Title Page

Title: Structural characterization of the Chlorophyllide *a* Oxygenase (CAO) enzyme through an *in silico* approach

Author information
Debayan Dey[1, 2], Ryouichi Tanaka[2], Hisashi Ito[2]
[1]Graduate School of Life Science, Hokkaido University, N10 W8, Sapporo 060-0810, Japan
[2]Institute of Low Temperature Science, Hokkaido University, N19 W8, Sapporo 060-0819, Japan

Corresponding author
Hisashi Ito
Institute of Low Temperature Science, Hokkaido University, N19 W8, Sapporo 060-0819, Japan
E-mail: ito98@lowtem.hokudai.ac.jp
Tel: +81-11-706-5469
Fax: +81-11-706-5463

Keywords
Chlorophyll *b* biosynthesis, chlorophyllide *a* oxygenase, *Micromonas pusilla*, computational prediction, molecular docking

Acknowledgments
We thank Prof. Ayumi Tanaka and Dr. Atsushi Takabayashi of Hokkaido University for the helpful discussions. This study was supported by funding from the Japan Society for the Promotion of Science through the KAKENHI Grant numbers 20H03017 to RT.

Abstract

Chlorophyllide *a* oxygenase (CAO) is responsible for converting chlorophyll *a* to chlorophyll *b* in a two-step oxygenation reaction. CAO belongs to the family of Rieske mononuclear iron oxygenases. Although the structure and reaction mechanism of other Rieske monooxygenases have been described, a member of plant Rieske non-heme iron-dependent monooxygenase has not been structurally characterized. The enzymes in this family usually form a trimeric structure and electrons are transferred between the non-heme iron site and the Rieske center of the adjoining subunits. CAO is supposed to form a similar structural arrangement. However, in Mamiellales such as *Micromonas* and *Ostreococcus*, CAO is encoded by two genes where non-heme iron site and Rieske cluster localize on the distinct polypeptides. It is not clear if they can form a similar structural organization to achieve the enzymatic activity. In this study, the tertiary structures of CAO from the model plant *Arabidopsis thaliana* and the Prasinophyte *Micromonas pusilla* were predicted by deep learning-based methods, followed by energy minimization and subsequent stereochemical quality assessment of the predicted models. Furthermore, the chlorophyll *a* binding cavity and the interaction of ferredoxin, which is the electron donor, on the surface of *Micromonas* CAO were predicted. The electron transfer pathway was predicted in *Micromonas* CAO and the overall structure of the CAO active site was conserved even though it forms a heterodimeric complex. The structures presented in this study will serve as a basis for understanding the reaction mechanism and regulation of the plant monooxygenase family to which CAO belongs.

1 Introduction

Light energy is captured by photosynthetic pigments in light harvesting complexes (LHC), which consist of core and peripheral antenna systems (Green and Durnford 1996). In addition to chlorophyll *a*, the peripheral antenna complex in land plants also contains chlorophyll *b*, which helps in absorbing a diverse range of light spectra for photosynthesis (Caffarri et al. 2001; Chen 2014). These antenna complexes exhibit controlled changes in size by altering the chlorophyll *a* to *b* ratio, allowing the optimal utilization of available light. For example, plants growing under low light conditions have a low chlorophyll *a* to *b* ratio and large antenna size (Bailey et al. 2001).

Chlorophyll *b* is synthesized from chlorophyll *a*, through conversion of a methyl group at the C7 position to a formyl group catalyzed by the enzyme, chlorophyllide *a* oxygenase (CAO) (Oster et al. 2000; Tanaka et al. 1998). The reaction is supposed to be sequential monooxygenation (Liu et al. 2022; Porra et al. 1994). A methyl group is oxidized to a hydroxymethyl group, the latter being subsequently oxidized by CAO and not by a dehydrogenase commonly observed in the hydroxymethyl–formyl group interconversion. The dihydroxylated intermediate is spontaneously dehydrated to produce a formyl group. CAO is the sole enzyme responsible for chlorophyll *b* synthesis from chlorophyll *a*. Almost all land plants use both chlorophyll *a* and *b*, the ratio of which is usually 3.0 – 3.5. Although Liu et al., 2022 have provided insights into the stereoselectivity and substrate range of the CAO protein, detailed structural information of the enzyme remain elusive till now (Liu et al. 2022; Oster et al. 2000).

Interestingly, unlike other chlorophyll metabolic enzymes, the structural organization of CAO varies among photosynthetic organisms (Nagata et al. 2004). Eukaryotic CAOs, except that in Mamiellales, are composed of three domains which are termed A, B, and C domains in order from the N-terminus (Nagata et al. 2004). Mamiellales, an order of green algae, includes some of the most ecologically important groups of marine photosynthetic picoeukaryotes (Leconte et al. 2020; Not et al. 2004). The conserved A domain, unique to land plants and most green algae, has a regulatory function that prevents the accumulation of the CAO protein in response to the chlorophyll *b* levels (Sakuraba et al. 2009; Yamasato et al. 2005). The B domain, which is less conserved even among land plants, probably serves as a linker between the A and C domains. The C domain, conserved in chlorophytes as well as prochlorophytes, is the catalytic domain possessing a Rieske center and a mononuclear iron-binding motif (Nagata et al. 2004). It is shown that the C domain is sufficient for chlorophyll *b* biosynthesis (Yamasato et al. 2005).

Surprisingly, in Mamiellales, which include *Micromonas* and *Ostreococcus*, the CAO sequence appears to lack the A and B domains, and its C domain is split into two polypeptides (Tanaka and Tanaka 2019). The first half of the enzyme is encoded by the *MpCAO1* gene which includes the Rieske motif, and the second half

of the enzyme is encoded by the *MpCAO2* gene including the mononuclear iron-binding motif in *Micromonas pusilla* CAO (MpCAO). It was demonstrated that simultaneous incorporation of both *MpCAO1* and *MpCAO2* into a chlorophyll *b*-less *Arabidopsis* mutant (*ch1-1*) compliments its chlorophyll *b* deficiency, indicating that coordination between the two subunits as a heterodimeric complex is required to form chlorophyll *b* (Kunugi et al. 2013). While CAO, a member of the Rieske-mononuclear iron oxygenase family, usually assumes a homotrimeric organization, the *Micromonas* CAO may be the first example of an evolutionary structural innovation for a Rieske oxygenase that forms a heterodimer (D'Ordine et al. 2009; Kunugi et al. 2016; Kunugi et al. 2013).

This study provides the first report of the detailed structural elucidation of the CAO protein, which enhances our understanding of the enzyme reaction mechanism. In this study, the tertiary, as well as quaternary structure of CAO, were predicted using a deep neural-network based method. In addition, the probable binding cavity for ligand interaction, the putative ferredoxin binding site, and residues of structural and functional importance have been elucidated.

2 Materials and methods

2.1 Sequence alignment and phylogenetic analysis

Protein sequences of CAO were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/protein/). Proteins encompassing land plants, streptophytes, prasinophytes, core chlorophytes, and cyanobacteria were considered for the multiple sequence alignment **(Table 1)**. Additionally, the amino acid sequence of another Rieske monooxygenase – dicamba (2-methoxy-3,6-dichlorobenzoic acid) O-demethylase (NCBI Accession ID: Q5S3I3.1), alternatively known as dicamba monooxygenase or DMO, was also used for comparison (D'Ordine et al. 2009). The protein sequences were aligned using Clustal Omega with the default settings (Sievers and Higgins 2018). Visualization and marking of the conserved residues in the multiple sequence alignment were implemented in Jalview v2.11.1.4 (Waterhouse et al. 2009). An unrooted maximum likelihood phylogenetic tree was determined using IQ-TREE v1.6.12 in the ultrafast mode with 1000 bootstrap replicates (Hoang et al. 2017; Trifinopoulos et al. 2016). The best-fitting amino acid substitution model – LG + G4 was applied automatically for phylogeny construction in IQ-TREE (Kalyaanamoorthy et al. 2017), and iTOL v6 was used for both visualization and figure representation (Letunic and Bork 2021).

2.2 Tertiary structure modelling and validation

In silico modelling of the two CAO subunits from *Micromonas pusilla* was performed using D-I-TASSER (Distance-guided Iterative Threading ASSEmbly Refinement) pipeline (Zheng et al. 2021), which is an extension of the I-TASSER method for highly accurate protein structure and function prediction. Furthermore, the 'Using D-I-TASSER-AF2 pipeline' option, which combines the potentials of both D-I-TASSER and AlphaFold2 programs, was selected during tertiary structure predictions of both *Micromonas* CAO proteins. In addition, tertiary structure of the single subunit CAO protein from the model plant *Arabidopsis thaliana* (AtCAO; Accession ID: AAD54323.1) was modelled using the RoseTTAFold tool (Baek et al. 2021). The CAO protein structure of *Prochlorococcus marinus* subsp. pastoris str. CCMP1986 (*Prochlorococcus marinus* MED4; Accession ID: CAE19267.1) was predicted by the SWISS-MODEL server (Waterhouse et al. 2018) using the *Arabidopsis* CAO predicted structure as the template. *Arabidopsis* and *Prochlorococcus* CAO (PmCAO) were compared, and the root mean square deviation (RMSD) was observed using PyMOL. For each predicted protein structure, the model with the best confidence, as appraised by the template modelling score (TM-score), was considered for further analyses (Xu and Zhang 2010). Further, the three-dimensional structure of ferredoxin from *Micromonas pusilla* (MpFd; Accession ID: XP_003064135.1) was determined using the SWISS-MODEL server (Waterhouse et al. 2018), following significant similarity with template sequence (PDB ID: 5AUK). Each protein model was structurally refined using the GalaxyRefine server (Heo et al. 2016). The stereochemical quality of the refined structures was assessed by Verify3D (Luthy et al. 1992), PROCHECK (Laskowski et al. 1993), and ERRAT (Colovos and Yeates 1993) in the Structural Analysis and Verification Server (SAVES) v. 5.0 server (https://servicesn.mbi.ucla.edu/SAVES/). The 3D models were also validated using the ProSA-web server (https://prosa.services.came.sbg.ac.at/prosa.php) (Wiederstein and Sippl 2007). Graphic modifications, visualization, and preparation of final illustrations were performed in

PyMOL v. 2 (Delano W 2002). The potential binding cavity on the protein structure was detected using the CavityPlus web server (http://www.pkumdl.cn/cavityplus) (Xu et al. 2018). Since CAO possesses mononuclear iron and Rieske binding domains, the metal ion binding site in the modelled structure was determined by the MIB server (Lin et al. 2016). The Rieske-bound conformation of the CAO protein as well as ferredoxin was predicted using the COACH server (Yang et al. 2013). The HDOCK server (http://hdock.phys.hust.edu.cn/) was used to predict the interaction of ferredoxin with the *Micromonas* oxygenase subunits (Yan et al. 2020). Furthermore, the interaction was cross-validated using the Local 3D Zernike descriptor-based protein Docking (LZerD) program (Christoffer et al. 2021; Venkatraman et al. 2009) and the ClusPro protein-protein docking server (Kozakov et al. 2017). ConSeq v. 1.1 was used to identify the functionally and structurally important amino acids in the primary sequence of CAO (Berezin et al. 2004).

