Doctoral Dissertation

Structural characterization of magnesium dechelatase and chlorophyllide *a* oxygenase involved in the chlorophyll metabolic pathway

(クロロフィル代謝系に含まれるマグネシウム脱離酵素とクロロフィリド *a* オキシゲナーゼの構造解析)

Debayan Dey

Institute of Low Temperature Science
Graduate School of Life Science
Hokkaido University

September 2022

Summary

In land plants, green algae and some cyanobacteria, chlorophyll *a* and chlorophyll *b* form the principal components of the photosynthetic machinery that play crucial role in absorption, transmission, and transformation of light energy. The difference between the two chlorophyll species is the presence of a formyl group at the C7 position in chlorophyll *b* while a methyl group occurs at the same position in chlorophyll *a*. Both chlorophylls possess distinct absorption spectra in the blue and red regions, which allows this combination of pigments to utilize a wide range of light spectra for photosynthesis. The light harvesting complexes (LHCs) of photosynthetic organisms are composed of core and peripheral antenna complexes. While chlorophyll *a* is present in the core antenna of photosystems I and II as chlorophyll-protein complexes, chlorophyll *b* mainly resides in the peripheral antenna complexes along with other pigments like fucoxanthin. Moreover, chlorophyll *a* is vital for photochemistry in oxygenic photosynthetic organisms whereas chlorophyll *b* is necessary for stabilizing the major light-harvesting chlorophyll-binding proteins and also in regulating the photosynthetic antenna size by altering the chlorophyll *a*/*b* ratio.

Chlorophyll biosynthesis must be finely regulated for efficient photosynthetic performance during the formation of photosystems at the greening stage and also during adaptation to various environmental conditions. Not only chlorophyll biosynthesis but also chlorophyll degradation needs to be regulated because the latter plays a crucial role in mobilizing resources from chloroplast to developing organs. In addition, chlorophyll breakdown forms a key part of nitrogen recycling and is important in avoiding cellular photodamage. Before degradation, chlorophyll *b* must be converted to chlorophyll *a* because chlorophyll *b* derivatives are not catalyzed in the later steps of the chlorophyll degradation pathway. The interconversion pathway between chlorophyll *a* and chlorophyll *b* is referred to as the chlorophyll cycle.

Chlorophyll *a* is converted to chlorophyll *b* in two successive steps by chlorophyll(ide) *a* oxygenase (CAO). In the first step of chlorophyll *b* conversion, the enzyme chlorophyll *b* reductase (CBR) reduces the formyl group of chlorophyll *b* to produce 7-hydroxymethyl chlorophyll *a*. In the final step, chlorophyll *a* is formed by the enzyme 7-hydroxymethyl chlorophyll *a* reductase (HCAR), the structure of which resembles an archaeal F_{420}-reducing [NiFe] hydrogenase. Chlorophyll *a* is then converted to a primary fluorescent Chl catabolite by four continuous steps. First, central magnesium (Mg) ion in chlorophyll *a* is extracted by a Mg-dechelatase enzyme encoded by the *Stay-Green* (SGR) gene to form pheophytin *a*, which is then hydrolyzed to become pheophorbide *a* and phytol by pheophytinase (PPH). As the porphyrin of pheophorbide *a* is cleaved by pheophorbide *a* oxygenase (PAO), the green color completely fades in chlorophyll catabolite, leading to the formation of red chlorophyll

catabolite. Subsequently, it is turned to the primary fluorescent chlorophyll catabolite by red chlorophyll catabolite reductase (RCCR) which is transferred out of the chloroplasts and isomerized to non-fluorescent products by acidic pH in the vacuole. My PhD study provides insights into the structural characteristics of two chlorophyll metabolic pathway enzymes – SGR and CAO.

Chapter 1 deals with the Mg-dechelatase enzyme which catalyzes Mg^{2+} dechelation from chlorophyll *a*. This reaction is the first committed step of chlorophyll degradation pathway in plants and is thus indispensable for the process of leaf senescence. There is no structural information available for this or its related enzymes. This chapter provides insight into the structure and reaction mechanism of the enzyme through biochemical and computational analysis of an SGR homolog from the Chloroflexi *Anaerolineae* (AbSGR-h). Recombinant AbSGR-h with its intact sequence and those with mutations were overexpressed in *Escherichia coli* and their Mg-dechelatase activity was compared. Two aspartates – D34 and D62 were found to be essential for catalysis, while R26, Y28, T29, and D114 were responsible for structural maintenance. Gel filtration analysis of the recombinant AbSGR-h revealed the formation of a homo-oligomer. The three-dimensional structure of AbSGR-h was predicted by a deep learning-based method, which was evaluated by protein structure quality evaluation programs while structural stability of wild-type and mutant forms were investigated through molecular dynamics simulations. Furthermore, in concordance with the results of the enzyme assay, molecular docking concluded the significance of D34 in ligand interaction. By combining biochemical analysis and computational prediction, the study unveils the detailed structural characteristics of the enzyme, including the probable pocket of interaction and the residues of structural and functional importance.

Chapter 2 also deals with the in-depth analysis of the structure of Mg-dechelatase enzyme. The crystal structure of a highly active SGR homolog from *Anaerolineae* (AbSGR-h) bacterium at 1.75 Å resolution has been reported. A previous study revealed the catalytic significance of D34 residue in AbSGR-h protein for interaction with the central Mg of chlorophyll *a*. Therefore, recombinant WT AbSGR-h and three mutants (D34E, D34N, and D34Q) were overexpressed in *E. coli* and purified by nickel column and size exclusion chromatography. Gel filtration profiles of the WT and three mutant proteins were found to be similar thus confirming the role of D34 to be solely catalytic rather than maintaining the multimeric conformation of the protein. Activity analysis revealed substantial decrease of Mg-dechelation level for the D34E mutant and loss of activity for the D34N and D34Q mutants. The kinetic parameters of WT and D34E mutant AbSGR-h were elucidated by Michaelis-Menten analysis. Furthermore, molecular docking analysis showed stable interaction of the central Mg ion of chlorophyll *a* with the carbonyl oxygen atom of D34 residue in the crystal structure of AbSGR-h monomer within a distance of 4.4 Å. Besides, the catalytic triad found in AbSGR-h was found to show high resemblance with those

observed in hydrolases. This study enhances the existing knowledge about the reaction mechanism of Mg-dechelatase and also provides the first crystal structure of a homolog from the SGR family.

Chapter 3 highlights the structural characteristics of the CAO enzyme, that is responsible for converting chlorophyll *a* to chlorophyll *b*. CAO belongs to the family of Rieske mononuclear iron oxygenases. Here, the tertiary structures of CAO from the Prasinophyte *Micromonas pusilla* (MpCAO) and model plant *Arabidopsis thaliana* (AtCAO) were predicted by deep learning-based methods, followed by energy minimization and subsequent stereochemical quality assessment of the predicted models. Although plant CAO structure exhibits the three-fold symmetric homotrimer form, like most other Rieske non-heme iron oxygenases, *Micromonas* CAO exist as two distinct polypeptides (MpCAO1 and MpCAO2). Thus, its heterodimeric association was computationally investigated. Furthermore, the chlorophyll *a* binding cavity on the surface of MpCAO2 was predicted and molecular docking analysis revealed presence of the substrate at the vicinity of the mononuclear iron center. This study enables the structural visualization of the electron transfer pathway between the two distinct subunits of MpCAO.

Mg-dechelatase or SGR plays an indispensable role in chlorophyll metabolism because it catalyzes the committed step of the chlorophyll degradation pathway where it removes Mg^{2+} from chlorophyll *a* to produce pheophytin *a*. Despite such importance, neither the three-dimensional nor the reaction mechanism has been elucidated until now. Combining the information from the tertiary protein structure, obtained by computational prediction as well as X-ray crystallography, and biochemical analysis, the reaction mechanism of the enzyme was proposed. There are two classes of metal dechelatase known to date – heme oxygenase and Mg-dechelatase. The former enzyme cleaves a porphyrin ring to extract Fe^{2+} in a totally different mechanism from that of Mg-dechelatase. This study proposes a novel reaction mechanism for a metal dechelatase enzyme based on structural analysis. Unexpectedly, my structural model suggests that deprotonated side chain of D34 may coordinate stably with Mg of chlorophyll. This coordination can be supposed to destabilize Mg-tetrapyrrole ring interaction, resulting in extraction of Mg from chlorophyll. This study will become a basis for further studies on this enzyme, such as those for substrate specificity, screening for inhibitors and evolutionary analysis. Furthermore, the tertiary and quaternary structure of CAO was also predicted computationally with special emphasis on the heterodimeric association between the two polypeptides of *Micromonas* CAO, leading to the prediction of a reaction mechanism for the enzyme. These structure-based enzyme studies will provide the clue to understand enzyme properties such as substrate specificity or regulatory mechanism of its activity, which facilitates understanding of plant life.

Chapter 1

Insights into the structure and function of the rate-limiting enzyme of chlorophyll degradation through analysis of a bacterial Mg-dechelatase homolog

1.1 Introduction

Chlorophyll is a pigment that plays a crucial role in absorption, transmission and transformation of light energy during photosynthesis. Land plants and green algae have chlorophyll *a* and *b* as their photosynthetic pigments. Chlorophyll biosynthesis must be finely regulated for efficient photosynthetic performance during the formation of photosystems at the greening stage and also during adaptation to various environmental conditions (R. Tanaka & Tanaka, 2007). Not only chlorophyll biosynthesis but also chlorophyll degradation needs to be regulated because the latter plays a crucial role in mobilizing resources from chloroplast to developing organs (Hörtensteiner & Kräutler, 2011). In addition, chlorophyll breakdown forms a key part of nitrogen recycling and is important in avoiding cellular photodamage. The major pathway and enzymes involved in chlorophyll degradation have been determined (Kuai et al., 2018). The first step of the degradation process is the extraction of magnesium (Mg) ion from chlorophyll *a* to form pheophytin *a* by Mg-dechelatase encoded by the *Stay-Green* (SGR) gene, which is also responsible for Mendel's green-cotyledon peas (Y. Sato et al., 2007). Furthermore, this reaction is strictly regulated to prevent the formation of detrimental photoreactive chlorophyll intermediates, thus serving as the rate-limiting step of the chlorophyll breakdown pathway (Hirashima et al., 2009; Shimoda et al., 2016). The *Arabidopsis* SGR-less mutants showed substantial retardation of chlorophyll degradation during senescence (Y. Chen et al., 2021), while overexpression of SGR resulted in the early promotion of chlorophyll degradation (Sakuraba et al., 2014; Shimoda et al., 2016).

Chlorophyll biosynthesis shares the common biosynthetic pathway with other tetrapyrroles such as heme, siroheme and phycobilins (Bryant et al., 2020). Chlorophyll is structurally similar to heme with regard to the tetrapyrrole macrocycle ring but contains a central magnesium ion instead of iron. Crystal structures of enzymes involved in heme biosynthesis were extensively investigated (Heinemann et al., 2008). In contrast, structural aspects of the enzymes related to the chlorophyll metabolic pathway remained unknown until recently, when the structures of Mg-chelatase and light-dependent protochlorophyllide oxidoreductase were reported (Adams et al., 2020; Zhang et al., 2019). Both of these enzymes catalyze the regulatory steps of chlorophyll biosynthesis. For chlorophyll breakdown, chlorophyll *b* must be converted to chlorophyll *a* by two successive reduction reactions because chlorophyll *b* derivatives are not catalyzed in the later steps of the chlorophyll degradation pathway (Hörtensteiner et al., 1995). In other words, the removal of Mg^{2+} from chlorophyll *b* leads to the formation of toxic pheophorbide *b* molecule which cannot be converted into another

metabolite and induces a cell-death phenotype (Shimoda et al., 2012). In the first step of chlorophyll *b* conversion, the enzyme chlorophyll *b* reductase (CBR) reduces the formyl group of chlorophyll *b* to produce 7-hydroxymethyl chlorophyll *a* (Meguro et al., 2011). In the final step, chlorophyll *a* is formed by the enzyme 7-hydroxymethyl chlorophyll *a* reductase (HCAR), the structure of which resembles an archaeal F_{420}-reducing [NiFe] hydrogenase (Wang & Liu, 2016). Furthermore, chlorophyll *a* is turned into a primary fluorescent Chl catabolite (pFCC) by four continuous steps **(Figure 1)**. Among the enzymes catalyzing these four steps of chlorophyll degradation, the crystal structure of only red chlorophyll catabolite reductase has been reported (Sugishima et al., 2010) while catalytic and structural properties of pheophytinase was investigated *in silico* (Guyer et al., 2018). However, the structure of Mg-dechelatase, catalyzing the committed step of the chlorophyll degradation pathway, is still unavailable.

Among eukaryotic photosynthesizing organisms, SGR is present only in green plants and Glaucophyta. On the other hand, SGR homologs are widely distributed in non-photosynthetic bacteria and Archaea (Obata et al., 2019). Despite the high sequence similarity between SGR and its homologs, their catalytic activity and substrate specificity vary considerably among species. According to Obata *et al.* (2019), a few bacterial SGR homologs, which are phylogenetically close to eukaryotic SGRs, show high Mg-dechelating activity and broad substrate specificity, suggesting the horizontal transfer of bacterial SGR homolog to photosynthetic eukaryotes.

SGR not only catalyzes the committed step of chlorophyll degradation, but also removes a metal ion from an organic compound in a chemically rare event. Therefore, extensive study on SGR will provide insights into a new type of enzymatic reaction mechanism. Besides, the substrate condition of SGR is different from conventional enzymes as it remains bound to the surface of chlorophyll-protein complexes instead of commonly free and small molecules that can fit into the active site of an enzyme. Thus, the active site of SGR needs to be exposed to the surface of the enzyme.

Preparation and purification of recombinant SGR with high yields is essential for performing detailed biochemical analyses of the enzyme. However, expression of recombinant SGR in large amounts without compromising the activity remains challenging and invariably results in insolubility. Even if it becomes soluble, the expressed enzyme precipitates on increasing the concentration making purification and further enzymatic analyses difficult. Though SGR from *Chlamydomonas reinhardtii* had been expressed in *E. coli* (Matsuda et al., 2016), SGRs from Streptophytes were not expressed before. The first recombinant plant SGR was obtained using a cell-free protein expression system (Shimoda et al., 2016). Mutant *Arabidopsis* SGR-Like and rice SGR were also obtained using the same *in vitro* expression system (Shin et al., 2020; Xie et al., 2019). Later, using an expression vector with a tag containing an

unstructured and hyper-acidic module (Sangawa et al., 2013), *Arabidopsis* SGR and bacterial SGR homolog expression in *E. coli* were reported (Obata et al., 2019). However, the expression levels for both the proteins did not increase substantially. In this study, I report the first successful overexpression of a soluble and highly active SGR homolog from *Anaerolineae* bacterium SM23_63 in *E. coli* using its general expression vector.

The evolutionary aspects of SGR coupled with its functional importance in plant senescence make it a molecule of utmost significance, the structural information of which emerges out to be absolutely vital. Therefore, I elucidated the structural characteristics of the protein through biochemical experiments and subsequently justified my observations using computational approaches.

1.2 Materials and methods

1.2.1 Multiple sequence alignment and phylogenetic analysis of green plant SGRs and bacterial SGR homologs

Protein sequences of SGRs and homologs were downloaded from the Phytozome v12.1 (Goodstein et al., 2012) (https://phytozome.jgi.doe.gov/pz/portal.html) and NCBI databases (https://www.ncbi.nlm.nih.gov/protein/). A total of 40 sequences were retrieved including 14 green plant SGRs, 3 archaeal and 23 bacterial SGR homologs representing a wide array of species with diverse homology **(Table 1)**. The protein sequences of the green plant SGRs and their homologs were aligned using Clustal Omega with the default settings for multiple sequence alignment (MSA) (Sievers & Higgins, 2018). Identification and marking of the conserved residues in the MSA were performed in Jalview v2.11.1.4 (Waterhouse et al., 2009). The maximum likelihood phylogenetic tree was inferred using IQ-TREE v1.6.12 with 1000 bootstrap replicates in the ultrafast mode (Hoang et al., 2018; Trifinopoulos et al., 2016). The best-fitting amino acid substitution model for the dataset – WAG+I+G4 was applied automatically in the IQ-TREE server for phylogeny construction (Kalyaanamoorthy et al., 2017). iTOL v6 was used for both visualization as well as generation of the figure (Letunic & Bork, 2019).

1.2.2 Cloning of bacterial SGR homolog

Anaerolineae bacterium SM23_63 SGR homolog (AbSGR-h) encoded by KPK94580 with optimized codon usage for *E. coli* was artificially synthesized as previously reported (Obata et al., 2019). AbSGR-h was amplified from the artificially synthesized DNA using the primer sets shown in **Table 2**. Amplified DNA fragments are cloned into pET 30a (+) vectors

(Novagen) containing a histidine-tag at the C terminus using the *Nde*I and *Xho*I sites through an in-fusion cloning system (Clontech). Several point mutations were introduced by PCR using primers as shown in **Table 2**.

1.2.3 Expression and detection of recombinant proteins

The constructed plasmids for protein expression were introduced into *E. coli* BL21 (DE3). *E. coli* was grown and recombinant protein was expressed in an auto-induction medium (6 g Na_2HPO_4, 3 g KH_2PO_4, 20 g tryptone, 5 g yeast extract, 5 g NaCl, 6 mL glycerol, 0.5 g glucose, 2 g lactose, 100 mg kanamycin in 1 L) at 37ºC for 16 h with 120 rpm shaking (Studier, 2005). After incubation, 1 mL of the cell was harvested by centrifuge at 20,000 g for 1 min. The pellet was suspended with 500 µL of BugBuster Protein Extraction Reagent (Millipore) with 0.1% Benzonase nuclease (Millipore). Crude supernatant was obtained by centrifugation at 20,000 g for 1 min. The whole cell lysate and crude supernatant were mixed with the same volume of sample buffer containing 125 mM Tris-HCl, pH 6.8, 4% (w/v) SDS, 10% (w/v) sucrose, and 5% (v/v) 2-mercaptoethanol. Mixtures were incubated at 95ºC for 1 min and 10 µL of the mixture was subjected to SDS-PAGE as previously reported (Obata et al., 2019). Proteins were visualized by staining with Coomassie Brilliant Blue.

1.2.4 Recombinant protein purification

After induction of the recombinant protein, 100 mL of culture cells were harvested by centrifugation at 7,000 g for 5 min. The harvested cells were resuspended in buffer A (20 mM Na-phosphate pH 7.4, 100 mM NaCl, 20 mM imidazole) and disrupted by sonication (Branson Sonifier SFX250: output 8, duty cycle 20%) for 6 min in an ice bath. After sonication, dodecyl β-maltoside (βDM) was added to the final concentration of 0.05% (w/v) and incubated for 5 min at 25ºC. The cleared supernatant of cell lysate was obtained by centrifugation at 20,000 g for 10 min and then loaded onto a 5 mL HisTrap HP column (Cytiva) equilibrated with buffer A containing 0.05% βDM using an ÄKTAprime plus system (Cytiva). The recombinant proteins were eluted by buffer B (20 mM Na-phosphate pH 7.4, 100 mM NaCl, 500 mM imidazole, 0.05% βDM). To examine the purity of the protein, elution was mixed with the same volume of the sample buffer and 10 µL of the mixture was used for SDS-PAGE. Purified protein was analyzed by size exclusion chromatography using Sephacryl S-400R (Cytiva) equilibrated with buffer C (20 mM Na-phosphate pH 7.4, 100 mM NaCl, 0.05% βDM). The protein elution profile was monitored by absorbance at 280 nm. The molecular weight of AbSGR-h was evaluated by comparison to protein standards (Gel Filtration Calibration Kit LMW, Cytiva).

1.2.5 Enzymatic assay

The cell lysate prepared by suspending with BugBuster Protein Extraction Reagent as described above was used to perform enzymatic assay. One µL of chlorophyll *a* dissolved in DMSO (1 nmol μL^{-1}) was mixed with 50 µL of cell lysate and incubated for 10 min at 25^{o}C. The reaction was stopped by adding 200 µL of acetone. After centrifuging at 20,000 g for 10 min, the supernatant was subjected to HPLC equipped with a fluorescence detector (RF- 20A, Shimadzu). The pigments were separated through a Symmetry C8 column (4.6 × 150 mm, Waters) with an eluent [methanol : acetonitrile : acetone (1:3:1 v:v:v)] at a flow rate of 1 mL min^{-1} at 40^{o}C. The elution profiles were monitored at fluorescence excitation/emission wavelengths of 410/680 nm.

1.2.6 Tertiary structure prediction and validation

De novo protein modelling of SGR protein from *Arabidopsis thaliana* (AtSGR1; Accession ID: AT4G22920.1) and an SGR homolog from *Anaerolineae* bacterium SM23_63 (AbSGR-h; Accession ID: KPK94580.1) were performed using trRosetta, which builds the tertiary structure based on direct energy minimizations with a restrained Rosetta (Yang et al., 2020). The restraints include inter-residue distance and orientation distributions, predicted by a deep residual neural network. Out of the five models predicted for each protein, only the model with the best confidence as judged by the template modeling score (TM-score), developed by Xu and Zhang (2010), was selected for further evaluation. TM-score can be used as an approximate but quantitative criterion for protein topology classification (J. Xu & Zhang, 2010). The stereochemical quality of the predicted models after energy minimization in GROMACS 2018.1 (Abraham et al., 2015) was assessed through Verify3D (Lüthy et al., 1992), PROCHECK (Laskowski et al., 1993) and ERRAT (Colovos & Yeates, 1993) in the Structural Analysis and Verification Server (SAVES) v. 5.0 server (https://servicesn.mbi.ucla.edu/SAVES/). The ProSA-web server was also used to validate the 3D models (Wiederstein & Sippl, 2007). ProSA evaluates model quality in terms of a 'Z' score and provides a 'Z' score plot, where the predicted model is placed within an experimental NMR and X-ray structure of equal residue length. PyMOL v. 2 was used for graphic modifications, visualization, and preparing final illustrations (Delano, 2002). Protein cavity detection was implemented in the CavityPlus web server (http://www.pkumdl.cn/cavityplus), which utilizes the 3D structural information as input to detect potential binding sites on the surface of a given protein structure (Y. Xu et al., 2018). Metal ion-binding site in the protein was predicted using the MIB server (Lin et al., 2016). The ConSurf server (http://consurf.tau.ac.il/) was used to determine the functional regions in the modeled protein. It is a tool which analyzes the evolutionary dynamics of amino acid substitutions among homologous sequences and maps them onto the structure of the query

protein (Ashkenazy et al., 2016). Additionally, ConSeq v. 1.1, integrated in the ConSurf server, was used to identify the functionally and structurally important residues in the amino acid sequence of AbSGR-h (Berezin et al., 2004). Furthermore, five mutated monomeric structures of AbSGR-h (T29A, H32A, D34N, D62N and R26D+D114R, referred as the double mutant) were generated in PyMOL using the Mutagenesis wizard by selecting the most probable rotamers for each amino acid substitution.

1.2.7 Molecular dynamics simulation

Molecular dynamics simulation was performed on the predicted structure of the wild-type and five mutated monomers of the AbSGR-h protein using the GROMACS simulation package 2018.1 (Abraham et al., 2015). OPLS-AA/L all-atom force field was used to model the intramolecular protein interactions and the intermolecular interactions between the protein and solvent molecules (Kaminski et al., 2001). Initially the energy of each system was minimized using 500 steps of the steepest descent algorithm followed by 20,000 steps of the Polak-Ribiere conjugate gradient method to remove the strain in the initial structure. The relaxed structure was immersed in a cubic box of extended simple point charge (SPC/E) water molecules with periodic boundary conditions in all directions. A minimum distance between the protein and the wall of the cell was set to 1 nm. Prior to energy minimization with periodic boundary conditions, each solvated system was neutralized by the addition of sodium and chloride ions.

MD simulation consists of equilibration and production phases. In the first stage of equilibration, the solutes (protein, counter ions) were fixed, and the solvent (water molecules) was equilibrated for 100 ps of MD at 200 K using an integral time step of 0.001 ps. During the equilibration phase, velocity was assigned to the atoms using Maxwell distribution. The system was coupled to the heat bath and heated to 300 K in a short run of 100 ps (0.001 ps time step) in which the system was allowed to relax in the new condition. This was followed by another short simulation of 100 ps with pressure coupling at 1 atm. Finally, the production phase of MD simulation was run keeping the temperature, pressure and number of molecules of the ensemble invariant. Production phase was continued up to 200 ns using 0.002 ps time step for each of the wild-type and mutated AbSGR-h proteins. The average structures for each monomer were obtained using the 200 ns trajectory of the MD production run. Subsequent analyses that include RMSD (Root Mean Square Deviation), RMSF (Root Mean Square Fluctuation) and radius of gyration (*Rg*) were performed using different programs of the GROMACS package on the 200 ns trajectory of the production run. The secondary structure content of the wild-type and mutant proteins along the production phase trajectory was computed using DSSP (Kabsch & Sander, 1983). The webPSN v. 2 server (Felline et al.,

2020) was utilized for the analysis of network of interacting amino acids wherein the average structure of the wild-type and mutant proteins were considered.

1.2.8 Protein-ligand docking study

The structure of chlorophyll *a* was retrieved from the KEGG LIGAND database (https://www.genome.jp/kegg/ligand.html) and was subjected to geometry optimization under the semi-empirical method in HyperChemTM 8.0.8 molecular modeling software (Hypercube Inc., Gainesville, FL, USA). 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 considering the average structures of wild-type and mutant AbSGR-h proteins obtained from respective MD simulations of 200 ns (Trott & Olson, 2010). 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). A grid box of 30 Å × 30 Å × 30 Å with grid spacing of 1 Å was set for docking. Hydrophilic and hydrophobic interactions in the docked conformations were visualized using PyMOL.

1.3 Results

1.3.1 Sequence comparison and phylogenetic analysis

Amino acid variability among 40 SGR and its homologs was observed from the analysis of the multiple sequence alignment that included 14 SGR sequences from green plants, 3 archaeal and 23 bacterial sequences as SGR homologs. **Figure 2** shows the alignment of the conserved region which depicts high sequence similarity between SGRs of photosynthetic eukaryotes and SGR homologs of archaea and bacteria. A total of 26 amino acid residues with conservation score above 90% were identified from the multiple sequence alignment. I also observed the presence of a motif similar to an incomplete metal-ion-dependent adhesion site (MIDAS) motif at 31-36 (T-H-S-D-S-T) for the SGR homolog from *Anaerolineae* bacterium SM23_63 shown in boldface in the multiple sequence alignment. A complete MIDAS motif usually consists of a consensus sequence (D-x-S-x-S… T… D) while an imperfect one is characterized either by the presence of conserved region 1 (D-x-S-x-S), without one or both of T and D, or those with conservative changes in region 1 with and without conservation of T and D (Whittaker & Hynes, 2002). Although the last T remains conserved in SGRs considered here, but for typical MIDAS motifs the T comes many amino acids after the conserved region 1. Structural studies of proteins containing this motif indicate

that an imperfect MIDAS motif is also capable of binding metal ions. Although the presence of this motif had been reported in prokaryotic and plant chelatases (Axelsson et al., 2006; Whittaker & Hynes, 2002), the existence and role of the same in metal dechelatase are yet to be determined.

A Maximum-likelihood phylogenetic tree revealed a distinct clading pattern of SGR proteins across different life forms – ranging from archaea to green plants **(Figure 3)**. Though SGR proteins possess significant sequence similarity in their respective domains, SGRs did not intermix with its homologs in the phylogeny besides a clade including *Anaerolineae*.

1.3.2 Protein expression and purification

Among all other bacterial SGR homologs studied here, SGR of *Anaerolineae* bacterium SM23_63 (AbSGR-h) is phylogenetically closer to that of the green plants. High expression level of the gene can be observed in *E. coli*. Interestingly, AbSGR-h shows much higher Mg-dechelating activity than *Arabidopsis* SGRs (Obata et al., 2019). Furthermore, the genome of *Anaerolineae* hosts a single gene of SGR unlike genomes of other species that accommodate several homologous SGR genes.

