B dimer is slightly more compact than AtHOZ1 (Figure S4A), and four metal-binding sites are observed in the dimer, which have not been found in AtHOZ1 (Figure 7B). Metal site 1A and 1B are at the dimer interface with residues from the split-barrel domains, while metal site 2A and 2B are in the DRI regions (Figure 7A). X-ray fluorescence energy scan (data not shown) of purified protein and crystals showed the presence of nickel ions (likely from the Ni-resin used for purification). Metal sites 1A and 1B have a distorted octahedral geometry formed by coordination with His61 and His90 of one subunit and His119 and Gln163 of another subunit (Figure S4B). The fifth coordination is from Glu279 of the intermolecularly interacting (symmetry related) molecule of subunit 1, and this coordinating side chain could be due to an effect of crystal packing (Figure S4B). A water molecule completes the coordination sphere (Figure S4B). Metal sites 2A and 2B consist of His201, His206 and four water molecules located at the surface of the DRI regions (Figure S4C).

Although we were unable to acquire a crystal structure of heme bound to PtHOZ1B, these metal-binding sites may be heme-binding sites based on structural homology with heme-bound crystal structures of HugZ, ChuZ and AtHOZ1. As observed previously (8), the heme bound at the dimer interface of HugZ and the heme bound at the dimer interface of AtHOZ1 align well (Figure 7C and 7D). A conserved histidine residue, His135 in AtHOZ1 and His136 in HugZ, interact with the propionic side chains (Figure 7D). This histidine residue is also conserved in PtHOZ1B (His90), but its side chain imidazole moiety is rotated, coordinating the metal ion found at the dimer interface (Figure 7D). To better understand whether these amino acid side chains may function in heme degradation by PtHOZ1B, we generated point mutations in metal binding site 1 and in the predicted heme-binding pocket and tested their impact on the ability of the PtHOZ1B to rescue the H_2O_2 sensitivity of a *hmx1*Δ yeast strain. Unlike WT Pt*HOZ1B*, mutants with mutated residues that line the predicted heme-binding pocket or are in metal site 1 could not rescue the H_2O_2-sensitivity of the *hmx1*Δ mutant (Figure S5).

Although not observed in AtHOZ1 and HugZ, the crystal structure of ChuZ from *Campylobacter jejuni* contains an extra heme bound to the DRI region (21). This heme is referred to as “extra”, because it is an additional heme not previously observed in the HugZ structure from *H. pylori*. The extra heme of ChuZ

aligns well with the metal site 2 of PtHOZ1B, with the two heme-binding histidine residues of ChuZ aligning with the two metal-binding residues of PtHOZ1B (Figure 7E-G). Although there is no electron density after the CG atom of the Lys263 side chain (Figure S4C), this conserved lysine could provide another ligand for binding the extra heme.

Discussion

Leveraging a combination of sequence similarity clustering, phylogenetics, domain analysis, and identification of conserved gene neighbors, this study has provided a more complete understanding of the functional diversification among members of the split-barrel superfamily, broadly, and among plant and algae, specifically. Functional divergence has resulted in a wide-array of possible heme-utilization proteins in prokaryotes and distinct families of heme-degrading proteins and heme sensors in plants and algae, combined with additional subfamilies (many unique to land plants and/or algae) of unknown function. Using similarity to the SUPERFAMILY hidden Markov model SSF50475 (27) (Figure S1), which aligns to both the split-barrel domain and DRI region, we identified multiple separate protein subfamilies in analyzed land plant genomes (Figure 1). Three of these contain uncharacterized proteins (named "unknown-1" (e.g., AT3G49140, AT3G59300, and AT5G24060), "unknown-2" (e.g., AT3G04020), and "HOZ2" (e.g., AT1G51560 and AT3G21140)), while five subfamilies contain previously characterized proteins: PNPO (e.g., At5g49970/PDX3 (13, 28)), PNPO2 (e.g., AT2G46580 (29)), CREG (*e.g.*, CREG from animals (14)), GBP (e.g., AT3G21200 (15, 16, 30–32)), and HOZ (named HOZ1; *e.g.*, AT3G03890 (8)) (Figure 1). Chlorophyte algal genomes encode members of these previously characterized subfamilies plus FLVA/FLVB (e.g., Cre12.g531900 and Cre16.g691800 (33)) and five subfamilies of uncharacterized proteins (unknown-2, HOZ2, unknown-3 (Cre07.g327079), unknown-4 (Cre03.g155350), and unknown-5 (Cre06.g297600)) (Figure 1). The plant and green algal PNPO subfamily is distinguished by the presence of an N-terminal YjeF-like domain (defined by IPR004443), which encodes an epimerase involved in repair of NAD (13) (Figure 2A). The YjeF-like domain is not present in PNPO2 homologs (Figure 2A). Like PNPO2, CREG is also a single-domain protein (Figure 2A). CREG has not been characterized in plants and is related to the Cellular Repressor of E1A-stimulated Genes (*i.e.*, CREG), a secreted glycoprotein in animals, which interacts with the cation-independent mannose 6-phosphate/insulin-like growth factor II receptor to inhibit cellular proliferation (34). The other eight *A. thaliana* proteins all contain the heme-binding domain, DRI (22) (previously named DUF2470; defined by IPR037119) that is found in the heme-degradation proteins HugZ, ChuZ, and MSMEG_6519 (Figure 2A-E). Three of these DRI-containing proteins (AT3G49140, AT3G59300, and AT5G24060) match the SSF50475 model for belonging to the split-barrel superfamily (Figure S1) but do not match the split-barrel domain defined by IPR012349 (Figure 2A). However, the split-barrel domain is evident in the computationally predicted structures with large unstructured loops inserted in the domain (Figure S2). The other *A. thaliana* homologs, including the previously characterized heme-sensor GBP (15, 32) and heme-degradation protein HOZ (8), all contain the split-barrel domain followed by DRI (Figure 2A), which is evident from the sequence and structure analysis (Figure 2A and 2C).

The closely related GBP, HOZ1, and HOZ2 subfamilies likely existed in the last common ancestor of Viridiplantae because of the presence of conserved Chlorophyte and Streptophyte orthologs in each subfamily (Figure 2F). They may have also already existed in the last common ancestor of Archeaplastida, since we identified close homologs of HOZ1 and GBP in the glaucophyte alga *Cyanophora paradoxa* and red algae (Figure 2F). HOZ2, however, appears to be specific to Viridiplantae and was not identified in the available genomes of either red algae or protist algae that have chloroplasts derived from secondary endosymbiotic events (Figure 2F). Whether because of a horizontal gene transfer event or because a HOZ-like protein was present in a more ancient eukaryotic ancestor, we also identified HOZ-like proteins in some fungal genomes, but IPR037119/DRI was consistently not detected in these fungal proteins (Figure 2F). IPR037119/DRI is often not detected among orthologous eukaryotic split-barrel proteins, and this absence could be a result of sequence divergence (and therefore have poor similarity scores to IPR037119/DRI),

since some of these proteins tend to be as large or larger than the *A. thaliana* two-domain proteins (Figure 2F).

Our analysis also reveals a large number of uncharacterized HOZ-like proteins in prokaryotes. The previously characterized HugZ-like and HutZ-like proteins represent only a small part of the larger heme-binding split-barrel family (Figure 2) that includes uncharacterized proteins from *Pseudomonas aeruginosa* (PA4388) and *Cupriavidus metallidurans* (Rmet_5353). Based on sequence similarity, domain architecture, and conserved gene neighbors putatively involved in heme/iron assimilation (Figure 3), the prokaryotic HOZ homologs are likely also involved in heme utilization like the HugZ proteins. Intriguingly, structural diversification has resulted in three distinct heme-utilization subtypes within the heme-binding split-barrel family: small single-domain HutZ-like (split-barrel domain only), two domain HugZ-like (DRI followed by split-barrel domain), and two domain HOZ-like (split-barrel domain followed by DRI). Based on our phylogenetic analysis of the split-barrel domain, the plant/algal HOZ family and bacterial HOZ-like proteins, such as the heme-degradation protein MSMEG_6519 from *Mycobacterium smegmatis* (12), represent an ancestral protein conformation: the split-barrel domain followed by DRI. The presence of HOZ-like and HutZ-like proteins at the base of the HutZ-HugZ clade (Figure 2F, dashed line box) suggest that HutZ evolved via loss of DRI, and HugZ subsequently evolved via an independent "re-fusion" event involving the split-barrel domain with DRI added to the N-terminus. Combined with experimental characterization of Dri1 from cyanobacteria, which contains DRI but lacks the split-barrel domain, DRI is likely a regulatory domain and is not required for enzymatic function. Coincidentally, in Actinomycetota, a group of canonical heme oxygenases containing DRI have also evolved (22), suggesting that the regulatory function provided by DRI may be portable across convergently evolved heme-degrading enzymes.

In addition to the three structural subtypes, published studies suggest that a small number of amino acid changes can alter substrate preference and/or function. There are single-domain HutZ-like proteins that, like HutZ, contain a C-terminal loop containing a conserved histidine residue important for heme *b* degradation, whereas in Pden_1323 the C-terminal loop is missing, and the protein degrades heme *c* (19). HupZ from *S. pyogenes* is also missing the histidine-containing loop but is hypothesized to be a heme chaperone (20). For the two-domain HOZ-like proteins, there are homologs that degrade heme *b* (*i.e.*, HOZ1 (8)) and paralogs that have lost that ability and function as heme sensors (*i.e.*, GBP). Further diversification can and has occurred with respect to subcellular location, as we experimentally verified for two closely related HOZ1 paralogs in the tree *P. trichocarpa*. Gene duplication followed by a change in localization of the encoded protein is a common adaptive mechanism (35–37). The duplication leading to the evolution of the cytosolic HOZ1 likely occurred in the *Saliceae* ancestor, because of the presence of orthologs in *Salix* and *Populus* species but absence in other *Salicaceae* members, suggesting a relatively recent (~48 million years ago (38)) neofunctionalization event. The cytosolic paralog, PtHOZ1B, has retained *in vitro* heme-degradation activity, but the *in vivo* function and the reason for maintaining this cytosolic protein in these trees has yet to be determined.