2.3 Oligomeric structure prediction

The heterodimeric complex consisting of two subunits of MpCAO was derived using the GalaxyHeteromer server (Park et al. 2021) whereas the homo-oligomeric structure of the CAO protein from *A. thaliana* was predicted using the GalaxyHomomer server (Baek et al. 2017). Both complex forms were predicted utilizing a similarity-based approach. Furthermore, the model accuracy was improved by refinement of the predicted complexes in GalaxyRefineComplex (Heo et al. 2016). Besides, the binding affinity of the protein-protein complexes was determined using the PRODIGY web server (Xue et al. 2016).

2.4 Molecular docking analysis

The KEGG LIGAND database (https://www.genome.jp/kegg/ligand.html) was used to retrieve the structure of the substrate, chlorophyll *a*, followed by geometry optimization under the semiempirical method in HyperChem 8.0.8 molecular modelling software (Hypercube). Steepest descent followed by the Polak-Ribiere conjugate gradient algorithm was performed for energy optimization of chlorophyll *a* until convergence was reached. Open Babel was used for the interconversion of structures with different file formats (O'Boyle et al. 2011). Protein-ligand docking studies were carried out using AutoDock Vina v1.1.2 (Trott and Olson 2010) considering the energy minimized structure of the MpCAO2 and AtCAO monomeric protein. The pre-docking parameters were set using AutoDock Tools v4 with the addition of polar hydrogen atoms and Gasteiger charges to the protein molecule (Morris et al. 2009). No constraints or solvation were considered in this procedure. A grid box of 30 Å × 30 Å × 30 Å with a grid spacing of 1Å was set for docking. Interactions in the docked conformations were visualized using PyMOL.

3 Results

3.1 Multiple sequence alignment and phylogenetic analysis

Variability in the amino acid sequences of CAO across life forms, ranging from cyanobacteria to land plants, was observed from the analysis of the multiple sequence alignment **(Supplementary Figure S1)**. Additionally, the protein sequence of DMO, a Rieske-mononuclear iron oxygenase, which shares high sequence homology with CAO sequences was also considered for the comparison. The protein sequences of CAO are highly conserved except for *Micromonas* where CAO is composed of two subunits – MpCAO1 and MpCAO2 that exclusively possesses a Rieske center motif and a mononuclear iron-binding motif, respectively. However, this conservation is restricted to the catalytic domain (C domain) of CAO sequences only and not to the regulatory domain (A domain), the latter showing considerable sequence variations between vascular plants and green algae. The conserved regions in the alignment mainly constitute the Rieske binding motif, mononuclear iron binding motif, and ligand binding site residues, with conservation score above 90%.

A maximum likelihood phylogenetic tree demonstrated a distinct clading pattern of CAO proteins across different life forms – ranging from cyanobacteria to land plants (**Supplementary Figure S2**). Despite possessing significant sequence similarity in their functional domains, CAO proteins did not intermix in the phylogeny except for *Micromonas* and *Prochlorococcus*, thus maintaining a unique spatial arrangement according to taxonomic forms. The separation of subunits and use of different substrate (8-vinyl chlorophyll)

are probably the reasons for the different spatial positions of *Micromonas* and *Prochlorococcus* in the phylogeny, respectively.

3.2 Predicted tertiary structure of CAO

Understanding the spatial distribution of amino acid residues in the predicted three-dimensional structure of CAO might provide insight into their reaction mechanism. An *in silico* approach was adopted for modelling owing to the absence of any experimentally derived structure for CAO proteins. Therefore, the structure of CAO was modelled using the D-I-TASSER tool (Zheng et al. 2021), which integrates the potentials of both D-I-TASSER and AlphaFold2 programs under the D-I-TASSER-AF2 pipeline. First, we applied this protocol for AtCAO modelling, however, we obtained comparatively low confidence for AtCAO using either the D-I-TASSER-AF2 pipeline or the available model in the AlphaFold database as observed from the low TM-score (eTM-score = 0.56) and low per-residue confidence score, respectively. Therefore, the tertiary structure of AtCAO, excepting the signal peptide region, was determined with the RoseTTAFold server **(Figure 1)** with a confidence score of 0.78. The structure of the two subunits of CAO from *Micromonas pusilla* was successfully modelled using the D-I-TASSER tool. The models with best confidence as appraised by the estimated TM-score (eTM-score for MpCAO1 = 0.82 and MpCAO2 = 0.80) were selected for further analysis. Compared to the available models in the AlphaFold Protein Structure Database (Varadi et al. 2022), these tertiary structures of CAO proteins predicted using RoseTTAFold and D-I-TASSER shared least RMSD value with the monomeric structures of other Rieske oxygenases such as carbazole 1,9a dioxygenase (CARDO; PDB ID: 1WW9) (Ashikawa et al. 2006) and dicamba monooxygenase (DMO; PDB ID: 3GB4) (D'Ordine et al. 2009).
Different protein structure quality assessment programs such as PROCHECK, ERRAT and Verify 3D available online on the SAVES server, were used to evaluate the stereochemical quality of the energy minimized modelled structures of AtCAO, MpCAO1, and MpCAO2. Ramachandran plots revealed that the predicted models follow all the stereochemical properties with favorable phi (Φ) and psi (Ψ) values. Besides, ERRAT and Verify3D confirmed the high global quality of the structural models **(Table 2)**. The ProSA analysis of MpCAO1, MpCAO2, and AtCAO showed a Z-score of -6.36, -6.54, and -8.78, respectively, accommodating the predicted structures in the X-ray zone, hence confirming their reliability.
The CavityPlus tool was used to identify the potential ligand binding site on the surface of AtCAO and MpCAO2 proteins. Since MpCAO1 contains solely the Rieske binding motif, it was not considered for the ligand cavity detection analysis. The amino acid residues constituting the predicted ligand binding cavity of AtCAO and MpCAO2 have been marked in the multiple sequence alignment with asterisk **(Supplementary Figure S1)**. It is to be noted that the majority of the residues were found to be conserved among CAOs. Out of the 21 conserved residues comprising the protein cavity, 12 residues were found to be substituted in case of MpCAO1 rendering it unsuitable for substrate binding. Furthermore, ConSeq analysis depicted the level of conservation as well as residues of structural and functional importance along the sequence of CAO proteins **(Supplementary Figure S3)**.
Four conserved amino acids – C28, H30, C47, H50 (for MpCAO1) and C262, H264, C281, and H284 (for AtCAO) were found to interact with the Rieske [2Fe-2S] cluster in which one iron is coordinated by two histidines and the other one by two cysteine residues **(Figure 1a)**. Among the two CAO subunits of *M. pusilla*, MpCAO2 only contains the mononuclear non-heme iron binding motif along with the chlorophyll *a* binding site. Therefore, docking of Fe^{2+}/Fe^{3+} to the energy-minimized structure of MpCAO2 using the MIB server displayed interaction of Fe ion with four residues: N173, H179, H184, and D328. Similarly, in case of AtCAO, these conserved residues – N361, H367, H372, and D487 are responsible for interaction with the iron molecule **(Figure 1b)**.
When chlorophyll *b* is produced sufficiently, CAO is degraded to suppress chlorophyll *b* overproduction (Yamasato et al. 2005). During this process, the A domain is believed to monitor chlorophyll *b* levels through an unidentified mechanism, allowing CAO for proteolysis (Sakuraba et al. 2009). In the predicted structure, A domain is structurally separated from the catalytic C domain, which may be advantageous for monitoring chlorophyll *b* levels.

3.3 Oligomeric structure of CAO

Biochemical experiments have demonstrated that the CAO protein usually exists as a trimer in order to facilitate inter-subunit electron transfer from a Rieske cluster of one subunit to a mononuclear iron of an adjacent subunit for carrying out its catalytic reaction (Kunugi et al. 2013). Indeed, recombinant AtCAO is found to exist in oligomeric forms, such as single, double, or triple trimers, under non-denaturing conditions (Kunugi et al. 2013). Recombinant *Prochlorothrix hollandica* CAO was shown to have a trimeric architecture (Liu et al. 2022). In this study, the trimeric organization of AtCAO was predicted using the GalaxyHomomer tool **(Figure 1c)**. Further refinement of the predicted trimer with GalaxyRefineComplex showed Ramachandran outliers to be less than one percent, confirming the accuracy of the structure. Within the monomer of AtCAO, the Rieske cluster and the mononuclear iron are present at a distance of ~43.7 Å apart. Furthermore, in the 3-fold symmetric arrangement of AtCAO trimer, the distance between the Rieske cofactor of one subunit and the non-heme iron center of the neighboring subunit is ~12.0 Å. These positioning and distances are adequate for electron transfer and catalysis and are also in agreement with those observed in other oxygenases (Furusawa et al. 2004; Gakhar et al. 2005; Martins et al. 2005; Nojiri et al. 2005).