The molecular weight and solubility of the expressed AbSGR-h protein was checked using SDS-PAGE **(Figure 4A)**. In the lysate of the AbSGR-h-expressing cells, the protein appeared as a single and prominent band corresponding to a molecular size of approximately 18 kDa. The band is absent in the cell lysate prepared from *E. coli* having empty vector indicating recombinant AbSGR-h to have been successfully expressed in *E. coli*. The soluble nature of AbSGR-h was confirmed from the presence of a clear band in the crude supernatant fraction. The protein was purified using a nickel column where the eluate in fractions 5–7 showed the maximum concentrations of the purified protein **(Figure 4B)**. The CBB-stained gel showed that the purity of AbSGR-h in elution is high.

1.3.3 Analysis of SGR mutants

Several mutants of AbSGR-h were constructed in this study to understand the effect of mutations on the structure and function of the protein. The mutations were carried out at conserved amino acid positions as found out from the multiple sequence alignment in **Figure 2**. Altogether, eight charged amino acid residues (Arg26, His32, Asp34, Arg61, Asp62, Glu63, Arg95 and Asp114) and three threonine residues (Thr29, Thr31 and Thr36) were mutated. One conserved aromatic amino acid (Tyr28) was also changed. According to the modeled AbSGR-h structure (discussed later), an intramolecular electrostatic interaction between Arg26 and Asp114 was predicted, so I included an additional double mutation

(R26D+D114R) by swapping their positions. Furthermore, to ensure that no steric repulsion occurs after the swapping, I substituted a non-conserved arginine at the 115th position with alanine, thereby, creating a triple mutant (R26D+D114R+R115A).

After expression of SGR mutants in *E. coli*, solubility of the wild-type and mutant proteins was analyzed by running the soluble fraction in SDS-PAGE **(Figure 5)**. Expressed protein bands were observed in the crude cell lysate derived from *E. coli* having mutant constructs. Since solubility is intrinsically linked to the structural integrity of a protein, mutation in the sequence might disrupt the interaction network between the amino acids resulting in strong destabilization of the structure ultimately leading to the loss of solubility. The solubility of the SGR mutants – R26D, Y28A, T29A and D114N was significantly decreased when compared with the wild-type. The same was observed with the double mutant (R26D+D114R) with exchanged residue position and the triple mutant (R26D+D114R+R115A) with an additional mutation to avoid positively charged arginine in consecutive positions. Thus, I suggest that these residues play a key role in maintaining the conformation of the protein.

1.3.4 Enzymatic activity

Enzymatic activity of AbSGR-h results in the removal of Mg^{2+} from chlorophyll *a* and produces pheophytin *a*. To assess the effect of the mutations on enzymatic activity of the wild-type and mutant proteins, cell lysates were incubated with chlorophyll *a*. Activity levels were evaluated based on the amounts of the product, pheophytin *a*, on HPLC profiles **(Figure 6)**. It may be emphasized that appearance of the pheophytin *a* peak is associated with the concomitant disappearance of the substrate chlorophyll *a*. Mutants R61A, E63Q and R95A along with the wild-type AbSGR-h showed high activity (major peak corresponding to pheophytin *a*), suggesting that these mutations does not affect the Mg-dechelating activity of the protein at all. AbSGR-h mutants T31A and T36A exhibited moderate activity whereas T29A and H32A mutation made the protein weakly active. Rest of the mutations *i.e.*, R26D, Y28A, D34N, D62N, D114N as well as double and triple mutations rendered the protein inactive. Both D34N and D62N showed no activity despite being soluble, implying their potential role in catalyzing the Mg-dechelatase reaction. The inactive nature of the remaining mutations *i.e.*, R26D, Y28A, D114N, R26D + D114R, and R26D + D114R + R115A, can be attributed to their insolubility due to disruption of the protein structure.

1.3.5 Complex formation of wild-type and mutant AbSGR-h

Some of the chlorophyll metabolizing enzymes form oligomeric complexes (Kunugi et al., 2013; Wang & Liu, 2016). Under the assumption that SGR may also form an oligomeric

complex, size exclusion chromatography was performed. The apparent molecular weight of wild-type AbSGR-h was evaluated by comparison to globular protein standards with known molecular weights. Comparison of the single, major AbSGR-h peak with the calibration curve yielded a molecular weight of approximately 110 kDa **(Figure 7A)**. The apparent molecular weight of AbSGR-h monomer obtained from SDS-PAGE analysis revealed the apparent molecular weight of AbSGR-h monomer to be 18 kDa, implying the eluted wild-type protein to possibly exist as a hexameric complex.

Further, in order to determine the effect of amino acid substitutions on complex forming ability, size exclusion chromatography was carried out for individual soluble AbSGR-h mutants **(Figure 7B)**. Along with the wild-type protein, similar major peaks were observed for H32A, D34N, T36A and D62N mutants, suggesting presence of hexameric complexes in these mutants. Absence of a major peak at that position in case of the other mutants like T31A, R61A, E63Q and R95A indicates loss of the hexameric form upon mutation. However, the major peak positions differed slightly among proteins, probably because the system used for the analysis was not very stable and the flow rate of the solution could not be completely regulated. Interestingly, though R61A, E63Q and R95A mutants lost their hexameric form, they remain highly active suggesting that the formation of multimeric complex is not indispensable for the catalytic activity of SGR. However, the reason behind the multimeric conformation and its implication on the function of SGR remains elusive.

1.3.6 Predicted 3D structure of SGR and its homolog

Spatial location of the amino acid residues in the predicted protein tertiary structure might be insightful for a better understanding of the functioning of the dechelatase. A *de novo* approach for protein modelling was adopted due to lack of any experimentally determined structure for SGR or its homologs. I, therefore, modeled the structure of *Anaerolineae* SGR homolog (AbSGR-h) using trRosetta **(Figure 8A)**. The method is based on a deep residual-convolution network that is trained on native proteins to predict inter-residue distance and orientation. Among the five predicted models of AbSGR-h, the model with best confidence (estimated TM-score = 0.870) was selected for further analysis. Additionally, I also derived the tertiary structure of *Arabidopsis thaliana* SGR (AtSGR1) **(Figure 8B)** using the same algorithm. The structure of AtSGR1 was modeled considering the amino acid sequence of its SGR domain only, as the entire sequence appears be too long to obtain a high confidence score. The TM-score for the modeled AtSGR1 domain was 0.840 whereas the same for the whole protein was 0.476.

Different quality evaluation programs such as PROCHECK, ERRAT and Verify 3D available online on the SAVES server, were used to assess the quality of the energy

minimized modeled structures of AbSGR-h and AtSGR1. Ramachandran plots for both the protein structures demonstrated that the predicted models follow all the stereochemical properties with favourable phi (ϕ) and psi (ψ) values **(Figure 8C and 8D)**. Furthermore, protein quality assessment by ERRAT and Verify3D confirmed that the structures are highly accurate **(Table 3)**. The ProSA analysis of AbSGR-h and AtSGR1 revealed a Z-score of -5.48 and -4.3, respectively, accommodating the modeled structures in the NMR zone and thus confirming their reliability **(Figure 9)**.

The potential binding site on the surface of AbSGR-h protein was determined by CavityPlus. The residues constituting the predicted cavity are: T31, H32, S33, D34, S35, T36, E38, L39, F40, W55, R58, F59, M60, R61, D62 and R95 **(Figure 10)**. Majority of the residues comprising the binding cavity were found to be conserved among SGRs. Additionally, I used ConSurf to find the evolutionary conservation score of each amino acid residue in AbSGR-h, where a score of less than 3 and more than 7 indicate variable and conserved residues, respectively **(Figure 11A)**. The result of ConSeq analysis, which shows the degree of conservation as well as the structurally and functionally important residues along the sequence of AbSGR-h, is also depicted in **Figure 11B**.

1.3.7 Molecular dynamics simulation analysis

Molecular dynamics (MD) simulations were performed on wild-type AbSGR-h and five mutants (T29A, H32A, D34N, D62N and the R26D+D114R double mutant) for 200 ns to assess the structural stability and conformational dynamics of the predicted protein structure in its wild-type and mutated form. The time-dependent changes of RMSD of the backbone atoms for each protein was estimated considering the respective input structure for MD production run as reference. The RMSD plot of AbSGR-h and its mutants reflected convergence of the simulation, indicating the overall structural stability of all the monomers **(Figure 12)**. The RMSF for individual residues for wild-type AbSGR-h and mutated monomers were computed to infer the residue specific flexibility, taking into account the respective input structure for MD production run as reference **(Figure 13)**. It is evident from the RMSF plot that the D34N mutant appear to be more rigid when compared to the wild-type and other mutated monomers. However, few residues pertaining to the N-terminus of the D34N and H32A mutant exhibited higher flexibility than the native protein. On the other hand, though the N-terminal and C-terminal ends of the T29A mutant showed more flexibility than the wild-type, both the wild-type and T29A monomers displayed similar fluctuation patterns from the 29^{th} residue onwards. I also calculated the radius of gyration (*Rg*) which is a measure of the compactness of a protein. It is evident from the invariant *Rg* values **(Figure 14)** that none of the mutations grossly disrupt the monomeric structure of the protein. Further information on the structural flexibility for both the wild-type and mutant

protein models are offered by the analysis of time-dependent secondary structure fluctuations **(Figure 15)**, calculated using the DSSP algorithm in GROMACS. It is interesting to note that once more the mutant D34N shows variation in the time evolution of the secondary structural elements. Out of the seven β-strands, the one at the N-terminal end gets disrupted after 100 ns of simulation. Amino-acid network analysis revealed R26 to be a hub residue interacting with six other residues, a reason that can be attributed to why the R26D mutant became insoluble. It is to be noted that even if the ionic interaction remains undisturbed in the double mutant (where the residue pair has been swapped), the replacement of R26 with D (an amino acid with shorter side chain) renders it incapable of being a hub residue thus disrupting all interactions with other amino acids, leading to insolubility. Likewise, the network analysis also revealed Y28 and T29 to interact with two and three other residues of AbSGR-h respectively, implying their importance in imparting the native structure of the protein. Mutation (Y28A or T29A) disturbs these interactions that in turn disrupts the native structure which probably is the cause of the observed insolubility.

1.3.8 Molecular docking analysis

For docking analysis, energy-optimized free chlorophyll *a* was used as ligand of SGR. The actual substrate of SGR in the chloroplast is chlorophyll *a* bound to chlorophyll-protein complexes that are embedded in the thylakoid membrane (Nelson & Yocum, 2006). Docking studies of chlorophyll *a* with the wild-type and mutated AbSGR-h proteins were carried out using two settings. In the first instance, the grid box was set large enough to cover the entire protein structure while in the second setting, I opted for a specific grid box with a size of 30 Å × 30 Å × 30 Å that was centered around the active site of the protein. I presumed D34 and D62 residues as the key components of the active site since biochemical experiments revealed that mutations of these two amino acids rendered the protein catalytically inactive despite being soluble. Furthermore, protein cavity prediction analysis by CavityPlus implicated the importance of these two aspartate residues. In addition, D34 was found to be a part of an incomplete MIDAS motif as observed from the multiple sequence alignment. The docked conformation of wild-type AbSGR-h revealed interaction of the ligand with H32 and D34 **(Figure 16A)**. However, despite being in the vicinity of the ligand, D62 did not show any interaction with the ligand. In both docking settings, chlorophyll *a* did not bind to the active site of the mutants T29A, D34N, D62N and double mutant. For the H32A mutant, although the ligand was found to interact with D34, the binding pose was different from that of the wild-type complex **(Figure 16B)**. Interestingly, the enzymatic assay showed that H32A mutation made the protein weakly active, a fact that can be attributed to the altered binding state of the ligand with this mutated monomer. Furthermore, the result of docking analysis with the D62N mutant is consistent with the corresponding observation of the biochemical study **(Figure 6)**, which revealed that this residue is essential for the catalytic activity of the

enzyme. The involvement of D62 in binding and/or activity is shown from the docked D62N-chlorophyll *a* complex **(Figure 16C)**, wherein the ligand orientation has changed from its wild-type counterpart **(Figure 16A)**.

1.4 Discussion

1.4.1 Computational prediction of AbSGR-h structure

Despite technological advancement and sincere attempts, there remains a huge gap between the number of known sequences and experimentally derived structures available in the Protein Data Bank (PDB), highlighting the difficulties of structure elucidation by experimental methods like X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy (Seffernick & Lindert, 2020). In recent times, protein tertiary structure prediction by deep learning-based methods made it possible to generate complete and accurate models of proteins that lack homologs in PDB. Our study provides the first three-dimensional structure of SGR, predicted by deep learning method using trRosetta. The quality and stability of the predicted AbSGR-h structure was probed by different protein quality evaluation programs and molecular dynamics simulation was also carried out with the model. I also compared the predicted structure with that obtained from the very recently developed RoseTTAFold tool (Baek et al., 2021) and found the RMSD (based on Cα-atoms) to be 1.2 Å **(Figure 17A)**.

1.4.2 Biochemical analysis of AbSGR-h

The physiological relevance of bacterial SGR homologs remains unknown. Therefore, common and distinct characters between plant SGRs and bacterial SGR homologs have not been determined yet. Nevertheless, I considered working on a bacterial SGR homolog as a representative of the chlorophyll degrading enzyme of green plants for several reasons. The *Anaerolineae* SGR homolog (AbSGR-h) shares substantial functional similarity with the *Arabidopsis thaliana* SGR-Like (AtSGRL) protein, as evident from Obata et al., 2019, where both proteins catalyze Mg dechelation with similar efficacy. The high sequence similarity among plant and bacterial SGR homologs, as seen in **Figure 2**, indicates that the structurally and functionally important residues determined for a bacterial SGR are of equal importance to that of the green plant SGRs. Preliminary analysis with mutations in Asp107 and Asp132 residues of AtSGRL exhibited loss of catalytic activity. Interestingly, Asp107 and Asp132 of AtSGRL protein correspond to Asp34 and Asp62 of AbSGR-h, respectively. Mutation of these aspartates in both the organisms exhibited loss of activity, indicating their catalytic role in the Mg dechelation reaction. Additionally, phylogenetic analysis revealed that AbSGR-h

is close to plant SGRs (**Figure 3**), suggesting that they are functionally related. Furthermore, our study showed that the predicted structures of AbSGR-h and *Arabidopsis* SGR1 are highly similar (**Figure 8**). Therefore, it can be stated that the reaction mechanism of plant SGRs and bacterial SGR homologs in terms of Mg extraction from chlorophyll may be similar, despite differences in their substrate specificity (Obata et al., 2019).

SDS-PAGE analysis of the expressed AbSGR-h protein revealed the molecular weight of the monomer to be ~18 kDa and size exclusion chromatography indicated that the recombinant protein may exist as a hexamer. Several AbSGR-h mutants were created by substituting conserved amino acid residues to determine their structural and functional significance **(Table 4)**. Single mutations at R26, Y28, T29 and D114 made the protein insoluble, suggesting them to play an important role in the structural maintenance of the protein **(Figure 17B)**. These amino acids were found relatively close to each other spatially and remained buried in the predicted structure. In the model, R26 and D114 form an ionic bond that is lost when these amino acids were exchanged. Supporting data from amino acid network analysis of the protein structure revealed R26 to be a hub residue, that is engaged in different kinds of interaction with six other residues including D114. Exchange of R26 and D114 leads to the disruption of the other five interactions that appear to be essential for the ionic interaction to occur. Thus, for all the three mutations, R26D, D114N and R26D-D114R, disruption of the ionic bond in the native protein structure leaves the molecule insoluble. Similarly, network analysis showed that Y28 and T29 interact with two and three other residues respectively in the wild-type structure, all of which get disrupted upon mutation. This interruption probably leads to insolubility of the mutant forms, as evident from the biochemical analysis. Mutations of two specific aspartates – D34N and D62N made the protein inactive without affecting its solubility and ability to form multimeric complex **(Figure 17B)**. Incidentally, D34 is present within an incomplete MIDAS motif, the functional role of which is to dechelate ions. Docking of the Mg^{2+} to the predicted wild-type structure using the MIB server, displayed interaction of the ion with D62 suggesting potential catalytic role of D62 in the Mg-dechelatase enzyme. Mutation in T31, R61, E63, and R95 resulted in the destruction of the multimeric complex.

1.4.3 Molecular dynamics and docking simulation

Conformational dynamics at the monomer level was analyzed through molecular dynamics simulations of 200 ns carried out for the wild-type and five mutant proteins. Time-dependent changes of RMSD and invariant Rg values revealed the structural stability of the wild-type and mutant protein forms. Interestingly, the D34N mutant showed overall less flexibility than the wild-type and distortion of an N-terminal β-strand in the time evolution of secondary

structural elements when analyzed by RMSF and DSSP, respectively. Molecular docking analysis displayed interaction of D34 with chlorophyll *a*, implying its importance once more in the catalytic activity of the enzyme. It is to be noted that the side chain of D34 is exposed to the surface of the predicted structure and residues surrounding D34 form a hydrophobic patch, an environment appropriate for interaction of chlorophyll *a* with the protein **(Figure 17C)**.

1.4.4 Probable reaction mechanism of the enzyme

Central Mg^{2+} of chlorophyll *a* is held by two N atoms of the tetrapyrrole structure. SGR catalyzes Mg extraction from chlorophyll *a*, resulting in incorporation of two protons into the chlorin ring to produce pheophytin *a*. Since the catalytic mechanism of SGR remains unknown, two hypotheses can be proposed. The first one is similar to the reactions observed under acidic condition. The electrophilic attack of protons to the core N atoms of chlorophyll remove Mg^{2+}, leading to the formation of pheophytin. As acidic amino acid residues can serve as proton donors, D34 can be considered as a potential residue involved in the dechelation reaction. On the other hand, formation of coordinate bond between an electronegative atom and central Mg^{2+} of chlorophyll destabilizes the Mg-N (pyrrole) interaction. Once this complex is formed, the Mg ion may be readily replaced with protons. Considering the optimum pH of SGR to be neutral (Matsuda et al., 2016) and the D34 residue to be present at the surface of the predicted structure, its de-protonated side chain can serve as a candidate to provide electronegative O for coordination with Mg^{2+}.

In conclusion, by combining biochemical analysis and structural prediction of the SGR homologue from *Anaerolineae*, I provide the first structural insights into the SGR protein family. It will serve as a basis for further investigation of its reaction mechanism, functional analysis and other aspects such as inhibitor screening and/or evolutionary studies.

1.5 Tables and Figures

Table 1: Details of SGRs and its homologs used for multiple sequence alignment and phylogenetic tree construction

Organism	Accession ID	Database	Taxon
Arabidopsis thaliana	AT4G22920.1	Phytozome	Green plant
Arabidopsis thaliana	AT4G11910.1	Phytozome	Green plant
Arabidopsis thaliana	AT1G44000.1	Phytozome	Green plant
Oryza sativa	LOC_Os09g36200.1	Phytozome	Green plant
Oryza sativa	LOC_Os04g59610.1	Phytozome	Green plant
Physcomitrella patens	Pp3c17_23030V3.1	Phytozome	Green plant
Physcomitrella patens	Pp3c3_28140V3.1	Phytozome	Green plant
Physcomitrella patens	Pp3c20_6290V3.1	Phytozome	Green plant
Physcomitrella patens	Pp3c8_17510V3.1	Phytozome	Green plant
Chlamydomonas reinhardtii	Cre12.g487500.t1.1	Phytozome	Green plant
Coccomyxa subellipsoidea **C-169**	XP_005650112.1	NCBI	Green plant
Micromonas pusilla **CCMP1545**	XP_003060977.1	NCBI	Green plant
Ostreococcus tauri	XP_022840207.1	NCBI	Green plant
Chlorella variabilis	XP_005850408.1	NCBI	Green plant
miscellaneous Crenarchaeota group-1 archaeon SG8-32-1	KON32503.1	NCBI	Archaea
Candidatus Thorarchaeota **archaeon SMTZ-45**	KXH70069.1	NCBI	Archaea
miscellaneous Crenarchaeota group-6 archaeon AD8-1	KON32872.1	NCBI	Archaea
Anaerolineae **bacterium SM23_63**	KPK94580.1	NCBI	Bacteria
Alicyclobacillus ferrooxydans	WP_054971308.1	NCBI	Bacteria
Clostridium novyi	WP_039245490.1	NCBI	Bacteria
Bacillus aquimaris	WP_071617569.1	NCBI	Bacteria
Cytobacillus oceanisediminis	WP_019379967.1	NCBI	Bacteria
Sporosarcina globispora	WP_053435952.1	NCBI	Bacteria
Mesobacillus selenatarsenatis	WP_041965262.1	NCBI	Bacteria
Mycobacteroides abscessus **subsp.** ***abscessus***	SHT39353.1	NCBI	Bacteria
Clostridium botulinum **sp. CDC54075**	WP_024931542.1	NCBI	Bacteria
Bacillus wudalianchiensis	WP_065409936.1	NCBI	Bacteria
Gottschalkia acidurici	WP_041701757.1	NCBI	Bacteria
Bacillus megaterium	WP_013082933.1	NCBI	Bacteria

Clostridium sporogenes	WP_058008982.1	NCBI	Bacteria
Bacillus massiliogorillae	WP_042348482.1	NCBI	Bacteria
Anoxybacillus tepidamans	WP_027408343.1	NCBI	Bacteria
Thalassobacillus cyri	WP_093042251.1	NCBI	Bacteria
Anaerotignum neopropionicum	WP_066089690.1	NCBI	Bacteria
Intestinibacter bartlettii	WP_082421703.1	NCBI	Bacteria
Mesobacillus campisalis	WP_046522531.1	NCBI	Bacteria
Peribacillus psychrosaccharolyticus	WP_040373082.1	NCBI	Bacteria
Peptoniphilus phoceensis	WP_062552496.1	NCBI	Bacteria
Hungatella xylanolytica	WP_104438939.1	NCBI	Bacteria
Paraclostridium bifermentans	WP_021432481.1	NCBI	Bacteria

Table 2: Primer sequences for cloning

	Forward (5' → 3')	Reverse (5' → 3')
WT	AAGGAGATATACATATGGATCATCT GAAACCGGAG	GGTGGTGGTGCTCGATTTTGTAATC GCGCAGAATG
R26D	TTATCCGGATTGCTATACCCTGACC CACAG	TATAGCAATCCGGATAACGCGGAC CATCGT
Y28A	TCCGCGCTGCGCCACCCTGACCCAC AGCGA	GGGTCAGGGTGGCGCAGCGCGGAT AACGCG
T29A	GCGCTGCTATGCCCTGACCCACAGC GACAG	TGGGTCAGGCCATAGCAGCGCGGA TAACGC
T31A	CTATACCCTGGCCCACAGCGACAGC ACCGG	TGTCGCTGTGGCCCAGGGTATAGC AGCGCG
H32A	CCTGACCGCCAGCGACAGCACCGGT GAACT	TGTCGCTGCCGGTCAGGGTATAGC AGCGCG
D34N	CCACAGCAATAGCACCGGTGAACTG TTTCT	CGGTGCTATTGCTGTGGGTCAGGG TATAGC
T36A	CAGCGACAGCGCCGGTGAACTGTTT CTGAC	GTTCACCGCCGCTGTCGCTGTGGGT CAGGG
R61A	CTTTATGGCCGATGAAGTGCTGGCC GTGTG	CTTCATCGCCCATAAAGCGGGTGT ACCAGC
D62N	TTTATGCGCAATGAAGTGCTGGCCG TGTGG	GCACTTCATTGCGCATAAAGCGGG TGTACC
E63Q	GCGCGATCAGGTGCTGGCCGTGTGG GAGAT	CCAGCACCTGATCGCGCATAAAGC GGGTGT
R95A	CAAATGGGCCGATAAAATCTTCCGC CAGCA	TTTTATCGCCCCATTTGGCGCTGCC CAGAA
D114N	TTATGGCAATCGCGAGCTGGTGAAG AAGTA	GCTCGCGATTGCCATAACGAAAGG CTTCCA
D114R	TTATGGCCGCCGCGAGCTGGTGAAG AAGTA	GCTCGCGGCGGCCATAACGAAAGG CTTCCA
D114R-R115A	TTATGGCCGCGCCGAGCTGGTGAAG AAGTATCC	CCAGCTCGGCGCGGCCATAACGAA AGGCTTCCA

Table 3: Model evaluation analyses for AbSGR-h and AtSGR1, obtained from the SAVES server

	***Anaerolineae* SGR homolog (AbSGR-h)**	***Arabidopsis* SGR (AtSGR)**
	Ramachandran Plot Analysis	
Residues in most favoured regions	90.9%	91.2%
Residues in additional allowed regions	9.1%	8.0%
Residues in generously allowed regions	0.0%	0.8%
Residues in disallowed regions	0.0%	0.0%
	ERRAT	
Overall quality factor	81.618%	86.923%
	Verify3D	
Residues with averaged 3D-1D score ≥ 0.2	94.77%	86.18%

Table 4: Summary of the activity level and solubility of the wild-type and mutant AbSGR-h proteins

AbSGR-h	Activity level	Solubility
Wild-type	Highly active	Soluble
R26D	Inactive	Insoluble
Y28A	Inactive	Insoluble
T29A	Inactive	Insoluble
T31A	Moderately active	Soluble
H32A	Weakly active	Soluble
D34N	Inactive	Soluble
T36A	Moderately active	Soluble
R61A	Highly active	Soluble
D62N	Inactive	Soluble
E63Q	Highly active	Soluble
R95A	Highly active	Soluble
D114N	Inactive	Insoluble
R26D + D114R (Double mutant)	Inactive	Insoluble
R26D + D114R + R115A (Triple mutant)	Inactive	Insoluble

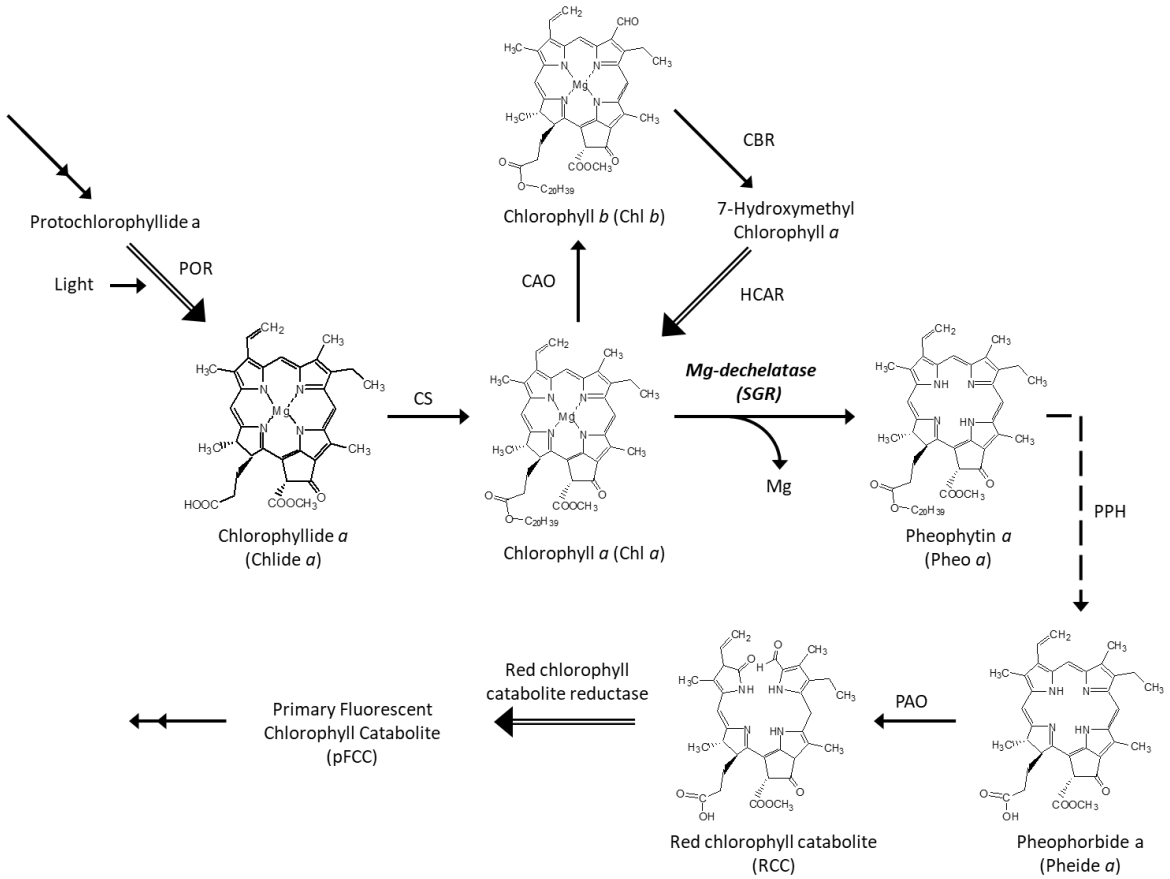

Figure 1: Chlorophyll degradation pathway in green plants. The enzymes that catalyze the reaction represented by double lined arrows have entry in the Protein Data Bank. Dash lined arrow indicates enzyme whose structure had been predicted using computational approaches. Structural information of enzymes involving reactions represented by normal arrows is awaited till when the work was done. CAO, chlorophyllide *a* oxygenase; CBR, chlorophyll *b* reductase; CS, chlorophyll synthase; HCAR, 7-hydroxymethyl chlorophyll *a* reductase; PPH, pheophytin pheophorbide hydrolase; POR, NADPH: protochlorophyllide oxidoreductase.