Like AtHOZ1 and HugZ, PtHOZ1B forms a homodimer where the interaction interface involves the split-barrel domain. There is very little difference between the previously determined crystal structure of the AtHOZ1 dimer and the PtHOZ1B structure presented here (Figure S4A). The most striking distinction between the two structures (other than the presence/absence of heme) is the metal-bound sites found in the split-barrel domain of PtHOZ1B and in the DRI region. The imidazole of His90 in PtHOZ1B is rotated roughly 100° compared to the corresponding residue (His135) in the AtHOZ1 structure where it interacts with the propionic side chains of heme. One hypothesis is that this metal site is a regulatory site, and when a metal ion is bound, His90 is prevented from binding the propionic side chains of heme. However, we found that mutation of any one of the amino acid residues in metal site 1 inhibited the ability of PtHOZ1B to rescue the yeast *hmx1*Δ strain, which does not support a role of this site in inhibiting heme degradation. Although we cannot rule out the possibility that these mutations lead to misfolding or other unintended

impacts on activity, it seems likely that metal site 1 may represent a crystallization artifact. Metal site 2, composed of two histidine residues that are almost absolutely conserved across DRI regions in homologous proteins (22), may also be a heme-binding, rather than metal ion binding, site based on comparison to the "extra" heme site of ChuZ. However, additional experimental work is needed to understand the occupancy of these sites *in plantae* and how they relate to function.

Focusing on the plastid-localized HOZ subfamilies that are unique to plants and algae, we further determined that these proteins likely play an important role in chloroplast physiology. At*HOZ1* and At*HOZ2A* can both rescue the H_2O_2 sensitivity of the yeast *hmx1*Δ mutant, while mutation in *A. thaliana* leads to significant developmental delays. These results are in contrast to the paralog At*HOZ2B* that cannot rescue the H_2O_2 sensitivity of the yeast *hmx1*Δ mutant, and the T-DNA line did not have any observable growth defects under the conditions tested. Why loss of At*HOZ1* and At*HOZ2A* impacts growth is presently unknown, especially as multiple plastid-localized heme oxygenases have been described in *A. thaliana* (39). *In vitro*, AtHOZ1 produces biliverdin and therefore could contribute to retrograde signaling and bilin biosynthesis like the canonical heme oxygenases (8). If this is the case *in vivo*, then there may be specialization in terms of gene/protein regulation and/or their protein-interaction networks compared to other heme oxygenases.

Although inconclusive, we do note that compared to the land plant proteins, the *C. reinhardtii* HOZ1 ortholog exhibited relatively poor *in vitro* heme-binding and -degradation activity under the conditions tested, and the gene failed to rescue the H_2O_2-senstivity of the yeast *hmx1*Δ strain, in contrast to the *HOZ1* genes from *A. thaliana*, *P. trichocarpa*, and *S. bicolor*. Therefore, further work is needed to determine whether the molecular function of the *C. reinhardtii* protein has diverged compared to the land plant orthologs or the protein performed poorly in these assays (due to mis-folding or some other issue). Although we were unable to provide conclusive evidence as to whether *C. reinhardtii* HOZ1 has heme-degradation activity, we did observe that, like in *A. thaliana*, loss of *HOZ1* impacts growth when a reduced carbon source is not available, pointing to a physiological function related to the chloroplast. These results suggest an ancient function of the HOZ1 family tied to the evolution of photosynthesis in eukaryotes. The recent identification of a suppressor mutation in *C. reinhardtii*, which compensates for loss of the canonical heme oxidase HMOX1, supports the existence of an alternative bilin biosynthetic pathway (40). Although the genetic mutation responsible for the published phenotype rescue was not determined, a mutation, such as disruption of a repressor, that leads to constitutive expression of CrHOZ1 (if CrHOZ1 functions in heme-degradation) is one hypothesis.

Out of the eight DRI-containing proteins in *A. thaliana*, only those encoded by At*HOZ1* and At*HOZ2A* were able to complement the loss of the heme oxidase gene in *S. cerevisiae* (Figure 4B). This implies that in addition to HOZ1, HOZ2 proteins can potentially degrade heme, whereas other homologs, such as GBP, may have lost enzymatic activity and could function as regulators. However, there are many other viable reasons why these genes did not complement the *hmx1*Δ mutation, and further biochemical studies would be needed to rule out *in vitro* heme-degradation activity. In the case of GBP, the inability to rescue the H_2O_2 sensitivity is expected, since this homolog is a relatively well-characterized heme sensor and regulator of tetrapyrrole biosynthesis (15, 31). Intriguingly, the evolution of non-enzymatic family members has happened in the canonical heme oxygenase family as well; proteins belonging to the HO-1 subfamily are able to bind and degrade heme, while HO-2 subfamily members are unable to do so, but rather bind to the precursor protoporphyrin IX (41). Like GBP, the function of the HO-2 subfamily has been hypothesized to be involved in the regulation of tetrapyrrole biosynthesis. Therefore, in parallel to the canonical heme oxygenase family, the presence of multiple HOZ-like protein families in plants and algae might serve to fine tune tetrapyrrole metabolism: preventing heme-induced toxicity, facilitating iron salvaging, and regulating the production of heme, chlorophyll, and intermediates. Nonetheless, whereas the expansion of the canonical heme oxygenase family is specific to land plants, HOZ1 and HOZ2 represent two distinct families that were already present in the last common ancestor shared between chlorophytes and

streptophytes. As such, there may also be some biochemical differences that underly distinct biological roles in the chloroplast, which have been conserved during evolution of Viridiplantae. However, at this point, other than complementation of the *hmx1*Δ mutation in yeast by At*HOZ2A* and a growth defect in *A. thaliana*, very little is known about this HOZ-like family.

Experimental procedures

Bioinformatic analyses

All sequence similarity networks were constructed using the EFI-EST tool (http://efi.igb.illinois.edu/efi-est/) (42) and visualized with Cytoscape v3.9.1 using the Prefuse Force Directed Layout incorporating the alignment scores. The FMN-binding split-barrel superfamily sequence similarity network (in Figure 1) was constructed by searching the UniProt (43), Phytozome (44) and PhycoCosm databases (45) for the split-barrel family defined by a match to SSF50475. An alignment score of 20 was used to build the network, and the nodes were collapsed based on a sequence identity of 90%. Node information is available in Table S2. To construct the HOZ family sequence similarity network (in Figure 3A), the split-barrel family members identified with SSF50475 from *A. thaliana*, which contain the DRI region (IPR037119), were used to search against UniProt using BLASTp. The network was constructed using an alignment score of 20; sequences were trimmed to the domain boundaries set by IPR037119; nodes were collapsed based on 95% identity. Node information is available in Table S3. Domain cartoons were drawn based on domain predictions in InterPro (46). TargetP was used to predict transit peptides (47). Gene neighborhoods were collected and visualized using the EFI-GNT webtool (42); predicted protein names were based on searches of NCBI Conserved Domain Database. The following structural models were used for visualizing 3D domains and were downloaded from the AlphaFold protein structure database (AF; https://alphafold.ebi.ac.uk/) or from the Protein Data Bank (PDB; https://www.rcsb.org/) (48): GBP, PDB: 5YJL (49); HOZ1, PDB: 6M0A; PNPO1, AF: Q9LTX3; PNPO2, AF: Q9ZPY1; HOZ2A, AF: Q8VXY3; HOZ2B, AF: Q8L637; AT3G49140, AF: Q0WMN5; AT3G59300, AF: Q949V7; AT5G24060, AF: F4KFP8; CREG, AF: Q8RY62; HutZ, PDB: 3TGV; HugZ, PDB: 3GAS; MSMEG_6519, PDB: 5BNC (12); Dri1, PDB: 8GDW (22). When required, unstructured N-terminal sequence was removed for better visualization. For the tree containing proteins similar to the HOZ family (in Figure 2F), subsequences representing IPR012349 were collected and aligned with MAFFT (50) and then used to construct a phylogenetic tree using FastTREE (51) with default parameters on CIPRES Science Gateway (52); AtPNPO1/PDX3 was used as an outgroup. Annotation information for each leaf (taxonomy, domain and domain order, and protein length) was downloaded from UniProt. For the plant HOZ tree (in Figure 4A), the phylogenetic tree was constructed using IQ-TREE webserver (53) with an alignment from COBALT (54) that was manually edited to remove sequences flanking the split-barrel and DRI regions. Trees were annotated and visualized with iTOL (55). For both trees, branches with a score less than 50% for bootstrap support were deleted. Sequence information, multiple sequence alignments and Newick trees can be found in Tables S4 and S5.

Localization

The subcellular localizations of proteins of interest were tested in poplar leaf mesophyll protoplasts as previously described (56). The cDNAs of Potri.013G057700, Potri.019G035200, and Sobic.002G349100 were synthesized and cloned into the pTwist ENTR Kozak plasmid using the clonal gene service of Twist Bioscience and then subcloned into the transient expression vector pUC-pGWB505 via LR reaction (Invitrogen) for C-terminal YFP fusion (56). Three transient expression plasmids were used to express organelle markers: PM-mCherry is the plasmid membrane marker constructed by cloning the expression cassette of CD3-1007 (57) plasmid into pUC19, plastid-mCherry is the plastid marker constructed by cloning the expression cassette of CD3-1000 (57) plasmid into pUC19, the nuclear marker nucleus-mCherry is the mCherry-VirD2NLS plasmid (58). The pUC-pGWB505 construct (8 µg) was co-transfected with 2 µg of corresponding organelle marker plasmid into 100 µl of protoplasts (~2 x 10^4 cells) to determine

their subcellular localizations. After 16 h incubation under weak light at room temperature, protoplasts were collected and resuspended in cold W5 solution (2 mM MES pH 5.7, 154 mM NaCl, 125 mM $CaCl_2$, and 5 mM KCl) to subject to microscopy. Images were collected using a Leica TCS SP5 confocal microscope, equipped with 514 and 543 nm laser lines for excitation of YFP and mCherry, respectively. The emission bandwidth for YFP and mCherry was 500-530 nm and 580-620 nm, respectively. Images were processed using LAS X software (Leica).