Unlike the presence of homotrimer CAO forms in most organisms, heterodimeric association between the two subunits of *Micromonas* CAO (MpCAO1 and MpCAO2) is indispensable for the synthesis of chlorophyll *b*. The heterodimeric complex of MpCAO1 and MpCAO2 proteins were derived using the GalaxyHeteromer program **(Figure 1c)**. Here also, less than one percent residues of the refined heterodimer complex was found to be in the outlier region of the Ramachandran plot. The distance between the Rieske cluster of MpCAO1 subunit and the non-heme iron center of the adjacent MpCAO2 subunit is ~12.2 Å. Interestingly, the distance between the amino acids responsible for electron transfer from the Rieske cluster to mononuclear iron of adjoining subunit and C7 position of chlorophyll *a* was found to be within ~4 Å, thus ensuring an efficient electron transfer pathway for the formation of chlorophyll *b*.

The estimated binding affinities (ΔG) for the AtCAO homotrimer and MpCAO heterodimer, as evaluated from the PRODIGY analysis, are -16.7 and -13.8 kcal mol^{-1}, respectively. The highly negative value of the binding free energies (ΔG) is indicative of the stable interaction between the protein-protein complexes for both AtCAO and MpCAO. Additionally, the strength of protein-protein interactions can also be measured by the dissociation constant (K_d), where the low values for AtCAO ($K_d = 5.9 \times 10^{-13}$ M) and MpCAO ($K_d = 7.0 \times 10^{-11}$ M) suggested formation of stable oligomeric complexes.

3.4 Ferredoxin and its interaction with CAO

In Rieske oxygenases, the iron-sulfur cluster serves as the initial acceptor of electrons from the partner ferredoxin or reductase, whereas the mononuclear iron is the downstream receptor of electrons that reductively activates molecular oxygen for interaction with the substrate (Costas et al. 2004; Kovaleva and Lipscomb 2008). Therefore, we derived the tertiary structure of ferredoxin from *Micromonas pusilla* using homology modelling in the SWISS-MODEL server. A GMQE (Global Model Quality Estimate) and QMEANDisCo global score of 0.86 and 0.82 ± 0.09, respectively, was obtained for the predicted structure. Furthermore, absence of any residues in the disallowed region of the Ramachandran plot hints at the accuracy of the protein model. Additionally, ProSA analysis provided a Z-score of -7.15, placing the model within the category of experimental NMR structure of equivalent residue length **(Table 2)**. Furthermore, COACH analysis revealed the Rieske [2Fe-2S] cluster of ferredoxin to be coordinated by four cysteine residues – C39, C44, C47, and C77 **(Figure 2a)**.

The interaction of ferredoxin with the CAO subunits of *Micromonas* was predicted by protein-protein docking using the HDOCK server **(Figure 2b)**. In addition, the LZerD and ClusPro docking algorithms revealed identical conformations as the HDOCK program. Ferredoxin was found to be docked at the interface of MpCAO1-MpCAO2 heterodimer and on the opposite face of the catalytic site in the MpCAO2 subunit. Furthermore, the amino acid residues located at the component interface were found to form an ideal hydrophobic atmosphere for interaction, a characteristic feature also observed for other Rieske oxygenases (Ashikawa et al. 2006). The distance between the Rieske clusters of ferredoxin and MpCAO1 subunit was found to be approximately 13 Å, which is in concordance with other Rieske oxygenases like carbazole 1,9a-dioxygenase and within the 14 Å threshold defining the limit of electron tunneling in a protein medium (Page et al. 1999). Interestingly, the predicted ferredoxin binding site in the MpCAO2 subunit includes the region

V345–R361, which is structurally conserved with other Rieske oxygenases such as V351–V363 in CARDO and I312–V325 in DMO (Ashikawa et al. 2006; D'Ordine et al. 2009). The interacting residues between MpFd and MpCAO heterodimer are listed in **Table 3**. Additionally, the binding affinity (ΔG) and dissociation constant (K_d) between ferredoxin and MpCAO heterodimer is -10.8 kcal mol^{-1} and 1.1×10^{-8} M, respectively, as revealed from PRODIGY analysis. The value of both the parameters is indicative of a feasible interaction between ferredoxin and the MpCAO subunits.

3.5 Protein-ligand docking

For molecular docking analysis, free chlorophyll *a*, that has been subjected to energy minimization procedure, was considered as the substrate for CAO proteins. While chlorophyllide was shown to be the substrate of CAO by the biochemical analysis (Liu et al. 2022; Oster et al. 2000), it was suggested that CAO not only catalyzes free chlorophyll *a*, but also chlorophyll *a* bound to apoproteins as almost all chlorophyll molecules remain attached to proteins *in vivo* (Jia et al. 2016). The molecular docking was performed with the refined monomers of AtCAO and MpCAO2 using a specific grid box centered around the predicted ligand binding cavity by CavityPlus. The docked chlorophyll molecule was observed to fit properly into the substrate pocket for both the proteins, with the methyl group at the C7 position of chlorophyll *a* located at close proximity to the mononuclear iron unit **(Figure 3)**. The phytol group was found to be outside the substrate pocket, suggesting that chlorophyllide also binds to the active site in the same manner as chlorophyll. It is worth mentioning that the docking results corroborate with the observations of Liu et al, 2022 (Liu et al. 2022), that the active site of CAO protein accommodates only the chlorin scaffold of the substrate with a central metal ion and not the hydrophobic tail. The lowest energy conformations for each protein-ligand docked pair were considered. The close arrangement of all the moieties implies an effective electron transfer pathway. The high degree of conservation for majority of residues comprising the ligand-binding cavity demonstrates a common chlorophyll *a* binding mode for all CAOs from different organisms.

4 Discussion

4.1 Structural comparison with Rieske mononuclear iron oxygenases

Rieske mononuclear iron oxygenase catalyzes a variety of complex oxidation reactions (Bugg and Ramaswamy 2008; Perry et al. 2018). They bind substrates with chemical structures of varying complexity, and their versatility have been extensively described in the recent work (Brimberry et al. 2022). Their structures show the α_3 or $\alpha_3\beta_3$ forms, which in turn depend on their subunit sub-domain organization (Ferraro et al. 2005). Plant CAO is a member of this group and is also supposed to possess the three-fold symmetric form. In Mamiellales, unicellular small algae such as *Micromonas* and *Ostreococcus*, CAO is formed with two polypeptides. In this study, MpCAO1 and MpCAO2 derived from *Micromonas* were examined through a computational approach. The 3-fold symmetric arrangement of AtCAO trimer was also analyzed computationally, where the distance between the Rieske cofactor of one subunit and the non-heme iron center of the neighboring subunit is ~12.0 Å. The orientations and distances of the residues involved are feasible for electron transfer and catalysis and are also in agreement with other Rieske oxygenases.

CAO sequences are highly conserved in the green lineage and cyanobacteria possessing chlorophyll *b*. However, there are exceptions, such as Mamiellales and *Prochlorococcus* CAO, that have been presented in this study. *Prochlorococcus* belongs to a marine picophytoplankton clade (Partensky and Garczarek 2010). The homology of their sequences to those of the green lineage is very low (Satoh and Tanaka 2006). Intriguingly, the 8-ethyl group of chlorophyll is replaced with vinyl group (3,8-divinyl chlorophyll) in *Prochlorococcus* chlorophyll because 8-vinyl chlorophyll absorbs light more efficiently in the open ocean light conditions where *Prochlorococcus* dominates. Although the overall structure of PmCAO is not very similar to that of general CAO, the active site and core structure are conserved (Supplementary Figure S2). This suggests that the CAO structure is modified depending on the light environment. Therefore, CAO structures are probably not changed through evolutionary processes, but they adapt to their growing environments individually. It is worth mentioning that the structural features predicted in this study for *Micromonas* and *Arabidopsis* CAO proteins correlates with the mutagenesis experiment done in barley *fch2* encoding chlorophyllide *a* oxygenase (Mueller et al. 2012) and almost all mutated residues therein are

conserved among the CAO protein family **(Table 4, Supplementary Figure S4)**. *Arabidopsis* chlorophyll *b*-deficient mutant *ch1-2* possesses the mutation V274E within the Rieske-binding site that is closely located to the G280D mutation site in barley (Espineda et al. 1999). Therefore, it is certain that these mutations will also have the same effect on the activity of *Arabidopsis* and *Micromonas* CAOs, since their three-dimensional structures (particularly active site arrangements) are identical.

During the catalytic process, an electron is transferred from ferredoxin to the Rieske cluster initially and further downstream to the mononuclear iron where it oxidizes the substrate, probably through the presence of an intermediate water molecule, which needs further experimental investigation (Oster et al. 2000). Though the Rieske cluster and the mononuclear iron are bound to separate polypeptides – MpCAO1 and MpCAO2, respectively, their arrangements and spatial proximities remain well conserved with other Rieske oxygenases. The distance between the Rieske cluster and the non-heme iron is found to be ~12.2 Å in the predicted MpCAO structures, allowing electron transfer between the sites as observed in other Rieske-mononuclear iron oxygenases (Ferraro et al. 2005). The conserved aspartate (D176 in MpCAO2) is involved in gating electron transport between the two centers (Parales et al. 1999). The distance of this aspartate to the two interacting histidine residues in different subunits is less than 4 Å and the arrangement of this electron transport system is well conserved among Rieske oxygenases **(Figure 3b)**. Taken together, the structure for substrate oxidation is conserved even though it is formed with two distinct polypeptides.