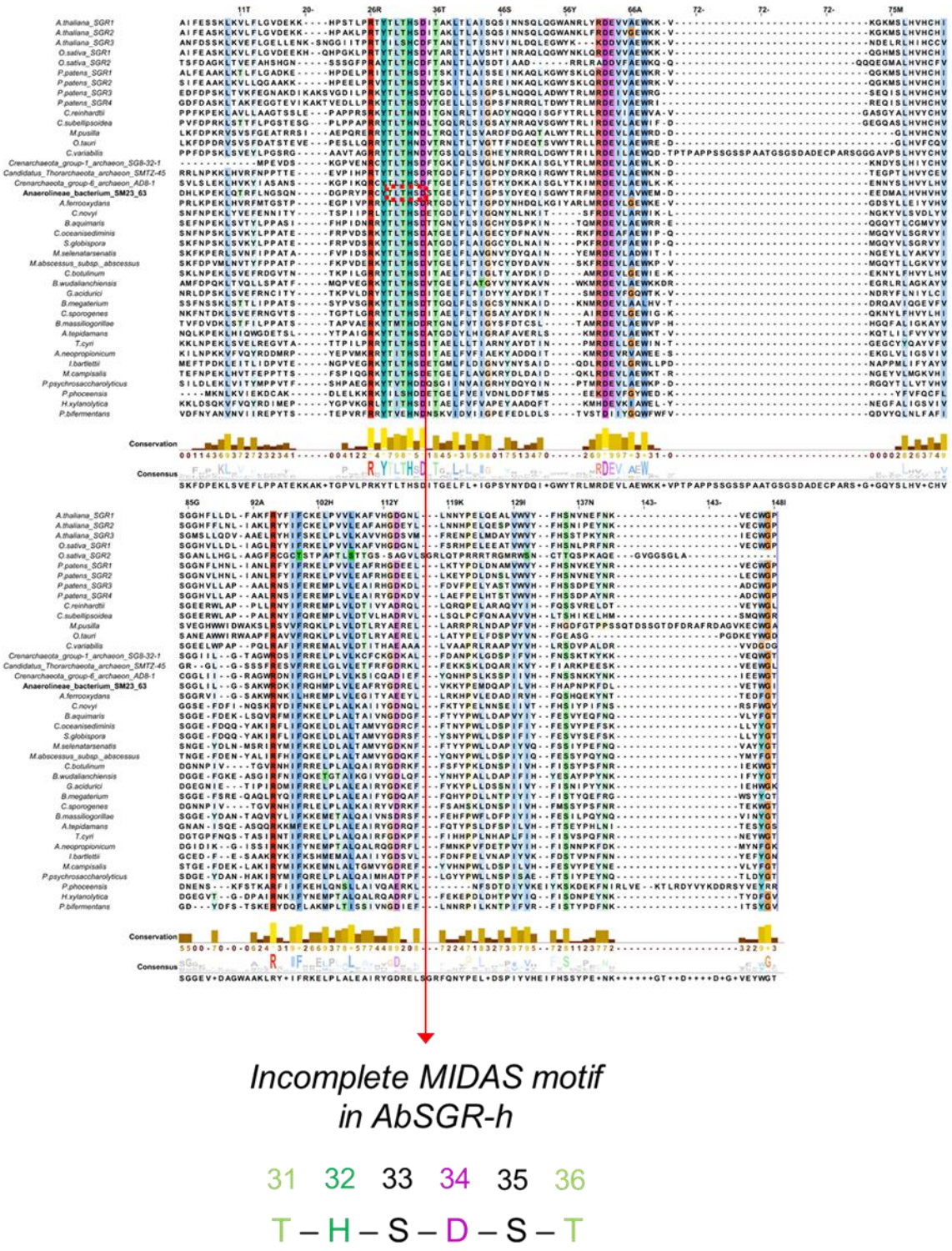

Figure 2: Multiple sequence alignment of green plant SGRs, archaeal and bacterial SGR homologs performed using Clustal omega. The alignment is colored in the Clustalx format. Solid triangles denote residues that have been mutated in the subsequent biochemical

experiments. Portion of the alignment displaying only conserved stretch of residues have been shown. The incomplete metal-ion-dependent adhesion site (MIDAS) motif for the SGR homolog from *Anaerolineae* bacterium SM23_63 is also shown.

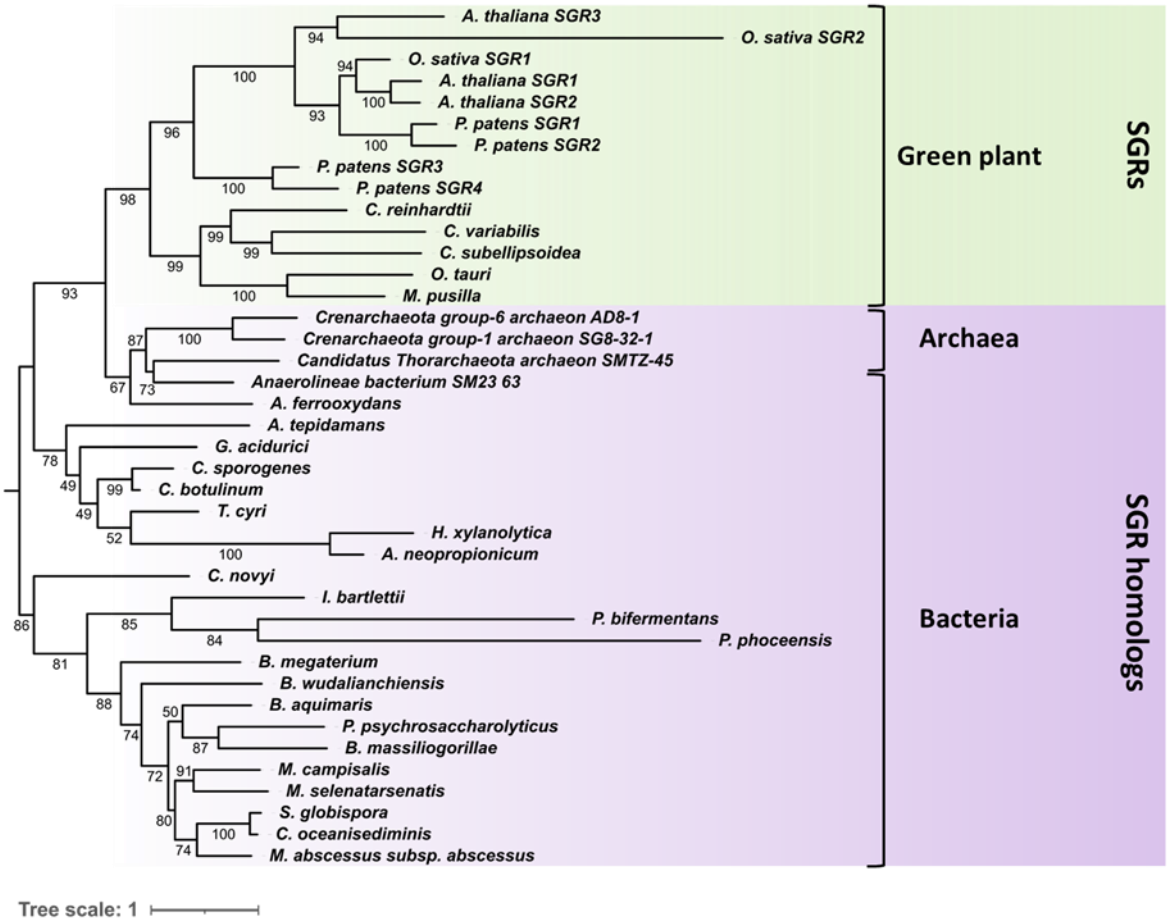

Figure 3: Maximum likelihood phylogeny of protein sequences of SGRs and its homologs. Bootstrap values, based on 1000 replicates, have been shown on branch nodes.

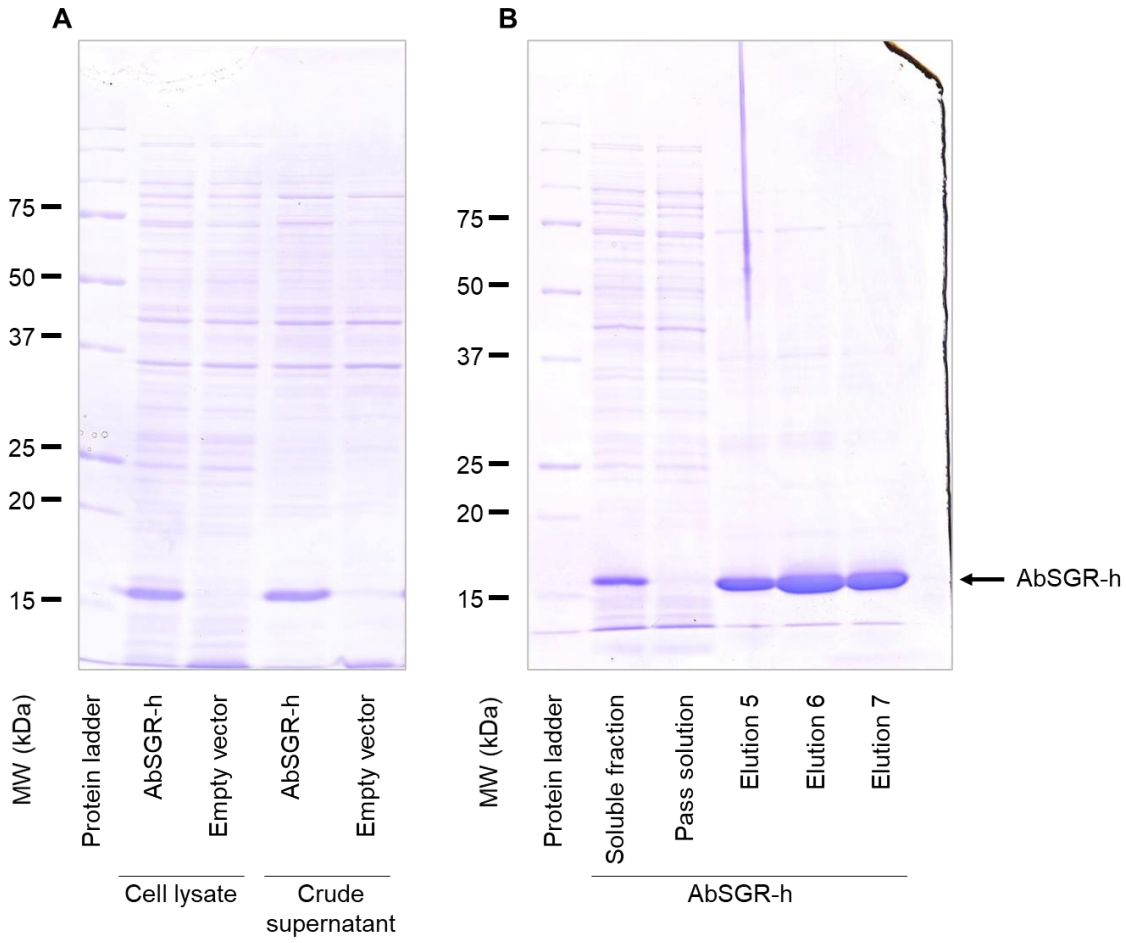

Figure 4: (A) Expression and (B) purification of the recombinant AbSGR-h. Histidine-tagged AbSGR-h was expressed in *E. coli* and purified by the nickel column.

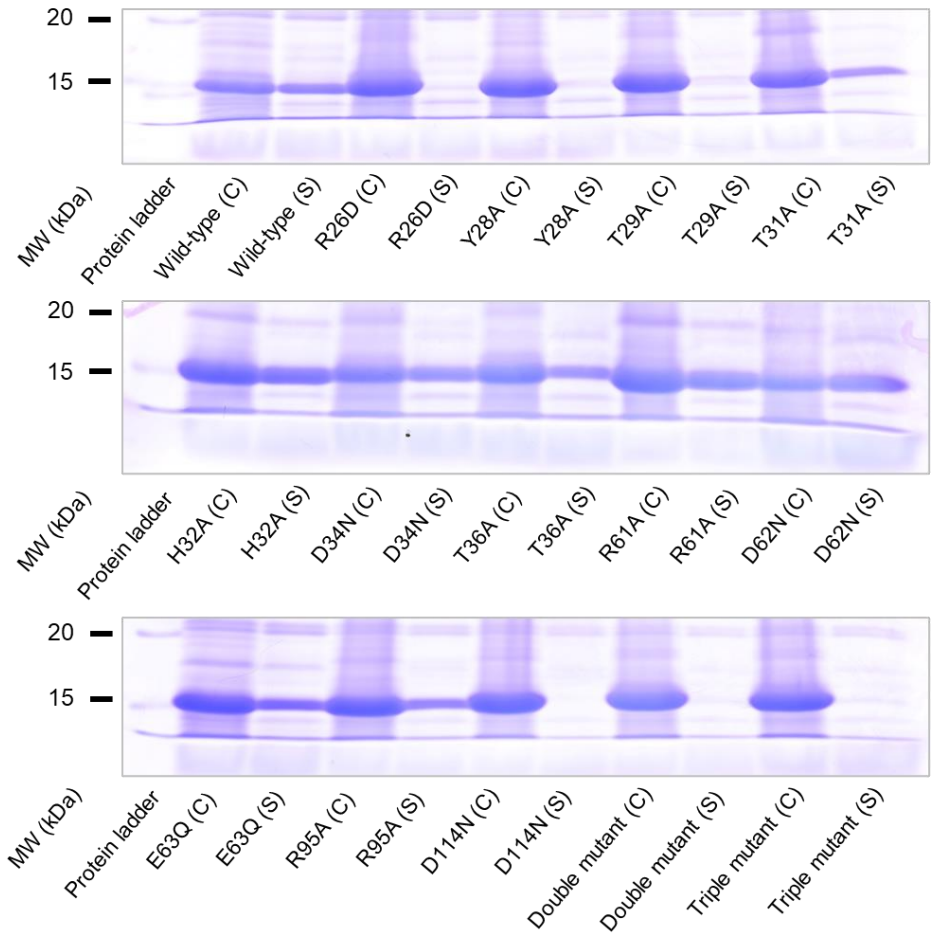

Figure 5: Examination of solubility of the expressed proteins. Wild-type and mutant AbSGR-h were expressed in *E. coli*. After lysis of *E. coli*, crude cell lysate (C) and the soluble fraction of cell lysate (S) were applied on SDS-PAGE.

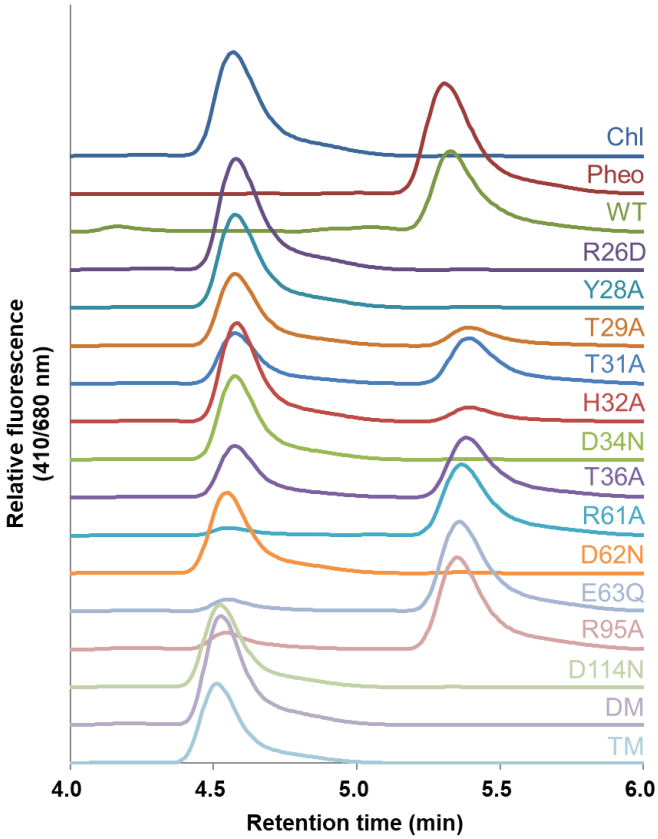

Figure 6: Determination of AbSGR-h activity. Chlorophyll *a* was incubated with crude cell lysate of *E. coli* expressing wild-type or mutant AbSGR-h. After incubation, pigments were extracted and analyzed by HPLC. Chlorophyll *a* and pheophytin *a* peaks have also been shown.

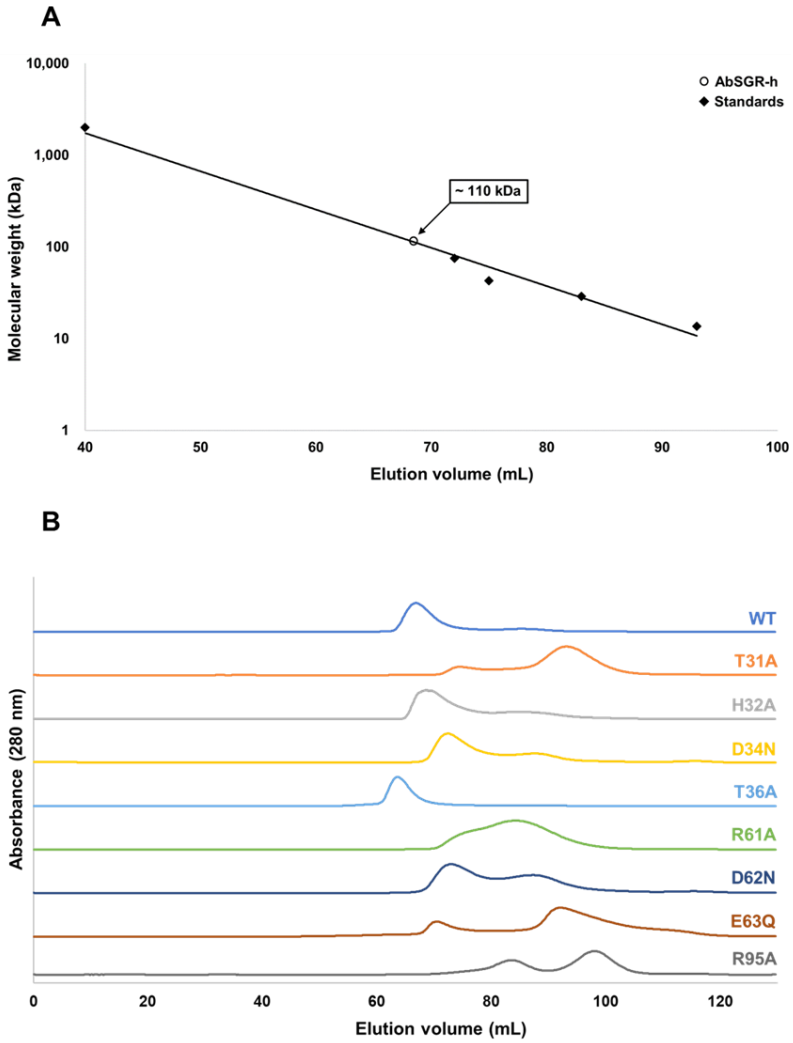

Figure 7: Determination of the molecular size of AbSGR-h complex. (A) Calibration curve of logarithm of molecular mass as a function elution volume. (B) Size exclusion chromatography profiles of AbSGR-h. Protein was monitored by the absorbance at 280 nm.

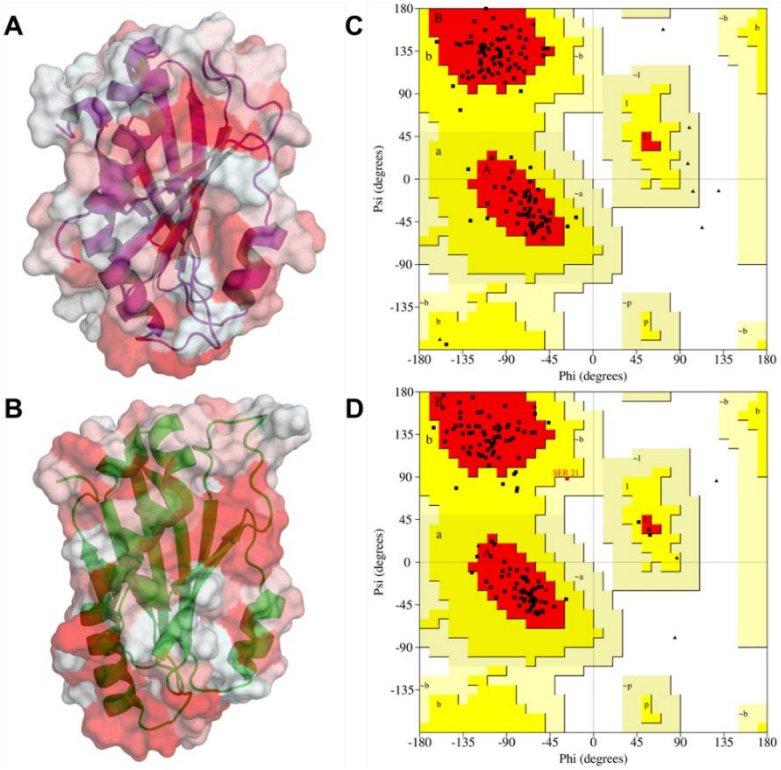

Figure 8: Hydrophobic surface and cartoon representation of the predicted three-dimensional structure of (A) AbSGR-h and (B) AtSGR1 where white and red color indicate hydrophilic and hydrophobic regions, respectively. Ramachandran plot showing the dihedral angle values for (C) AbSGR-h and (D) AtSGR1 is also given.

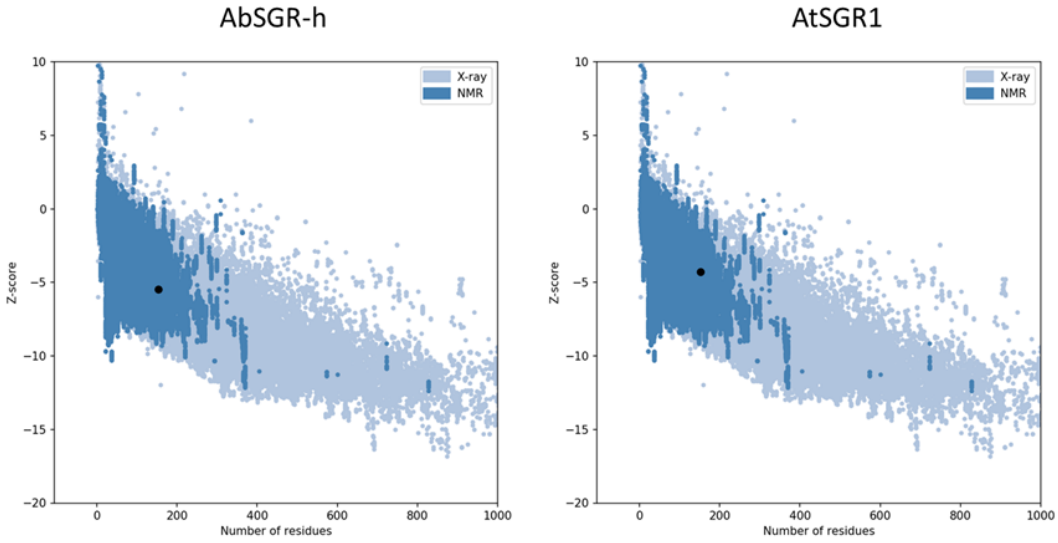

Figure 9: ProSA plot of the two proteins showing Z-score. The regions with blue and light blue colors are about the groups of similar structures of protein, determined from NMR and X-Ray methods, respectively. The black dot in the plots represent the predicted SGR models.

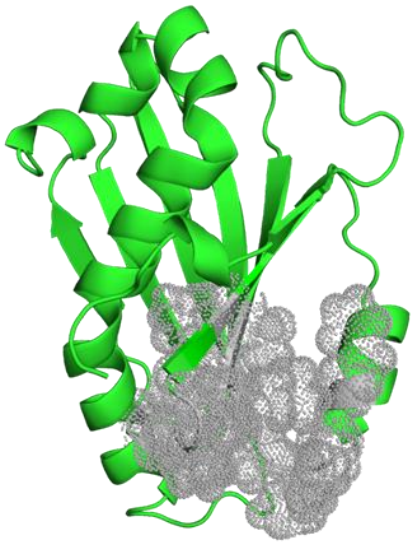

Figure 10: Predicted binding cavity of AbSGR-h, as detected by the CavityPlus tool. The residues constituting the predicted binding site have been shown in dot representation.

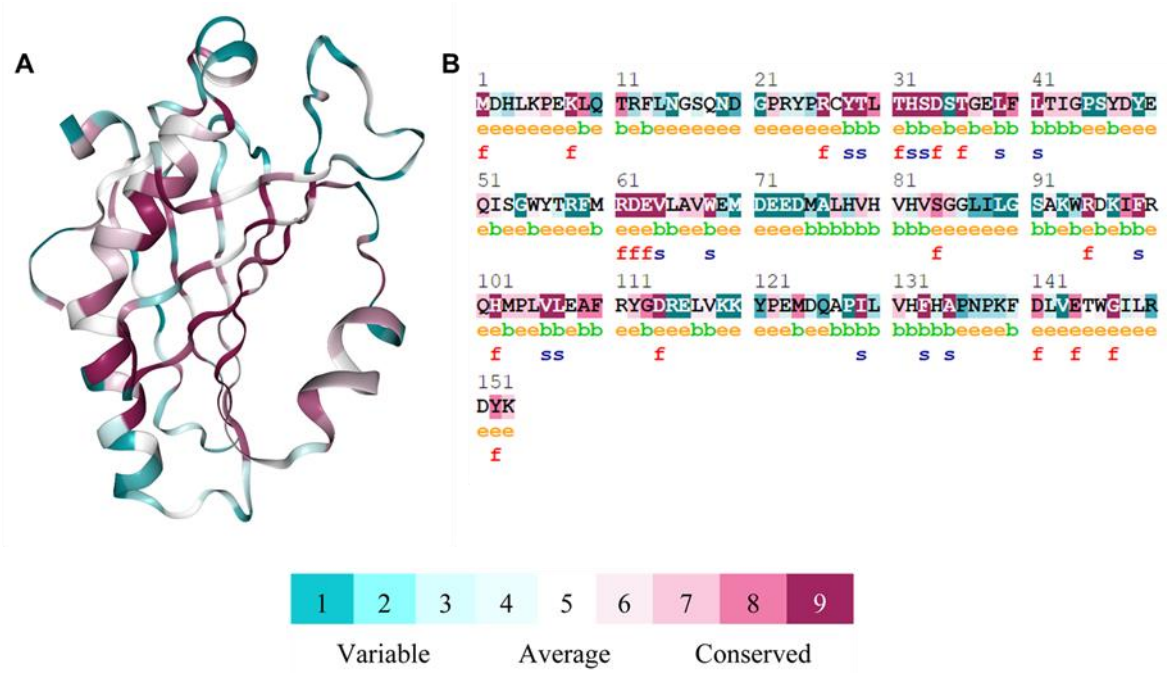

Figure 11: Evolutionary conservation of amino acid residues in the three-dimensional structure of AbSGR-h protein by ConSurf analysis (A) and in the primary sequence of AbSGR-h by ConSeq analysis (B). ‘e’ refers to an exposed residue according to the neural-network algorithm; ‘b’ refers to a buried residue 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).