Protein expression and purification

The *E. coli*-codon-optimized synthetic CDS (Twist Bioscience) of Potri.013G057700, Potri.019G035200, and Sobic.002G349100 were cloned into the *Nde*I and *Xho*I sites of the pET29b vector carrying C-terminal TEV cleavage site followed by 6 × His-Tag sequence. The Cre02.g098250 gene construct was designed and assembled by the Joint Genome Institute (JGI) and cloned into the pET11e vector (22). The coding sequences of the N-terminal transit peptides were not included in the DNA synthesis. The plasmids encoding the plant HOZs were transformed into competent *E. coli* BL21(DE3) cells. Multiple colonies from the transformed plate were picked and incubated in LB medium containing 100 µg/ml kanamycin overnight at 37 °C. The cells were inoculated into 500 ml of auto-induction ZYM-50521 medium and grown at 37 °C to an OD600 of about 0.6. The cells were then allowed to auto-induce overnight at 20 °C (59). The cells were harvested by centrifugation at 5,000 rpm for 20 min, lysed at ice-cold temperature using bacterial protein extraction agent (B-PER, Thermo Fisher Scientific) in the presence of 300mM NaCl, 30mM Tris pH 8.0, lysozyme and benzonase. The soluble and insoluble fractions were separated by centrifugation at 18,000 rpm for 20 min. The resulting cell-free supernatant was allowed to bind for 20 min at 20 °C with Ni–NTA agarose (Thermo scientific) resin that had earlier been equilibrated with buffer A (40 mM Tris, 400 mM NaCl, 5% glycerol, 10 mM imidazole, pH 8.0). This mixture was then poured into a column and the resin was washed with a 50 ml buffer A. The protein was eluted using a step gradient with increasing concentration of imidazole (50, 100 and 250 mM). Fractions of the eluate were analyzed on 4–10% SDS–PAGE gel. The histidine tag was cleaved by TEV protease followed by reverse nickel-affinity chromatography. Further purification was achieved with a size-exclusion column (Superdex increase-200) that had previously been equilibrated with buffer that consisted of 40 mM HEPES, 2 mM TCEP, 3% glycerol, 150 mM NaCl, pH 7.0. Purified proteins were concentrated to ~12 mg/mL and stored at -80 °C. The selenomethionine of PtHOZ1B was expressed in a BASM SeMet medium and purified in the same way as native.

Heme-binding and degradation assays

Heme-binding assays were conducted following the protocol from Leung *et al.* (60). Briefly, a concentration of 5 µM of purified proteins from *P. trichocarpa* (Potri.013G057700, Potri.019G035200), *S. bicolor* (Sobic.002G349100) and *C. reinhardtii* (Cre02.g098250) were used for titration in 850 µL of buffer (50 mM HEPES, 200 mM NaCl, pH 7.5). Fresh hemin stock solutions were prepared by dissolving hemin with a few drops of 0.1 M NaOH, and 1 mL ultrapure water. Following filtration through a 0.22 µm filter, the hemin concentration was quantified at 385 nm ($\varepsilon_{385\ nm}$ = 58.4 mM^{-1} cm^{-1}). Spectra were recorded in quartz cuvettes (pathlength 1 mm) using a NanoDrop One^c Spectrophotometer (Thermo Scientific). Heme titration was performed by incrementally adding hemin (from 0 µM to 14.8 µM) to the protein, incubating for 3 minutes, and collecting the spectra. The formation of a characteristic Soret peak (~405 nm) indicates heme binding. Heme degradation was assayed by monitoring the UV-Vis absorption spectra kinetics of the protein (5 µM) in the presence of hemin (5 µM) and ascorbate (10 mM) as electron donor in 50 mM HEPES, 200 mM NaCl, pH 7.5. Kinetics were started by addition of ascorbate (t0) and monitored every 3 min for 170 min.

Crystallization, data collection and structure determination of PtHOZ1B

The SeMet PtHOZ1B crystals were grown at 20 °C by the sitting drop vapor diffusion method, using a 1:1 ratio of protein: reservoir solution containing 20% PEG monomethyl ether 2000, 10 mM nickel chloride and 100 mM sodium acetate pH 5.0. Before data collection, crystals were transferred into mother liquor

containing 20% glycerol and then flash cooled in liquid nitrogen. The X-ray diffraction data were collected at the FMX (17-ID-2) beamline of NSLS-II, Brookhaven National Laboratory (BNL), Upton, NY, United States. Crystals of PtHOZ1B were diffracted to 1.8 Å resolution and the data were processed with HKL20002 (61). The Matthews coefficient (VM) was calculated as 2.2 $Å^3$ Da^{-1}, which corresponds to one molecule per asymmetric unit with an estimated solvent content of 45%. The structure of PtHOZ1B was determined by SAD technique with SAD phasing pipeline of PHASER in the CCP4 program suite (62, 63). The structure was refined by several rounds of iterative model building with COOT5 and refinement in CCP4 module, REFMAC (64, 65). A summary of the data-collection and refinement statistics are shown in Table S1. The final model contains one PtHOZ1B molecule, two nickel ions, and 97 waters. The N-terminal 20 amino acids were not included in the model due to lack of electron density. Also, the regions Ala144 to Phe149 and Val168 to Gly174 were disordered and exhibited poor electron density. The model was evaluated and deposited in the PDB (PDB id: 9BL1). ChimeraX-1.8 was used to visualize and align structures (66).

Plant Materials

A. thaliana Columbia-0 (Col-0) seeds were surface sterilized with 20% bleach and dH_2O. T-DNA insertion lines SALK_025420C for HOZ2A (AT1G51560) and SALK_001793 for HOZ2B (AT3G21140) were obtained from the ABRC stock center (Figure S3). Primers GCCAGAAGAGCAGCAAGTGAG and TGCATGGTGCTCTCTTCATCA were used to verify T-DNA insertion in *HOZ2A*. Primers GAAGGATGGGCTGAAAACGAC and TTACTCCCGAGCTTTTCCTTCAC were used to verify T-DNA insertion in *HOZ2B*. The seeds were stratified in darkness at 4°C for three days. The greenhouse was maintained at 22°C with a light density of ~100 µmol $m^{-2}s^{-1}$ with 16 hr light / 8 hr dark cycle.

CRISPR-Cas9 mutagenesis of *HOZ1*

Two guide RNAs (ACATCTAGGAACTAATCTAG and TGAGAGGAGAAAATCGACGG) mapping to the first exon of At*HOZ1* were designed and checked for off-targets using the website CRISPR-P http://crispr.hzau.edu.cn/CRISPR2/ (67). CRISPR-Cas9 plasmids were generated according to the protocol and materials outlined in the Golden Gate Molecular Cloning Kit (68). The sgRNAs were amplified using pICH86966::AtU6p::sgRNA_PDS (AddGene) as a template. Level 1 guides were assembled with pICH47751-NOSpro:NPTII and pICH47742-35Spro:Cas9 into binary level 2 vector pAGM4723. The final binary vectors were transformed into *A. thaliana* Col-0 by *Agrobacterium tumefaciens* (GV3101), using floral dipping methods outlined by (69). T1 and T2 plants were selected by red fluorescence, *i.e.*, seeds with red seed coat color would be used for T1, and Cas9-free T2 plants would be seeds with no red seed coat (70). To increase the efficiency of mutagenesis, 10-day-old T1 plants were treated with cycles of 30 hours at 37°C, and 42 hours at 22°C (71). To identify the homozygous gene-edited lines, a 0.5 mm leave disc from each plant was used as a template for PCR using KAPA 3G Plant PCR Kit (Roche) with primers flanking the edited region (HOZCR forward primer: CTGCATCCCTCTCCTCCAACT, HOZCR reverse primer: CACTGTTGCCACTCCCGTAAG). PCR products were analyzed by sanger sequencing by both sequences and chromatograms (Figure S3). Identical procedures were carried out for two generations of plants. T3 plants with gene sequences that match wild-type *HOZ1* were used as the WT control, and independent deletions that caused a frame-shift or stop-gain were analyzed. The triple At*HOZ2A* At*HOZ2B* At*HOZ1* mutant was generated as outlined above, except that the *hoz2a hoz2b* mutant (generated by crossing SALK_025420C and SALK_001793) was used to transform the CRISPR-Cas9 construct. The sequenced mutation of the *hoz1* allele is identical to *hoz1-7*.

Yeast complementation assay

Yeast *hmx1* and plant HOZ coding sequences were PCR amplified and cloned into a yeast expression vector for the yeast complementation assay (Table S6). The yeast gene *hmx1* was amplified from genomic DNA of *S. cerevisiae* BY4742. All *A. thaliana* genes were PCR amplified from whole plant cDNA. *S. bicolor*, *P. trichocarpa* and *C. reinhardtii* genes were PCR amplified from the synthetic genes synthesized from Twist Bioscience, also used for heterologous protein purification. Site-directed mutagenesis on PtHOZ1B

was achieved using the megaprimers method (primers in Table S6). The PCR-amplified genes were cloned into *Xma*I/*Eco*RI digested P413-GPD plasmid using Gibson assembly. The p413-GPD plasmids containing the different genes and empty p413-GPD were transformed by lithium acetate method into *hmx1*Δ strain (from the Yeast matalpha collection; Horizon Discovery), and empty p413-GPD also transformed into WT strain BY4742. Complementation assay was performed by growing yeast cells in SD-His until saturation. Five microliters of ten-fold serial dilutions (10^0, 10^{-1}, 10^{-2}) were spotted on agar solidified YPD with or without 3 mM H_2O_2. Plates were incubated at 30 °C and imaged at 3 days post inoculation.

Chlamydomonas growth

Three independent *C. reinhardtii* cell lines mutated for Cr*HOZ1* (Cre02.g098250), LMJ.RY0402.068342, LMJ.RY0402.099542, LMJ.RY0402.105224, and the WT strain LMJ.RY0402 were obtained from the Chlamydomonas Library Project (CLiP) (72). Chlamydomonas cells were pre-cultured in liquid TAP media at 24 °C and shaken at 250 rpm in an Innova® 44/44R shaker (New Brunswick) for 5 days with a light intensity of 100 µmol m^{-2} s^{-1} photons. Cells were washed twice with TP media, and inoculated in TP or TAP media containing different concentrations of Fe (0.1 µM, 1µM, 20 µM and 75 µM) as indicated, at an OD_{750nm} of 0.05 and grown for 6 days in 24-well plates in triplicates (n=3).

Data availability

All data supporting this study are reported within this manuscript and/or supplemental files. Raw data are available from the corresponding author upon reasonable request. Atomic coordinates and structure factors for the PtHOZ1 structure were deposited into the Protein Data Bank as 9BL1. Requests for resources and reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

Supporting information

This article contains supporting information.