4.2 Separation and unification of the components involved in electron transport

In this study, we examined separation of the components involved in electron transport of MpCAO structurally. Similar subunit separation has been observed in *Halomicronema hongdechloris* BciB, which reduces 8-vinyl group during chlorophyll biosynthesis (Bryant et al. 2020). *Halomicronema hongdechloris* is a cyanobacterium having chlorophyll *f* in addition to chlorophyll *a*. BciB usually possesses two Fe-S clusters for its functioning. In this cyanobacterium, Fe-S clusters are constructed in separate polypeptides (NCBI Accession ID: ASC70450.1 and ASC70451.1). Additionally, Fe-S cluster of BciB has another uncommon feature. BciB is homologous to the β subunit of F_{420}-reducing [NiFe]-hydrogenase complex from *Methanothermobacter marburgensis* (Vitt et al. 2014). Flavin adenine dinucleotide (FAD) is the terminal electron carrier for the substrate reduction, where Fe-S cluster transfers electrons to FAD. These components are found in the FrhB subunit of F_{420}-reducing hydrogenase. This Fe-S cluster is also reduced by a second Fe-S cluster labeled Cluster 2 in the FrhG subunit of F_{420}-reducing hydrogenase as previously discussed (Wang and Liu 2016), while BciB contains the same Fe-S cluster on its own polypeptide labeled Cluster 2 **(Figure 4)**. It is not known how BciB obtained Fe-S cluster at a similar position to that of F_{420}-reducing hydrogenase. The variation of these electron transfer component sites suggests the flexibility in arrangement through assembly of the different subunits. Along with the results obtained for MpCAO, the aforementioned examples also suggest its flexible subunit construction for electron transport. However, the physiological importance of this diversity remains to be elucidated.

5 Conclusion

Although the structure and reaction mechanism of other Rieske monooxygenases have been described, this study provides the first report of structural characterization for a member of plant Rieske non-heme iron dependent monooxygenase *i.e.*, CAO. The high degree of conservation for majority of residues comprising the ligand binding cavity demonstrates a common chlorophyll *a* binding mode for all CAOs from different organisms. In addition, the inter-residue distances and orientations of the amino acids involved in interaction with Rieske cluster and mononuclear iron-binding are well conserved among the members of Rieske monooxygenases that are distributed across various life forms and responsible for catalyzing a wide array of oxidative transformations in a range of catabolic and biosynthetic pathways. Though a feasible electron transfer pathway can be hypothesized from this computational analysis, experimental validation remains necessary for better understanding of the reaction mechanism of CAO.

Figure legends

Fig. 1 Cartoon representation of the predicted three-dimensional structures. a Rieske [2Fe-2S] cluster in MpCAO1 and AtCAO. The position of the domains in AtCAO has been shown. The residues involved in interaction with Rieske unit have been shown for each protein. b Non-heme iron center in MpCAO2 and AtCAO. The residues interacting with the Fe ion have been shown for each protein. c Cartoon representation of AtCAO homotrimer and MpCAO1 (green)-MpCAO2 (blue) heterodimer. The cofactors in MpCAO heterodimer are shown as spheres.

Fig. 2 Ferredoxin and its interaction with CAO heterodimer in *Micromonas pusilla*. a Predicted tertiary structure of MpFd with Rieske [2Fe-2S] cluster. The residues involved in interaction with Rieske unit have been shown. b Docked complex of MpFd (cartoon) with MpCAO1-MpCAO2 heterodimer (hydrophobic surface representation where light-color and deep-color indicate hydrophilic and hydrophobic regions, respectively).

Fig. 3 Substrate interaction with the predicted structure. a Docked structure of chlorophyll *a* with MpCAO2 and AtCAO monomers. The substrate is depicted as sticks (yellow) and the mononuclear iron is shown as sphere (orange). b Hypothetical electron transfer pathway between the two subunits of *Micromonas* CAO. The red dashed line indicates the interface between the two subunits. The Rieske unit (orange-red) and the ligand (yellow) are represented as sticks while the mononuclear iron (red) is shown as sphere.

Fig. 4 Different localization of the Fe-S cluster. Cyanobacterial BciB and FrhB-FrhG complex of methanobacterial F420-reducing hydrogenase are shown as cartoon. Fe-S cluster is shown as spheres inside circle. FAD is shown in sticks. Second Fe-S cluster marked Cluster 2 in F420-reducing hydrogenase localizes FrhG subunit.

References

Ashikawa Y, Fujimoto Z, Noguchi H, Habe H, Omori T, Yamane H, Nojiri H (2006) Electron Transfer Complex Formation between Oxygenase and Ferredoxin Components in Rieske Nonheme Iron Oxygenase System. Structure 14:1779

Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millán C, Park H, Adams C, Glassman CR, DeGiovanni A, Pereira JH, Rodrigues AV, van Dijk AA, Ebrecht AC, Opperman DJ, Sagmeister T, Buhlheller C, Pavkov-Keller T, Rathinaswamy MK, Dalwadi U, Yip CK, Burke JE, Garcia KC, Grishin NV, Adams PD, Read RJ, Baker D (2021) Accurate prediction of protein structures and interactions using a three-track neural network. Science 373:871

Baek M, Park T, Heo L, Park C, Seok C (2017) GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure. Nucleic Acids Res. 45:W320

Bailey S, Walters R, Jansson S, Horton P (2001) Acclimation of Arabidopsis thaliana to the light environment: the existence of separate low light and high light responses. Planta 213:794

Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N (2004) ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics 20:1322

Brimberry M, Garcia AA, Liu J, Tian J, Bridwell-Rabb J (2022) Engineering Rieske oxygenase activity one piece at a time. Curr. Opin. Chem. Biol. 72:102227

Bryant DA, Hunter CN, Warren MJ (2020) Biosynthesis of the modified tetrapyrroles-the pigments of life. J. Biol. Chem. 295:6888

Bugg TD, Ramaswamy S (2008) Non-heme iron-dependent dioxygenases: unravelling catalytic mechanisms for complex enzymatic oxidations. Curr. Opin. Chem. Biol. 12:134

Caffarri S, Croce R, Breton J, Bassi R (2001) The major antenna complex of photosystem II has a xanthophyll binding site not involved in light harvesting. J. Biol. Chem. 276:35924

Chen M (2014) Chlorophyll Modifications and Their Spectral Extension in Oxygenic Photosynthesis. Annu. Rev. Biochem. 83:317

Christoffer C, Chen S, Bharadwaj V, Aderinwale T, Kumar V, Hormati M, Kihara D (2021) LZerD webserver for pairwise and multiple protein-protein docking. Nucleic Acids Res. 49:W359

Colovos C, Yeates TO (1993) Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci. 2:1511

Costas M, Mehn MP, Jensen MP, Que L (2004) Dioxygen Activation at Mononuclear Nonheme Iron Active Sites: Enzymes, Models, and Intermediates. Chemical Reviews 104:939

D'Ordine RL, Rydel TJ, Storek MJ, Sturman EJ, Moshiri F, Bartlett RK, Brown GR, Eilers RJ, Dart C, Qi Y, Flasinski S, Franklin SJ (2009) Dicamba monooxygenase: structural insights into a dynamic Rieske oxygenase that catalyzes an exocyclic monooxygenation. J. Mol. Biol. 392:481

Delano W L (2002) The PyMOL Molecular Graphics System. http://www.pymol.org

Espineda CE, Linford AS, Devine D, Brusslan JA (1999) The AtCAO gene, encoding chlorophyll a oxygenase, is required for chlorophyll b synthesis in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. 96:10507

Ferraro DJ, Gakhar L, Ramaswamy S (2005) Rieske business: structure-function of Rieske non-heme oxygenases. Biochem. Biophys. Res. Commun. 338:175

Furusawa Y, Nagarajan V, Tanokura M, Masai E, Fukuda M, Senda T (2004) Crystal Structure of the Terminal Oxygenase Component of Biphenyl Dioxygenase Derived from Rhodococcus sp. Strain RHA1. J. Mol. Biol. 342:1041

Gakhar L, Malik ZA, Allen CCR, Lipscomb DA, Larkin MJ, Ramaswamy S (2005) Structure and Increased Thermostability of Rhodococcus sp. Naphthalene 1,2-Dioxygenase. J. Bacteriol. 187:7222

Green BR, Durnford DG (1996) The chlorophyll-carotenoid proteins of oxygenic photosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47:685

Heo L, Lee H, Seok C (2016) GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. Sci Rep 6:32153

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2017) UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35:518

Jia T, Ito H, Tanaka A (2016) Simultaneous regulation of antenna size and photosystem I/II stoichiometry in Arabidopsis thaliana. Planta 244:1041

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Meth. 14:587

Kovaleva EG, Lipscomb JD (2008) Versatility of biological non-heme Fe(II) centers in oxygen activation reactions. Nat. Chem. Biol. 4:186

Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S (2017) The ClusPro web server for protein-protein docking. Nat Protoc 12:255

Kunugi M, Satoh S, Ihara K, Shibata K, Yamagishi Y, Kogame K, Obokata J, Takabayashi A, Tanaka A (2016) Evolution of Green Plants Accompanied Changes in Light-Harvesting Systems. Plant Cell Physiol. 57:1231

Kunugi M, Takabayashi A, Tanaka A (2013) Evolutionary changes in chlorophyllide a oxygenase (CAO) structure contribute to the acquisition of a new light-harvesting complex in Micromonas. J. Biol. Chem. 288:19330