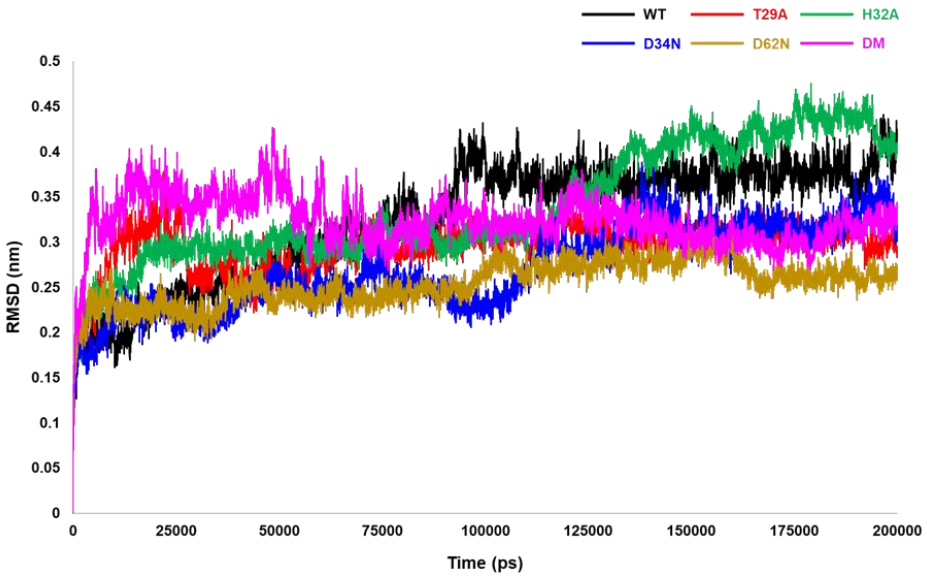

Figure 12: RMSD plot of the backbone atoms of wild-type and mutant AbSGR-h proteins over the 200 ns trajectory of the MD production run.

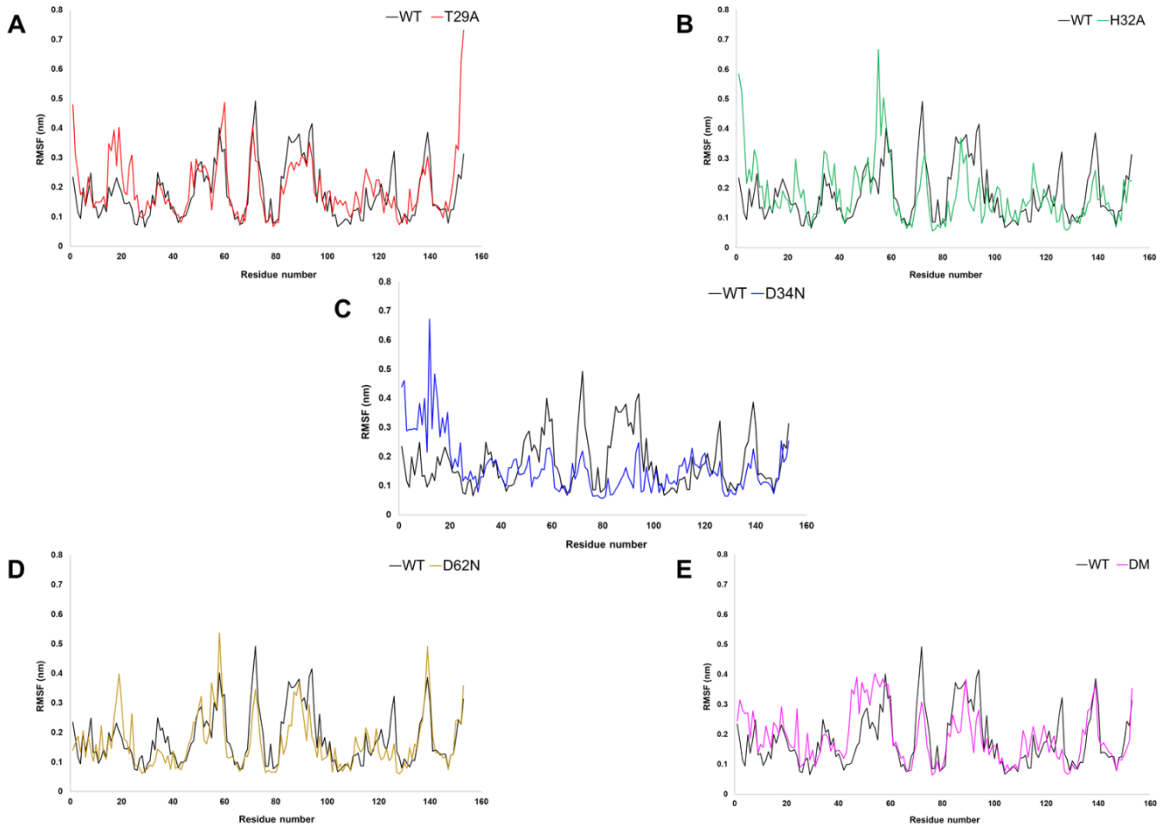

Figure 13: RMSF plots of each residue of wild-type and mutant AbSGR-h proteins over the 200 ns trajectory of the MD production run.

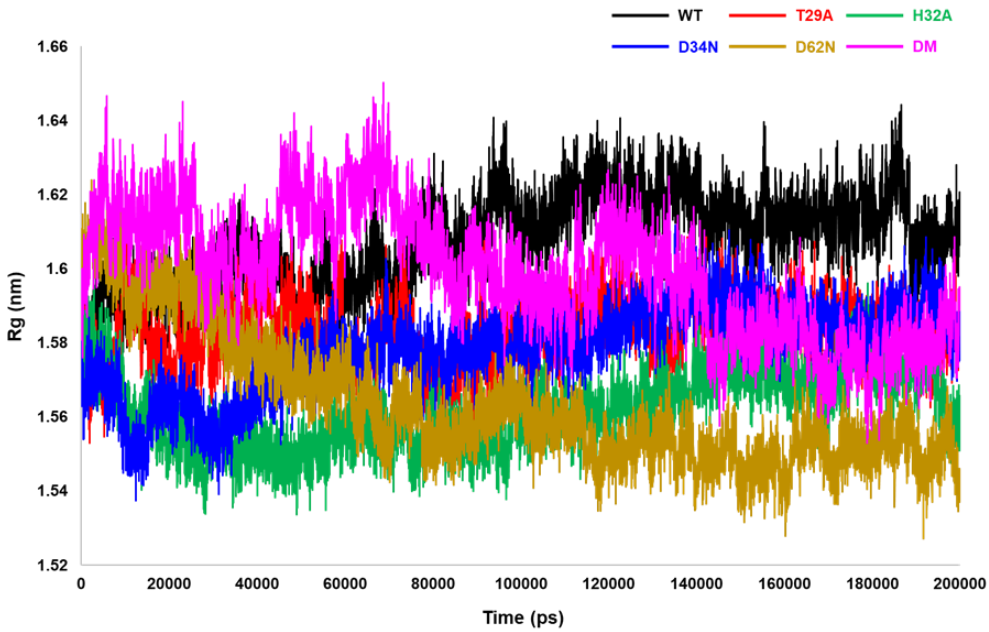

Figure 14: Radius of gyration of wild-type and mutant AbSGR-h proteins over the 200 ns trajectory of the MD production run.

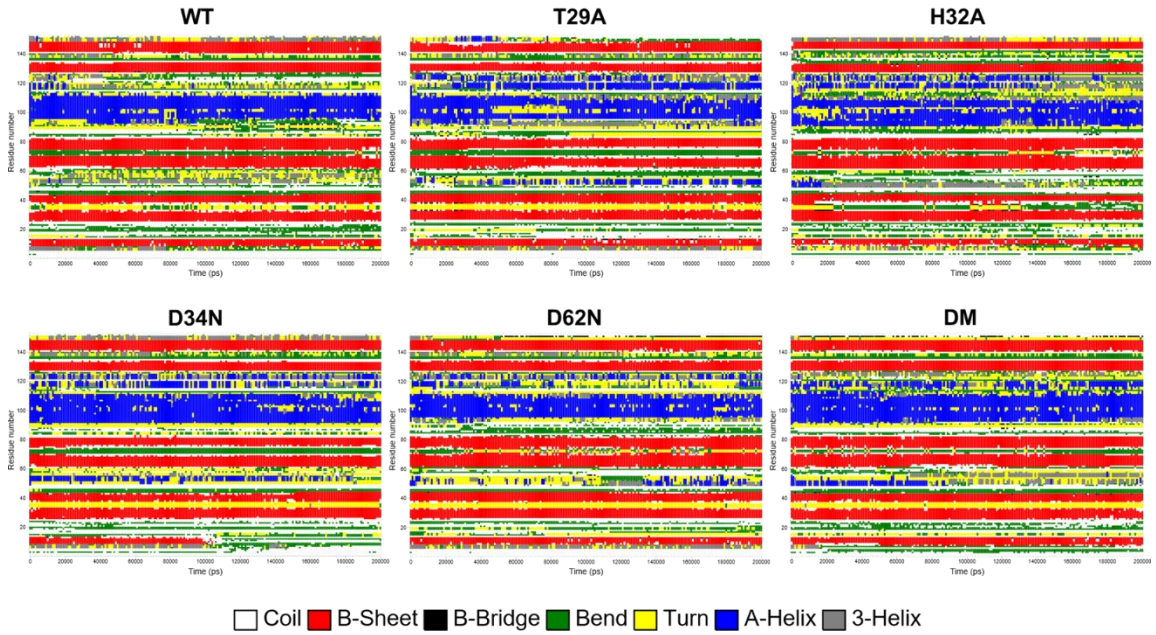

Figure 15: Time evolution of the secondary structures of wild-type and mutant AbSGR-h proteins over the 200 ns trajectory of the MD production run, determined using the DSSP method. In each plot, the vertical axis depicts the residue number while the horizontal axis shows time frame in picoseconds.

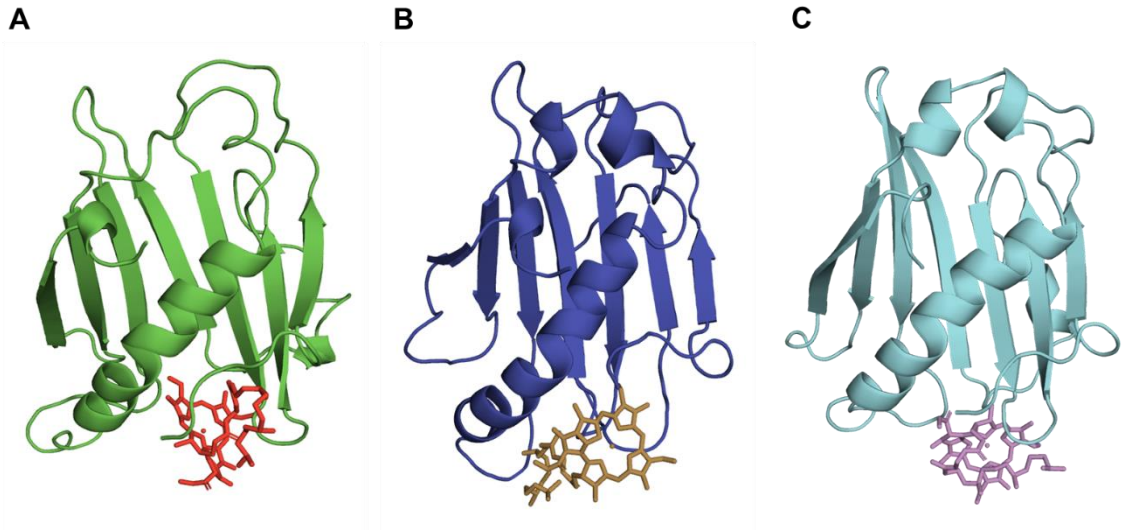

Figure 16: Docked structure of chlorophyll *a* with (A) wild-type, (B) H32A mutant and (C) D62N mutant AbSGR-h proteins. The average structure of wild-type and mutant monomers, obtained from the MD simulations of 200 ns, were docked with chlorophyll *a* using AutoDock Vina.

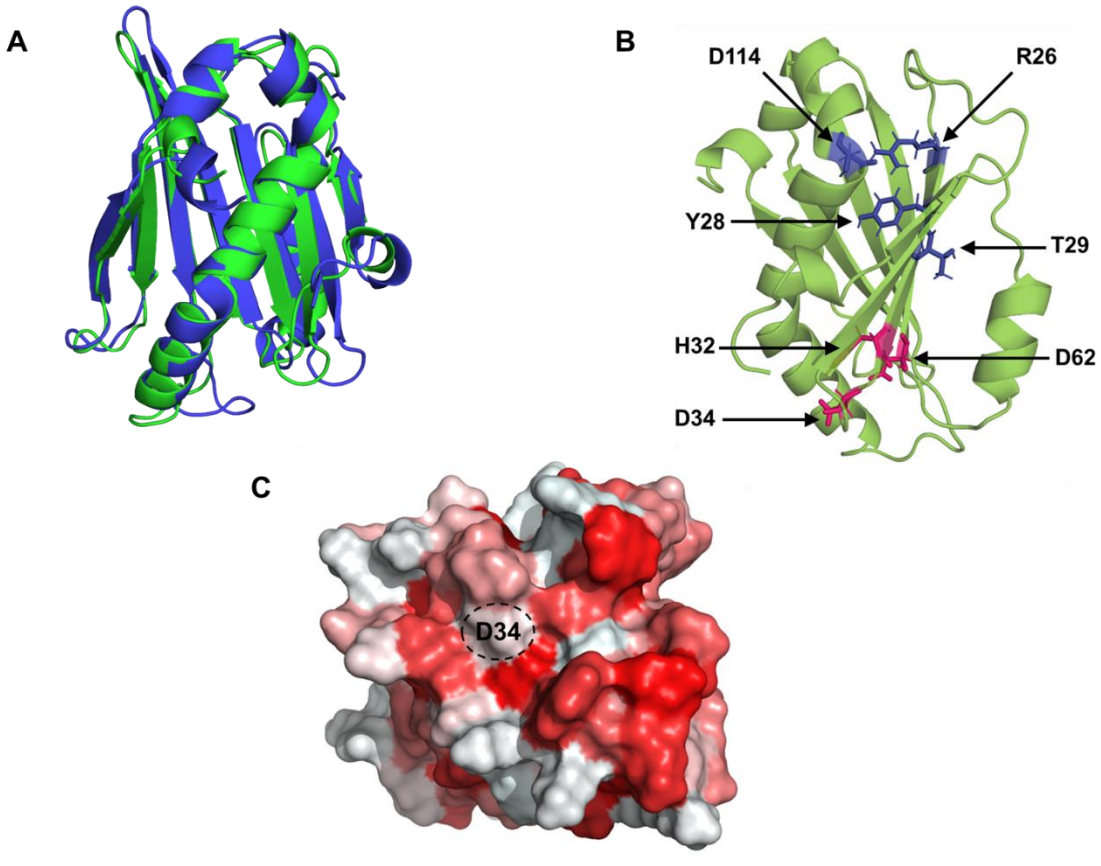

Figure 17: (A) Structural superimposition between AbSGR-h protein structures predicted by trRosetta (green) and RoseTTAFold (blue) (B) Structure of AbSGR-h showing structurally and functionally important residues. Side chain of amino acids responsible for maintaining structure of the protein are marked in blue while those involved in catalysis are marked in red. (C) Hydrophobic patch surrounding D34 shown in the hydrophobicity surface representation of AbSGR-h. White and red color indicate hydrophilic and hydrophobic regions, respectively.

Chapter 2

Crystal structure and reaction mechanism of a bacterial Mg-dechelatase homolog

2.1 Introduction

Chlorophyll (Chl), the most abundant pigment on earth, is an indispensable molecule involved in the conversion of light energy and driving electron transfer during photosynthesis (Hörtensteiner, 2009). Land plants and chlorophytes contain the two major types of chlorophyll – chlorophyll *a* (Chl *a*) and *b* (Chl *b*) as protagonist molecules (R. Tanaka & Tanaka, 2007). In order to ensure efficient photosynthesis and other critical biological processes, maintenance of a balance in Chl metabolism becomes essential (Mochizuki et al., 2010). Chl biosynthesis is important for effective photosynthetic performance during photosystem formation and adaptation to different environmental conditions (R. Tanaka & Tanaka, 2007). On the other hand, Chl degradation plays significant role in senescence, fruit ripening, and seed maturation (Jiao et al., 2020). Furthermore, Chl breakdown facilitates nutrient remobilization (Vom Dorp et al., 2015) and protects the plant against cellular photodamage (Hörtensteiner & Kräutler, 2011). Therefore, both synthesis and degradation of Chl must be strictly regulated during the greening and senescence stages for ensuring plant viability.

Chl breakdown, one of the most visually striking biochemical reactions on Earth, represents a physiological manifestation of leaf senescence (Aubry et al., 2021; Schumacher et al., 2021). The pathway of Chl degradation can be divided into: (1) a chloroplastic phase involving disassembly of the thylakoid leading to the opening of the porphyrin ring of Chl and (2) a cytosolic and vacuolar phase that includes detoxification and sorting of the linear tetrapyrrole (Kuai et al., 2018; Schumacher et al., 2021). Prior to the initiation of Chl breakdown, Chl *b* must be converted to Chl *a* in a two-step enzymatic process called the Chl cycle (Meguro et al., 2011; Y. Sato et al., 2009). This cycle is needed to finely regulate the Chl *a*/*b* ratio, a step important for acclimatizing plants to the light environment. Following conversion to Chl *a*, the pigment is eventually processed by four enzymes: (1) magnesium-dechelatase (Mg-dechelatase) (Shimoda et al., 2016); (2) pheophytinase (PPH) acting as a dephytylase (Schelbert et al., 2009); (3) pheophorbide *a* oxygenase (PAO) which catalyzes the irreversible opening of the porphyrin ring (Pruzinská et al., 2003) to form the first linear tetrapyrrole (red chlorophyll catabolite) and (4) red chlorophyll catabolite reductase (RCCR) producing the primary fluorescent Chl catabolite (pFCC) which are then exported from chloroplasts and isomerized to non-fluorescent products by the acidic pH in the vacuole (Hauenstein et al., 2016; Pružinská et al., 2007).

The most important enzyme of the Chl degradation pathway is Mg-dechelatase encoded by the *Stay-Green* (SGR) gene, which is also responsible for Mendel's green cotyledon peas (Y. Sato et al., 2007). It catalyzes the extraction of Mg from Chl *a* to form pheophytin *a* in a tightly regulated reaction so as to prevent the formation of damaging photosensitizing Chl metabolic intermediates (Hirashima et al., 2009; Shimoda et al., 2016). Therefore, SGR not only catalyzes the most crucial and committed step of chlorophyll degradation but also removes a metal ion from an organic moiety in a biochemically enigmatic reaction. Although high sequence similarity exists between SGR and its homologs, their catalytic activity and ligand specificity differ substantially among species (Obata et al., 2019). For instance, *Arabidopsis* SGR participate in Chl degradation whereas SGR in *Chlamydomonas reinhardtii* contributes to PSII formation (Chen et al., 2019). In addition to catalyzing the rate-limiting step of Chl breakdown, SGR plays a myriad of physiological roles in plant growth and development (Delmas et al., 2013), nodule senescence in legumes (Zhou et al., 2011), fruit maturation (Luo et al., 2013) and regulation of the expression of genes encoding other chlorophyll degradation enzymes (Sato et al., 2018).

Although the biochemical reactions and key enzymes involved in the Chl metabolic pathway have been identified (Kuai et al., 2018; Shimoda et al., 2016), information on the structural aspects of these enzymes remain limited. In the Chl biosynthesis pathway, the crystal structures of Mg-chelatase and light dependent protochlorophyllide oxidoreductase have been elucidated, both of which catalyzes regulatory steps in the process (Adams et al., 2020; Zhang et al., 2019). Among the enzymes participating in the four steps of Chl *a* conversion to pFCC, the only crystal structure determined till now is of RCCR (Sugishima et al., 2010). While the catalytic and structural properties of *Arabidopsis* pheophytinase was investigated *in silico* by Guyer and his colleagues (Guyer et al., 2018), I first reported the detailed structural characteristics of SGR deciphered using a combination of computational and biochemical approaches. In this study, I determined the crystal structure of highly active SGR homolog from *Anaerolineae* bacterium SM23_63 (AbSGR-h). Furthermore, the enzyme kinetics of wild-type and mutant AbSGR-h was evaluated. The reaction mechanism of the enzyme was also proposed.

2.2 Materials and methods

2.2.1 Cloning of bacterial SGR homolog

Bacterial SGR homolog from *Anaerolineae* bacterium SM23_63 (AbSGR-h) encoded by KPK94580 with optimized codon usage for *E. coli* was artificially synthesized according to a previously reported protocol (Obata et al., 2019). AbSGR-h was amplified from the artificially synthesized DNA using the primer sets provided in **Table 5**. PCR amplified DNA

fragments were cloned into pET 30a (+) vectors (Novagen) containing a histidine-tag at the C terminus using the *Nde*I and *Xho*I sites through an in-fusion cloning system (Clontech). Point mutations were introduced by PCR using primers as shown in **Table 5**.

2.2.2 Expression, detection and purification of recombinant proteins

The constructed plasmids for protein expression were introduced into *E. coli* BL21 (DE3). *E. coli* was grown and an auto-induction medium (6 g Na_2HPO_4, 3 g KH_2PO_4, 20 g tryptone, 5 g yeast extract, 5 g NaCl, 6 mL glycerol, 0.5 g glucose, 2 g lactose, 100 mg kanamycin in 1L) at 37ºC for 16 h with 120 rpm shaking was used for expression of the recombinant protein (Studier, 2005). After induction of the recombinant protein, 200 mL of culture cells were harvested by centrifugation at 7,000 g for 5 min. The harvested cells were resuspended in buffer A (20 mM Na-phosphate pH 7.4, 100 mM NaCl, 20 mM imidazole) and disrupted by sonication (Branson Sonifier SFX250: output 8, duty cycle 20%) for 6 min in an ice bath. After sonication, dodecyl β-maltoside (βDM) was added to the final concentration of 0.05% (w/v) and incubated for 5 min at 25ºC. The cleared supernatant of cell lysate was obtained by centrifugation at 15,000 g for 20 min in 4ºC and then loaded onto a 5 mL HisTrap HP column (Cytiva) equilibrated with buffer A containing 0.05% βDM using an ÄKTAprime plus system (Cytiva). The recombinant proteins were eluted by buffer B (20 mM Na-phosphate pH 7.4, 100 mM NaCl, 500 mM imidazole, 0.05% βDM). The purified protein was further analyzed by size exclusion chromatography using Sephacryl S-400R (Cytiva) equilibrated with buffer C (20 mM Na-phosphate pH 7.4, 100 mM NaCl, 0.05% βDM). The protein elution profile was monitored by absorbance at 280 nm. The molecular weight of AbSGR-h was evaluated by comparison to protein standards (Gel Filtration Calibration Kit LMW, Cytiva) as previously reported. To examine the purity of the protein, elution was mixed with the same volume of the sample buffer containing 125 mM Tris- HCl, pH 6.8, 4% (w/v) SDS, 10% (w/v) sucrose, and 5% (v/v) 2-mercaptoethanol. Mixtures were incubated at 95ºC for 1 min and 2 µL of it was used for SDS-PAGE followed by staining with Coomassie Brilliant Blue for visualization. Protein concentration in the purified solution was quantified using the Bradford protein assay.

2.2.3 Activity assay

Activity assays were performed with Chl *a* dissolved in dimethyl sulfoxide (DMSO) and at a final DMSO concentration of 2% (v/v) in the reaction mixture. Purified recombinant proteins (50 µM) were incubated with Chl *a* (20 µM) for 10 min at 25ºC in 50 µL of the reaction buffer comprising 20 mM Na-phosphate pH 7.4, 100 mM NaCl, 0.05% βDM. Reactions were stopped by adding 200 µL of acetone, followed by centrifugation at 20,000 g for 10 min. Reactions were analyzed by HPLC as previously reported. To determine kinetic

parameter, Chl *a* at different concentrations (2.5, 5, 10, 20, 40 μM) was incubated with WT (5 μM) or D34E (50 μM) for 20 min. The Michaelis-Menten curves were calculated by the least-squares method using Microsoft Excel Solver.

2.2.4 Crystallization

Crystals of SGR were obtained by sitting-drop vapor diffusion method at 298 K. A sitting drop was prepared by mixing equal volumes of SGR solution and reservoir solution containing 100 mM Tris-HCl (pH 8.5), 200 mM Ammonium phosphate monobasic and 50% (v/v) 2-Methyl-2,4-pentanediol (Crystal Screen Kit II No. 43). The initial crystals appeared for two days, and in the 1-2 weeks, these crystals grew to as large as 0.1 mm in their longest dimensions. Since the reservoir solution contained 50% (v/v) 2-Methyl-2,4-pentanediol (MPD), the crystals were picked up directly from the sitting drop and immediately frozen in liquid nitrogen. This crystal diffracted up to 1.75 Å resolution and belonged to space group $P6_122$ with cell dimensions of a=80.3 Å, b=80.3 Å, c=224.8 Å, β=120.0°.

Although not included in the current study, crystals of AbSGR-h bound to Zn-chlorophyllide *a* (ZnChlide) was obtained by sitting-drop vapor diffusion method at 298 K where the drop was prepared by mixing equal volumes of SGR and reservoir solution containing 0.1 M HEPES-NaOH (pH 7.5), 28% PEG 400 and 0.2 M $CaCl_2$. The protein (AbSGR-h) and substrate (ZnChlide) concentration used is 10mg/ml and 0.5 mg/ml, respectively.

2.2.5 X-ray intensity data collection and structure determination

X-ray intensity data was collected on BL44XU at SPring-8 (Harima, Hyogo, Japan) using EIGER X16M detector (Dectris) at cryogenic temperature (100 K). The diffraction data were processed and scaled using the program XDS (Kabsch, 2010). The initial phase was determined by the molecular replacement method with the program PHASER in CCP4 (McCoy et al., 2007) using the AlphaFold2 predicted structure of rice SGR (UniProt: Q652K1) as a starting model. The structure model was manually built using the program COOT (Emsley & Cowtan, 2004) in CCP4, and refinement was performed using phenix refine (Liebschner et al., 2019) in PHENIX. Secondary structure assignment was implemented using the STRIDE web-server (Heinig & Frishman, 2004). All figures showing the atomic coordinates were made with PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.).

2.2.6 Substrate docking analysis

The KEGG LIGAND database (https://www.genome.jp/kegg/ligand.html) was used for retrieving the structure of Chl *a*. It was then subjected to geometry optimization under the semiempirical method in HyperChemTM 8.0.8 molecular modeling software (Hypercube Inc., Gainesville, FL, USA). Steepest descent followed by the Polak-Ribiere conjugate gradient algorithm was performed for energy optimization of Chl *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 study was carried out using AutoDock Vina v1.1.2 considering chain A of the crystal structure of wild-type AbSGR-h (Trott & Olson, 2010). 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). A grid box of 30 Å × 30 Å × 30 Å with grid spacing of 1 Å was set and no solvation was considered for the docking procedure. Interactions in the docked conformations were visualized using PyMOL.

2.3 Results

2.3.1 Protein expression and purification

Among all the bacterial SGR homologs, AbSGR-h is ancestrally closer to land plant SGRs (Obata et al., 2019). Furthermore, the overexpression of a highly soluble as well as active form of this protein is possible in *E. coli* using its general expression vector. In fact, AbSGR-h exhibited higher Mg-dechelatase activity than *Arabidopsis* SGRs (Obata et al., 2019). Previous study has showed that three residues – H32, D34 and D62 in the AbSGR-h protein are critical for its catalytic activity, in which D34 is involved in direct interaction with Chl *a*. Additionally, the presence of D34 on the surface of protein surrounded by a hydrophobic patch of residues, provides an ideal environment for interaction with the substrate.