Funding and additional information

This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, as part of the Quantitative Plant Science Initiative at Brookhaven National Laboratory. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The work (Award DOI https://doi.org/10.46936/10.25585/60000750) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. The work at Cold Spring Harbor Laboratory was supported by U.S. Department of Agriculture, Agricultural Research Service awards 8062-21000-051-000D.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Kumar, S., and Bandyopadhyay, U. (2005) Free heme toxicity and its detoxification systems in human. *Toxicol Lett*. 10.1016/j.toxlet.2005.03.004
2. Chiabrando, D., Vinchi, F., Fiorito, V., Mercurio, S., and Tolosano, E. (2014) Heme in pathophysiology: A matter of scavenging, metabolism and trafficking across cell membranes. *Front Pharmacol*. 10.3389/fphar.2014.00061
3. Hirotsu, S., Chu, G. C., Unno, M., Lee, D. S., Yoshida, T., Park, S. Y., Shiro, Y., and Ikeda-Saito, M. (2004) The Crystal Structures of the Ferric and Ferrous Forms of the Heme

Complex of HmuO, a Heme Oxygenase of Corynebacterium diphtheriae. *Journal of Biological Chemistry*. **279**, 11937–11947

4. Schuller, D. J., Wilks, A., Ortiz De Montellano, P. R., and Poulos, T. L. (1999) Crystal structure of human heme oxygenase-1. *Nat Struct Biol*. 10.1038/12319
5. Haley, K. P., Janson, E. M., Heilbronner, S., Foster, T. J., and Skaar, E. P. (2011) Staphylococcus lugdunensis IsdG liberates iron from host heme. *J Bacteriol*. **193**, 4749–4757
6. Wu, R., Skaar, E. P., Zhang, R., Joachimiak, G., Gornicki, P., Schneewind, O., and Joachimiak, A. (2005) Staphylococcus aureus IsdG and IsdI, heine-degrading enzymes with structural similarity to monooxygenases. *Journal of Biological Chemistry*. 10.1074/jbc.M409526200
7. Lojek, L. J., Farrand, A. J., Wisecaver, J. H., Blaby-Haas, C. E., Michel, B. W., Merchant, S. S., Rokas, A., and Skaar, E. P. (2017) *Chlamydomonas reinhardtii* LFO1 Is an IsdG Family Heme Oxygenase. *mSphere*. **2**, e00176-17
8. Wang, J., Guo, Q., Li, X., Wang, X., and Liu, L. (2020) The Arabidopsis locus AT3G03890 encodes a dimeric β-barrel protein implicated in heme degradation. *Biochemical Journal*. 10.1042/BCJ20200712
9. Liu, X., Gong, J., Wei, T., Wang, Z., Du, Q., Zhu, D., Huang, Y., Xu, S., and Gu, L. (2012) Crystal structure of HutZ, a heme storage protein from Vibrio cholerae: A structural mismatch observed in the region of high sequence conservation. *BMC Struct Biol*. 10.1186/1472-6807-12-23
10. Hu, Y., Jiang, F., Guo, Y., Shen, X., Zhang, Y., Zhang, R., Guo, G., Mao, X., Zou, Q., and Wang, D. C. (2011) Crystal structure of HugZ, a novel heme oxygenase from Helicobacter pylori. *Journal of Biological Chemistry*. 10.1074/jbc.M110.172007
11. Kitamura, M., Kojima, S., Ogasawara, K., Nakaya, T., Sagara, T., Niki, K., Miura -, K. I., Akutsu, H., and Kumagai, I. (1994) Novel FMN-binding protein from *Desulfovibrio vulgaris* (Miyazaki F). Cloning and expression of its gene in *Escherichia coli*. *Journal of Biological Chemistry*. 10.1016/s0021-9258(17)37499-9
12. Ahmed, F. H., Carr, P. D., Lee, B. M., Afriat-Jurnou, L., Mohamed, A. E., Hong, N. S., Flanagan, J., Taylor, M. C., Greening, C., and Jackson, C. J. (2015) Sequence-Structure-Function Classification of a Catalytically Diverse Oxidoreductase Superfamily in Mycobacteria. *J Mol Biol*. **427**, 3554–3571
13. Colinas, M., Shaw, H. V., Loubéry, S., Kaufmann, M., Moulin, M., and Fitzpatrick, T. B. (2014) A pathway for repair of NAD(P)H in plants. *Journal of Biological Chemistry*. 10.1074/jbc.M114.556092
14. Ghobrial, G., Araujo, L., Jinwala, F., Li, S., and Lee, L. Y. (2018) The structure and biological function of CREG. *Front Cell Dev Biol*. 10.3389/fcell.2018.00136
15. Czarnecki, O., Hedtke, B., Melzer, M., Rothbart, M., Richter, A., Schröter, Y., Pfannschmidt, T., and Grimm, B. (2011) An Arabidopsis GluTR binding protein mediates spatial separation of 5-Aminolevulinic acid synthesis in chloroplasts. *Plant Cell*. 10.1105/tpc.111.086421
16. Sinha, N., Eirich, J., Finkemeier, I., and Grimm, B. (2022) Glutamate 1-semialdehyde aminotransferase is connected to GluTR by GluTR-binding protein and contributes to the rate-limiting step of 5-aminolevulinic acid synthesis. *Plant Cell*. **34**, 4623–4640

17. Guo, Y., Guo, G., Mao, X., Zhang, W., Xiao, J., Tong, W., Liu, T., Xiao, B., Liu, X., Feng, Y., and Zou, Q. (2008) Functional identification of HugZ, a heme oxygenase from Helicobacter pylori. *BMC Microbiol*. **8**, 226
18. Uchida, T., Sekine, Y., Matsui, T., Ikeda-Saito, M., and Ishimori, K. (2012) A heme degradation enzyme, HutZ, from Vibrio cholerae. *Chemical Communications*. 10.1039/c2cc31147j
19. Li, S., Isiorho, E. A., Owens, V. L., Donnan, P. H., Odili, C. L., and Mansoorabadi, S. O. (2021) A noncanonical heme oxygenase specific for the degradation of c-type heme. *Journal of Biological Chemistry*. 10.1016/j.jbc.2021.100666
20. Lyles, K. V., Thomas, L. S., Ouellette, C., Cook, L. C. C., and Eichenbaum, Z. (2022) HupZ, a Unique Heme-Binding Protein, Enhances Group A Streptococcus Fitness During Mucosal Colonization. *Front Cell Infect Microbiol*. 10.3389/fcimb.2022.867963
21. Zhang, R., Zhang, J., Guo, G., Mao, X., Tong, W., Zhang, Y., Wang, D. C., Hu, Y., and Zou, Q. (2011) Crystal structure of Campylobacter jejuni ChuZ: a split-barrel family heme oxygenase with a novel heme-binding mode. *Biochem.Biophys.Res.Commun.* **415**, 82–87
22. Grosjean, N., Yee, E. F., Kumaran, D., Chopra, K., Abernathy, M., Biswas, S., Byrnes, J., Kreitler, D. F., Cheng, J.-F., Ghosh, A., Almo, S. C., Iwai, M., Niyogi, K. K., Pakrasi, H. B., Sarangi, R., van Dam, H., Yang, L., Blaby, I. K., and Blaby-Haas, C. E. (2024) A hemoprotein with a zinc-mirror heme site ties heme availability to carbon metabolism in cyanobacteria. *Nat Commun*. **15**, 3167
23. Gerdes, S., El Yacoubi, B., Bailly, M., Blaby, I. K., Blaby-Haas, C. E., Jeanguenin, L., Lara-Núñez, A., Pribat, A., Waller, J. C., Wilke, A., Overbeek, R., Hanson, A. D., and de Crécy-Lagard, V. (2011) Synergistic use of plant-prokaryote comparative genomics for functional annotations. *BMC Genomics*. 10.1186/1471-2164-12-S1-S2
24. Collinson, E. J., Wimmer-Kleikamp, S., Gerega, S. K., Yang, Y. H., Parish, C. R., Dawes, I. W., and Stocker, R. (2011) The yeast homolog of heme oxygenase-1 affords cellular antioxidant protection via the transcriptional regulation of known antioxidant genes. *Journal of Biological Chemistry*. 10.1074/jbc.M110.187062
25. Lin, Q., Weis, S., Yang, G., Weng, Y. H., Helston, R., Rish, K., Smith, A., Bordner, J., Polte, T., Gaunitz, F., and Dennery, P. A. (2007) Heme oxygenase-1 protein localizes to the nucleus and activates transcription factors important in oxidative stress. *Journal of Biological Chemistry*. 10.1074/jbc.M607954200
26. Krause, K., Oetke, S., and Krupinska, K. (2012) Dual targeting and retrograde translocation: Regulators of plant nuclear gene expression can be sequestered by plastids. *Int J Mol Sci*. 10.3390/ijms130911085
27. Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol*. 10.1006/jmbi.2001.5080
28. Sang, Y., Barbosa, J. M., Wu, H., Locy, R. D., and Singh, N. K. (2007) Identification of a pyridoxine (pyridoxamine) 5′-phosphate oxidase from Arabidopsis thaliana. *FEBS Lett*. 10.1016/j.febslet.2006.12.028
29. Sang, Y., Goertzen, L. R., Tzou, Y. M., Locy, R. D., and Singh, N. K. (2011) Identification of a second pyridoxine (pyridoxamine) 5′-phosphate oxidase in Arabidopsis thaliana. *Acta Physiol Plant*. 10.1007/s11738-010-0579-6