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283

Leconte J, Benites LF, Vannier T, Wincker P, Piganeau G, Jaillon O (2020) Genome Resolved Biogeography of Mamiellales. Genes (Basel) 11
Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49:W293
Lin Y-F, Cheng C-W, Shih C-S, Hwang J-K, Yu C-S, Lu C-H (2016) MIB: Metal Ion-Binding Site Prediction and Docking Server. Journal of Chemical Information and Modeling 56:2287
Liu J, Knapp M, Jo M, Dill Z, Bridwell-Rabb J (2022) Rieske Oxygenase Catalyzed C–H Bond Functionalization Reactions in Chlorophyll b Biosynthesis. ACS Central Science
Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83
Martins BM, Svetlitchnaia T, Dobbek H (2005) 2-Oxoquinoline 8-Monooxygenase Oxygenase Component: Active Site Modulation by Rieske-[2Fe-2S] Center Oxidation/Reduction. Structure 13:817
Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30:2785
Mueller AH, Dockter C, Gough SP, Lundqvist U, von Wettstein D, Hansson M (2012) Characterization of Mutations in Barley fch2 Encoding Chlorophyllide a Oxygenase. Plant Cell Physiol. 53:1232
Nagata N, Satoh S, Tanaka R, Tanaka A (2004) Domain structures of chlorophyllide a oxygenase of green plants and Prochlorothrix hollandica in relation to catalytic functions. Planta 218:1019
Nojiri H, Ashikawa Y, Noguchi H, Nam J-W, Urata M, Fujimoto Z, Uchimura H, Terada T, Nakamura S, Shimizu K, Yoshida T, Habe H, Omori T (2005) Structure of the Terminal Oxygenase Component of Angular Dioxygenase, Carbazole 1,9a-Dioxygenase. J. Mol. Biol. 351:355
Not F, Latasa M, Marie D, Cariou T, Vaulot D, Simon N (2004) A single species, Micromonas pusilla (Prasinophyceae), dominates the eukaryotic picoplankton in the Western English Channel. Appl. Environ. Microbiol. 70:4064
O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: An open chemical toolbox. J Cheminform 3:33
Oster U, Tanaka R, Tanaka A, Rudiger W (2000) Cloning and functional expression of the gene encoding the key enzyme for chlorophyll b biosynthesis (CAO) from Arabidopsis thaliana. Plant J. 21:305
Page CC, Moser CC, Chen X, Dutton PL (1999) Natural engineering principles of electron tunnelling in biological oxidation–reduction. Nature 402:47
Parales RE, Parales JV, Gibson DT (1999) Aspartate 205 in the Catalytic Domain of Naphthalene Dioxygenase Is Essential for Activity. J. Bacteriol. 181:1831
Park T, Won J, Baek M, Seok C (2021) GalaxyHeteromer: protein heterodimer structure prediction by template-based and ab initio docking. Nucleic Acids Res. 49:W237
Partensky F, Garczarek L (2010) Prochlorococcus: advantages and limits of minimalism. Ann Rev Mar Sci 2:305
Perry C, de los Santos Emmanuel LC, Alkhalaf LM, Challis GL (2018) Rieske non-heme iron-dependent oxygenases catalyse diverse reactions in natural product biosynthesis. Nat. Prod. Rep. 35:622
Porra RJ, SchÄFer W, Cmiel E, Katheder I, Scheer H (1994) The derivation of the formyl-group oxygen of chlorophyll b in higher plants from molecular oxygen. Eur. J. Biochem. 219:671

Sakuraba Y, Tanaka R, Yamasato A, Tanaka A (2009) Determination of a chloroplast degron in the regulatory domain of chlorophyllide a oxygenase. J. Biol. Chem. 284:36689
Satoh S, Tanaka A (2006) Identification of Chlorophyllide a Oxygenase in the Prochlorococcus Genome by a Comparative Genomic Approach. Plant Cell Physiol. 47:1622
Sievers F, Higgins DG (2018) Clustal Omega for making accurate alignments of many protein sequences. Protein science : a publication of the Protein Society 27:135
Tanaka A, Ito H, Tanaka R, Tanaka NK, Yoshida K, Okada K (1998) Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. Proc. Natl. Acad. Sci. U. S. A. 95:12719
Tanaka A, Tanaka R (2019) Chapter Six - The biochemistry, physiology, and evolution of the chlorophyll cycle. In: Grimm B (ed) Adv. Bot. Res. Academic Press, pp. 183-212
Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44:W232
Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455
Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Žídek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50:D439
Venkatraman V, Yang YD, Sael L, Kihara D (2009) Protein-protein docking using region-based 3D Zernike descriptors. BMC Bioinformatics 10:407
Vitt S, Ma K, Warkentin E, Moll J, Pierik AJ, Shima S, Ermler U (2014) The F420-reducing [NiFe]-hydrogenase complex from Methanothermobacter marburgensis, the first X-ray structure of a group 3 family member. J. Mol. Biol. 426:2813
Wang X, Liu L (2016) Crystal Structure and Catalytic Mechanism of 7-Hydroxymethyl Chlorophyll a Reductase. J. Biol. Chem. 291:13349
Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46:W296
Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189
Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35:W407
Xu J, Zhang Y (2010) How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics 26:889
Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, Shen Y, Chen F, Lai L, Pei J (2018) CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. Nucleic Acids Res. 46:W374
Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A (2016) PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. Bioinformatics 32:3676
Yamasato A, Nagata N, Tanaka R, Tanaka A (2005) The N-terminal domain of chlorophyllide a oxygenase confers protein instability in response to chlorophyll b accumulation in Arabidopsis. Plant Cell 17:1585
Yan Y, Tao H, He J, Huang S-Y (2020) The HDOCK server for integrated protein–protein docking. Nat Protoc 15:1829

Yang J, Roy A, Zhang Y (2013) Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29:2588

Zheng W, Li Y, Zhang C, Zhou X, Pearce R, Bell EW, Huang X, Zhang Y (2021) Protein structure prediction using deep learning distance and hydrogen-bonding restraints in CASP14. Proteins 89:1734

Fig. 1

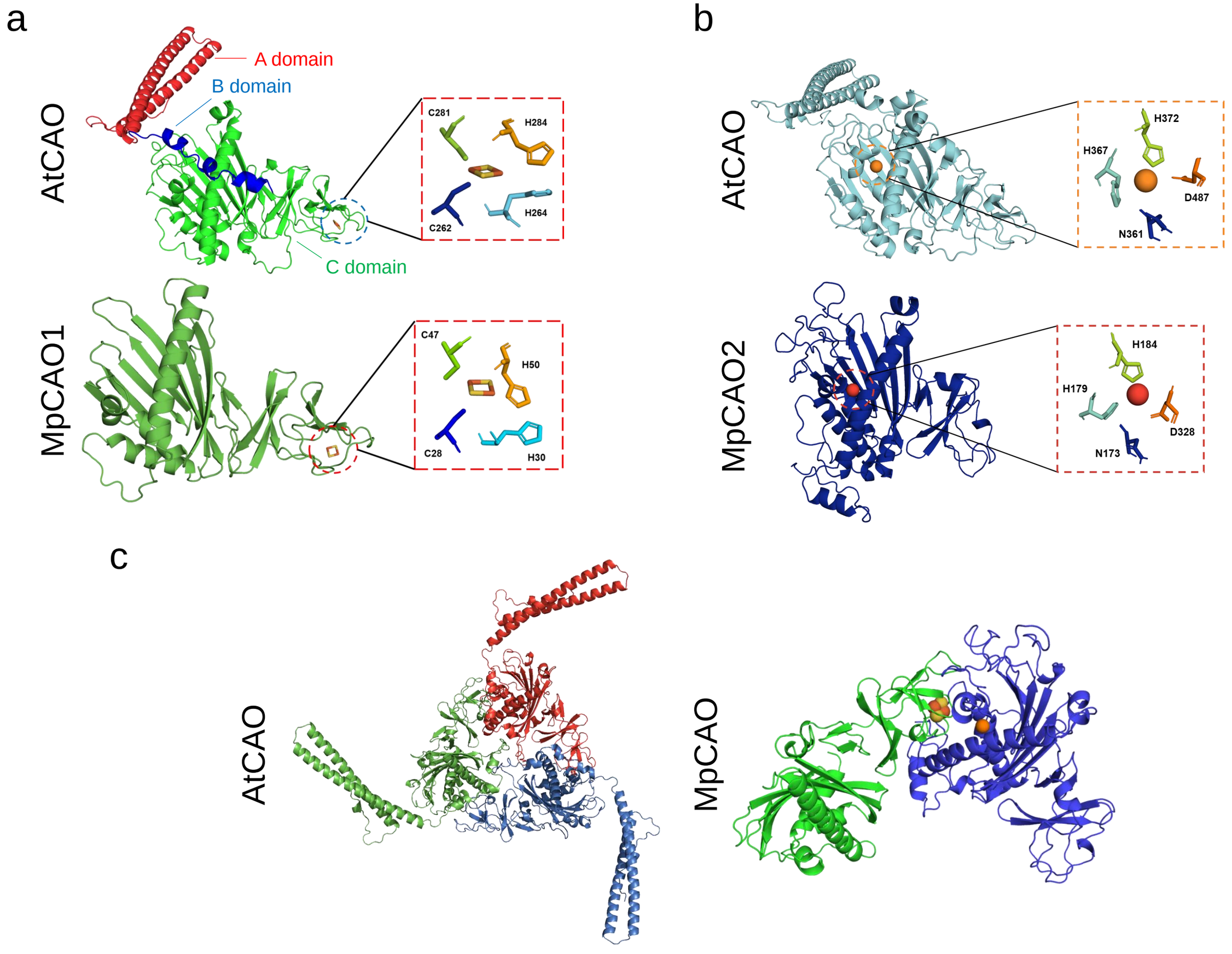
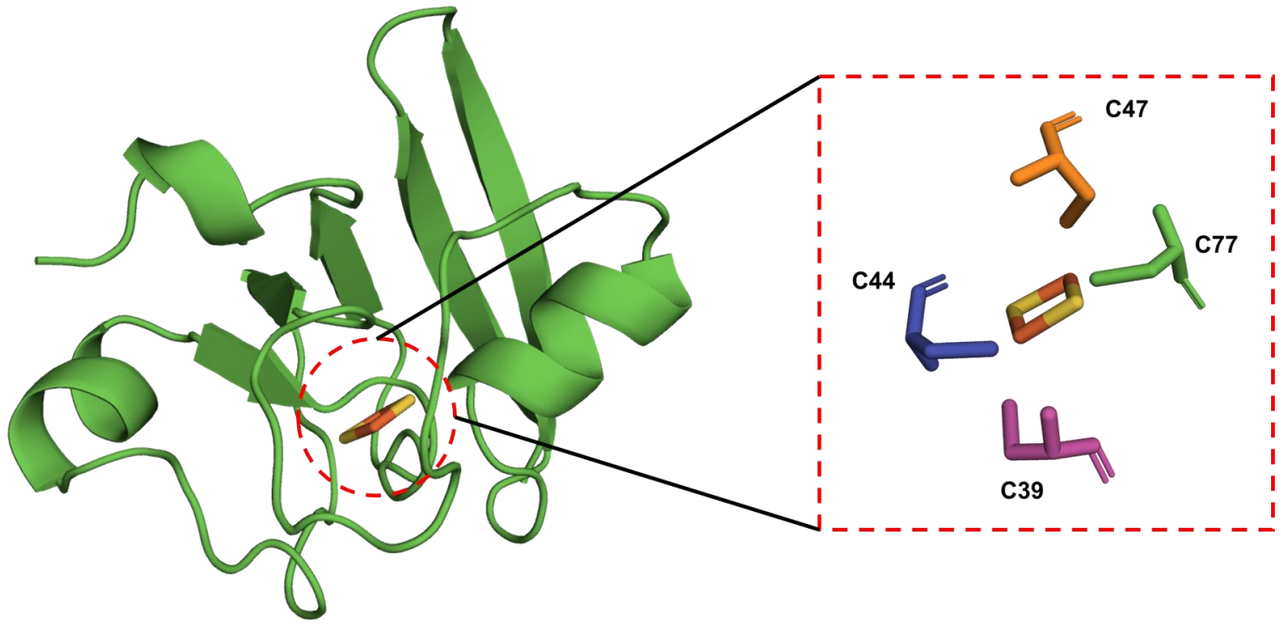