Since the main objective of this study is to delve deep into the reaction mechanism of Mg-dechelatase enzyme, three mutations on the aforementioned D34 residue were prepared to understand the effect of substitutions on the catalytic ability of AbSGR-h. D34 was changed to glutamate (D34E) to check whether replacement with a similar property amino acid altered the activity of AbSGR-h. It was also substituted with uncharged amino acids resulting in the formation of D34N and D34Q mutants. After expression of the wild-type and mutant SGRs in *E. coli*, the proteins were subjected to nickel column purification and subsequently to gel filtration analysis **(Figure 18)**.

According to a previous study, AbSGR-h exists as a hexameric complex in solution. Here also, major peaks were observed at the same position for the wild-type and three mutant AbSGR-h (D34E, D34N and D34Q), indicating the presence of the hexameric form of the

proteins. This observation further confirms the role of D34 to be solely catalytic rather than maintaining the multimeric conformation of the protein. The molecular weight of the purified proteins was analyzed by SDS-PAGE **(Figure 19)**. Both wild-type and mutant proteins appeared as single and distinct bands corresponding to a molecular size of approximately 18 kDa. The CBB stained gel also showed that the purity of these proteins is very high.

2.3.2 Mg-dechelating activity of SGR homolog

The Mg-dechelating activity of wild-type AbSGR-h and its mutants was tested *in vitro* using Chl *a* as the substrate. Enzymatic activity of AbSGR-h leads to the extraction of Mg^{2+} from Chl *a* to form pheophytin *a*. Activity levels were assessed based on the amounts of pheophytin *a* using HPLC **(Figure 20)**. It is worth mentioning that the increase in the pheophytin peak on the HPLC profile is associated with the concomitant disappearance of the substrate Chl *a*. The wild-type protein exhibited the highest Mg-dechelating activity resulting in 1:1 chlorophyll to pheophytin ratio. Interestingly, despite changing D34 with a similar kind of amino acid (Glu), the Mg-dechelating activity level of the mutant decreased substantially. The remaining mutations, *i.e.*, D34N and D34Q rendered the protein inactive in spite of being highly soluble, implying the potential role of this residue in catalysis.

2.3.3 Determination of kinetic parameter of SGR homolog

K_m and k_{cat} were determined for WT and D34E AbSGR-h with Mg-dechelating activity **(Figure 21)**. D34N and D34Q were inactive and hence not studied. Purified proteins through nickel column and size exclusion chromatography were used for the determination of kinetic parameters. As D34E showed lower activity, higher concentration (50 µM) of the protein was used compared with WT (5 µM). High concentrations of Chl *a* inhibited enzymatic activity, therefore the Michaelis-Menten curves were constructed with low Chl concentrations. The k_{cat} value of D34E (0.00031 min^{-1}) was significantly lower than that of WT (0.076 min^{-1}), suggesting that D34 is involved in catalysis. Though K_m value of D34E (6.00 µM) was lower than that of WT (15.91 µM), the difference was not so significant as that of k_{cat} **(Table 6)**. Amino acid residues other than D34 may determine the affinity of Chl to AbSGR-h more specifically.

2.3.4 Crystal structure of AbSGR-h

The crystal structure of AbSGR-h was solved using the molecular replacement method at 1.75 Å resolution **(Figures 22 & 23)**. Owing to the lack of homologous templates with SGR, the AlphaFold2 predicted structure of rice SGR (UniProt: Q652K1) was used as a starting model. The crystallographic data and refinement statistics are provided in **Table 7**. Crystals of AbSGR-h bound to ZnChlide were also obtained **(Figure 24)**, the analysis of which forms the scope of a future study.

AbSGR-h exists as a dimer in solution **(Figure 25A)**, as revealed by the X-ray crystal structure. The molecular weight of AbSGR-h monomer is estimated to be 18 kDa while the apparent molecular weight according to the size exclusion chromatography elution profile is 110 kDa, which suggests the presence of a hexameric complex probably in the form of a trimer of dimers. However, the multimeric form cannot be ascertained from this crystal analysis and requires further investigation. Each monomeric structure comprises two α-helices, four 3-10 helices, and seven β-strands, as revealed by STRIDE analysis **(Figure 25B)**. The β-strands constitute one β-sheet leading to the formation of a curved structure, inside which the larger α-helix (residues 92 – 120) is accommodated. These observations are almost in accordance with the secondary structure architecture of the computationally predicted structure of AbSGR-h. The Root Mean Square Deviation (RMSD), based on Cα-atoms, between the crystal and computationally predicted structure is 1.28 Å, where the difference remains restricted to the loop regions of the two structures.

2.3.5 Substrate docking analysis

Molecular docking analysis considering energy-minimized free Chl *a* as the substrate was carried out in AutoDock Vina. The actual substrate of SGR *in vivo* is Chl *a* bound to chlorophyll-protein complexes that are embedded in the thylakoid membrane (Nelson & Yocum, 2006). The grid box was set around the active site of the protein such that it covers three residues – H32, D34 and D62, all of which are catalytically important. The docking analysis revealed interaction of the central Mg ion of Chl *a* with the carbonyl oxygen atom of D34 residue in the crystal structure of AbSGR-h monomer **(Figure 26)**. The distance between the aforementioned atoms was found to be ~ 4.4 Å. The interaction between the two moieties was stable, as indicated from the binding affinity value (-7.8 kcal/mol).

2.4 Discussion

2.4.1 Comparison of the crystal structure with the computationally predicted structure

In this study, the crystal structure of AbSGR-h was solved using the molecular replacement method at 1.75 Å resolution. The crystal and computationally predicted structure were found to be almost identical. However, the arrangements of the active site were slightly different **(Figure 27)**. The turn between the 2nd and 3rd β-sheet was not tightly packed in the crystal structure. Thus, H32 and D34 residing in this region are more flexible than those in the computationally predicted structure, though the Cα-atoms of D62 were found in the same position for both the structures. However, the orientation of the side chain of D62 was different. In the crystal structure, the side chain faces outwards, which facilitate interaction of this side chain with H32 side chain. In the case of crystal structure, the active site was shown to be flexible which is difficult to predict computationally.

Proteins are not static but dynamic moieties that constantly undergo conformational changes in their natural environments. The observed fluctuations may include both local motions involving a few residues and global coordinated motions of several residues. Thus, there is always a small deviation in computational structural prediction. In the case of SGR, the computational prediction was almost precise and sufficient to provide the basis for the analysis of reaction mechanism, and the roles of the amino acid residues in maintaining the structure and function of the protein.

2.4.2 Enzymatic properties of AbSGR-h

The kinetic parameters of AbSGR-h were examined in this study where K_m was found to be 15.91 µM. This is in concordance with the reported values of plant, green algae and bacterial recombinant proteins (Obata et al., 2019). Pheophytinase and pheophorbide *a* oxygenase catalyzes the successive reactions of SGR. K_m values of the recombinant proteins of *Arabidopsis* pheophytinase and pheophorbide *a* oxygenase were 14.35 µM and 6.0 µM, respectively (Guyer et al., 2018; Pruzinská et al., 2003). These enzymes, involved in chlorophyll breakdown pathway, have similar substrate affinities. When D34 in AbSGR-h was substituted by glutamate, the K_m was 6.0 µM. This substitution did not have negative effect on the affinity of ligand to the protein, suggesting that D34 is indispensable for the activity but does not play a major role in the determination of substrate specificity. k_{cat} of WT AbSGR-h was 0.076 in 1 min. Physiologically, this catalytic rate is too low. Low reaction rates of the recombinant protein have also been reported in the bacteriochlorophyll anabolic pathway (Nomata et al., 2005) and the cyanobacterial chlorophyll catabolic pathway (Takatani et al., 2022). k_{cat} of D34E was 0.00031 in 1 min, which is much lower than that of WT AbSGR-h. D34N and D34Q resulted in inactivation of the enzyme. All these four recombinant proteins showed the same profiles in gel filtration, suggesting that mutation in D34 does not affect the quaternary structure. These observations suggest that the carbonyl group of this acidic amino acid residue is indispensable for activity and SGR is optimized to use aspartate for the reaction.

2.4.3 Proposed reaction mechanism of SGR

In a previous study, multiple sequence alignment showed that H32, D34 and D62 are conserved in SGR homologs. These three amino acid residues were essential for the enzymatic activity and docking simulation also suggested that D34 interacts with chlorophyll *a*. Additionally, these residues were close together in the tertiary structure of AbSGR-h. The observations, herein, show striking resemblance with the catalytic triad found in hydrolases. The catalytic triad in serine proteases is composed of serine, histidine and aspartate residues with similar spatial arrangements including some variations (Ekici et al., 2008). Since the catalytic triad of chymotrypsin (S195/H57/D102) show high resemblance with the proposed

catalytic triad of AbSGR-h (D34/H32/D62), I compared the catalytic triad organization between these two proteins to understand the probable reaction mechanism of SGR. The side chain of S195 is polarized and deprotonated with the support of H57 to act as a base catalyst. D102 is involved in the charge-relay mechanism. AbSGR-h has aspartate in the active site instead of serine. The distance between D34 side chain and H32 side chain is 3.5 Å and that between H32 side chain and D62 side chain is 2.6 Å **(Figure 28A)**. In the catalytic triad of chymotrypsin (PDB accession: 1AFQ) distance between S195 – H57 side chains and H57 – D102 side chains are both 2.8 Å apart **(Figure 28B)**. Though distance observed in AbSGR-h is longer, it is sufficient enough to form a hydrogen bond. These observations suggest that D34, H32 and D62 are arranged to deprotonate D34 side chain, which is exposed to the aqueous environment and may almost be deprotonated. Hence, cooperation among D34, H32 and D62 will be able to complete the deprotonation event. The distance between D34 and H32 of AbSGR-h is 3.5 Å, which is longer than that between S195 and H57 (2.6 Å) of chymotrypsin. H32 of AbSGR-h probably does not need to act as a base catalyst as strongly as H57 of chymotrypsin, because the aspartate carboxyl group is readily deprotonated than the serine hydroxyl group. Deprotonated side chain of D34 may coordinate stably with Mg of chlorophyll. This coordination can be supposed to destabilize Mg-tetrapyrrole ring interaction, resulting in extraction of Mg from chlorophyll. This is one possible catalytic mechanism of SGR.

When the catalytic mechanism of SGR is investigated, it can also be proposed that D34 side chain functions as an acid catalyst to donate protons to nitrogen of pyrroles coordinating Mg, resulting in exchange of Mg with protons to produce pheophytin. However, it is not likely in case of SGR. Acidic amino acid side chain can function as the acid catalyst as found in the lysozyme glutamate (Malcolm et al., 1989). Lysozyme glutamate is present in a protonated form because of localization in the hydrophobic environment, which is essential to keep the protonated form of the acidic amino acid residues. D34 of AbSGR-h stays in a hydrophilic environment and thus may not be able to act as the acidic catalyst. Altogether, protein structure and enzymatic analysis suggest that SGR removes Mg by coordinating with the deprotonated side chain of aspartate and destabilizing the bond between Mg and pyrrole nitrogen. However, co-crystallization of SGR with the pigment is essential for confirming this predicted reaction mechanism.

2.5 Tables and Figures

Table 5: Primer sequences used for cloning

	Forward (5' → 3')	Reverse (5' → 3')
WT	AAGGAGATATACATATGGATCATCTGAAACCGGAG	GGTGGTGGTGCTCGATTTTGTAATCGCGCAGAATG
D34E	CCACAGCGAAAGCACCGGTGAACTGTTTCT	CGGTGCTTTCGCTGTGGGTCAGGGTATAGC
D34N	CCACAGCAATAGCACCGGTGAACTGTTTCT	CGGTGCTATTGCTGTGGGTCAGGGTATAGC
D34Q	CCACAGCCAGAGCACCGGTGAACTGTTTCT	CGGTGCTCTGGCTGTGGGTCAGGGTATAGC

Table 6: Kinetic parameters of WT and D34E AbSGR-h protein

Enzyme	K_m **(μM)**	k_{cat} **(min^{-1})**
WT	15.91	0.076
D34E	6.00	0.00031

Table 7: Data collection and refinement statistics

Native Data	
X-ray source	SPring-8 BL44XU
Detector	EIGER X 16M
Wavelength (Å)	0.90000
Space group	$P6_122$
Unit cell parameters	
***a, b, c* (Å)**	80.3, 80.3, 224.8
β (°)	120.0
Resolution range (Å)	47.74 – 1.75 (1.84-1.75)
Total number of reflections	288,991 (45,936)
Number of unique reflections	44,094 (6,906)
Multiplicity	6.5 (6.7)
Completeness (%)	99.4 (98.7)
Mean *I*/sigma (*I*)	14.07 (1.59)
***R*-merge (*I*) (%)**	6.1 (115.8)
CC½ (%)	99.9 (82.9)
Refinement	
Resolution range (Å)	39.53 – 1.75
Reflection used	44,289
R_{work} (%)	23.5
R_{free} (%)	26.5
Total number of atoms	44,635
Averaged B-factor (Å ^{2})	148.0
RMSD from ideal values	
bond length (Å)	0.0084
bond angles (deg)	1.4577
Ramachandran plot	
Residues in most favored region (%)	96.7
Residues in allowed (%)	2.7
Residues in disallowed region (%)	0.6

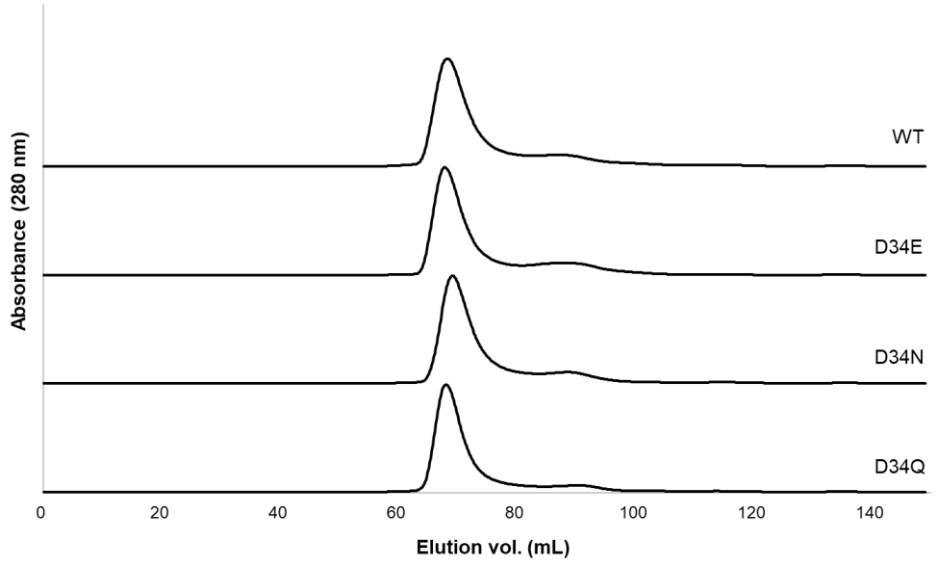

Figure 18: Size exclusion chromatography profiles of AbSGR-h and its mutants. Protein was monitored by the absorbance at 280 nm.

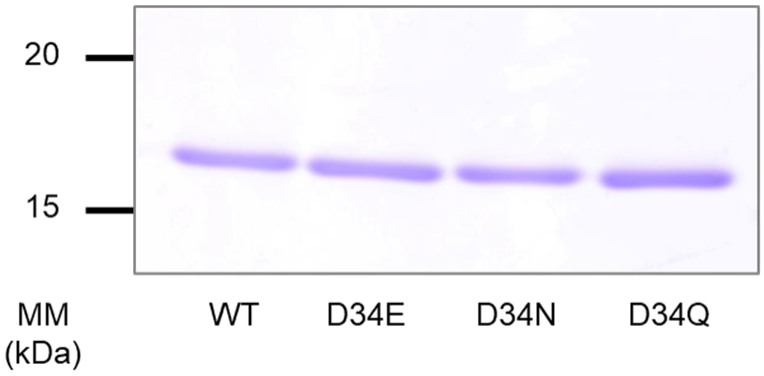

Figure 19: Analysis of purity and molecular weight of purified WT and mutated AbSGR-h proteins, obtained from size exclusion chromatography, by SDS-PAGE.

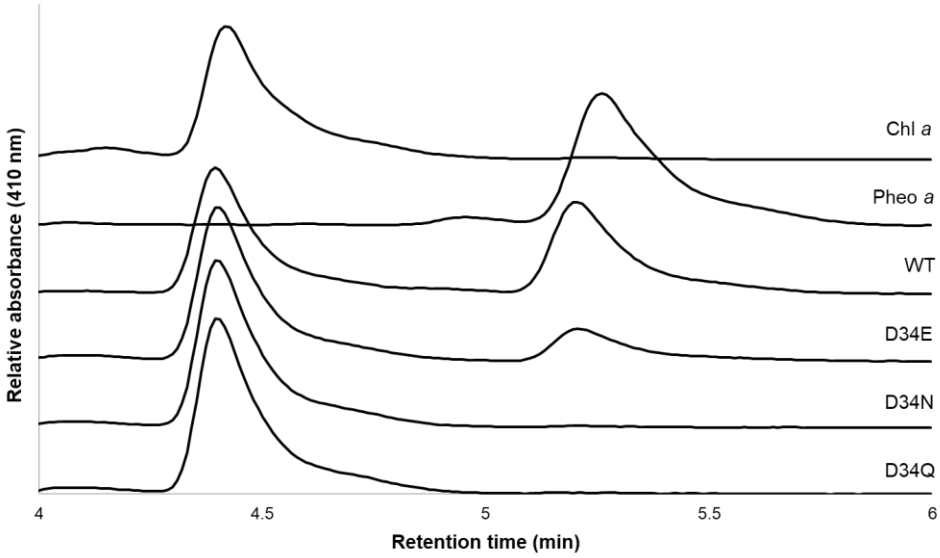

Figure 20: Detection of activity of AbSGR-h. Chlorophyll *a* was incubated with gel filtration derived purified proteins of wild-type or mutant AbSGR-h. After incubation, pigments were extracted and analyzed by HPLC. Chlorophyll *a* and pheophytin *a* peaks are also shown.

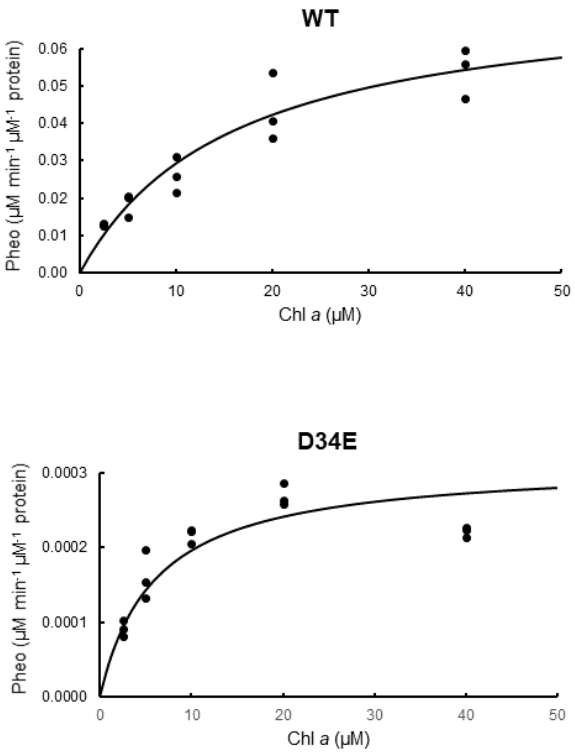

Figure 21: Michaelis-Menten analysis of WT and D34E mutant AbSGR-h protein. Different concentrations (2.5, 5, 10, 20, 40 µM) of Chl *a* was incubated with WT (5 µM) or D34E (50 µM) for 20 min and the accumulation of pheophytin *a* was monitored by HPLC.

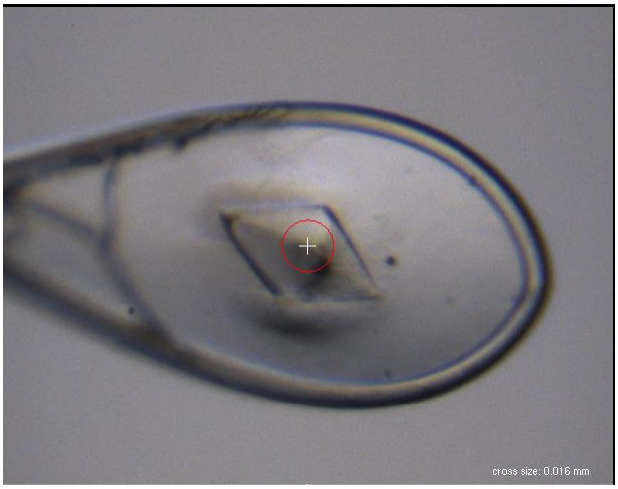

Figure 22: Photograph of SGR crystal. The crystal was about 0.2 mm × 0.1 mm × 0.05 mm, and belongs to the space group $P6_122$ with cell dimensions of a=80.3 Å, b=80.3 Å, c=224.8 Å, β=120.0°.

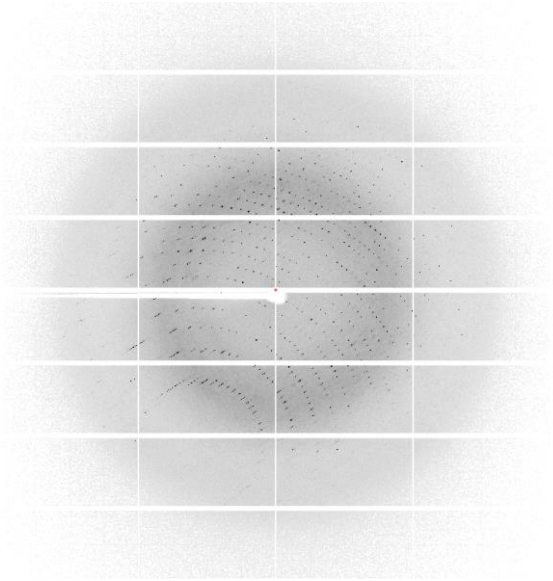

Figure 23: X-ray diffraction pattern of SGR crystal. X-ray experiments were performed at 100K with a EIGER X16M detector (Dectris) on BL44XU beamline at SPring-8 (Harima, Hyogo, Japan).

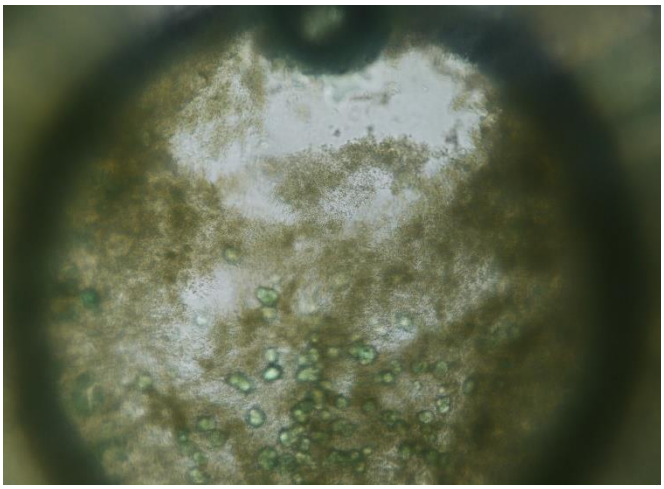

Figure 24: Potential crystals of AbSGR-h bound to Zn-chlorophyllide *a* (ZnChlide), obtained by sitting-drop vapor diffusion method at 298 K where the drop was prepared by mixing equal volumes of SGR and reservoir solution containing 0.1 M HEPES-NaOH (pH 7.5), 28% PEG 400 and 0.2 M $CaCl_2$. The protein (AbSGR-h) and substrate (ZnChlide) concentration used is 10mg/ml and 0.5 mg/ml, respectively.

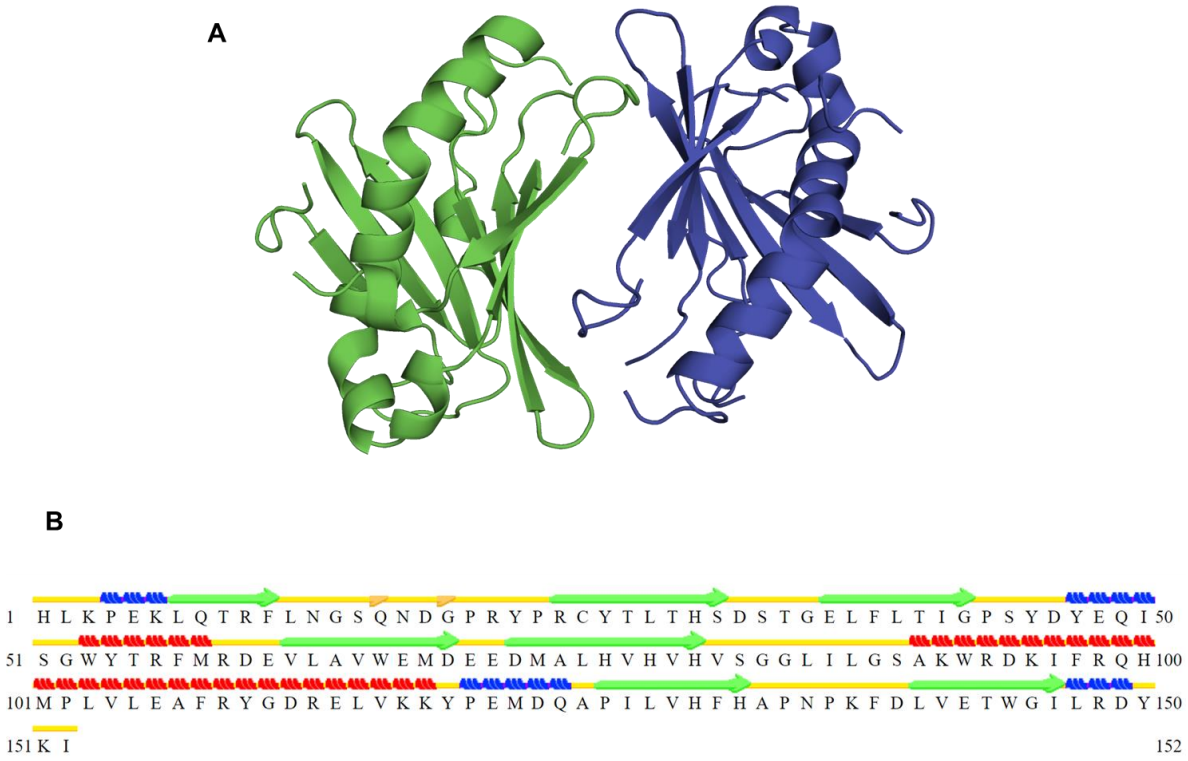

Figure 25: Overall structure of the AbSGR-h protein. (A) Cartoon representation of the AbSGR-h dimer showing chain A (green) and chain B (blue). (B) Secondary structure assignment of AbSGR-h with STRIDE. Red and blue helix shows α-helices and 3-10 helices, respectively. Green arrows indicate β-strands.

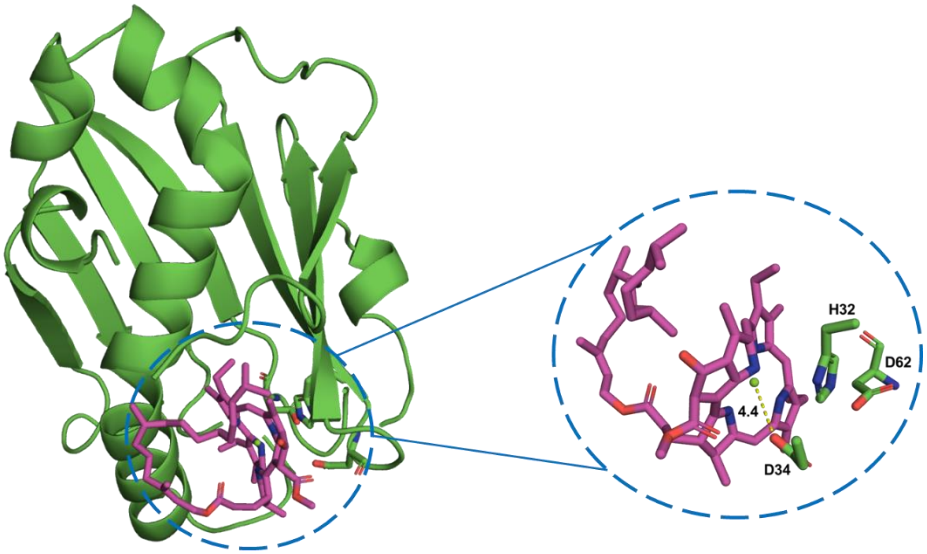

Figure 26: Molecular docking analysis of Chl *a* with AbSGR-h monomer, performed in AutoDock Vina.