30. Jung, H. S., Okegawa, Y., Shih, P. M., Kellogg, E., Abdel-Ghany, S. E., Pilon, M., Sjolander, K., Shikanai, T., and Niyogi, K. K. (2010) Arabidopsis thaliana PGR7 encodes a conserved chloroplast protein that is necessary for efficient photosynthetic electron transport. *PLoS One*. 10.1371/journal.pone.0011688
31. Apitz, J., Nishimura, K., Schmied, J., Wolf, A., Hedtke, B., van Wijk, K. J., and Grimm, B. (2016) Posttranslational Control of ALA Synthesis Includes GluTR Degradation by Clp Protease and Stabilization by GluTR-Binding Protein. *Plant Physiol*. **170**, 2040–2051
32. Richter, A. S., Banse, C., and Grimm, B. (2019) The GluTR-binding protein is the heme-binding factor for feedback control of glutamyl-tRNA reductase. *Elife*. 10.7554/eLife.46300
33. Chaux, F., Burlacot, A., Mekhalfi, M., Auroy, P., Blangy, S., Richaud, P., and Peltier, G. (2017) Flavodiiron proteins promote fast and transient O2 photoreduction in chlamydomonas. *Plant Physiol*. 10.1104/pp.17.00421
34. Di Bacco, A., and Gill, G. (2003) The secreted glycoprotein CREG inhibits cell growth dependent on the mannose-6-phosphate/insulin-like growth factor II receptor. *Oncogene*. 10.1038/sj.onc.1206670
35. Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016) Evolution of gene duplication in plants. *Plant Physiol*. 10.1104/pp.16.00523
36. Liu, S. L., Pan, A. Q., and Adams, K. L. (2014) Protein subcellular relocalization of duplicated genes in Arabidopsis. *Genome Biol Evol*. 10.1093/gbe/evu191
37. Marques, A. C., Vinckenbosch, N., Brawand, D., and Kaessmann, H. (2008) Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. *Genome Biol*. 10.1186/gb-2008-9-3-r54
38. Bell, C. D., Soltis, D. E., and Soltis, P. S. (2010) The age and diversification of the angiosperms re-revisited. *Am J Bot*. 10.3732/ajb.0900346
39. Emborg, T. J., Walker, J. M., Noh, B., and Vierstra, R. D. (2006) Multiple heme oxygenase family members contribute to the biosynthesis of the phytochrome chromophore in arabidopsis. *Plant Physiol*. **140**, 856–868
40. Zhang, W., Deng, R., Shi, W., Li, Z., Larkin, R. M., Fan, Q., and Duanmu, D. (2022) Heme oxygenase-independent bilin biosynthesis revealed by a hmox1 suppressor screening in Chlamydomonas reinhardtii. *Front Microbiol*. 10.3389/fmicb.2022.956554
41. Gisk, B., Yasui, Y., Kohchi, T., and Frankenberg-Dinkel, N. (2010) Characterization of the haem oxygenase protein family in Arabidopsis thaliana reveals a diversity of functions. *Biochemical Journal*. 10.1042/BJ20090775
42. Zallot, R., Oberg, N., and Gerlt, J. A. (2019) The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry*. 10.1021/acs.biochem.9b00735
43. Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. Da, Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Castro, L. G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C. S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M. R., Saidi, R.,

741 Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V.,
742 Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L.,
743 Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M. C.,
744 Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K. C.,
745 Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M. L.,
746 Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N.,
747 Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le
748 Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T. B.,
749 Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C.,
750 Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C. H.,
751 Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Garavelli, J. S., Huang, H., Laiho, K.,
752 McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q., Wang, Y., Yeh, L. S., and
753 Zhang, J. (2021) UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res*.
754 10.1093/nar/gkaa1100
755 44. Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T.,
756 Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D. S. (2012) Phytozome: A comparative
757 platform for green plant genomics. *Nucleic Acids Res*. 10.1093/nar/gkr944
758 45. Grigoriev, I. V, Hayes, R. D., Calhoun, S., Kamel, B., Wang, A., Ahrendt, S., Dusheyko, S.,
759 Nikitin, R., Mondo, S. J., Salamov, A., and others (2021) PhycoCosm, a comparative algal
760 genomics resource. *Nucleic Acids Res*. **49**, D1004--D1011
761 46. Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G.,
762 Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork,
763 P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D. A.,
764 Necci, M., Orengo, C. A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J. A., Sillitoe, I., Thanki,
765 N., Thomas, P. D., Tosatto, S. C. E., Wu, C. H., Bateman, A., and Finn, R. D. (2021) The
766 InterPro protein families and domains database: 20 years on. *Nucleic Acids Res*.
767 10.1093/nar/gkaa977
768 47. Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G.,
769 Elofsson, A., and Nielsen, H. (2019) Detecting sequence signals in targeting peptides
770 using deep learning. *Life Sci Alliance*. 10.26508/lsa.201900429
771 48. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I.
772 N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res*.
773 10.1093/nar/28.1.235
774 49. Zhao, A., and Han, F. (2018) Crystal structure of Arabidopsis thaliana glutamyl-tRNAGlu
775 reductase in complex with NADPH and glutamyl-tRNAGlu reductase binding protein.
776 *Photosynth Res*. 10.1007/s11120-018-0518-8
777 50. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software
778 version 7: improvements in performance and usability. *Mol Biol Evol*. **30**, 772–780
779 51. Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2 - Approximately maximum-
780 likelihood trees for large alignments. *PLoS One*. **5**, e9490
781 52. Miller, M., Pfeiffer, W., and Schwartz, T. (2010) Creating the CIPRES science gateway for
782 inference of large phylogenetic trees, p 11572--8. in *Proceedings of the Gateway*
783 *Computing Environments Workshop (GCE). Institute of Electrical and Electronics*
784 *Engineers, Piscataway, NJ. http://dx. doi. org/10.1109/GCE*

785 53. Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., and Minh, B. Q. (2016) W-IQ-TREE: a fast
786 online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. **44**, W232–
787 W235
788 54. Papadopoulos, J. S., and Agarwala, R. (2007) COBALT: Constraint-based alignment tool
789 for multiple protein sequences. *Bioinformatics*. **23**, 1073–1079
790 55. Letunic, I., and Bork, P. (2021) Interactive tree of life (iTOL) v5: An online tool for
791 phylogenetic tree display and annotation. *Nucleic Acids Res*. 10.1093/nar/gkab301
792 56. Xie, M., Muchero, W., Bryan, A. C., Yee, K., Guo, H. B., Zhang, J., Tschaplinski, T. J., Singan,
793 V. R., Lindquist, E., Payyavula, R. S., Barros-Rios, J., Dixon, R., Engle, N., Sykes, R. W.,
794 Davis, M., Jawdy, S. S., Gunter, L. E., Thompson, O., Difazio, S. P., Evans, L. M., Winkeler,
795 K., Collins, C., Schmutz, J., Guo, H., Kalluri, U., Rodriguez, M., Feng, K., Chen, J. G., and
796 Tuskan, G. A. (2018) A 5-enolpyruvylshikimate 3-phosphate synthase functions as a
797 transcriptional repressor in populus. *Plant Cell*. **30**, 1645–1660
798 57. Nelson, B. K., Cai, X., and Nebenführ, A. (2007) A multicolored set of in vivo organelle
799 markers for co-localization studies in Arabidopsis and other plants. *Plant Journal*.
800 10.1111/j.1365-313X.2007.03212.x
801 58. Lee, L. Y., Fang, M. J., Kuang, L. Y., and Gelvin, S. B. (2008) Vectors for multi-color
802 bimolecular fluorescence complementation to investigate protein-protein interactions in
803 living plant cells. *Plant Methods*. 10.1186/1746-4811-4-24
804 59. Studier, F. W. (2005) Protein production by auto-induction in high density shaking
805 cultures. *Protein Expr Purif*. 10.1016/j.pep.2005.01.016
806 60. Leung, G. C. H., Fung, S. S. P., Dovey, N. R. B., Raven, E. L., and Hudson, A. J. (2019)
807 Precise determination of heme binding affinity in proteins. *Anal Biochem*.
808 10.1016/j.ab.2019.02.021
809 61. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in
810 oscillation mode. *Methods Enzymol*. 10.1016/S0076-6879(97)76066-X
811 62. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan,
812 R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N.,
813 Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011)
814 Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol*
815 *Crystallogr*. 10.1107/S0907444910045749
816 63. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read,
817 R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr*.
818 10.1107/S0021889807021206
819 64. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta*
820 *Crystallogr D Biol Crystallogr*. 10.1107/S0907444904019158
821 65. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular
822 structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*.
823 10.1107/S0907444996012255
824 66. Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., and
825 Ferrin, T. E. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and
826 analysis. *Protein Science*. **27**, 14–25

67. Lei, Y., Lu, L., Liu, H. Y., Li, S., Xing, F., and Chen, L. L. (2014) CRISPR-P: A web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol Plant*. 10.1093/mp/ssu044
68. Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D. G., Patron, N. J., and Marillonnet, S. (2014) A Golden Gate modular cloning toolbox for plants. *ACS Synth Biol*. 10.1021/sb4001504
69. Clough, S. J., and Bent, A. F. (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal*. **16**, 735–743
70. Gao, X., Chen, J., Dai, X., Zhang, D., and Zhao, Y. (2016) An effective strategy for reliably isolating heritable and Cas9-free arabidopsis mutants generated by CRISPR/Cas9-mediated genome editing. *Plant Physiol*. **171**, 1794–1800
71. LeBlanc, C., Zhang, F., Mendez, J., Lozano, Y., Chatpar, K., Irish, V. F., and Jacob, Y. (2018) Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. *Plant Journal*. 10.1111/tpj.13782
72. Li, X., Patena, W., Fauser, F., Jinkerson, R. E., Saroussi, S., Meyer, M. T., Ivanova, N., Robertson, J. M., Yue, R., Zhang, R., Vilarrasa-Blasi, J., Wittkopp, T. M., Ramundo, S., Blum, S. R., Goh, A., Laudon, M., Srikumar, T., Lefebvre, P. A., Grossman, A. R., and Jonikas, M. C. (2019) A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. *Nat Genet*. 10.1038/s41588-019-0370-6

Figure legends

Figure 1 The split-barrel superfamily contains functionally distinct plant proteins. A sequence similarity network is shown of proteins from *Arabidopsis thaliana* (At) and *Chlamydomonas reinhardtii* (Cr) that belong to the split-barrel superfamily (defined as inclusion in SSF50475) and their homologs. Nodes are colored by taxonomy according to the color key. Clusters are named based on members with characterized or predicted function. Clusters with *A. thaliana* and/or *C. reinhardtii* homologs of unknown function are labeled unknown "1" through unknown "5". Pink background shading is used to highlight separate clusters, and dotted red lines are used to delineate subclusters. A grey box is used to indicate those clusters containing proteins that are a fusion between the split-barrel domain and DRI. The location of nodes representing *A. thaliana* and *C. reinhardtii* proteins are indicated with a white arrow and label, as is the CREG protein from *Homo sapiens*. Abbreviations and locus IDs for protein names: DRI: domain related to iron, PNPO: pyridoxine 5'-phosphate oxidase, CREG: cellular repressor of E1A-stimulated genes, FLV: flavodiiron, HOZ: homolog of HugZ, GBP: GluTR-binding protein, AtPPOX2: AT2G46580, AtGBP: AT3G21200, AtHOZ1: AT3G03890, AtHOZ2A: AT1G51560, AtHOZ2B: AT3G21140, AtPPOX: At5g49970, AtPPOX2: At2g46580 , CrGTPBP1: Cre03.g156600, CrFLVB: Cre16.g691800, CrFLVA: Cre12.g531900, CrPPO1: Cre08.g378200, CrPNPO: Cre02.g095100, CrHOZ2: Cre12.g520200, CrHOZ1: Cre02.g098250.