a
AtCAO
A domain
B domain
C domain
C281
H284
C262
H264
MpCAO1
C47
H50
C28
H30
b
AtCAO
H372
H367
D487
N361
MpCAO2
H184
H179
D328
N173
c
AtCAO
MpCAO

Fig. 2

a



b

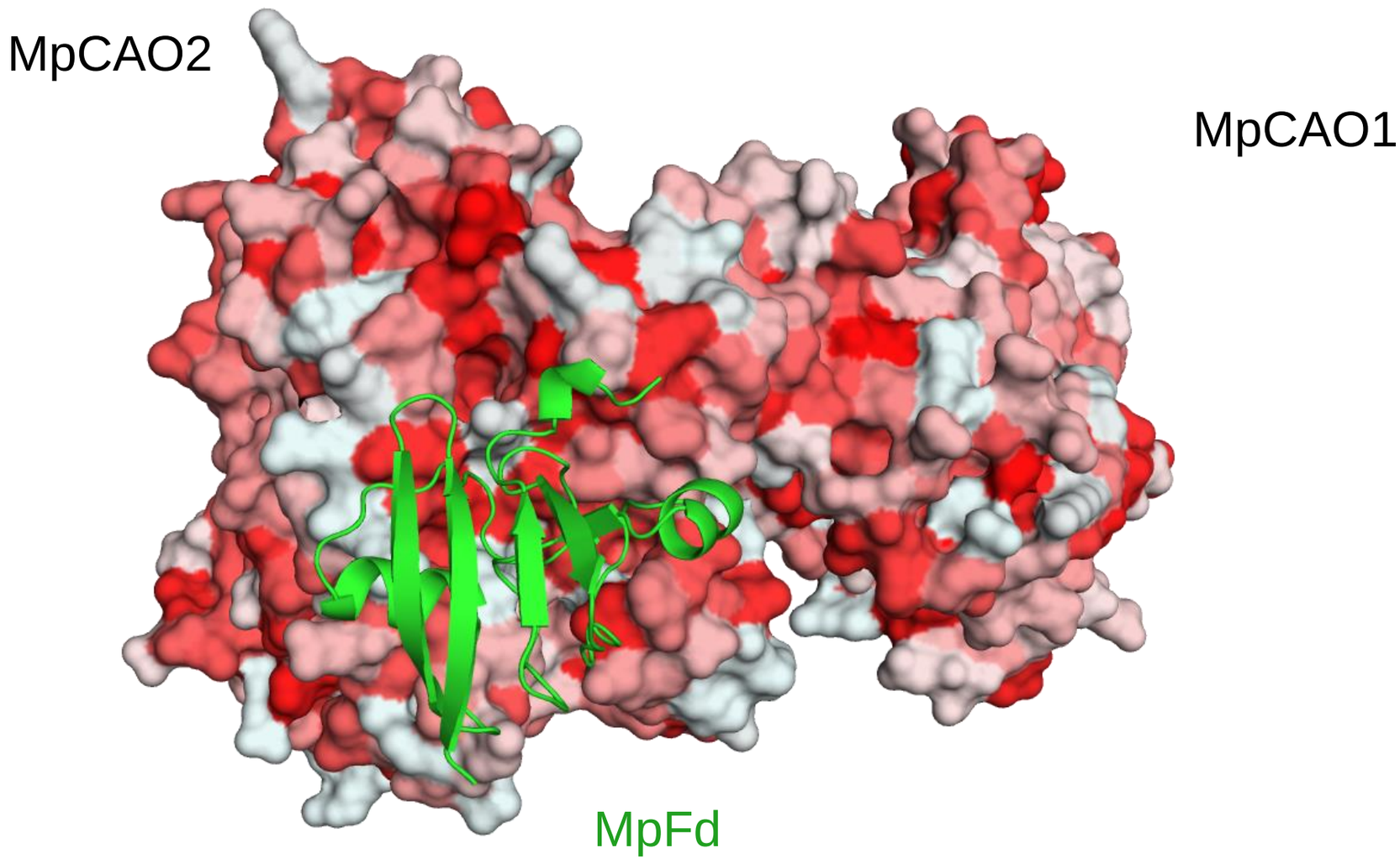


Fig. 3