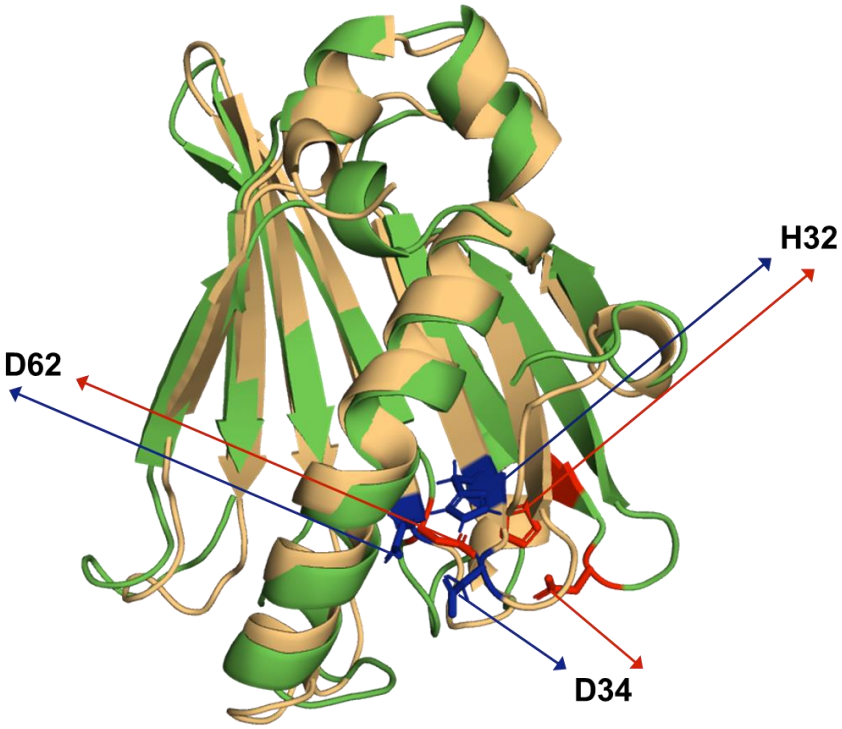

Figure 27: Structural superimposition of the crystal (green) and computationally predicted structure (orange) of AbSGR-h protein. Side chain of H32, D34 and D62 residues for crystal and predicted structure are colored in red and blue, respectively.

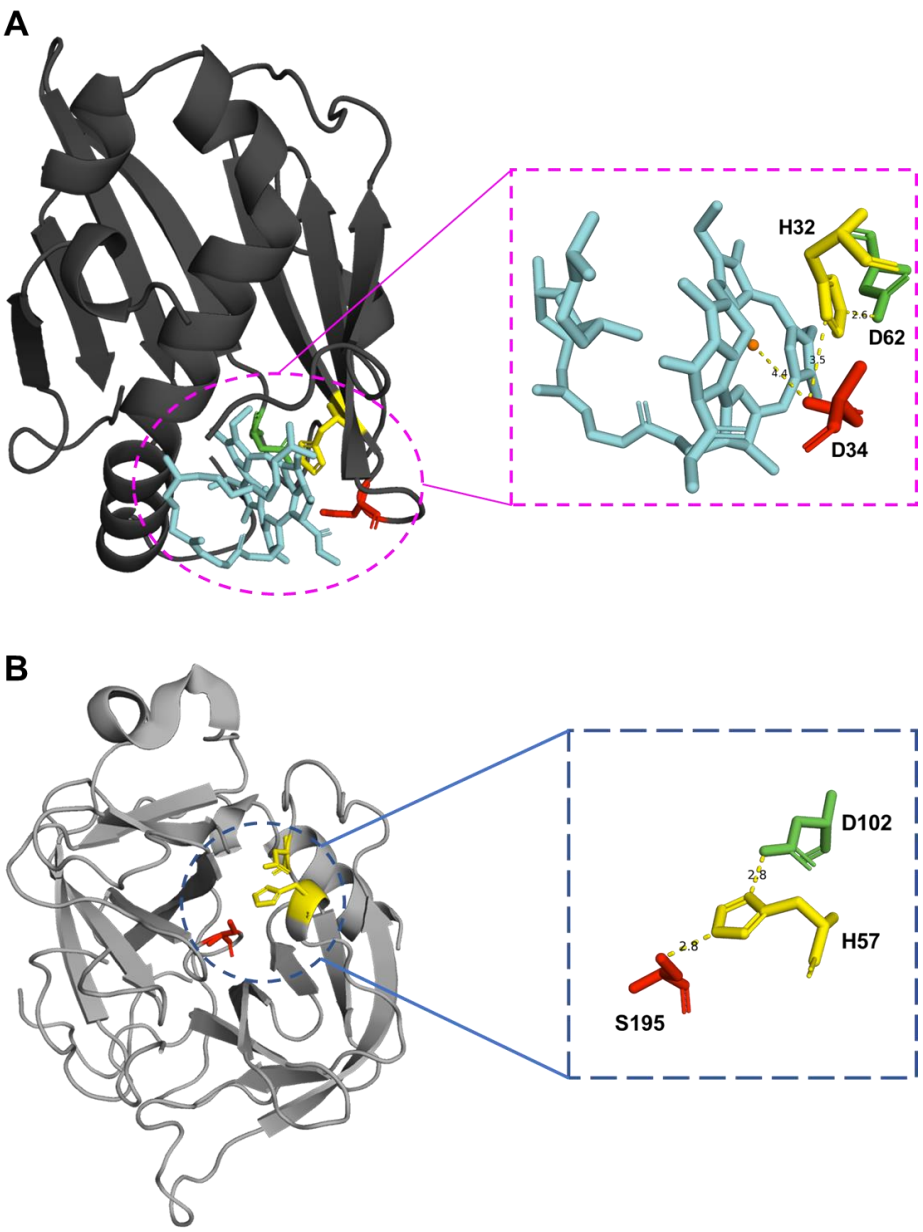

Figure 28: Catalytic triad in the crystal structure of (A) AbSGR-h and (B) Chymotrypsin. The distance between the residues constituting the catalytic triad of each protein have been shown.

Chapter 3

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

3.1 Introduction

Photosynthesis involves two processes – (1) harvesting light energy for the production of ATP and NADPH, and (2) fixation of CO_2 to form sugar using ATP and NADPH (Neilson & Durnford, 2010). Light energy is captured by photosynthetic pigments in light harvesting complexes (LHC), which consists of core and peripheral antenna systems (Green & Durnford, 1996). The core antenna complexes of photosystem I (PSI) and II (PSII) contain P700-chlorophyll *a*-protein complex (CP1) and CP43/CP47, respectively, possessing chlorophyll *a* and β-carotene as the major photosynthetic pigments (Alfonso et al., 1994; Ben-Shem et al., 2003). In addition to chlorophyll *a*, the peripheral antenna complex in land plants also contain chlorophyll *b*, lutein, violaxanthin and neoxanthin as accessory pigments, 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 present not only in land plants and green algae, but also in some cyanobacterial lineages (Tomitani et al., 1999). Although the light-harvesting machineries are quite different among these organisms, they all contain chlorophyll *b* as an important light-harvesting pigment (Kunugi et al., 2013). Additionally, it does not exist as a free pigment but as part of LHCs in the chloroplast (Yamasato et al., 2005). Previous studies have showed that chlorophyll *b* biosynthesis is critical for LHC formation and regulation (A. Tanaka & Tanaka, 2019). Chlorophyll *b* is synthesized from chlorophyll *a*, through conversion of a methyl group at the C7 position to a formyl group via a hydroxymethyl intermediate, in two successive oxygenation reactions catalyzed by the enzyme, chlorophyllide *a* oxygenase (CAO) (Ito et al., 1996; A. Tanaka et al., 1998). CAO is a unique Rieske-mononuclear iron oxygenase that is not structurally related to bciD, the latter being involved in the C7-formylation in bacteriochlorophyll *e* synthesis (Harada et al., 2013).

Interestingly, unlike other chlorophyll metabolic enzymes, the structural organization of CAO varies among photosynthetic organisms (Nagata et al., 2004). Eukaryotic CAOs, excepting that in Mamiellales, are composed of A, B, and C domains in order from the N-terminus (Nagata et al., 2004). The conserved A domain, unique to land plants and most

green algae, has regulatory function that prevents excessive accumulation of the CAO protein (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). Although *Prochlorothrix hollandica* CAO (PhCAO) shares high sequence similarity to land plant CAOs, it consists of only the sequence pertaining to the catalytic domain without any regulatory domain in the N-terminus (Hirashima et al., 2006). However, in Mamiellales such as *Micromonas* and *Ostreococcus*, CAO is encoded by two genes, which appear to have evolved from a single ancestral CAO gene (A. Tanaka & Tanaka, 2019). For instance, *MpCAO1* and *MpCAO2* genes in *Micromonas pusilla* CAO (MpCAO) possess Rieske and mononuclear iron-binding motifs, respectively. 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).

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. Despite the significance and omnipresence of chlorophyll *b* in photosynthetic organisms, it is surprising that the reaction mechanism of CAO still remains poorly investigated. To the best of our knowledge, only one study succeeded in determining the recombinant protein activity of CAO *in vitro* (Oster et al., 2000). Therefore, its structural characterization will help in understanding the enzyme reaction mechanism. In this study, the tertiary as well as quaternary structure of CAO were predicted using deep neural-network based method. In addition, the probable binding cavity for ligand interaction and residues of structural and functional importance have been elucidated.

3.2 Materials and methods

3.2.1 Sequence alignment and phylogenetic analysis

Protein sequences of CAO were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/protein/). Proteins ranging from land plants, green algae and cyanobacteria were considered for the multiple sequence alignment **(Table 8)**. Since Prasinophytes CAO consists of two different polypeptides (MpCAO1 & MpCAO2), two datasets were created separately with other CAO sequences for phylogeny construction. The alignment for individual datasets were performed using Clustal Omega with the default settings (Sievers & Higgins, 2018). Visualization and marking of the conserved residues in the MSA were implemented in Jalview v2.11.1.4 (Waterhouse et al., 2009). The maximum

likelihood phylogenetic trees were constructed using IQ-TREE v1.6.12 with 1000 bootstrap replicates in the ultrafast mode (Hoang et al., 2018; Trifinopoulos et al., 2016), considering dicamba (2-methoxy-3,6-dichlorobenzoic acid) O-demethylase (alternatively known as dicamba monooxygenase or DMO) protein (NCBI Accession ID: Q5S3I3.1) as an outgroup in each case (D'Ordine et al., 2009). The best-fitting amino acid substitution model – WAG + G4, obtained for each dataset, was applied automatically in the IQ-TREE server for deriving the phylogeny (Kalyaanamoorthy et al., 2017). iTOL v6 was used for both visualization and generating the tree figures (Letunic & Bork, 2019).

3.2.2 Tertiary structure modelling and validation

In silico modelling of the two CAO subunits from *Micromonas pusilla* were 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. In addition, tertiary structures of the single subunit CAO protein from the model plant *Arabidopsis thaliana* (AtCAO; Accession ID: AAD54323.1) was also modelled using the RoseTTAFold tool (Minkyung et al., 2021). For each predicted protein structure, the model with the best confidence, as appraised by the template modeling score (TM-score), was considered for further analyses (J. Xu & Zhang, 2010). Each protein model was structurally refined using the GalaxyRefine server (Heo et al., 2013). The stereochemical quality of the refined structures were assessed by Verify3D (Lüthy et al., 1992), PROCHECK (Laskowski et al., 1993) and ERRAT (Colovos & 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 & Sippl, 2007). Graphic modifications, visualization and preparation of final illustrations were performed in PyMOL v. 2 (Delano, 2002). The potential binding cavity on the protein structure was detected using the CavityPlus web server (http://www.pkumdl.cn/cavityplus) (Y. 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 protein was predicted using the COACH server (Yang et al., 2013). 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).

3.2.3 Oligomeric structure prediction

The heterodimeric complex consisting of two subunits of MpCAO were 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), followed by stereochemical quality assessment through Ramachandran plot analysis (Laskowski et al., 1993). Besides, the binding affinity of the protein-protein complexes was determined using the PRODIGY web server (Xue et al., 2016).

3.2.4 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 HyperChemTM 8.0.8 molecular modeling software (Hypercube Inc., Gainesville, FL, USA). 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 & 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.

3.3 Results

3.3.1 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 **(Figure 29)**. Additionally, the protein sequence of DMO, a Rieske-mononuclear iron oxygenase, which shares high sequence homology with CAOs was also considered for the comparison. The protein sequences of CAOs 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 CAOs 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, non-heme iron binding motif and ligand binding site residues, with conservation score above 90%.

Two rooted maximum-likelihood phylogenetic trees revealed distinct clading pattern of MpCAO1 and MpCAO2 proteins with other CAOs in accordance with their taxonomic groups **(Figure 30A and B)**. DMO was treated as an outgroup to establish the clustering pattern of CAOs in the presence of a non-CAO monooxygenase. In spite of possessing significant sequence similarity among different organisms and with another monooxygenase, CAOs did not intermix with its homologs thus maintaining a marked spatial arrangement in the phylogenies.

3.3.2 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 the two subunits of CAO from *Micromonas pusilla* were modelled using the D-I-TASSER tool. The D-I-TASSER protocol integrates deep convolutional neural network-based distance and hydrogen-bonding network prediction to drive the assembly of template fragments into full-length model by Replica-Exchange Monte Carlo (REMC) simulations. 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 **(Figure 31)**. Furthermore, the tertiary structure of *Arabidopsis* CAO (AtCAO) was determined with the RoseTTAFold server as I obtained comparatively low confidence for AtCAO using the D-I-TASSER tool. However, I modelled the entire sequence of AtCAO, excepting the signal peptide region, and a confidence of 0.78 was obtained.

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 MpCAO1, MpCAO2 and AtCAO. Ramachandran plots revealed that the predicted models follow all the stereochemical properties with favourable phi (Φ) and psi (Ψ) values. Besides, ERRAT and Verify3D confirmed the high global quality of the structural models **(Table 9)**. 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 MpCAO2 and AtCAO 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 MpCAO2 and AtCAO have been marked in the multiple sequence alignment with asterisk **(Figure 29)**. It is to be noted that 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 **(Figure 32)**.

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 non-heme iron **(Figure 33)**. On the other hand, four conserved amino acids – C28, H30, C47, H50 (for MpCAO1) and C262, H264, C281, 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 34)**.

3.3.3 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 adjacent subunit for carrying out its catalytic reaction. Indeed, recombinant AtCAO are found to exist in oligomeric forms, such as single, double or triple trimers, under non-denaturing conditions (Kunugi et al., 2013). In this study, the trimeric organization of AtCAO were predicted using the GalaxyHomomer tool **(Figure 35A)**. 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 non-heme iron center of the neighboring subunit is ~12.0 Å. These positioning and distances are adequate for electron transfer and catalysis and 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*. A previous study has shown that simultaneous expression of both proteins in a chlorophyll *b*-less *Arabidopsis* mutant compliments its chlorophyll *b*

deficiency, while transformation with either of them does not compliment the phenotype (Kunugi et al., 2013). The heterodimeric complex of MpCAO1 and MpCAO2 proteins were derived using the GalaxyHeteromer program **(Figure 35B)**. Here also, Ramachandran plot analysis of the refined heterodimer complex, carried out in PROCHECK server, revealed less than one percent residues to be in the outlier region. 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 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.

3.3.4 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. It is known 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 MpCAO2 and AtCAO 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 a close proximity to the mononuclear iron unit **(Figure 36A and B)**. The lowest energy conformations for each protein-ligand docked pair was 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 demonstrate a common chlorophyll *a* binding mode for all CAOs from different organisms.

3.4 Discussion

3.4.1 Structural comparison with Rieske mononuclear iron oxygenases

Rieske mononuclear iron oxygenase catalyzes a variety of complex oxidation reactions (Bugg & Ramaswamy, 2008; Perry et al., 2018). 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 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 analysed computationally, where the distance between the Rieske cofactor of one subunit and 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 also in agreement with other Rieske oxygenases.

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 investigation **(Figure 37)**. 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 monooxygenases. The distance between 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 are less than 4 Å and the arrangement of this electron transport system is well conserved among Rieske oxygenases **(Figure 37)**. Taken together, the structure for substrate oxidation is conserved even though it is formed with two distinct polypeptides.

3.4.2 Separation and unification of the components involved in electron transport

In this study, I 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. *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 electron 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 in the FrhG subunit of F_{420}-reducing hydrogenase as previously discussed (Wang & Liu, 2016), while BciB contains the same Fe-S cluster on its own polypeptide **(Figure 38)**. It is not known how BciB obtained Fe-S cluster at a similar position like 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.

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 demonstrate 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.

3.5 Tables and Figures

Table 8: Details of the species containing the CAO protein.

Organism	Abbreviation	Accession ID	Taxon
Arabidopsis thaliana	AtCAO	AAD54323.1	Land plant
Oryza sativa subsp. japonica	OsCAO	Q8S7E1.1	Land plant
Zea mays	ZmCAO	NP_001142071.2	Land plant
Physcomitrium patens	PpCAO	XP_024404225.1	Land plant
Chlamydomonas reinhardtii	CrCAO	BAA33964.1	Green algae
Volvox carteri f. nagariensis	VcCAO	EFJ44919.1	Green algae
Micromonas pusilla CCMP1545	MpCAO1	EEH58153.1	Green algae
Micromonas pusilla CCMP1545	MpCAO2	EEH54524.1	Green algae
Prochlorothrix hollandica	PhCAO	BAD02269.1	Cyanobacteria
Acaryochloris sp. RCC1774	Asp. CAO	WP_110987895.1	Cyanobacteria

Table 9: Model evaluation analyses for MpCAO1, MpCAO2, and AtCAO obtained from the SAVES and ProSA-web server.

Protein	ERRAT	Verify3D	Ramachandran plot analysis				ProSA Z-score
			Most favored	Additionally allowed	Generously allowed	Disallowed	
MpCAO1	89.4309	82.35	86.3	12.0	0.5	1.2	-6.36
MpCAO2	78.877	89.49	85.8	12.0	1.8	0.4	-6.54
AtCAO	86.0515	79.71	87.7	11.1	1.0	0.2	-8.78

```
AtCAO       PYSPHLKNFWYPVAFTADLKH-DTMVPIECFEQPWVIFRGEDGKPGCVRNTCAHRACPLD  270
OsCAO       PYNPSLKNFWYPVAFSSDLKD-DTMVPIDCFEEQWVIFRGKDGRPGCVMNTCAHRACPLH  269
ZmCAO       PYNPSLKNFWYPVAFSSDLKAPDTMVPIDCFEEQWVIFRGKDGRPGCVQNTCAHRACPLH  263
PpCAO       PYPAQLKNYWYPVAFSADIDD-KTMVPFNSFEEAWVIFRGKDGRPGCVRDSCAHRACPLS  331
CrCAO       EIEEGLRNFWYPAEFSARLPK-DTLVPFELFGEPWVMFRDEKGQPSYIRDECAHRGCPLS  172
VcCAO       DFEPGLRNFWYPAEFSAKLGQ-DTLVPFELFGEPWVLFRDEKGQPACIKDECAHRACPLS  364
MpCAO1      ----------------------MIPFDLFNVPWVAFRDQDGMAGCIKDECAHRACPIS  36
MpCAO2      SDVGAIRNYWYPIHFISKLNKGDAATSFVLFGERWELVADDDAAVAAAKTAVGVFGPE--  99
PhCAO       DLANGLRNFWYPVEFSKNLGM-ADPLGFELFDQCWVLFRDDQGTAACILDECAHRACPLS  70
Asp.CAO     LLASGLKDYWYAVEFSSKLQD-ATLISFELFDQPWVLFRDRQGQVGCIQDECAHRACPLS  81
DMO         --MTFVRNAWYVAALPEELSE--KPLGRTILDTPLALYRQPDGVVAALLDICPHRFAPLS  56
                                                              o o

AtCAO       LGTVNEGRIQCPYHGWEYSTDGECKKMPSTKL-L-K-VKIKSLPCLEQEGMIWIWPGDEP  327
OsCAO       LGSVNEGRIQCPYHGWEYSTDGKCEKMPSTKM-L-N-VRIRSLPCFEQEGMVWIWPGNDP  326
ZmCAO       LGSVNEGRIQCPYHGWEYSTDGKCEKMPSTKM-L-D-VRIQSLPCFEQEGMVWIWPGDDP  320
PpCAO       LGKVEEGRIQCPYHGWEYNTSGKCEKMPSTRF-V-N-AKLDSLPCIEQDGMVWIWPGNET  388
CrCAO       LGKVVEGQVMCPYHGWEFNGDGACTKMPSTPF-C-RNVGVAALPCAEKDGFIWVWPGDGL  230
VcCAO       LGKVVEGQVVCAYHGWEFNGDGHCTKMPSTPH-C-RNVGVSALPCAEKDGFIWVWPGDGL  422
MpCAO1      LGKVVEGRVQCPYHGWEYTSGGECKKMPSIKNLL-PNVYVDAAPIVERDGLLYVWAGVWE  95
MpCAO2      YAETQAHLVDGAAQRWTCRSRD---------------DATRFLPIGLQDGLV-------M  137
PhCAO       LGKVIQGRIQCPYHGWEYDRQGECVHMPSCQA-I-S-NPILTLPVMEQGGMIWVWPGTDE  127
Asp.CAO     LGQVVDGTVQCGYHGWQYDASGSCTHMPSCQH-I-Q-VQIKSLPCQEQNGMIWVWPGSAQ  138
DMO         DGILVNGHLQCPYHGLEFDGGGQCVHNPHGNGARPASLNVRSFPVVERDALIWIWPGDPA  116
                      o  o

AtCAO       PAPIL----------PSL-QPPSGFLIHAELVM-DLPVEHGLLLDNLLDLAHAPFTHTS  374
OsCAO       PKSTI----------PSL-LPPSGFTIHAEIVM-ELPVEHGLLLDNLLDLAHAPFTHTS  373
ZmCAO       PKATI----------PSL-LPPSGFTVHAEIVM-ELPVEHGLLLDNLLDLAHAPFTHTS  367
PpCAO       PSTNL----------PCL-NPPSHYTIHAQITM-ELPVEHGLLVENLLDLAHAPFTHTT  435
CrCAO       PAETL----------PDFAQPPEGFLIHAEIMV-DVPVEHGLLIENLLDLAHAPFTHTS  278
VcCAO       PAQTL----------PDFARPPEGFQVHAEIMV-DVPVEHGLLMENLLDLAHAPFTHTT  470
MpCAO1      PERAE---EILSELPPSAATAPPSGFAAMAEVTV-DVPLDAPAILSRLMDENKVPFTRVD  151
MpCAO2      PDVAL----------PTTFTPPAGYTTHAELIIEDVPVEHGLLMENLLDLAHAPFTHTG  186
PhCAO       PGALP----------SLAPTLPDNFTLQAELVM-DLEVEHGLMLENLLDLAHAPFTHTG  175
Asp.CAO     PTELS----------EHIYQLPEGFQLHAEVAM-ELPVEHGLLLENLLDLAHAPFTHTG  186
DMO         LADPGAIPDFGCRVDP--AYRTVGGY--------GHVDCNYKLLVDNLMDLGHAQYVHRA  166
                                                  □***▲* □*    □
```

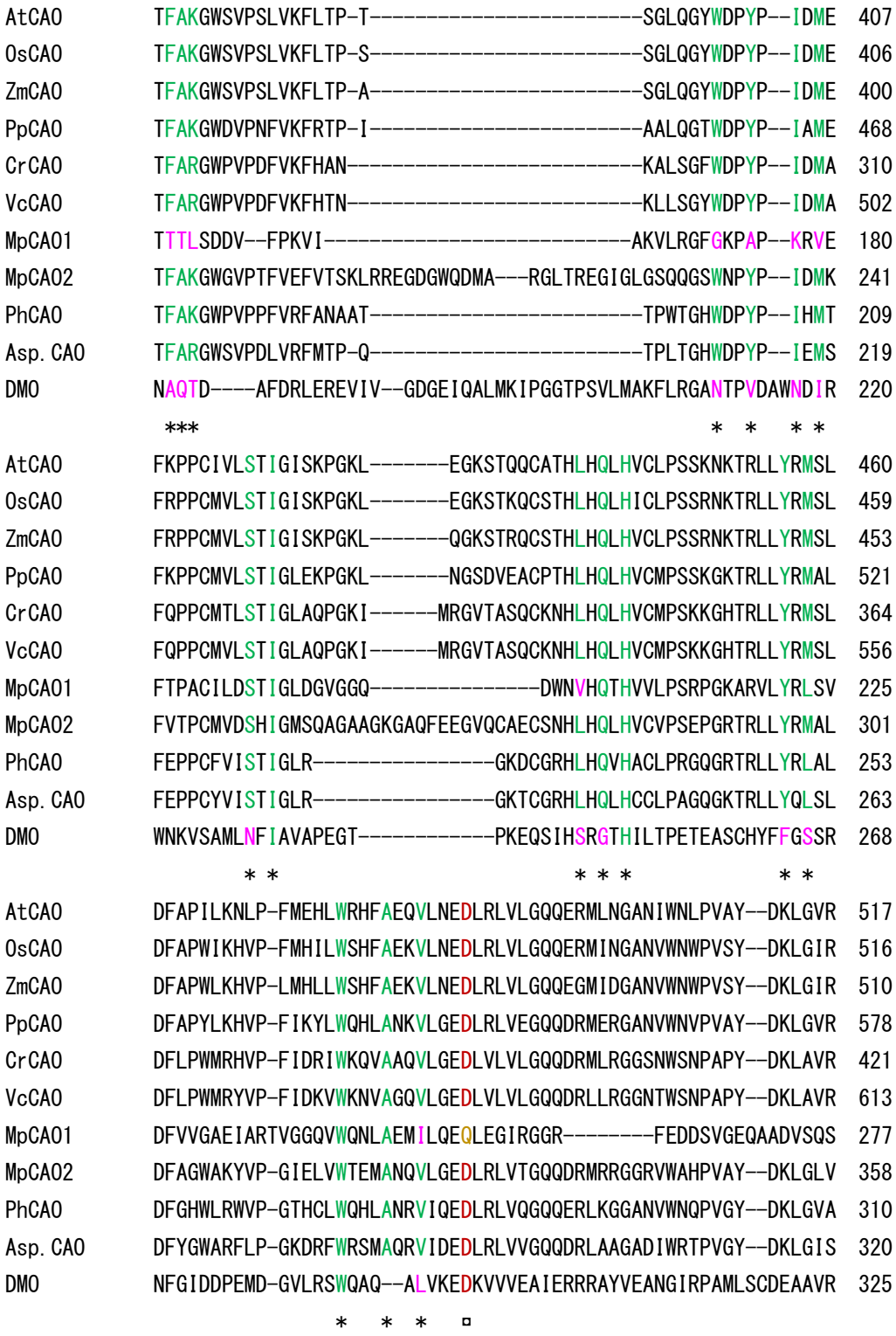

Figure 29: Multiple sequence alignment of CAO proteins from different organisms and dicamba monooxygenase (DMO) using Clustal Omega. The MSA has been marked with a user defined color code. Conserved residues involved in interaction with Rieske [2Fe-2S]

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 essential ‘gatekeeper’ role by transferring electrons through CAO subunit interface is colored in red.