Figure 2 The split-barrel domain is found in a large family of prokaryotic and eukaryotic proteins linked to heme utilization. *A*, scaled cartoons of split-barrel superfamily members from *A. thaliana* depicting presence and location of transit peptides, the YjeF domain, the split-barrel domain (defined by IPR012349), and DRI. Locus IDs: AtPNPO1: At5g49970, AtPNPO2: AT2G46580, AtCREG: At2g04690, AtGBP: AT3G21200, AtHOZ1: AT3G03890, AtHOZ2A: AT1G51560, AtHOZ2B: AT3G21140. Numbers in parentheses refer to the numbered clusters in Fig. 1. *B*, scaled cartoons depicting representative bacterial proteins related to the eukaryotic HOZ family. *C*, experimental (PDB ID given) or computationally predicted structural models for *A. thaliana* proteins containing IPR012349. Corresponding pLDDT confidence measures for the computationally predicted structures can be found in Figure S1. *D*,

computationally predicted structural models of *A. thaliana* proteins from cluster unknown "1". Corresponding pLDDT confidence measures for the computationally predicted structures can be found in Figure S1. *E*, experimentally determined structural models of bacterial proteins that serve as representatives for the 4 common types of identified domain architectures. For panels *A-E*, the split-barrel domain is colored orange, the YjeF domain is colored blue, and DRI is colored yellow. Stretches of sequences that do not match to an available domain model are colored grey. *F*, approximate maximum likelihood phylogenetic tree of the split-barrel domain sequences from the HOZ family and related homologs. Leaves representing *A. thaliana* proteins are labeled, as are proteins shown in panel *B*. Taxonomic information (T), domain architecture (D), and protein length (L) for each leaf is given at the bottom according to the color key.

Figure 3 Prokaryotic homologs of the eukaryotic HOZ family are linked to Fe and heme utilization. *A*, protein sequence similarity network of HOZ-family proteins and homologs. Nodes are colored according to the color key. Nodes representing *Arabidopsis thaliana* (At) and *Chlamydomonas reinhardtii* (Cr) HOZ proteins are labelled, as are some uncharacterized proteins from well-studied bacteria. Pink background shading is used to highlight separate clusters, and dotted red lines are used to delineate subclusters. *B*, representative gene neighborhoods where genes encoding prokaryotic HOZ-like homologs are in close proximity to genes encoding proteins predicted to be involved in Fe assimilation based on presence of conserved domains (colored according to the color key). Genes encoding proteins that are not obviously related to Fe or heme utilization are grey. The number in parentheses is the UniProt ID for the HOZ-like protein. Protein name labels are derived from searches against NCBI's Conserved Domain Database. *C*, representative proteins from each named prokaryotic cluster showing domain architecture.

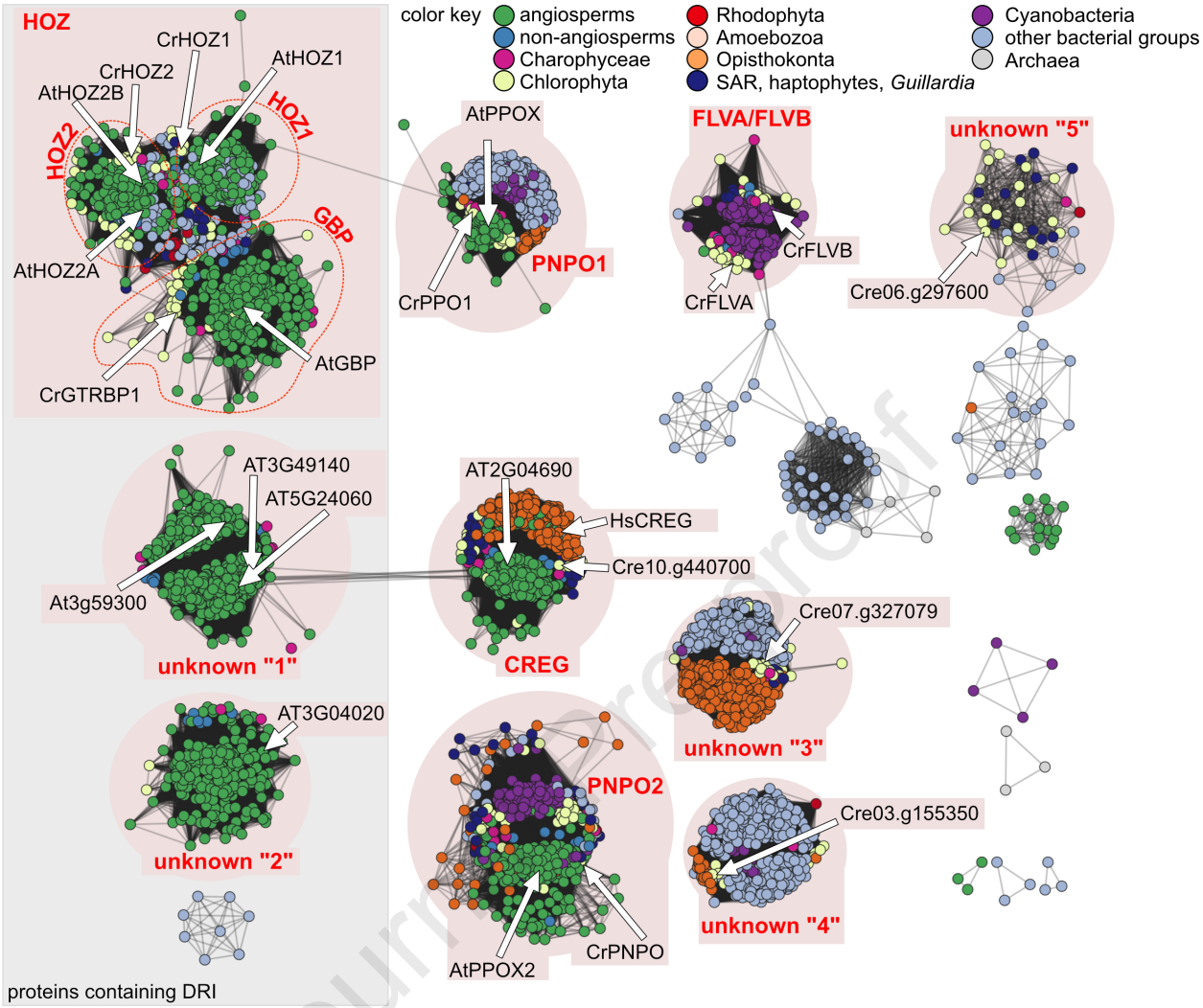
Figure 4 The HOZ family in plants may contain multiple heme-degradation proteins. *A*, maximum-likelihood phylogenetic tree of HOZ family proteins from Viridiplantae and selected bacteria. Tree leaves are colored according to the color key. Proteins from *Arabidopsis thaliana* (At), *Populus trichocarpa* (Pt), *Sorghum bicolor* (Sb), *Vitis vinifera* (Vv), *Zea mays* (Zm) and *Chlamydomonas reinhardtii* (Cr) are labeled, and the presence of a plastid transit peptide in these sequences is indicated with a green pentagon. Protein names in bold font were experimentally tested for heme-degradation in this study. Locus IDs starting with "Potri" are from *P. trichocarpa*, "VIT" are from *V. vinifera*, "Sobic" are from *S. bicolor*, and "Zm" are from *Z. mays*. *B*, yeast complementation to assess the functionality of *HOZ1* from *A. thaliana, P. trichocarpa, S. bicolor*, and *C. reinhardtii* and HOZ-like proteins from *A. thaliana*. HOZ1 and HOZ-like proteins were overexpressed in the yeast mutant *hmx1*Δ lacking the canonical heme oxidase Hxm1p and assayed for their growth on YPD media in the absence or presence of 3 mM H_2O_2. AtGBP and AtCREG are used as negative controls. Original plate images can be found in Figure S6. From left to right, cultures were plated without diluting, diluting 10-fold or diluting 100-fold. The images were captured after 3 days at 30°C. Proteins names and corresponding locus IDs: AtHOZ1: AT3G03890, PtHOZ1B: Potri.013G057700, PtHOZ1A: Potri.019G035200, SbHOZ1: Sobic.002G349100, CrHOZ1: Cre02.g098250, AtHOZ2B: AT3G21140, AtHOZ2A: AT1G51560, AtGBP: AT3G21200, AtCREG: AT2G04690, VvHOZ1: VIT_208s0007g07350, ZmHOZ1: Zm00001d021881, CrHOZ2: Cre12.g520200, AtHOZ2B, VvHOZ2: VIT_209s0054g00600, PtHOZ2A: Potri.008G006400, PtHOZ2B: Potri.010G252300, ZmHOZ2A: Zm00001d031540, ZmHOZ2B: Zm00001d015274, ZmHOZ2C: Zm00001d037710, SbHOZ2A: Sobic.007G200700, SbHOZ2B: Sobic.007G200900, ZmHOZ2B: Zm00001d015274, ZmHOZ2C: Zm00001d015274, ZmHOZ2C: Zm00001d037710, SbHOZ2C: Sobic.004G040600, CrGTRBP1: Cre03.g156600, ZmGBPA: Zm00001d049343, ZmGBPB: Zm00001d021251, SbGBP: Sobic.007G090400