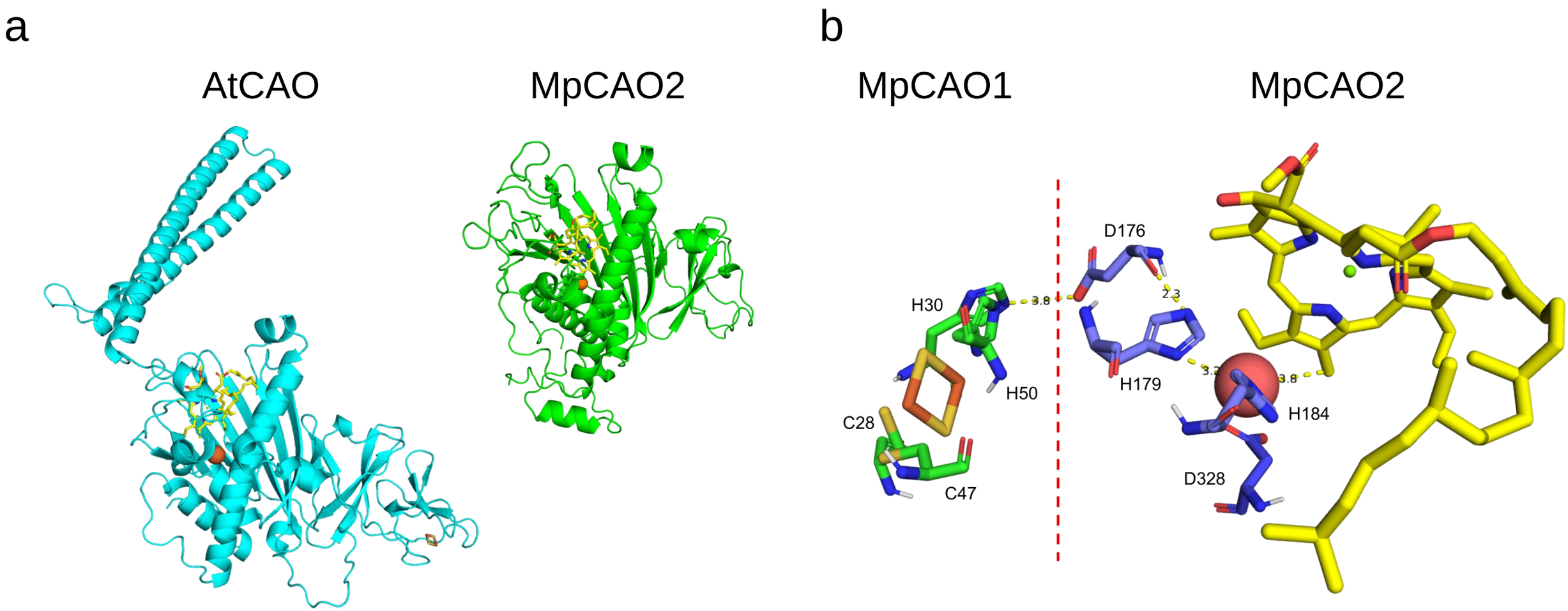
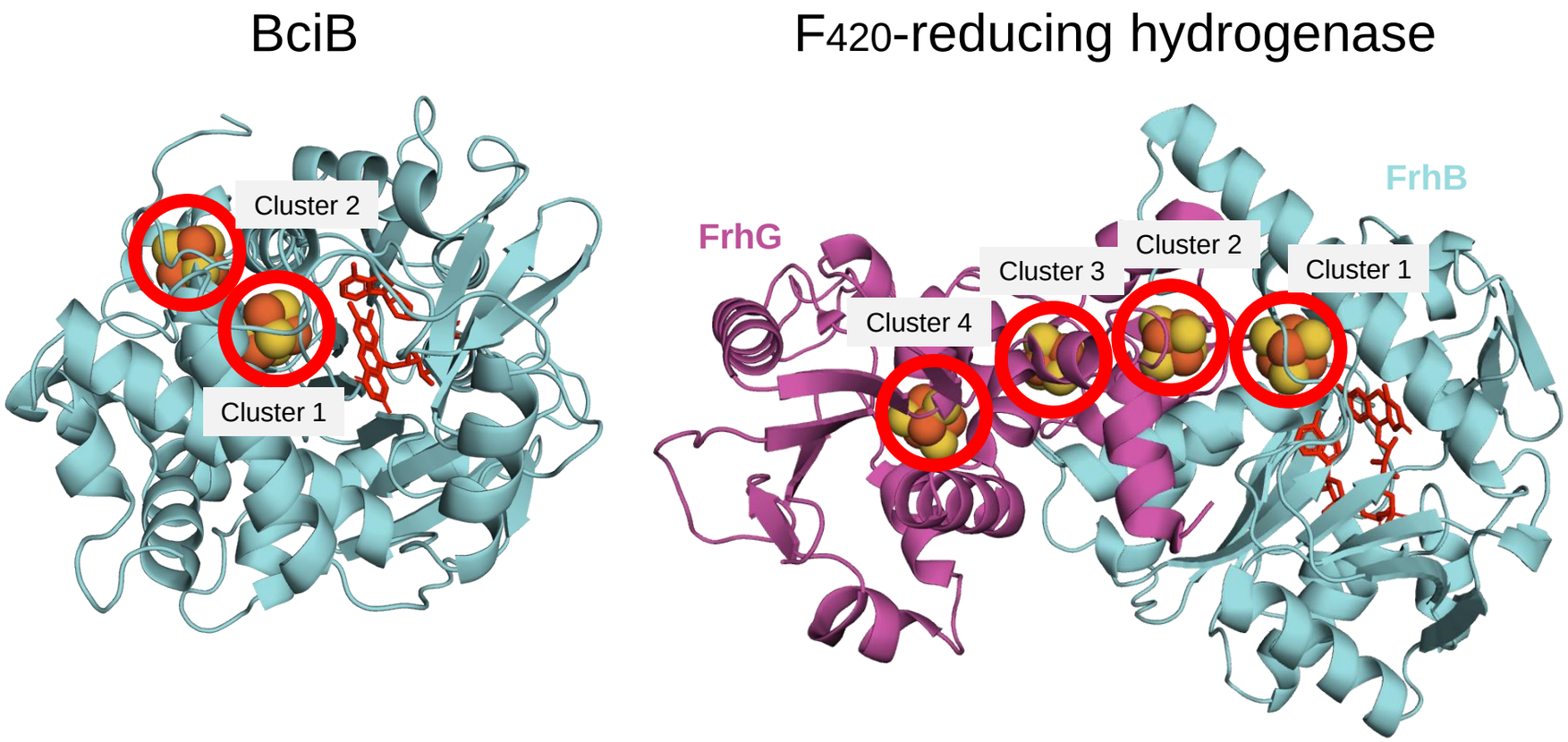
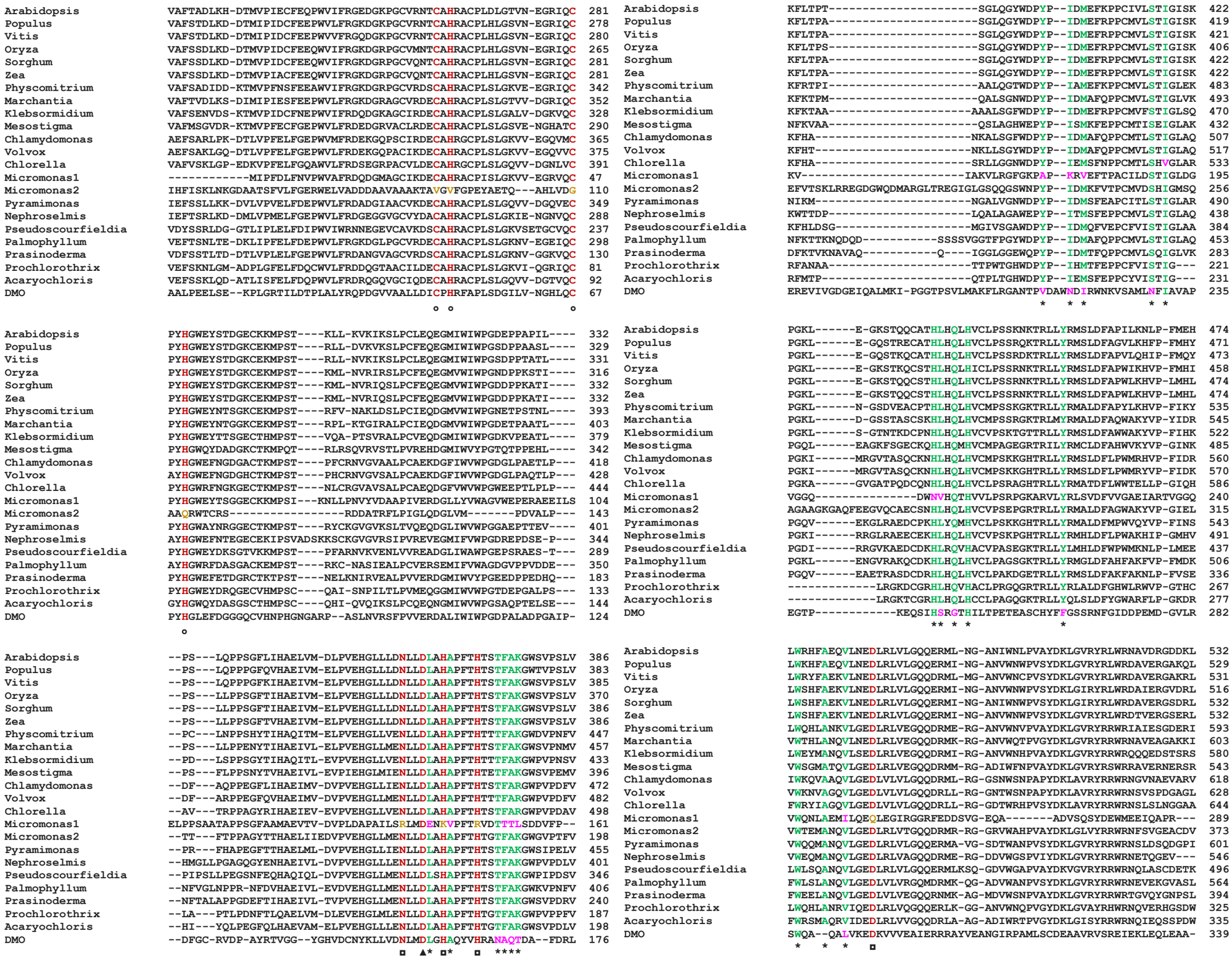
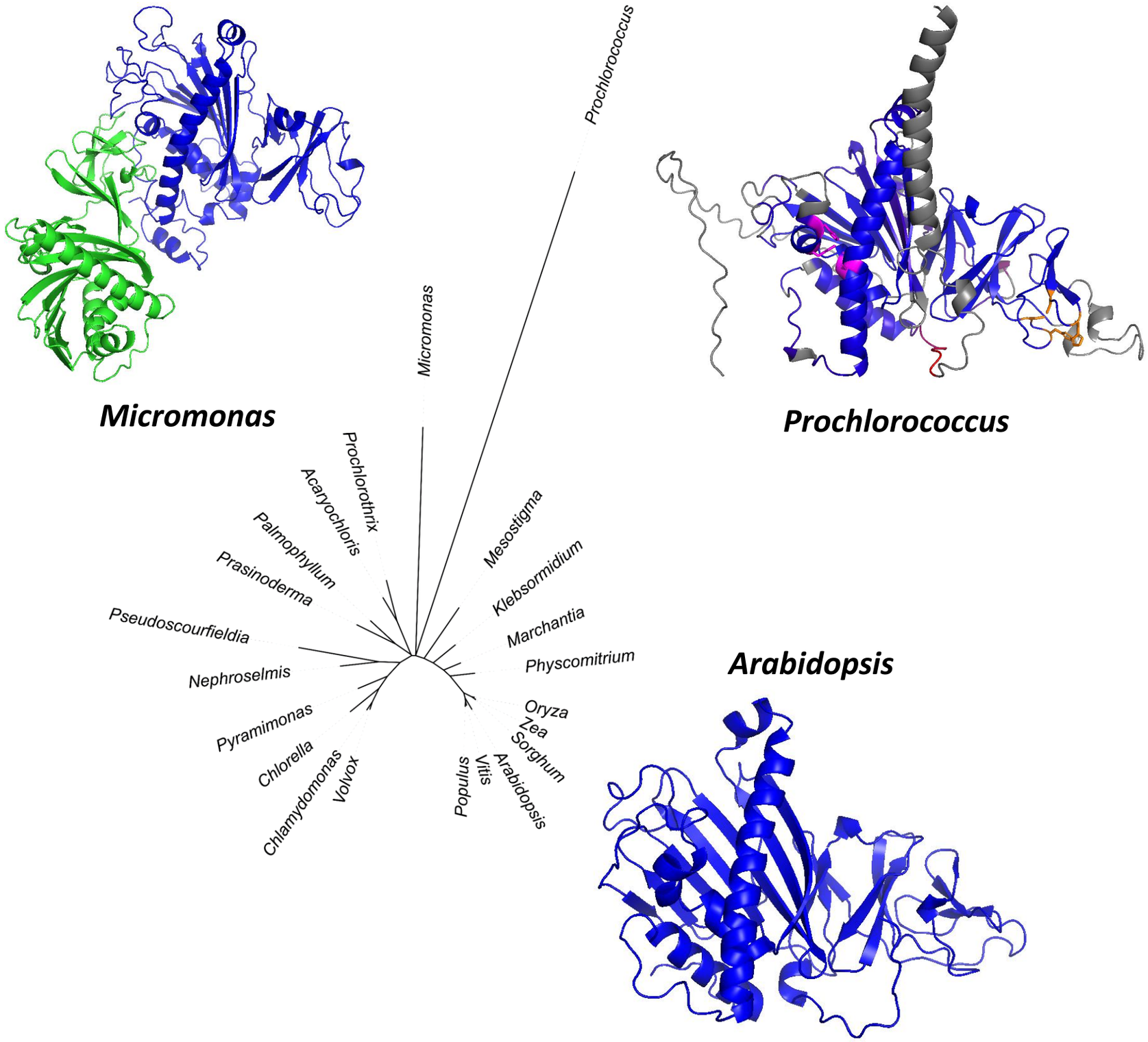

a
AtCAO
MpCAO2
b
MpCAO1
MpCAO2
H30
H50
C28
C47
D176
H179
H184
D328
3.8
2.3
3.2
3.8