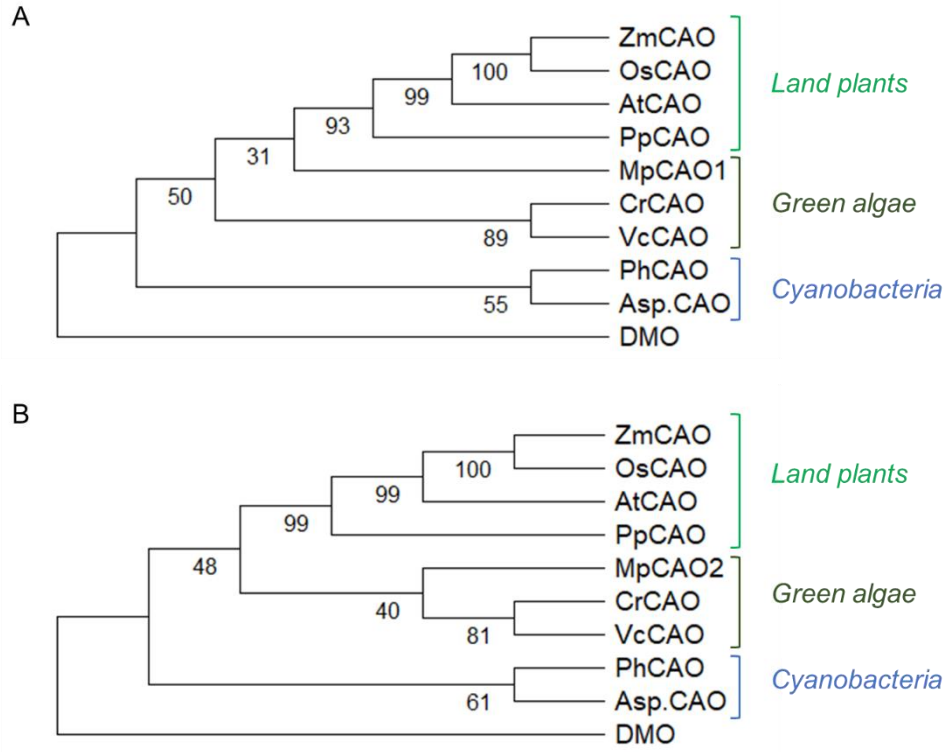

Figure 30: Rooted maximum likelihood trees showing phylogenetic relationship of (A) MpCAO1 and (B) MpCAO2 with other CAO protein sequences. Bootstrap values, based on 1000 replicates, have been shown on branch nodes. Dicamba monooxygenase (DMO) was used as outgroup in both phylogenies.

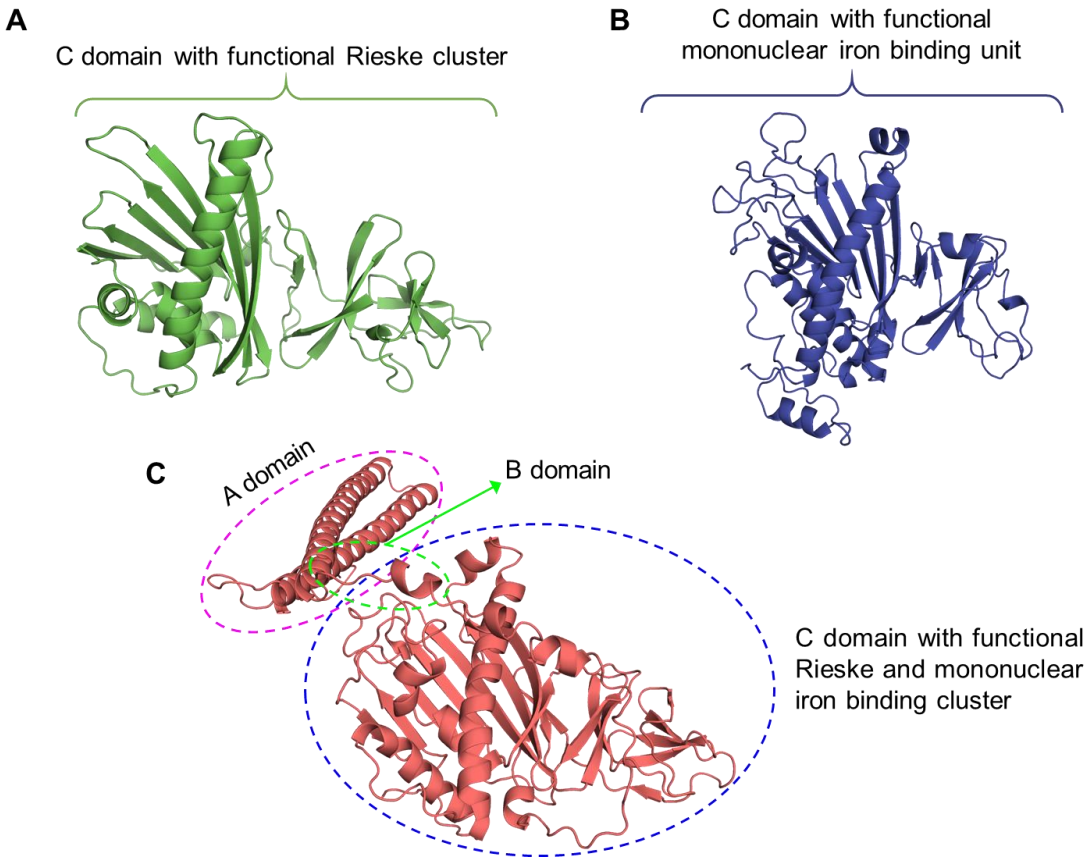

Figure 31: Cartoon representation of the predicted three-dimensional structures of (A) MpCAO1, (B) MpCAO2, and (C) AtCAO. The position of the domain(s) in individual structures has been shown.

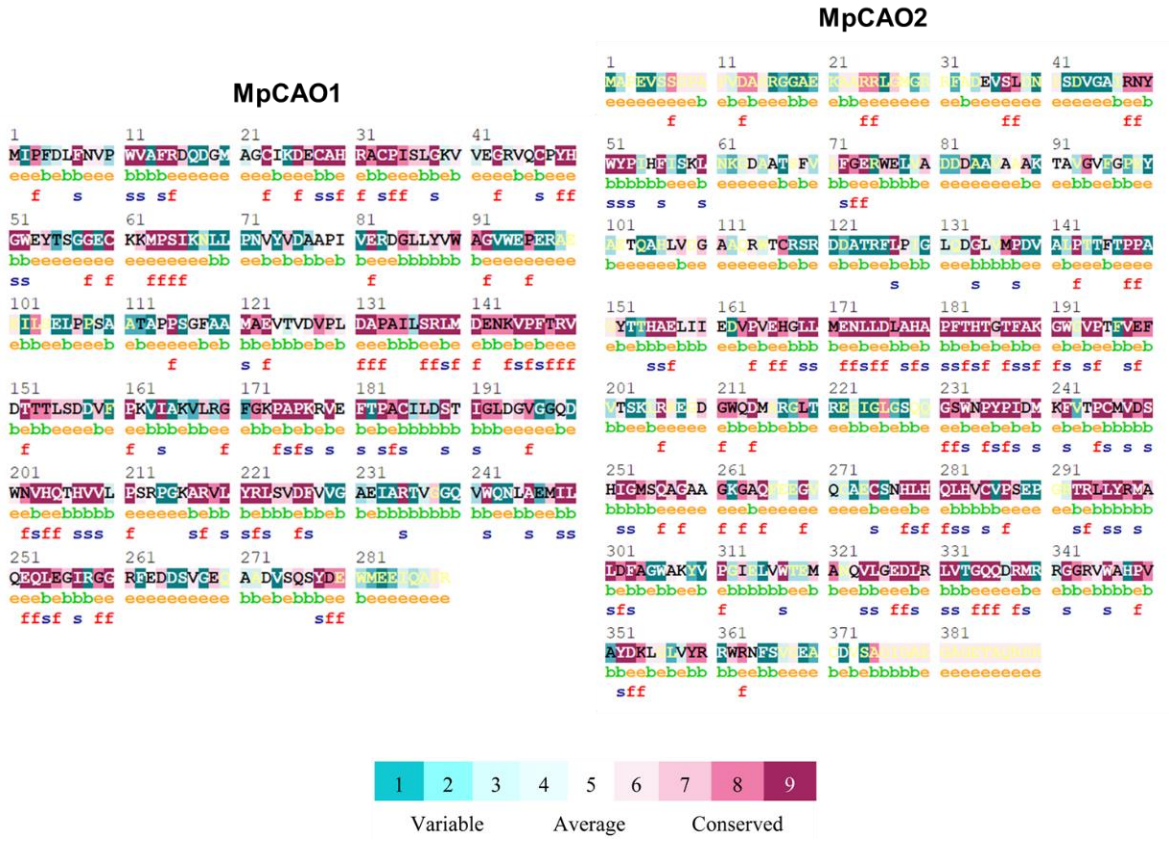

Figure 32: Evolutionary conservation of amino acid residues in the primary sequence of MpCAO1, MpCAO2 and AtCAO by ConSeq analysis. Note: ‘e’ refers to an exposed residue according to the neural-network algorithm; ‘b’ refers to a buried residue according to the neural-network algorithm; ‘f’ refers to a predicted functional residue (highly conserved and exposed); ‘s’ refers to a predicted structural residue (highly conserved and buried).

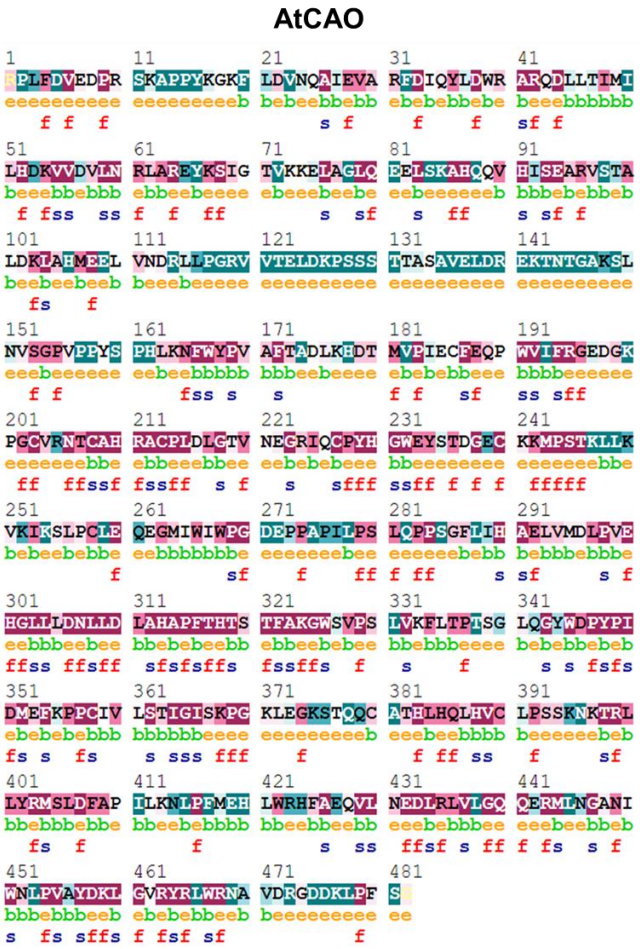

Figure 32 (continued)

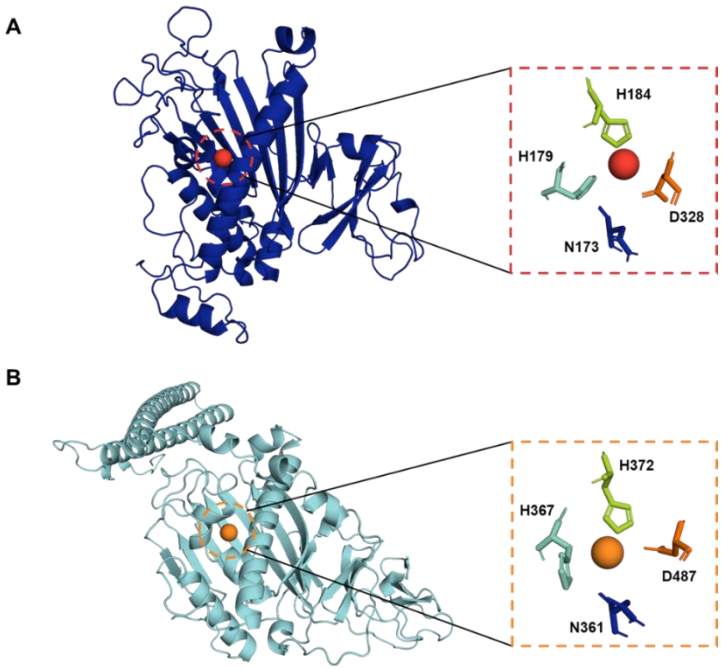

Figure 33: Non-heme iron center in (A) MpCAO2 and (B) AtCAO. The residues interacting with the Fe ion have been shown for each protein.

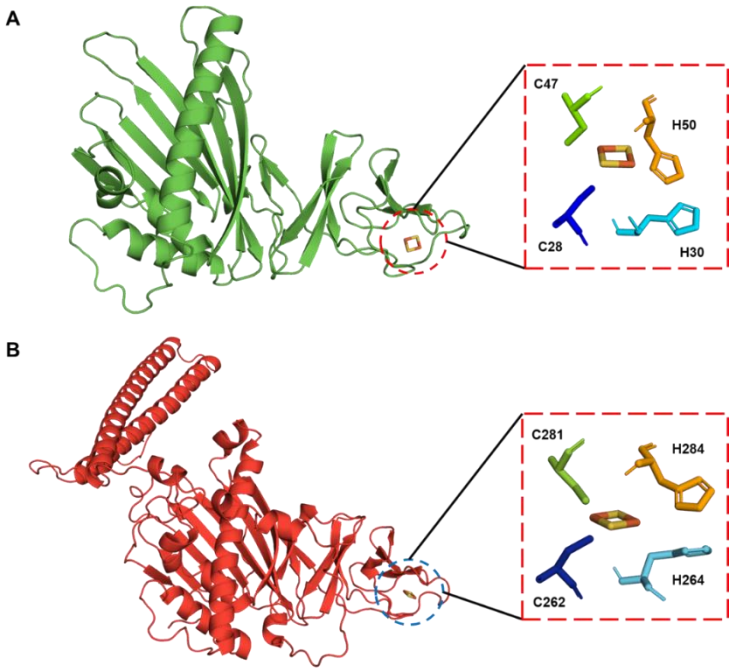

Figure 34: Rieske [2Fe-2S] cluster in (A) MpCAO1 and (B) AtCAO. The residues involved in interaction with Rieske unit have been shown for each protein.

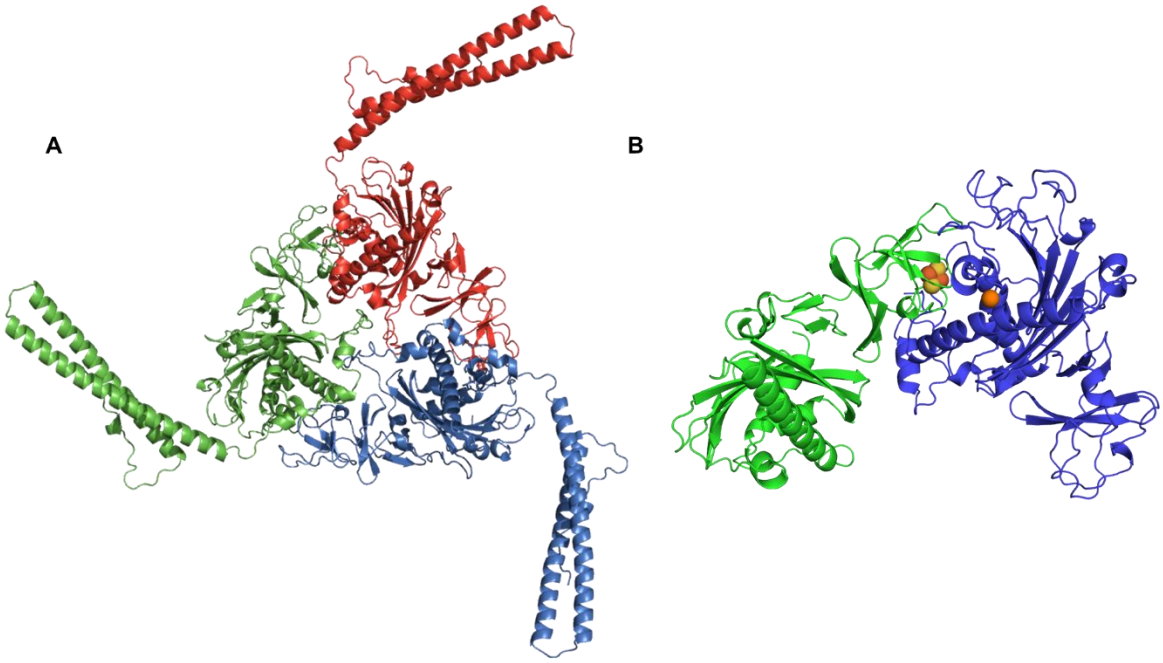

Figure 35: Cartoon representation of (A) AtCAO homotrimer and (B) MpCAO1 (green)-MpCAO2 (blue) heterodimer. The cofactors in MpCAO heterodimer are shown as spheres.

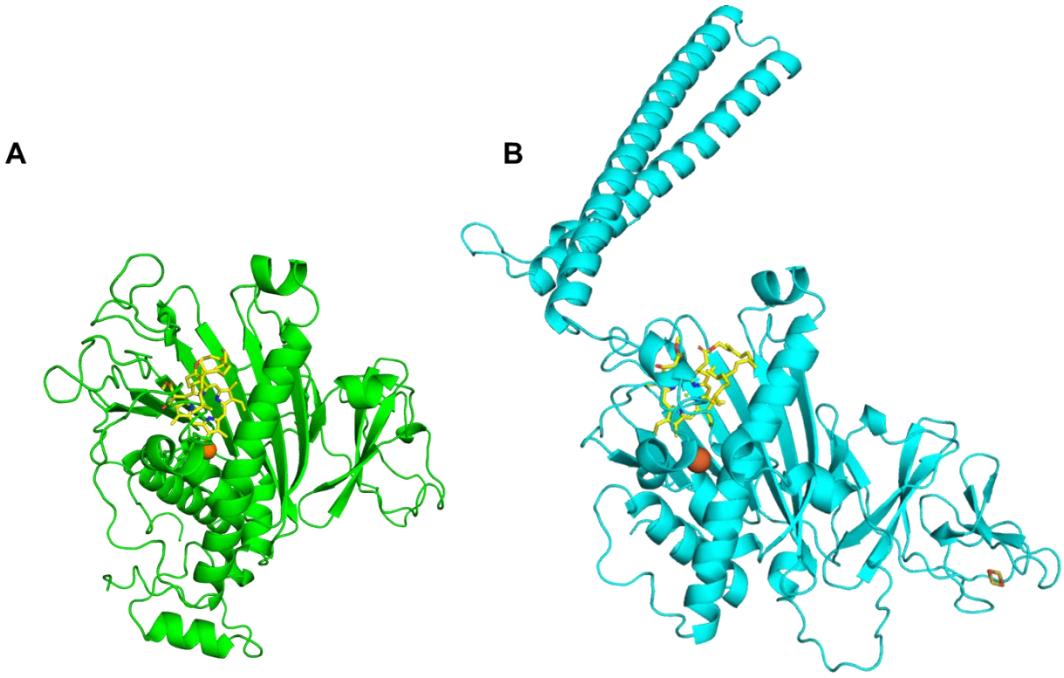

Figure 36: Docked structure of chlorophyll *a* with (A) MpCAO2 (green) and (B) AtCAO (cyan) monomers. The substrate is depicted as sticks (yellow) and the mononuclear iron is shown as sphere (orange).

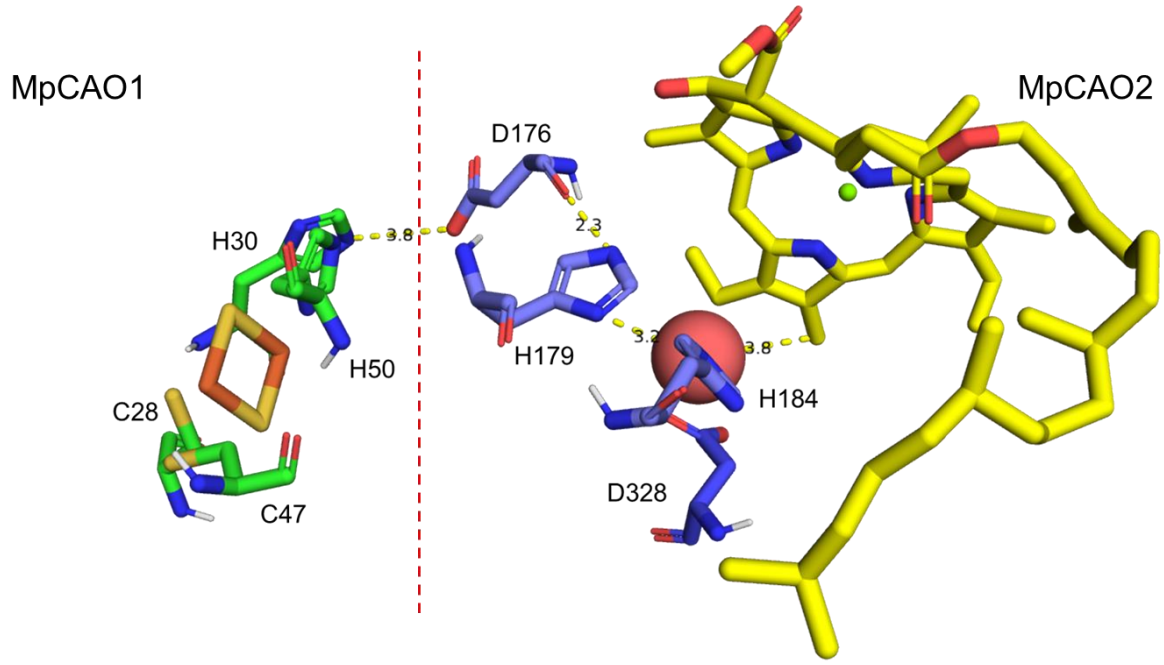

Figure 37: Hypothetical electron transfer pathway between the two subunits of *Micromonas* CAO. The red dashed line indicates the interface between the two subunits. The Rieske unit (orange-red) and the ligand (yellow) are represented as sticks while the mononuclear iron (red) is shown as sphere.

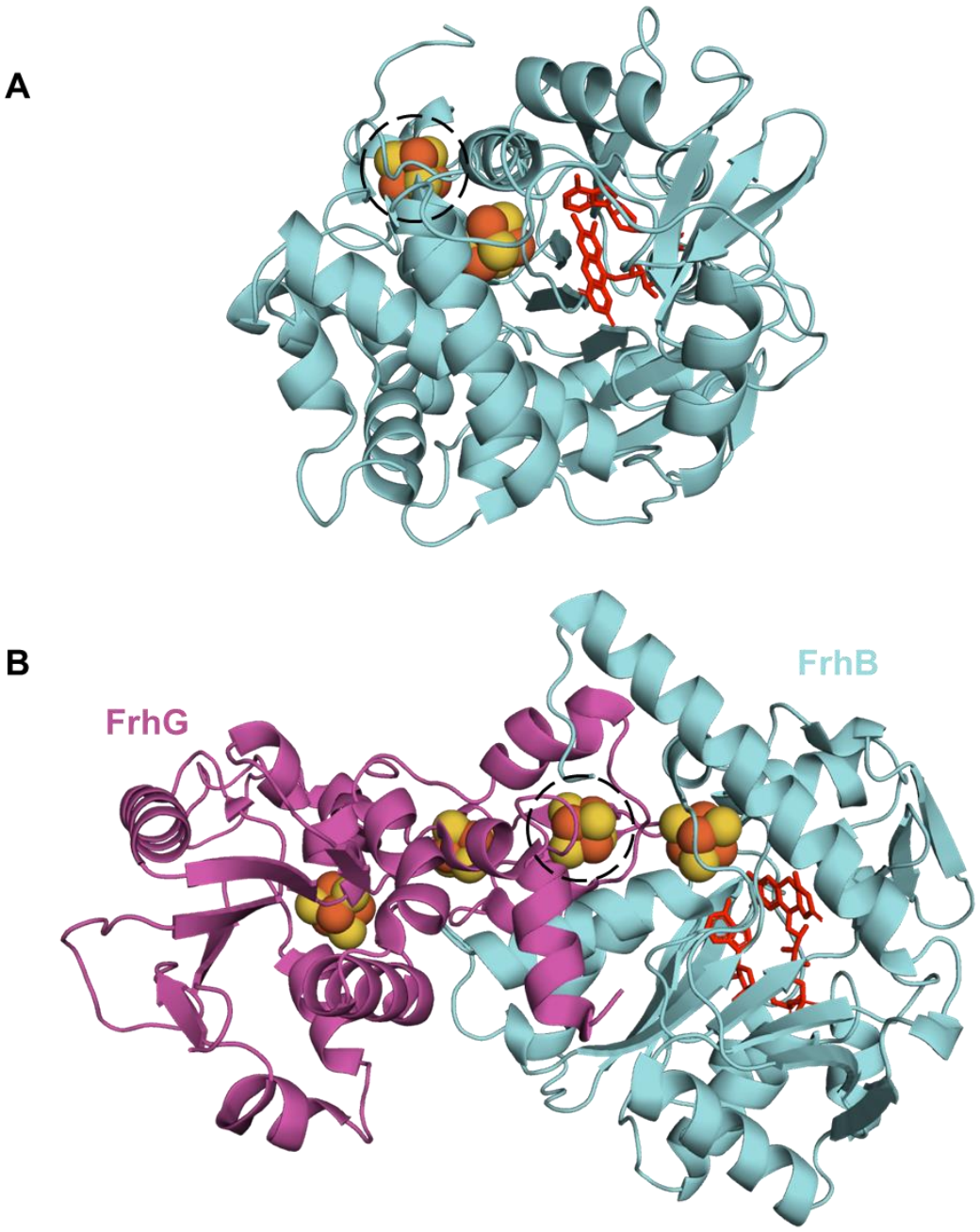

Figure 38: Different localization of the second Fe-S cluster (shown as spheres inside dotted lines) between (A) BciB and (B) F_{420}-reducing hydrogenase. FAD is shown in red as sticks.

Acknowledgements

I would like to express my sincere gratitude to Prof. Ryouichi Tanaka, my supervisor, for his continuous support and guidance from my very first day in Sapporo. You have been an excellent guardian, Ryouichi sensei! I am very much indebted to Dr. Hisashi Ito for his constant guidance and cooperation with me. Thank you, Ito sensei for always supporting and motivating me. My heartfelt thanks to Prof. Ayumi Tanaka for being an inspiration. I am grateful to Dr. Atsushi Takabayashi and Dr. Kiyomi Ono for their support. I want to acknowledge Ms. Junko Kishimoto for always helping me. Besides, I would like to thank Prof. Fujita and Prof. Katsu for reviewing my thesis with their insightful comments and suggestions. My sincere thanks to Prof. Genji Kurisu and Dr. Hideaki Tanaka from the Protein Crystallography Lab of Osaka University for their assistance and collaboration in my PhD study. Lastly, thank you very much Hokkaido University and the Plant Adaptation Biology Lab for providing me this opportunity to conduct my doctoral study.

Furthermore, I am very grateful to The Ministry of Education, Culture, Sports, Science and Technology (MEXT), for granting me the MEXT scholarship for three and a half years (April 2019 – September 2022). I am thankful to the members of Graduate School Educational Affairs Section of Hokkaido University for their constant help during my PhD tenure.

I would like to express my heartiest gratitude to my parents (Mr. Narendranath Dey & Mrs. Sreeparna Dey and Mr. Bachchu Dhar & Mrs. Mili Dhar) for their immeasurable contributions. Thank you won't be adequate to describe the affection and support you have always shown me maa! This PhD would have been incomplete without your constant help and blessings.