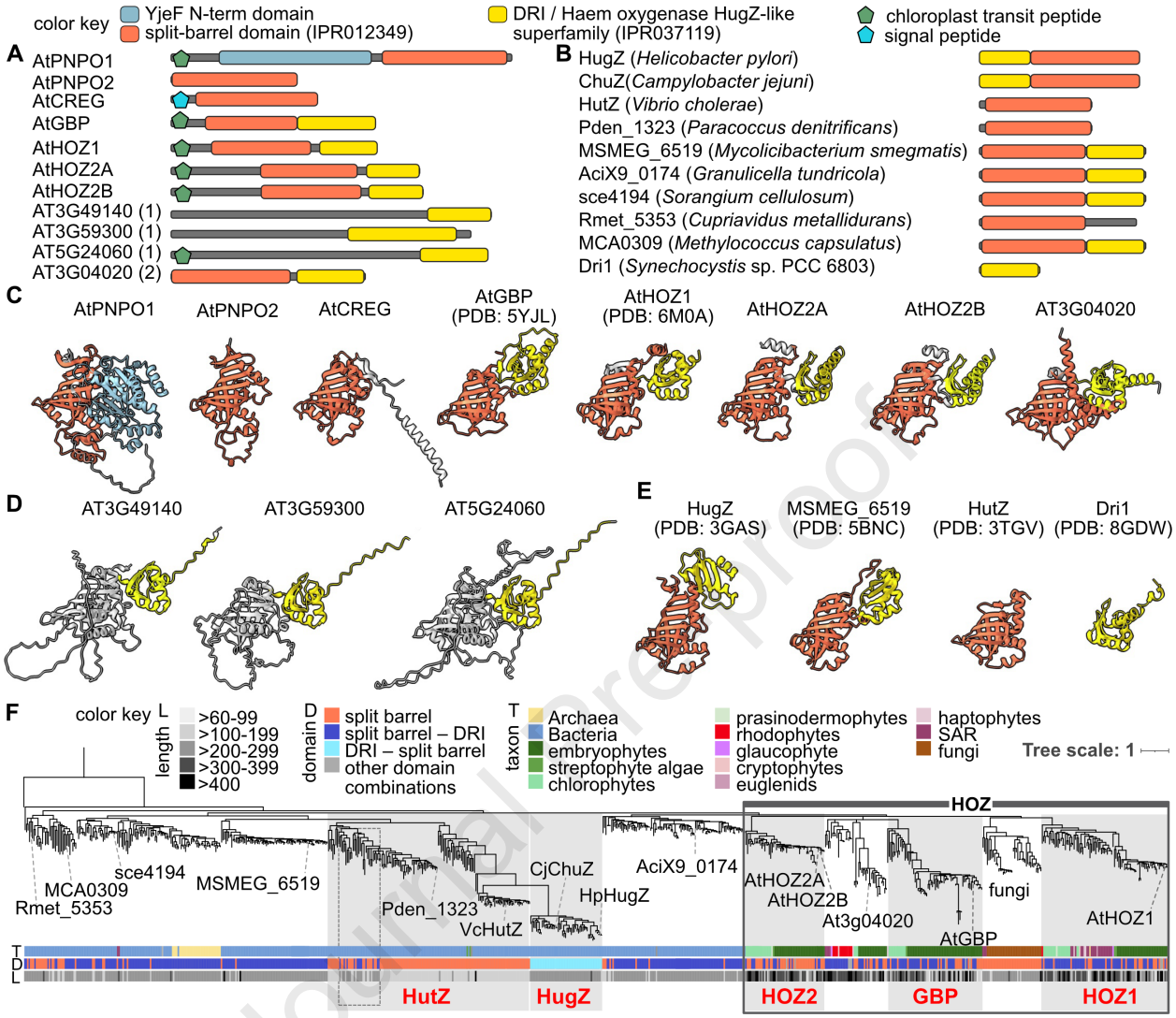
Figure 5 The *in vitro* ability to degrade heme is conserved in the cytosolic *Populus trichocarpa* paralog. *A-C*, subcellular localization of HOZ1 homologs in *P. trichocarpa* (PtHOZ1B: Potri.013G057700 and PtHOZ1A: Potri.019G035200) and *Sorghum bicolor* (SbHOZ1: Sobic.002G349100). Proteins were fused with YFP at the C-terminus and transiently expressed in *P. trichocarpa* mesophyll protoplasts (green). The mCherry-fused organelle markers are co-expressed and shown in red. PM-mCherry, plasmid membrane

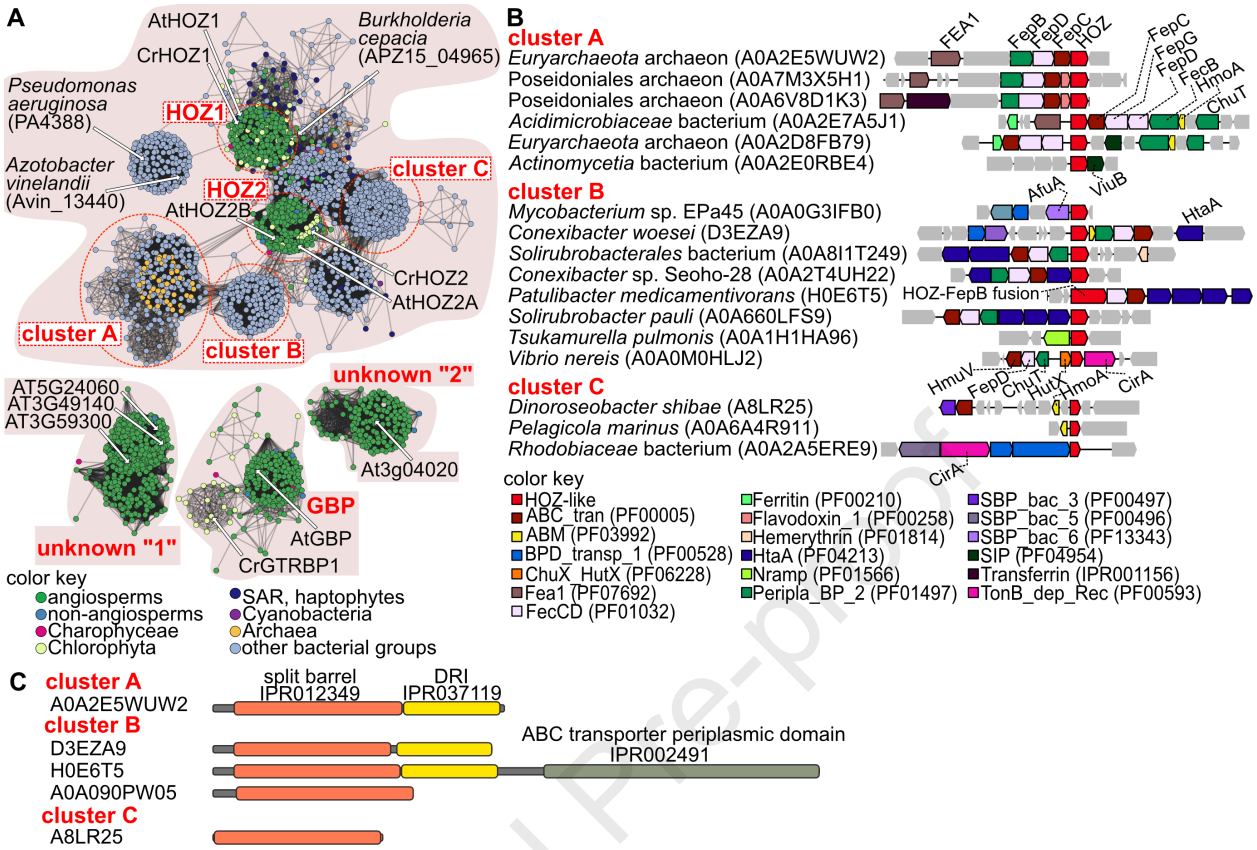
marker cloned from CD3-1007 plasmid; Plastid-mCherry, plastid marker cloned from CD3-1000 plasmid; Nucleus-mCherry, mCherry-VirD2NLS plasmid. Scale bar = 5 µm. *D*, UV-Vis absorption spectra of heme binding by *Chlamydomonas reinhardtii* (Cre02.g098250), *P. trichocarpa* and *S. bicolor* HOZ1 homologs showing a characteristic Soret peak at 405 nm indicating heme binding. Red spectra correspond to free hemin. Spectra gradient from light blue to dark blue correspond to the protein spectra with increasing heme concentration. *E* and *F*, UV-Vis absorption spectra kinetics of heme-bound HOZs in the absence (panel *E*) or presence (panel *F*) of ascorbate as the electron donor. Blue-colored spectrum represents t = 0 min, and red-colored spectrum represents t = 170 min. *G* and *H*, *C. reinhardtii* mutants and parental strain were grown in photo-heterotrophic (TAP) (panel *G*) or photo-autotrophic (TP) (panel *H*) conditions for 6 days with various concentrations of Fe, as indicated. *I*, growth curve of *C. reinhardtii* cells in Fe-replete TP media (20 µM Fe) for 6 days. For panels *G-I*, each circle represents the average of three separate biological replicates and the error bars represent the standard deviation. Cr*hoz-1*, Cr*hoz*-2, *and* Cr*hoz*-refer to LMJ.RY0402.068342, LMJ.RY0402.099542, LMJ.RY0402.105224, respectively, three independent *C. reinhardtii* cell lines with mutations in Cr*HOZ1*. The WT strain used is LMJ.RY0402.