Fig. 4

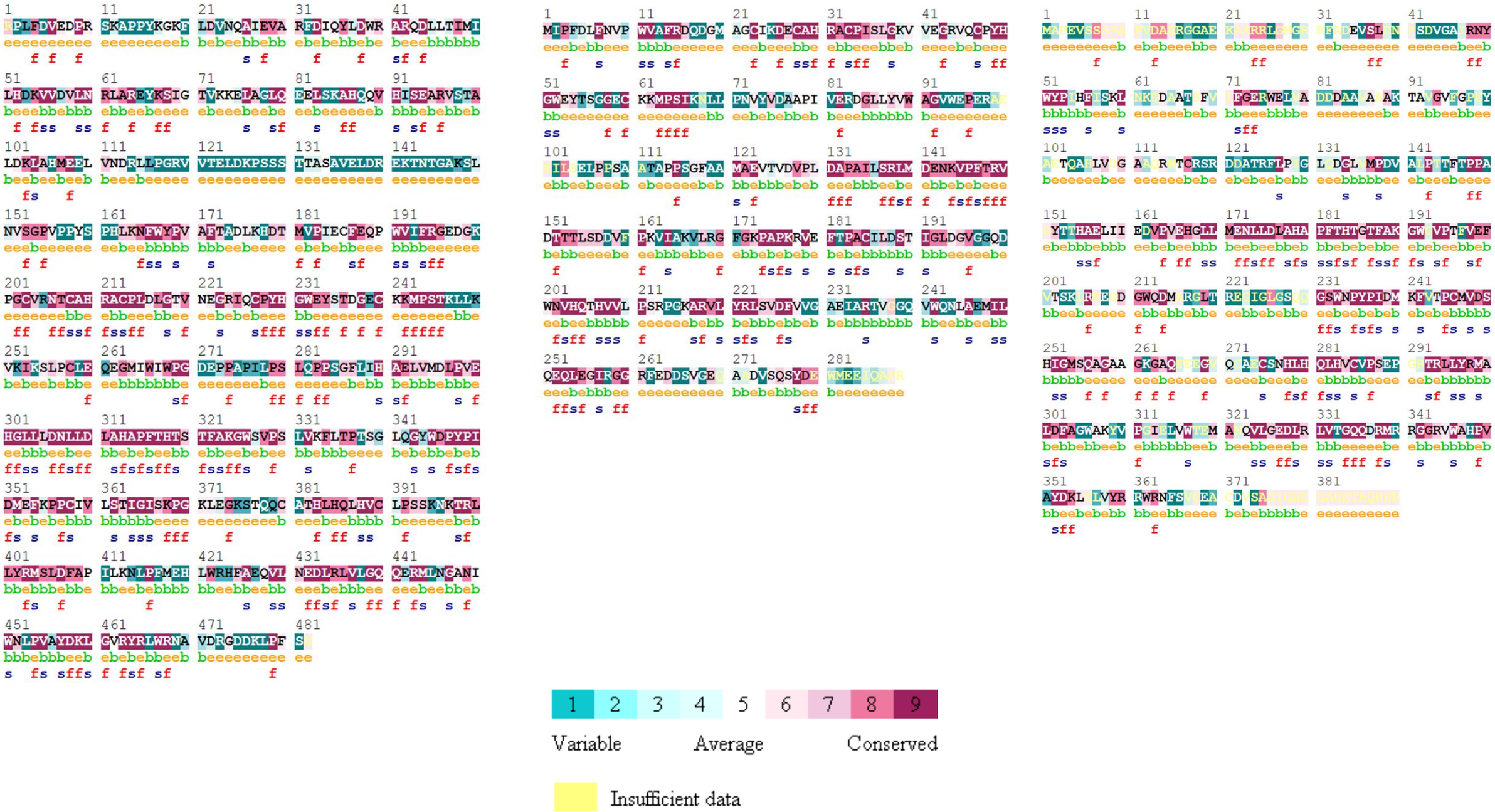

BciB
Cluster 2
Cluster 1
F420-reducing hydrogenase
FrhG
FrhB
Cluster 4
Cluster 3
Cluster 2
Cluster 1



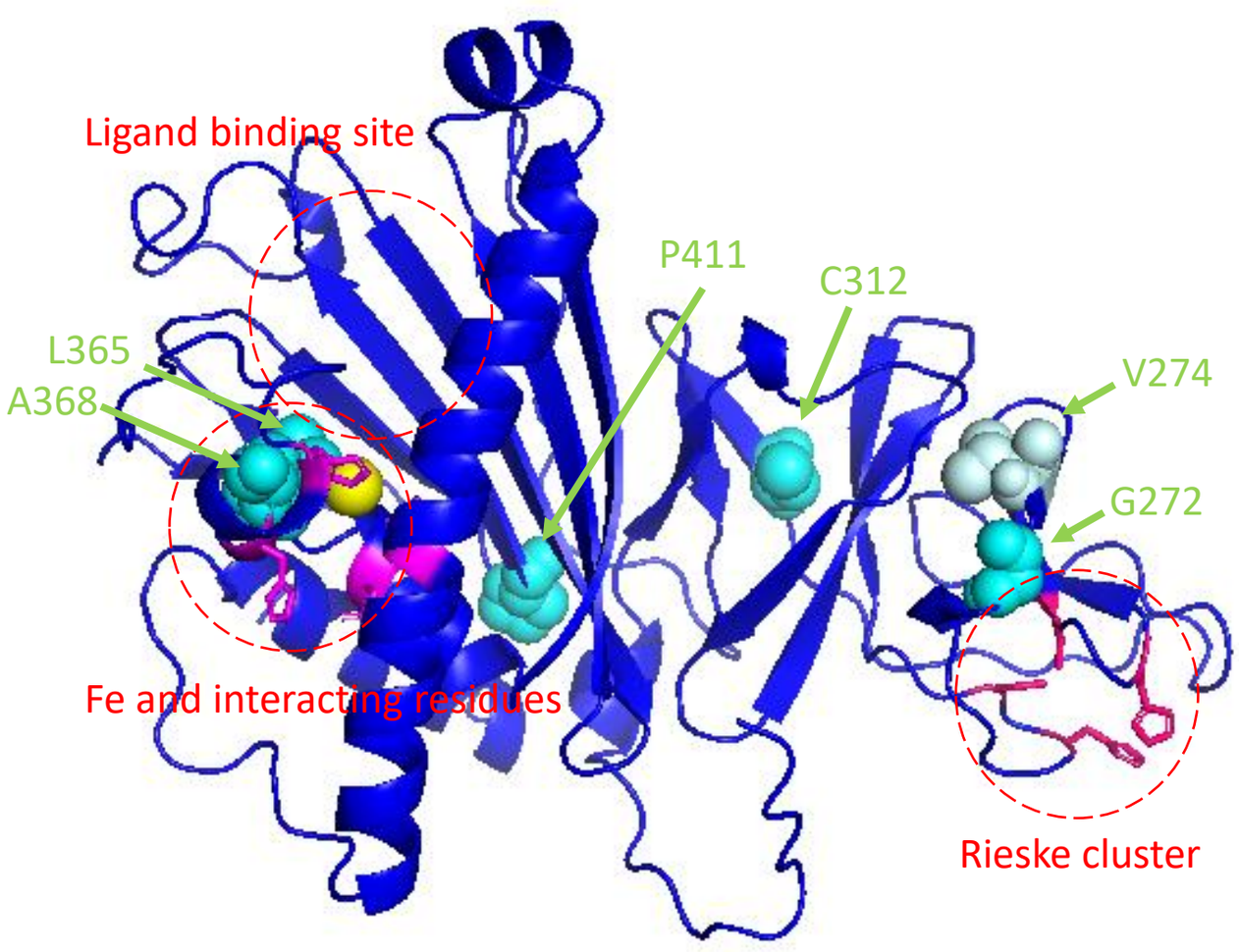
Supplementary Figure S1. Multiple sequence alignment of CAO proteins from different organisms and dicamba monooxygenase (DMO) using Clustal Omega. The multiple sequence alignment has been marked with a user defined color code. Conserved residues involved in interaction with Rieske [2Fe-2S] cluster (○) and non-heme iron (▫) are colored in red while mutation in the otherwise conserved column is marked in gold. Residues constituting the ligand binding cavity (*) are colored in green whereas mutations therein are coded in magenta. A conserved aspartate (▲) that plays an essential ‘gatekeeper’ role by transferring electrons through CAO subunit interface is colored in red.



Supplementary Figure S2. A maximum likelihood phylogenetic tree involving CAO protein sequences is determined using IQ-TREE v 1.6.12. *Arabidopsis*, *Micromonas*, and *Prochlorococcus* CAO structures are shown. Separated MpCAO1 and MpCAO2 are shown in green and blue. PmCAO is colored based on RMSD to AtCAO. Blue indicates the minimum pairwise RMSD while red shows the maximum. Gray shows unaligned residues. Non-heme iron-interacting residues and Rieske cluster are shown with side chains in magenta and orange, respectively.



Supplementary Figure S3. Evolutionary conservation of amino acid residues in the primary sequence of AtCAO, MpCAO1, and MpCAO2 by ConSeq analysis. Note: ‘e’ refers to an exposed residue according to the neural-network algorithm; ‘b’ refers to a buried residue according to the neural-network algorithm; ‘f’ refers to a predicted functional residue (highly conserved and exposed); ‘s’ refers to a predicted structural residue (highly conserved and buried).



Supplementary Figure S4. Mutation positions in CAO. AtCAO C-domain is shown. Ligand and non-heme iron (yellow) binding sites, interacting residues (magenta), and Rieske cluster binding site (light magenta) are marked with red dash circles. Mutated residues in barley CAO (cyan) and *Arabidopsis ch1-2* (light cyan) are shown as spheres.