My deepest gratitude to Dr. Dipanjana Dhar for always being there by me. A simple thanks is not enough to describe your contributions. The stay in Japan and my PhD won't have been possible without you. A big thank you for all your love, care, support and most importantly, your homemade meals!

My sincerest gratitude to Dr. Soumalee Basu, Head of the Department of Microbiology, University of Calcutta for her unparalleled support. Thank you so much Ma'am for being my true guardian. Without you, nothing would have been possible.

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/https://doi.org/10.1016/j.softx.2015.06.001

Adams, N. B. P., Bisson, C., Brindley, A. A., Farmer, D. A., Davison, P. A., Reid, J. D., & Hunter, C. N. (2020). The active site of magnesium chelatase. *Nature Plants*, *6*(12), 1491–1502. https://doi.org/10.1038/s41477-020-00806-9

Alfonso, M., Montoya, G., Cases, R., Rodríguez, R., & Picorel, R. (1994). Core antenna complexes, CP43 and CP47, of higher plant photosystem II. Spectral properties, pigment stoichiometry, and amino acid composition. *Biochemistry*, *33*(34), 10494–10500. https://doi.org/10.1021/bi00200a034

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, *44*(W1), W344–W350. https://doi.org/10.1093/nar/gkw408

Aubry, S., Christ, B., Kräutler, B., Martinoia, E., Thomas, H., & Zipfel, C. (2021). An evergreen mind and a heart for the colors of fall. *Journal of Experimental Botany*, *72*(13), 4625–4633. https://doi.org/10.1093/jxb/erab162

Axelsson, E., Lundqvist, J., Sawicki, A., Nilsson, S., Schröder, I., Al-Karadaghi, S., Willows, R. D., & Hansson, M. (2006). Recessiveness and Dominance in Barley Mutants Deficient in Mg-Chelatase Subunit D, an AAA Protein Involved in Chlorophyll Biosynthesis. *The Plant Cell*, *18*(12), 3606–3616. https://doi.org/10.1105/tpc.106.042374

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V, van Dijk, A. A., Ebrecht, A. C., … Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, eabj8754. https://doi.org/10.1126/science.abj8754

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 Research*, *45*(W1), W320–W324. https://doi.org/10.1093/nar/gkx246

Bailey, S., Walters, R. G., 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*(5), 794–801. https://doi.org/10.1007/s004250100556

Ben-Shem, A., Frolow, F., & Nelson, N. (2003). Crystal structure of plant photosystem I.

Nature, *426*(6967), 630–635. https://doi.org/10.1038/nature02200

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 (Oxford, England)*, *20*(8), 1322–1324. https://doi.org/10.1093/bioinformatics/bth070

Bryant, D. A., Hunter, C. N., & Warren, M. J. (2020). Biosynthesis of the modified tetrapyrroles-the pigments of life. *The Journal of Biological Chemistry*, *295*(20), 6888–6925. https://doi.org/10.1074/jbc.REV120.006194

Bugg, T. D. H., & Ramaswamy, S. (2008). Non-heme iron-dependent dioxygenases: unravelling catalytic mechanisms for complex enzymatic oxidations. *Current Opinion in Chemical Biology*, *12*(2), 134–140. https://doi.org/https://doi.org/10.1016/j.cbpa.2007.12.007

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. *The Journal of Biological Chemistry*, *276*(38), 35924–35933. https://doi.org/10.1074/jbc.M105199200

Chen, M. (2014). Chlorophyll modifications and their spectral extension in oxygenic photosynthesis. *Annual Review of Biochemistry*, *83*, 317–340. https://doi.org/10.1146/annurev-biochem-072711-162943

Chen, Y., Shimoda, Y., Yokono, M., Ito, H., & Tanaka, A. (2019). Mg-dechelatase is involved in the formation of photosystem II but not in chlorophyll degradation in Chlamydomonas reinhardtii. *The Plant Journal : For Cell and Molecular Biology*, *97*(6), 1022–1031. https://doi.org/10.1111/tpj.14174

Chen, Y., Yamori, W., Tanaka, A., Tanaka, R., & Ito, H. (2021). Degradation of the photosystem II core complex is independent of chlorophyll degradation mediated by Stay-Green Mg2+ dechelatase in Arabidopsis. *Plant Science*, *307*, 110902. https://doi.org/https://doi.org/10.1016/j.plantsci.2021.110902

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science : A Publication of the Protein Society*, *2*(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

D'Ordine, R. L., Rydel, T. J., Storek, M. J., Sturman, E. J., Moshiri, F., Bartlett, R. K., Brown, G. R., Eilers, R. J., Dart, C., Qi, Y., Flasinski, S., & Franklin, S. J. (2009). Dicamba monooxygenase: structural insights into a dynamic Rieske oxygenase that catalyzes an exocyclic monooxygenation. *Journal of Molecular Biology*, *392*(2), 481–497. https://doi.org/10.1016/j.jmb.2009.07.022

DELANO, W. L. (2002). The PyMOL Molecular Graphics System. *Http://Www.Pymol.Org*. https://ci.nii.ac.jp/naid/10020095229/en/

Delmas, F., Sankaranarayanan, S., Deb, S., Widdup, E., Bournonville, C., Bollier, N., Northey, J. G. B., McCourt, P., & Samuel, M. A. (2013). ABI3 controls embryo

degreening through Mendel's I locus. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(40), E3888-94. https://doi.org/10.1073/pnas.1308114110

Ekici, O. D., Paetzel, M., & Dalbey, R. E. (2008). Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Science : A Publication of the Protein Society*, *17*(12), 2023–2037. https://doi.org/10.1110/ps.035436.108

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography*, *60*(Pt 12 Pt 1), 2126–2132. https://doi.org/10.1107/S0907444904019158

Felline, A., Seeber, M., & Fanelli, F. (2020). webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules. *Nucleic Acids Research*, *48*(W1), W94–W103. https://doi.org/10.1093/nar/gkaa397

Ferraro, D. J., Gakhar, L., & Ramaswamy, S. (2005). Rieske business: structure-function of Rieske non-heme oxygenases. *Biochemical and Biophysical Research Communications*, *338*(1), 175–190. https://doi.org/10.1016/j.bbrc.2005.08.222

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. *Journal of Molecular Biology*, *342*(3), 1041–1052. https://doi.org/10.1016/j.jmb.2004.07.062

Gakhar, L., Malik, Z. A., Allen, C. C. R., Lipscomb, D. A., Larkin, M. J., & Ramaswamy, S. (2005). Structure and increased thermostability of Rhodococcus sp. naphthalene 1,2-dioxygenase. *Journal of Bacteriology*, *187*(21), 7222–7231. https://doi.org/10.1128/JB.187.21.7222-7231.2005

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(Database issue), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Green, B. R., & Durnford, D. G. (1996). THE CHLOROPHYLL-CAROTENOID PROTEINS OF OXYGENIC PHOTOSYNTHESIS. *Annual Review of Plant Physiology and Plant Molecular Biology*, *47*, 685–714. https://doi.org/10.1146/annurev.arplant.47.1.685

Guyer, L., Salinger, K., Krügel, U., & Hörtensteiner, S. (2018). Catalytic and structural properties of pheophytinase, the phytol esterase involved in chlorophyll breakdown. *Journal of Experimental Botany*, *69*(4), 879–889. https://doi.org/10.1093/jxb/erx326

Harada, J., Mizoguchi, T., Satoh, S., Tsukatani, Y., Yokono, M., Noguchi, M., Tanaka, A., & Tamiaki, H. (2013). Specific gene bciD for C7-methyl oxidation in bacteriochlorophyll e biosynthesis of brown-colored green sulfur bacteria. *PloS One*, *8*(4), e60026. https://doi.org/10.1371/journal.pone.0060026

Hauenstein, M., Christ, B., Das, A., Aubry, S., & Hörtensteiner, S. (2016). A Role for TIC55 as a Hydroxylase of Phyllobilins, the Products of Chlorophyll Breakdown during Plant Senescence. *The Plant Cell*, *28*(10), 2510–2527. https://doi.org/10.1105/tpc.16.00630

Heinemann, I. U., Jahn, M., & Jahn, D. (2008). The biochemistry of heme biosynthesis. *Archives of Biochemistry and Biophysics*, *474*(2), 238–251. https://doi.org/10.1016/j.abb.2008.02.015

Heinig, M., & Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Research*, *32*(Web Server issue), W500–W502. https://doi.org/10.1093/nar/gkh429

Heo, L., Lee, H., & Seok, C. (2016). GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. *Scientific Reports*, *6*, 32153. https://doi.org/10.1038/srep32153

Heo, L., Park, H., & Seok, C. (2013). GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Research*, *41*(Web Server issue), W384–W388. https://doi.org/10.1093/nar/gkt458

Hirashima, M., Satoh, S., Tanaka, R., & Tanaka, A. (2006). Pigment shuffling in antenna systems achieved by expressing prokaryotic chlorophyllide a oxygenase in Arabidopsis. *The Journal of Biological Chemistry*, *281*(22), 15385–15393. https://doi.org/10.1074/jbc.M602903200

Hirashima, M., Tanaka, R., & Tanaka, A. (2009). Light-independent cell death induced by accumulation of pheophorbide a in Arabidopsis thaliana. *Plant & Cell Physiology*, *50*(4), 719–729. https://doi.org/10.1093/pcp/pcp035

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hörtensteiner, S. (2009). Stay-green regulates chlorophyll and chlorophyll-binding protein degradation during senescence. *Trends in Plant Science*, *14*(3), 155–162. https://doi.org/10.1016/j.tplants.2009.01.002

Hörtensteiner, S., & Kräutler, B. (2011). Chlorophyll breakdown in higher plants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1807*(8), 977–988. https://doi.org/https://doi.org/10.1016/j.bbabio.2010.12.007

HÖRTENSTEINER, S., VICENTINI, F., & MATILE, P. (1995). Chlorophyll breakdown in senescent cotyledons of rape, Brassica napus L.: Enzymatic cleavage of phaeophorbide a in vitro. *New Phytologist*, *129*(2), 237–246. https://doi.org/https://doi.org/10.1111/j.1469-8137.1995.tb04293.x

Ito, H., Ohtsuka, T., & Tanaka, A. (1996). Conversion of chlorophyll b to chlorophyll a via 7-hydroxymethyl chlorophyll. *The Journal of Biological Chemistry*, *271*(3), 1475–1479. https://doi.org/10.1074/jbc.271.3.1475

Jia, T., Ito, H., & Tanaka, A. (2016). Simultaneous regulation of antenna size and photosystem I/II stoichiometry in Arabidopsis thaliana. *Planta*, *244*(5), 1041–1053. https://doi.org/10.1007/s00425-016-2568-5

Jiao, B., Meng, Q., & Lv, W. (2020). Roles of stay-green (SGR) homologs during chlorophyll degradation in green plants. *Botanical Studies*, *61*(1), 25. https://doi.org/10.1186/s40529-020-00302-5

Kabsch, W. (2010). XDS. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 2), 125–132. https://doi.org/10.1107/S0907444909047337

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577–2637. https://doi.org/10.1002/bip.360221211

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., & Jorgensen, W. L. (2001). Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *The Journal of Physical Chemistry B*, *105*(28), 6474–6487. https://doi.org/10.1021/jp003919d

Kuai, B., Chen, J., & Hörtensteiner, S. (2018). The biochemistry and molecular biology of chlorophyll breakdown. *Journal of Experimental Botany*, *69*(4), 751–767. https://doi.org/10.1093/jxb/erx322

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. *The Journal of Biological Chemistry*, *288*(27), 19330–19341. https://doi.org/10.1074/jbc.M113.462663

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. https://doi.org/https://doi.org/10.1107/S0021889892009944

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, *47*(W1), W256–W259. https://doi.org/10.1093/nar/gkz239

Liebschner, D., Afonine, P. V, Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V, … Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica. Section D, Structural Biology*, *75*(Pt 10), 861–877. https://doi.org/10.1107/S2059798319011471

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*(12), 2287–2291. https://doi.org/10.1021/acs.jcim.6b00407

Luo, Z., Zhang, J., Li, J., Yang, C., Wang, T., Ouyang, B., Li, H., Giovannoni, J., & Ye, Z. (2013). A STAY-GREEN protein SlSGR1 regulates lycopene and β-carotene accumulation by interacting directly with SlPSY1 during ripening processes in tomato. *The New Phytologist*, *198*(2), 442–452. https://doi.org/10.1111/nph.12175

Lüthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, *356*(6364), 83–85. https://doi.org/10.1038/356083a0

Malcolm, B. A., Rosenberg, S., Corey, M. J., Allen, J. S., de Baetselier, A., & Kirsch, J. F. (1989). Site-directed mutagenesis of the catalytic residues Asp-52 and Glu-35 of chicken egg white lysozyme. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(1), 133–137. https://doi.org/10.1073/pnas.86.1.133

Martins, B. M., Svetlitchnaia, T., & Dobbek, H. (2005). 2-Oxoquinoline 8-monooxygenase oxygenase component: active site modulation by Rieske-[2Fe-2S] center oxidation/reduction. *Structure (London, England : 1993)*, *13*(5), 817–824. https://doi.org/10.1016/j.str.2005.03.008

Matsuda, K., Shimoda, Y., Tanaka, A., & Ito, H. (2016). Chlorophyll a is a favorable substrate for Chlamydomonas Mg-dechelatase encoded by STAY-GREEN. *Plant Physiology and Biochemistry : PPB*, *109*, 365–373. https://doi.org/10.1016/j.plaphy.2016.10.020

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(Pt 4), 658–674. https://doi.org/10.1107/S0021889807021206

Meguro, M., Ito, H., Takabayashi, A., Tanaka, R., & Tanaka, A. (2011). Identification of the 7-Hydroxymethyl Chlorophyll a Reductase of the Chlorophyll Cycle in Arabidopsis . *The Plant Cell*, *23*(9), 3442–3453. https://doi.org/10.1105/tpc.111.089714

Minkyung, B., Frank, D., Ivan, A., Justas, D., Sergey, O., Rie, L. G., Jue, W., Qian, C., N., K. L., Dustin, S. R., Claudia, M., Hahnbeom, P., Carson, A., R., G. C., Andy, D., H., P. J., V., R. A., A., van D. A., C., E. A., … David, B. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *373*(6557), 871–876. https://doi.org/10.1126/science.abj8754

Mochizuki, N., Tanaka, R., Grimm, B., Masuda, T., Moulin, M., Smith, A. G., Tanaka, A., & Terry, M. J. (2010). The cell biology of tetrapyrroles: a life and death struggle. *Trends in Plant Science*, *15*(9), 488–498. https://doi.org/https://doi.org/10.1016/j.tplants.2010.05.012

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with

selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. https://doi.org/10.1002/jcc.21256

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*(6), 1019–1025. https://doi.org/10.1007/s00425-003-1181-6

Neilson, J. A. D., & Durnford, D. G. (2010). Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes. *Photosynthesis Research*, *106*(1–2), 57–71. https://doi.org/10.1007/s11120-010-9576-2

Nelson, N., & Yocum, C. F. (2006). Structure and function of photosystems I and II. *Annual Review of Plant Biology*, *57*, 521–565. https://doi.org/10.1146/annurev.arplant.57.032905.105350

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. *Journal of Molecular Biology*, *351*(2), 355–370. https://doi.org/10.1016/j.jmb.2005.05.059

Nomata, J., Swem, L. R., Bauer, C. E., & Fujita, Y. (2005). Overexpression and characterization of dark-operative protochlorophyllide reductase from Rhodobacter capsulatus. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1708*(2), 229–237. https://doi.org/https://doi.org/10.1016/j.bbabio.2005.02.002

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, *3*(1), 33. https://doi.org/10.1186/1758-2946-3-33

Obata, D., Takabayashi, A., Tanaka, R., Tanaka, A., & Ito, H. (2019). Horizontal Transfer of Promiscuous Activity from Nonphotosynthetic Bacteria Contributed to Evolution of Chlorophyll Degradation Pathway. *Molecular Biology and Evolution*, *36*(12), 2830–2841. https://doi.org/10.1093/molbev/msz193

Oster, U., Tanaka, R., Tanaka, A., & Rüdiger, W. (2000). Cloning and functional expression of the gene encoding the key enzyme for chlorophyll b biosynthesis (CAO) from Arabidopsis thaliana. *The Plant Journal : For Cell and Molecular Biology*, *21*(3), 305–310. https://doi.org/10.1046/j.1365-313x.2000.00672.x

Parales, R. E., Parales, J. V, & Gibson, D. T. (1999). Aspartate 205 in the catalytic domain of naphthalene dioxygenase is essential for activity. *Journal of Bacteriology*, *181*(6), 1831–1837. https://doi.org/10.1128/JB.181.6.1831-1837.1999

Park, T., Won, J., Baek, M., & Seok, C. (2021). GalaxyHeteromer: protein heterodimer structure prediction by template-based and ab initio docking. *Nucleic Acids Research*, *49*(W1), W237–W241. https://doi.org/10.1093/nar/gkab422

Perry, C., de los Santos, E. L. C., Alkhalaf, L. M., & Challis, G. L. (2018). Rieske non-

heme iron-dependent oxygenases catalyse diverse reactions in natural product biosynthesis. *Natural Product Reports*, *35*(7), 622–632. https://doi.org/10.1039/C8NP00004B

Pružinská, A., Anders, I., Aubry, S., Schenk, N., Tapernoux-Lüthi, E., Müller, T., Kräutler, B., & Hörtensteiner, S. (2007). In Vivo Participation of Red Chlorophyll Catabolite Reductase in Chlorophyll Breakdown. *The Plant Cell*, *19*(1), 369–387. https://doi.org/10.1105/tpc.106.044404

Pruzinská, A., Tanner, G., Anders, I., Roca, M., & Hörtensteiner, S. (2003). Chlorophyll breakdown: pheophorbide a oxygenase is a Rieske-type iron-sulfur protein, encoded by the accelerated cell death 1 gene. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(25), 15259–15264. https://doi.org/10.1073/pnas.2036571100

Sakuraba, Y., Lee, S.-H., Kim, Y.-S., Park, O. K., Hörtensteiner, S., & Paek, N.-C. (2014). Delayed degradation of chlorophylls and photosynthetic proteins in Arabidopsis autophagy mutants during stress-induced leaf yellowing. *Journal of Experimental Botany*, *65*(14), 3915–3925. https://doi.org/10.1093/jxb/eru008

Sakuraba, Y., Tanaka, R., Yamasato, A., & Tanaka, A. (2009). Determination of a chloroplast degron in the regulatory domain of chlorophyllide a oxygenase. *The Journal of Biological Chemistry*, *284*(52), 36689–36699. https://doi.org/10.1074/jbc.M109.008144

Sangawa, T., Tabata, S., Suzuki, K., Saheki, Y., Tanaka, K., & Takagi, J. (2013). A multipurpose fusion tag derived from an unstructured and hyperacidic region of the amyloid precursor protein. *Protein Science : A Publication of the Protein Society*, *22*(6), 840–850. https://doi.org/10.1002/pro.2254

Sato, T., Shimoda, Y., Matsuda, K., Tanaka, A., & Ito, H. (2018). Mg-dechelation of chlorophyll a by Stay-Green activates chlorophyll b degradation through expressing Non-Yellow Coloring 1 in Arabidopsis thaliana. *Journal of Plant Physiology*, *222*, 94–102. https://doi.org/10.1016/j.jplph.2018.01.010

Sato, Y., Morita, R., Katsuma, S., Nishimura, M., Tanaka, A., & Kusaba, M. (2009). Two short-chain dehydrogenase/reductases, NON-YELLOW COLORING 1 and NYC1-LIKE, are required for chlorophyll b and light-harvesting complex II degradation during senescence in rice. *The Plant Journal : For Cell and Molecular Biology*, *57*(1), 120–131. https://doi.org/10.1111/j.1365-313X.2008.03670.x

Sato, Y., Morita, R., Nishimura, M., Yamaguchi, H., & Kusaba, M. (2007). Mendel's green cotyledon gene encodes a positive regulator of the chlorophyll-degrading pathway. *Proceedings of the National Academy of Sciences*, *104*(35), 14169 LP – 14174. https://doi.org/10.1073/pnas.0705521104

Schelbert, S., Aubry, S., Burla, B., Agne, B., Kessler, F., Krupinska, K., & Hörtensteiner, S. (2009). Pheophytin pheophorbide hydrolase (pheophytinase) is involved in chlorophyll breakdown during leaf senescence in Arabidopsis. *The Plant Cell*, *21*(3),

767–785. https://doi.org/10.1105/tpc.108.064089

Schumacher, I., Menghini, D., Ovinnikov, S., Hauenstein, M., Fankhauser, N., Zipfel, C., Hörtensteiner, S., & Aubry, S. (2021). Evolution of chlorophyll degradation is associated with plant transition to land. *The Plant Journal : For Cell and Molecular Biology*. https://doi.org/10.1111/tpj.15645

Seffernick, J. T., & Lindert, S. (2020). Hybrid methods for combined experimental and computational determination of protein structure. *The Journal of Chemical Physics*, *153*(24), 240901. https://doi.org/10.1063/5.0026025

Shimoda, Y., Ito, H., & Tanaka, A. (2012). Conversion of chlorophyll b to chlorophyll a precedes magnesium dechelation for protection against necrosis in Arabidopsis. *The Plant Journal*, *72*(3), 501–511. https://doi.org/https://doi.org/10.1111/j.1365-313X.2012.05095.x

Shimoda, Y., Ito, H., & Tanaka, A. (2016). Arabidopsis STAY-GREEN, Mendel's Green Cotyledon Gene, Encodes Magnesium-Dechelatase. *The Plant Cell*, *28*(9), 2147–2160. https://doi.org/10.1105/tpc.16.00428

Shin, D., Lee, S., Kim, T.-H., Lee, J.-H., Park, J., Lee, J., Lee, J. Y., Cho, L.-H., Choi, J. Y., Lee, W., Park, J.-H., Lee, D.-W., Ito, H., Kim, D. H., Tanaka, A., Cho, J.-H., Song, Y.-C., Hwang, D., Purugganan, M. D., … Nam, H. G. (2020). Natural variations at the Stay-Green gene promoter control lifespan and yield in rice cultivars. *Nature Communications*, *11*(1), 2819. https://doi.org/10.1038/s41467-020-16573-2

Sievers, F., & Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Science : A Publication of the Protein Society*, *27*(1), 135–145. https://doi.org/10.1002/pro.3290

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–234. https://doi.org/https://doi.org/10.1016/j.pep.2005.01.016

Sugishima, M., Okamoto, Y., Noguchi, M., Kohchi, T., Tamiaki, H., & Fukuyama, K. (2010). Crystal Structures of the Substrate-Bound Forms of Red Chlorophyll Catabolite Reductase: Implications for Site-Specific and Stereospecific Reaction. *Journal of Molecular Biology*, *402*(5), 879–891. https://doi.org/https://doi.org/10.1016/j.jmb.2010.08.021

Takatani, N., Uenosono, M., Hara, Y., Yamakawa, H., Fujita, Y., & Omata, T. (2022). Chlorophyll and Pheophytin Dephytylating Enzymes Required for Efficient Repair of PSII in Synechococcus elongatus PCC 7942. *Plant and Cell Physiology*, *63*(3), 410–420. https://doi.org/10.1093/pcp/pcac006

Tanaka, A., Ito, H., Tanaka, R., Tanaka, N. K., Yoshida, K., & Okada, K. (1998). Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(21), 12719–12723. https://doi.org/10.1073/pnas.95.21.12719

Tanaka, A., & Tanaka, R. (2019). Chapter Six - The biochemistry, physiology, and evolution of the chlorophyll cycle. In B. B. T.-A. in B. R. Grimm (Ed.), *Metabolism, Structure and Function of Plant Tetrapyrroles: Introduction, Microbial and Eukaryotic Chlorophyll Synthesis and Catabolism* (Vol. 90, pp. 183–212). Academic Press. https://doi.org/https://doi.org/10.1016/bs.abr.2019.03.005

Tanaka, R., & Tanaka, A. (2007). Tetrapyrrole Biosynthesis in Higher Plants. *Annual Review of Plant Biology*, *58*(1), 321–346. https://doi.org/10.1146/annurev.arplant.57.032905.105448

Tomitani, A., Okada, K., Miyashita, H., Matthijs, H. C., Ohno, T., & Tanaka, A. (1999). Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts. *Nature*, *400*(6740), 159–162. https://doi.org/10.1038/22101

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, *44*(W1), W232–W235. https://doi.org/10.1093/nar/gkw256

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334

Vitt, S., Ma, K., Warkentin, E., Moll, J., Pierik, A. J., Shima, S., & Ermler, U. (2014). The F_{420}-reducing [NiFe]-hydrogenase complex from Methanothermobacter marburgensis, the first X-ray structure of a group 3 family member. *Journal of Molecular Biology*, *426*(15), 2813–2826. https://doi.org/10.1016/j.jmb.2014.05.024

Vom Dorp, K., Hölzl, G., Plohmann, C., Eisenhut, M., Abraham, M., Weber, A. P. M., Hanson, A. D., & Dörmann, P. (2015). Remobilization of Phytol from Chlorophyll Degradation Is Essential for Tocopherol Synthesis and Growth of Arabidopsis. *The Plant Cell*, *27*(10), 2846–2859. https://doi.org/10.1105/tpc.15.00395

Wang, X., & Liu, L. (2016). Crystal Structure and Catalytic Mechanism of 7-Hydroxymethyl Chlorophyll a Reductase. *The Journal of Biological Chemistry*, *291*(25), 13349–13359. https://doi.org/10.1074/jbc.M116.720342

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Whittaker, C. A., & Hynes, R. O. (2002). Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Molecular Biology of the Cell*, *13*(10), 3369–3387. https://doi.org/10.1091/mbc.e02-05-0259

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(suppl_2), W407–W410. https://doi.org/10.1093/nar/gkm290

Xie, Z., Wu, S., Chen, J., Zhu, X., Zhou, X., Hörtensteiner, S., Ren, G., & Kuai, B. (2019). The C-terminal cysteine-rich motif of NYE1/SGR1 is indispensable for its function in chlorophyll degradation in Arabidopsis. *Plant Molecular Biology*, *101*(3), 257–268. https://doi.org/10.1007/s11103-019-00902-1

Xu, J., & Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics (Oxford, England)*, *26*(7), 889–895. https://doi.org/10.1093/bioinformatics/btq066

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 Research*, *46*(W1), W374–W379. https://doi.org/10.1093/nar/gky380

Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*, *32*(23), 3676–3678. https://doi.org/10.1093/bioinformatics/btw514

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. *The Plant Cell*, *17*(5), 1585–1597. https://doi.org/10.1105/tpc.105.031518

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., & Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, *117*(3), 1496 LP – 1503. https://doi.org/10.1073/pnas.1914677117

Yang, J., Roy, A., & Zhang, Y. (2013). Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics (Oxford, England)*, *29*(20), 2588–2595. https://doi.org/10.1093/bioinformatics/btt447

Zhang, S., Heyes, D. J., Feng, L., Sun, W., Johannissen, L. O., Liu, H., Levy, C. W., Li, X., Yang, J., Yu, X., Lin, M., Hardman, S. J. O., Hoeven, R., Sakuma, M., Hay, S., Leys, D., Rao, Z., Zhou, A., Cheng, Q., & Scrutton, N. S. (2019). Structural basis for enzymatic photocatalysis in chlorophyll biosynthesis. *Nature*, *574*(7780), 722–725. https://doi.org/10.1038/s41586-019-1685-2

Zheng, W., Li, Y., Zhang, C., Zhou, X., Pearce, R., Bell, E. W., Huang, X., & Zhang, Y. (2021). Protein structure prediction using deep learning distance and hydrogen-bonding restraints in CASP14. *Proteins*, *89*(12), 1734–1751. https://doi.org/10.1002/prot.26193

Zhou, C., Han, L., Pislariu, C., Nakashima, J., Fu, C., Jiang, Q., Quan, L., Blancaflor, E. B., Tang, Y., Bouton, J. H., Udvardi, M., Xia, G., & Wang, Z.-Y. (2011). From model to crop: functional analysis of a STAY-GREEN gene in the model legume Medicago truncatula and effective use of the gene for alfalfa improvement. *Plant Physiology*,

157(3), 1483–1496. https://doi.org/10.1104/pp.111.185140