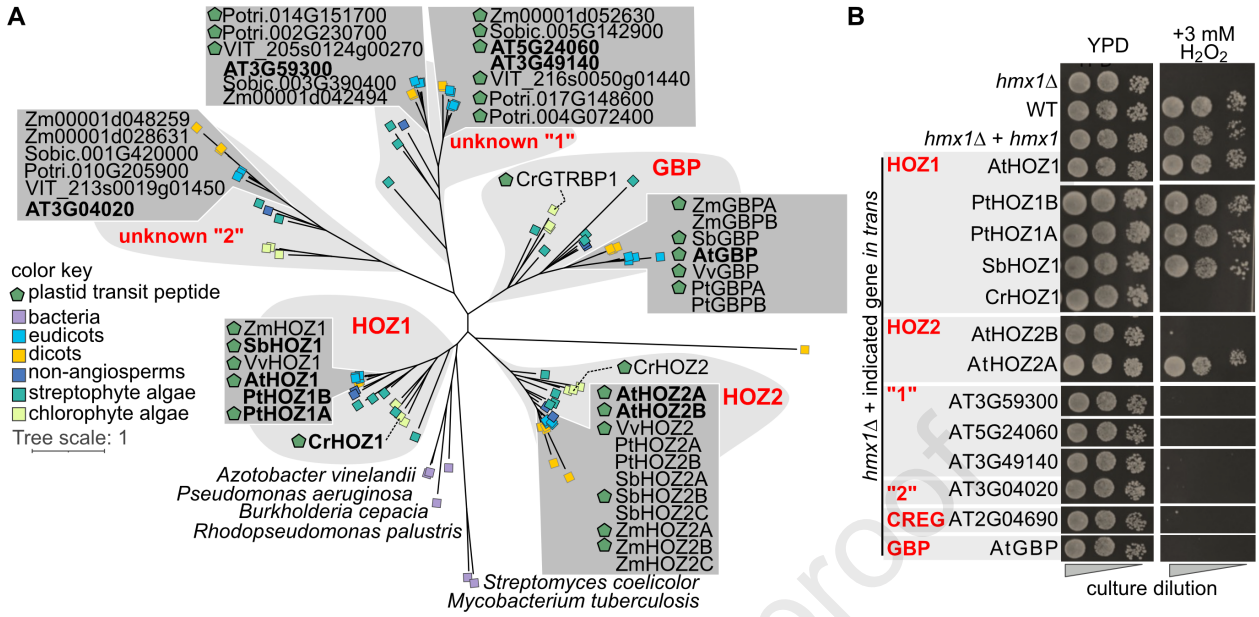
Figure 6 Mutation of At*HOZ1* and At*HOZ2A* led to a developmental delay in *Arabidopsis thaliana*. Images of soil-potted plants were taken 28 days after sowing (DAS). A scale bar corresponding to the height of plants is given on the left. WT-1 refers to Col-1 wild-type plants, WT-2 refers to plants acquired after transformation of Col-0 with the CRISPR-Cas construct and are WT at the At*HOZ1* gene, the *hoz1*-1 and *hoz1*-3 are independent mutants with different length deletions in the coding region of At*HOZ1*, *hoz2a* and *hoz2b* refer to T-DNA lines of At*HOZ2A* and At*HOZ2B*, respectively, *hoz2a hoz2b* refer to the double mutant, and *hoz1-1 hoz2a hoz2b* is a triple mutant. Pictures of plants from 31, 34, 36, and 40 days after sowing are available in (Figure S7).

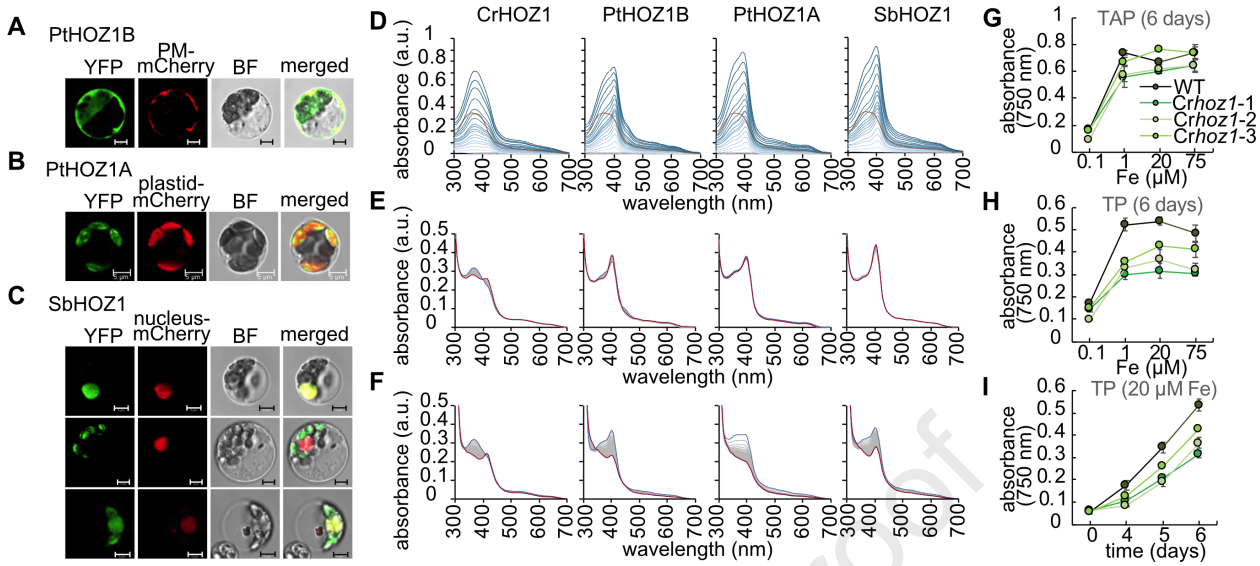
Figure 7 The structure of cytosolic PtHOZ1B. *A*, ribbon cartoon of the observed PtHOZ1B homodimer. Chain A is colored light grey, and chain B is colored based on whether the corresponding sequence matches to the split-barrel domain (orange) or DRI (yellow). Dark grey indicates the amino acid residues corresponding to the linker that remains after cleavage of the His-tag. The two metal-sites are labelled. On the right, only chain B is shown and is rotated by 90°. *B*, alignment of the PtHOZ1B and AtHOZ1 (PDB: 6M0A) structures using the matchmaker command implemented in ChimeraX-1.8. The root mean square deviation (RMSD) between 206 pruned atom pairs is 0.965 Å. Chains are colored according to the color key and the heme from AtHOZ1 is in red. *C*, alignment of the split-barrel domain dimers from PtHOZ1B, AtHOZ1 and HugZ (PDB: 3GAS) using the matchmaker command implemented in ChimeraX-1.8. For PtHOZ1B and HugZ, the root mean square deviation (RMSD) between 70 pruned atom pairs is 1.228 Å. The backbone and hemes are colored according to the key. A close-up view of the metal and heme sites, surrounded by a grey box, is given in panel D. *D*, close-up view of the metal and heme bindng sites in the split-barrel domain dimer, highlighted with a box in panel C. The close-up is rotated relative to panel C to give a better view of the sites. *E*, alignment of PtHOZ1B with the DRI region of ChuZ (PDB: 3swj) using the matchmaker command implemented in ChimeraX-1.8. RMSD between 49 pruned atom pairs is 0.891 Å. The "extra" heme found in the ChuZ structure is colored red. *F*, surface representation of PtHOZ1B with a heme modeled in the DRI region based on the alignment with ChuZ. *G*, close-up of metal site 2 that overlaps with the "extra" heme site of ChuZ. Top, surface representation of PtHOZ2A with metal-binding histidine residues of PtHOZ1B and heme-binding histidine residues of ChuZ shown as stick representations. A close-by asparagine in ChuZ that is in an equivalent position as an aspartic acid in PtHOZ1B is also shown as is the conserved lysine predicted to face toward the iron atom of the heme. The lysine in ChuZ is not modelled due to lack of electron density. Bottom, the same view, but the surface representation was removed to show alignment of the backbones. Throughout, "Pt" refers to *Populus trichocarpa*, "At" is *Arabidopsis thaliana*, and "Hp" is *Helicobacter pylori*.

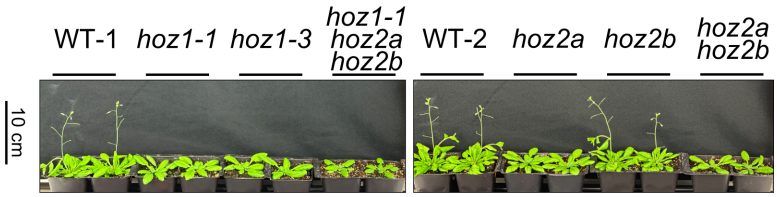


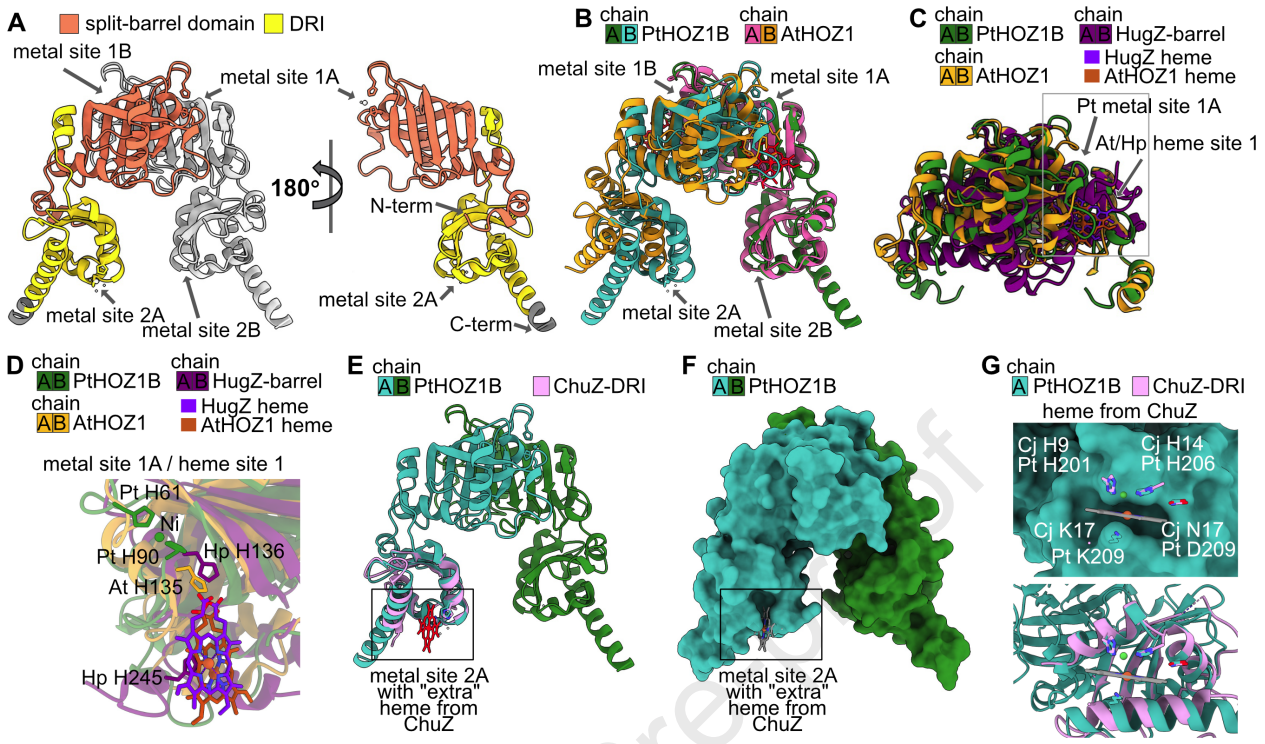












Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for *[Journal name]* and was not involved in the editorial review or the decision to publish this article.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: