Domain structures of chlorophyllide *a* oxygenase of green plants and *Prochlorothrix hollandica* in relation to catalytic functions

N. Nagata · S. Satoh · R. Tanaka · A. Tanaka

The Institute of Low Temperature Science, Hokkaido University, Sapporo 0600819, Japan

Corresponding author. Ayumi Tanaka

E-mail: ayumi@lowtem.hokudai.ac.jp

Fax: +81-11-706-5493

Abstract Chlorophyll *b* is a photosynthetic antenna pigment found in prochlorophytes and chlorophytes. In chlorophytes, its biosynthesis regulates the photosynthetic antenna size. Chlorophyll *b* is synthesized from chlorophyll *a* in a two-step oxygenation reaction by chlorophyllide *a* oxygenase (CAO). In this study, we first identified the entire sequence of a prochlorophyte CAO gene from *Prochlorothrix hollandica* to compare it with those from chlorophytes, and we examined the catalytic activity of the gene product. Southern blot analysis showed that the CAO gene is present in one copy in the *P. hollandica* genome. The *P. hollandica* CAO gene *(PhCAO*) has a coding capacity for 367 amino acids, which is much smaller than those of *Arabidopsis thaliana* (537 amino acids) and *Oryza sativa* (542 amino acids) CAO genes. In spite of the small size, *PhCAO* catalyzed the formation of chlorophyll *b*. By comparing these sequences, we classified the land plant sequences into four parts: the N-terminal sequence predicted to be a transit peptide, the successive conserved sequence unique in land plants (A-domain, 134 amino acids), a less-conserved sequence (B-domain, 30 amino acids) and the C-terminal conserved sequence common in chlorophytes and prochlorophytes (C-domain, 337 to 344 amino acids). We demonstrated that the C-domain is sufficient for catalytic activity by transforming the cyanobacterium *Synechosystis* sp. PCC6803 with the C-domain from *A. thaliana*. In

this paper, the role of the A-domain is discussed in relation to the formation of light-harvesting chlorophyll *a*/*b*-protein complexes in land plants.

Key words Chlorophyllide *a* oxygenase · *Prochlorothrix hollandica* · Chlorophyll *b* · Cyanobacterium

Abbrevations CAO: clorophyllide *a* oxygenase · CP: chlorophyll protein · HPLC: high-performance liquid chromatography · LHC: light-harvesting complex · PCR: polymerase chain reaction · PS: photosystem

Introduction

Photosynthetic organisms harvest light energy by core and peripheral antenna systems for photosynthesis (Grossman et al. 1995). The core antenna complexes of oxygenic

photosystems consists of CP43 and CP47 in the case of photosystem II (PSII) and P700-chlorophyll *a*-protein complexes (CP1) in the case of photosystem I (PSI) (Green 1996). The composition of these chlorophyll *a*-protein complexes is highly conserved and does not change under any light conditions. Chlorophyll *b* is a peripheral antenna pigment in chlorophytes (Grossman et al. 1995) and prochlorophytes (La Roche et al. 1996, Ting et al. 2002). In green plants, chlorophyll *b* is found only in light-harvesting complexes (LHC), and the amount of LHCII changes depending on the light conditions. Under conditions of low light intensity, a large amount of LHCII is associated with a core antenna complex, resulting in a large antenna size and low chlorophyll *a*/*b* ratios (Huner et al. 1998; Bailey et al. 2001). The synthesis of chlorophyll *b* is thought to play an important role in the regulation of antenna size, because stimulation of chlorophyll *b* synthesis by feeding 5-amino levulinic acid, a precursor of chlorophyll, increased the accumulation of LHCII (Tanaka et al. 1994) and because overexpression of the gene for chlorophyll *b* synthesis, chlorophyllide *a* oxygenase (CAO), enlarged the PSII antenna size in *Arabidopsis thaliana* (Tanaka et al. 2001).

CAO is encoded by a single gene in the nuclear genome (Tanaka et al. 1998), and its translation products are imported into chloroplasts. In chloroplasts, CAO catalyzes two-step oxygenation reactions and converts chlorophyllide *a* to chlorophyllide *b* via

7-hydroxymethyl chlorophyllide (Oster et al. 2000), and the newly synthesized chlorophyll *b* is assembled with apoproteins to form chlorophyll-protein complexes. Chlorophyll *b* is exclusively incorporated into LHC apoproteins, and very little chlorophyll *b* has been found in core antenna complexes of photosystems. However, core antenna complexes have been reported to have the ability to bind chlorophyll *b* functionally. Cyanobacteria accumulated chlorophyll *b* when they were transformed by CAO. In the transformant, chlorophyll *b* replaced 10% of chlorophyll *a* in CP1 and functioned as a photosynthetic pigment (Satoh et al. 2001). Furthermore, when CAO was introduced and expressed in a photosystem I-less strain of the cyanobacterium with *lhcb* encoding LHCII apoproteins, chlorophyll *b* accumulated to levels exceeding those of chlorophyll *a* and 75% of chlorophyll *a* in the PSII core complex was replaced by chlorophyll *b* (Vavilin et al. 2003). It is not clear why core antenna complexes of green plants have very little chlorophyll *b* in spite of their ability to bind chlorophyll *b* and active synthesis of chlorophyll *b*. One possible reason is that the specific binding of chlorophyll *b* to LHC apoproteins is achieved by the mechanisms of LHCII formation. It has been proposed that CAO exists in the chloroplast envelope and uses chlorophyllide *a* associated with the LHC apoproteins as a substrate (Hoober et al. 2001). These mechanisms may enable the preferential incorporation of chlorophyll *b*

into LHC but requiring interaction between CAO and LHC.

Interaction of CAO with LHC was also indicated by the finding that cyanobacteria transformed with CAO accumulated more chlorophyll *b* in cells that have *lhcb* genes (Xu et al. 2001). Lhcb should regulate CAO activity or function as a chlorophyll *b*-binding protein. However, Lhcb did not function as a chlorophyll *b*-binding protein in their experiments because synthesized chlorophyll *b* was incorporated into other chlorophyll-binding proteins and the level of Lhcb was undetectable. These results indicate that Lhcb influences CAO activity. Comparison of chlorohyte CAOs with prochlorophyte ones would help to elucidate the interaction of LHC with CAOs. In prochlorophytes, chlorophyll *b* is bound to 'prochlorophytes chlorophyll *b*-binding proteins' that are related to IsiA, also known as CP43' (La Roche et al. 1996), but not to proteins of the LHC superfamily. It is inferred from this fact that some differences exist between the structures of CAOs in prochlorohytes and chlorophytes. However, since only partial sequences of CAOs in prochlorophytes are available (Tomitani et al. 1999), this issue can not be addressed.

In this study, we isolated a full-length CAO gene from *Prochlorothrix hollandica* and compared it with those from *A. thaliana* and *Oriza sativa.* The size of *P. hollandica CAO* (*PhCAO*) was smaller than that of *A. thalinana CAO* (*AtCAO*), but it catalyzed the

formation of chlorophyll *b*. Alignment of *PhCAO* and green plant CAO indicated that green plant CAO consists of 3 domains but that *PhCAO* had only one domain. We discuss the functions of these domains in this paper.

Materials and methods

Plant materials

Prochlorothrix hollandica and *Synechocystis* sp. PCC6803 were grown at 22 ℃ in BG11 medium (Hihara et al. 1998) with 10 mM TES-KOH (pH 8.2) under continuous illumination (10 or 50 micro-einsteins $m^{-2} s^{-1}$). CAO-expressing transformant cyanobacteria were grown with 20 μg/ml kanamycin, bubbled with air.

Isolation of genomic DNA

Genomic DNA was isolated from *P. hollandica* according to the method of Rochaix (1980).

Southern blot analysis

P. hollandica genomic DNA (2.3 μg) was digested with two restriction enzymes (*Hpa*I and *Nhe*I), separated on 0.8% agarose gels, transferred onto Hybond-XL nylon membranes, and hybridized in 0.5 M sodium phosphate, 7% SDS, and 1 mM EDTA at 60 °C. ^{32}P-labeled probes were produced from PCR fragments for *P. hollandica CAO* (*PhCAO*). The blots were washed and exposed to phosphor-imaging plates (Fuji Film, Tokyo, Japan) and analyzed with Bas1500 (Fuji Film).

Cloning of a full-length *P. hollandica CAO* gene

P. hollandica genomic DNA was digested overnight using *Hpa*I at 37 °C. DNA fragments were self-ligated using a Takara DNA Ligation Kit Ver.2 (Takara Shuzo Co. Ltd., Shiga, Japan) and used as a template for inverted PCR. Amplified DNA fragments were sequenced using a Big-Dye terminator DNA Sequence kit (Applied Biosystems Japan Ltd., Tokyo, Japan) on a Genetic analyzer 310 (Applied Biosystems).

Alignment of amino acid sequences

The amino acid sequences listed below were retrieved from Gene Bank (http://www.genome.ad.jp/dbget-bin/www_bfind?genbank-today) and Kazusa DNA Research Institute (http://www.kazusa.or.jp/ja/index.html). Transit peptides of the proteins were predicted by TargetP (http://www.cbs.dtu.dk/services/TargetP/). The amino acid sequences were aligned using Clustal_X1.81 (Thompson et al. 1997).

Accession numbers

The following accession number was used for *Oriza sativa*: Chlorophyllide *a* oxygenase (CAO), AF284781. Accession numbers used for *Arabidopsis thaliana* were as follows: Chlorophyllide *a* oxygenase (CAO), BT002075; Glutamyl-tRNA reductase (*hemA*), NP_176125; Glutamate-1-semialdehyde aminotransferase (*hemL*), NP_390690; Porphobilinogen synthase (*hemB*), NP_215026; Porphobilinogen deaminase (*hemC*), NP_207035; Uroporphyrinogen III decarboxylase (*hemE*), NP_465736; Coproporphyrinogen III oxidase (*hemF*), NP_221228; Mg-protoporphyrin IX chelatase (*chlD*, *chlI*, *chlH*), AY063821, AB016870, S71288; Mg protoporphyrin IX methyl transferase (*chlM*), AL161562; Protochlorophyllide oxidoreductase A (*porA*), BT005080; Chlorophyll synthase (*chlG*, *chlP*), S60222, Y14044; Arabidopsis lethal leaf-spot1 homolog Lls1 (CAC03538), Arabidopsis lethal leaf-spot1-like gene (AL050400), Arabidopsis Tic55 homolog (AC006585) and Arabidopsis choline monooxgenase precursor (T08550). The following accession numbers was used for cyanobacteria: Glutamyl-tRNA reductase (*hemA*), D90908; Glutamate-1-semialdehyde aminotransferase (*hemL*), Q55665; Porphobilinogen synthase (*hemB*), AP003596; Porphobilinogen deaminase (*hemC*), AP003587; Uroporphyrinogen III decarboxylase (*hemE*), AP003594; Coproporphyrinogen III oxidase

(*hemF*), AP003583; Mg-protoporphyrin IX chelatase (*chlD*, *chlI*, *chlH*), X96599, U35144, U29131; Mg protoporphyrin IX methyl transferase (*chlM*), L47126; Protochlorophyllide oxidoreductase A (*porA*), AI2023; Chlorophyll synthase (*chlP*), AP003581 and Geranylgeranyl hydrogenase (*chlG*), AP003596.

Transformation of cyanobacteria

The plasmid vector pFS10 carries a promoter and a terminator of *psbA2* encoding D1 protein of cyanobacteria *Synechocystis* sp. PCC6803 (Jansson et al. 1998) and a kanamycin-resistance marker. The *A. thaliana CAO* (ABC, BC, C) or *P. hollandica CAO* gene was integrated into pFS10, and these plasmids were transformed into cyanobacteria according to the method of Williams (1988).

Analysis of chlorophylls by HPLC

Chlorophylls were extracted from *Synechocystis* sp. PCC6803 cells with 100% acetone and subjected to HPLC on a CLC-ODS column (6 × 150 mm) (SHIMADZU Co. Ltd., Kyoto, Japan). Methanol at a flow rate of 1.7 ml/min was used to elute the HPLC column. Chlorophylls were monitored by their absorbance at 650 nm.

Results

Isolation of CAO from *Prochlorothrix hollandica*

Recombinant *AtCAO* catalyzes two-step oxygenation reactions and converts chlorophyllide *a* into chlorophyllide *b*, and chlorophyll *b*-less mutants have defects in CAO in *Arabidopsis thaliana* (Espineda 1999) and *Chlamydomonas reinhardtii* (Tanaka 1998). Cyanobacteria, which originally have no chlorophyll *b*, can accumulate chlorophyll *b* when they acquired CAO (Satoh et al. 2001). The results of these studies clearly show that CAO is involved in chlorophyll *b* biosynthesis in green plants.

On the other hand, only partial sequences of CAO for two genera of prochlorophytes have been reported, and there has been no report on the enzymatic functions of prochlorophyte CAOs. Based on the partial sequence of *PhCAO* (Tomitani et al. 1999), we prepared a 720-bp *PhCAO* probe and carried out Southern blot analysis. The results suggested that CAO is most likely encoded by a single-copy gene in the *P. hollandica* genome as it is in *Chlamydomonas reinhardtii* (Tanaka et al. 1998) and *A. thaliana* (The Arabidopsis Genome Initiative 2000). We isolated a 2.2-kb genomic clone that covered the complete *PhCAO* locus by using PCR amplification. Sequence analysis showed that the clone had a 1.2-kb open reading frame for 367 amino acids, giving a calculated molecular weight of 40830. The predicted amino acid sequence had binding motifs for a Rieske-type [2Fe-2S] cluster and a mononuclear iron-binding site, which are thought to be necessary for CAO activity.

Expression of *PhCAO* in cyanobacteria

The predicted length of *PhCAO* (367 amino acids) was much smaller than that of

AtCAO (537 amino acids) and that of *OsCAO* (541 amino acids). Therefore, we examined whether *PhCAO* had a catalytic activity and could synthesize chlorophyll *b* by itself. In this study, we used cyanobacteria transformation systems to examine whether *PhCAO* has complete activity to synthesize chlorophyll *b*, because the use of such systems has been shown to be a reliable method to investigate CAO activity (Satoh et al. 2001). The complete *PhCAO* gene was introduced into cyanobacteria cells and the pigments were analyzed by HPLC (Fig. 1). We observed a new peak of chlorophyll *b*, indicating that the transformant cells accumulated chlorophyll *b*. Based on the results of these experiments, we concluded that *PhCAO* has complete activity by itself.

Alignment of the amino acid sequences of CAO

We aligned and compared the amino acid sequence of *PhCAO* with those of *AtCAO* and *OsCAO* (Fig. 2a). Alignment of these sequences clarified the unique structures of the CAOs. Computer analysis with TargetP predicted a transit sequence of 36 amino acids

in *AtCAO*. The predicted transit peptides of *AtCAO* and *OsCAO* showed low sequence similarities. Except for transit peptides, the mature sizes of *AtCAO* and *OsCAO* are 501 and 508, respectively. *PhCAO* is composed of 367 amino acids and is smaller than mature *AtCAO* and *OsCAO*, and it has an extension of 37 amino acids at the C-terminal end. One of the conspicuous features of *AtCAO* and *OsCAO* is the existence of a long extension of 164 amino acids at the N-terminal end. This extension does not exist in *PhCAO*. Comparison of these CAOs sequences clearly showed that mature CAOs of green plants could be divided into three domains. We named these domains A-, B- and C- domains (Fig.2b). The most-N-terminal part, the A-domain, in *AtCAO* is composed of 134 amino acids. High sequence similarities were found in this domain between *AtCAO* and *OsCAO*, but this domain has no significant sequence similarities to any known proteins in the database. The most-C-terminal part, the C-domain, has a Rieske-type [2Fe-2S] cluster and a mononuclear iron-binding site, indicating that it plays a catalytic role in chlorophyll *b* synthesis. This domain is conserved for all the known CAO sequences from prochlorophytes, green algae and land plants. The B-domain might be a linker between A- and C- domains, because the B-domain was found to consist of only 30 amino acids and showed no sequence homology between *AtCAO* and *OsCAO*. *PhCAO* has only a C-domain with a small C-terminal extension.

Analysis of the catalytic function of the *AtCAO* domain

In order to determine whether the C-domain of *AtCAO* is sufficient for catalytic activity or whether the A-domain and/or B- domain is also required, we introduced genes corresponding to the C-domain, BC-domain and ABC-domain of *AtCAO* into the cyanobacterial genome. Transformed cyanobacteria were harvested, and chlorophylls were subjected to HPLC. As we reported previously (Satoh et al. 2001), cyanobacteria transformed with a complete *AtCAO* gene (ABC-domain) accumulated chlorophyll *b*. We also found peaks corresponding to chlorophyll *b* in the case of cyanobacteria transformed with the BC-domain or C-domain (Fig. 3). These results indicate that the C-domain is sufficient for the synthesis of chlorophyll *b*. The heights of the peaks of chlorophyll *b* in the HPLC profiles were different among these transformant cells, but this does not necessary indicate the levels of enzymatic activities of the proteins. At present, it is not clear which construct has the highest level of CAO activity.

Comparison of the N-terminal extension sequences of enzymes for chlorophyll synthesis

Eukaryotic CAO sequences have N-terminal extensions compared to that of *PhCAO*. Next, we investigated whether these N-terminal extensions in eukaryote are specific for CAO or are a common feature of the enzymes for chlorophyll synthesis by comparing sequences of these enzymes in *A. thaliana* and a cyanobacterium (Fig. 4) whose complete genome sequences have been determined. In *A. thaliana*, all of the enzymes for chlorophyll biosynthesis function in chloroplasts and should have chloroplast target peptides. Their cleavage sites were predicted by TargetP. Lengths of the transit peptides for these enzymes were predicted to be between 35 and 86 amino acids. The N-terminal extensions of the mature enzymes were smaller than 58 amino acids, much smaller than those of CAOs. These results indicate that the N-terminal extensions are a unique structure of CAO among chlorophyll synthetic enzymes. Lengths of the C-terminal extensions were smaller than 17 amino acids (data not shown), indicating that there are no significant differences in C-terminal structures of the chlorophyll

biosynthesis enzymes in *A. thaliana* and cyanobacteria.

Comparison of amino acid sequences of CAO homologs of *A. thaliana*

We found a unique structure of CAO that did not exist in other chlorophyll synthesis enzymes. Next, we investigated whether the N-terminal extension is specific for CAO or a common feature for CAO homologous proteins. The complete genome sequence of *A. thaliana* shows that there are four proteins that have binding motifs for a Rieske-type [2Fe-2S] cluster and mononuclear iron-binding site and whose amino acid sequences have homology to CAO. We retrieved lethal leaf-spot1 homolog (Lls1), lethal leaf-spot1-like gene, putative Rieske iron-sulfur protein of TIC55 homolog and choline monooxgenase, and the alignment of these proteins is shown in Fig. 5. A significant sequence homology was found among these proteins, but the sequence corresponding to the A-domain of CAO was not found in other enzymes. The N-terminals of these proteins corresponded to the region of B-domain. These results showed that the N-terminal extension is a unique structure of CAO.

Discussion

Three genera of prochlorophytes have been reported, and CAOs of *Prochlorothrix hollandica* and *Prochloron didemni*, which have high sequence similarity with other CAOs, have been identified. However, *Prochlorococcus CAO* has not been identified, although the complete genome sequences of two *Prochlorococcus* strains have been determined. Hess et al. (2001) proposed genes encoding putative CAO based on the genome sequences. These genes contain binding domains for a Rieske-type [2Fe-2S] cluster and for a mononuclear iron, but their predicted amino acid sequences share low similarity with CAOs of other prochlorophytes. Furthermore, *PhCAO* encodes 367 amino acids while CAOs of the two strains of *Prochlorococcus* encode 440 and 436 amino acids. The question arises as to why the chlorophyll *b* synthesis gene of *Prochlorococcus* is quite different from other CAOs. One possible reason is that *Prochlorococcus* has divinyl chlorophylls instead of monovinyl chlorophyll and CAO might have evolved to use divinyl chlorophyll *a* as a substrate. However, this is not

likely because *AtCAO* can convert divinyl chlorophyll *a* to divinyl chlorophyll *b* with almost the same efficiency as monovinil chlorophyll (unpublished data). The gene for chlorophyll *b* synthesis in the *Prochlorococcus* genome has not yet been identified, and further enzymatic studies are required.

We found a unique structure of eukaryotic CAOs. Since the A-domain of *AtCAO* and *OsCAO* have high sequence similarity and since this domain is not involved in catalytic functions, the A-domain should play some roles in the regulation of catalytic activities or assembly of the pigment with apoproteins. It has been reported that some enzymes for chlorophyll biosynthesis are associated with other proteins and that this feature is important for the enzymatic activities. Glutamyl-tRNA reductase, which is involved in 5-amonolevulinic acid syntheses, has been suggested to form complexes with glutamate-1-semialdehyde-1, 2-mutase, an enzyme required for the subsequent step (Moser et al. 2001). It is thought that this structure promotes the efficient synthesis of 5-amonolebulinic acids. Recently, a gene for the negative regulator (FLU) of chlorophyll biosynthesis was isolated from *A. thaliana* (Meskauskiene et al. 2001). The gene product was tightly associated with chloroplast plastid membranes where chlorophylls are synthesized. It was speculated that FLU associates with the enzymes for chlorophyll biosynthesis and regulates their activities. It seems reasonable to

assume that chlorophyll biosynthesis is finely regulated at the level of the enzyme activity, because accumulation of chlorophyll intermediate molecules cause photodamage to cells. The N-terminal extension of *CAO*, the A-domain, might play an important role in regulation of chlorophyll *b* biosynthesis by interacting with the same proteins. Another possible role of the A-domain is interaction with LHC. Cyanobacteria synthesize chlorophyll *b* when they are transformed with the *AtCAO* gene (Satoh et al. 2001). The accumulation of chlorophyll *b* drastically increased when the strain was transformed with an *lhcb* gene encoding light-harvesting chlorophyll *a*/*b*-protein complexes (Xu et al. 2001). However, there was little accumulation of an *lhcb* product, and it could not be detected by antibodies against LHC apoproteins. This strongly suggests that LHC apoproteins do not function as chlorophyll *b*-binding proteins but as regulators for *AtCAO* activity. This idea is consistent with the fact that *PhCAO* has no N-terminal extension and LHC is not the chlorophyll *b*-binding complex in *P. hollandica*.

In this study, we found N-terminal extension in *AtCAO* and *OsCAO*. Partial sequences of CAOs are available for many photosynthetic organisms such as Zea mays and *C. reinhardtii*, and all of these CAOs have N-terminal extensions. It therefore seems reasonable to assume that N-terminal extension of CAO is acquired after the first

endosymbiotic event. In higher plants, chlorophyll *b* is bound to proteins belonging to LHC superfamily, which have 3 membrane-spanning α helixes. Cyanobacteria have HLIP, whose amino acid sequence has homology to that of LHC apoproteins (Dolganov et al. 1995), but do not have LHC. LHC may appear in the eukaryotic stage from HLIP-like proteins (Green et al. 1994). Thus, both the A-domain and LHC appear after primary endosymbiosis. This also suggests that there is a relationship between the A-domain and LHCII.

Acknowledgements This work was supported in part by grants-in-aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

Bailey S, Walters RG, Jansson S (2001) Acclimation of *Arabidopsis thaliana* to the light environment: the existence of separate low light and high light responses. Planta 213: 794-801

Dolganov NA, Bhaya D, Grossman AR (1995) Cyanobacterial protein with similarity to the chlorophyll a/b binding proteins of higher plants: evolution and regulation. Proc Natl Acad Sci USA 92: 636-40

Espineda CE, Linford AS, Devine D, Brusslan JA (1999) The *AtCAO* gene, encoding chlorophyll a oxygenase, is required for chlorophyll b synthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 96: 10507-10511

Green BR, Pichersky E (1994) Hypothesis for the evolution of three-helix Chl a/b and Chl a/c hight –harvesting antenna proteins from tow-helix and four-helix ancestors. Photosynth. Res. 39: 141-162

Green BR, Durnford DG (1996) The chlorophyll-carotenoid proteins of oxygenic photosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47: 685-714

Grossman AR, Bhaya D, Apt KE, Kehoe DM (1995) Light-harvesting complexes in oxygenic photosynthesis: Diversity, Control, and Evolution. Annu. Rev. Genetics. 29: 231-288

Hess WR, Rocap G, Ting CS, Larimer F, Stilwagen S, Lamerdin J, Chisholm SW (2001) The photosynthetic apparatus of *Prochlorococcus*: Insights through comparative genomics. Photosynth Res. 70: 53-71, 2001.

Hihara Y, Sonoike K, Ikeuchi M (1998) A novel gene, *pmgA*, specifically regulates photosystem stoichiometry in the cyanobacterium *Synechocysits* species PCC 6803 in response to high light.

Plant Physiol 117: 1205-1216

Hoober JK, Eggink LL (2001) A potential role of chlorophylls *b* and *c* in assembly of light-harvesting complexes. FEBS 489: 1-3

Huner NP, Öquist G, Sarhan F (1998) Energy balance and acclimation to light and cold. Physiol Plant 3: 1360-1385

Janson C, Salih G, Eriksson J, Wiklund R, Ghebramedhin H (1998) Use of *Synechocysits* 6803 to study expression of a *psbA* gene family. Methods Enzymol 297: 166-182

La Roche J, van der Staay GWM, Partensky F, Ducret A, Aebersold R, Li I, Golden SS, Hiller RG, Wrench PM, Larkum AWD, Green BR (1996) Independent evolution of the prochlorophytes and green plant chlorophyll *a/b* light-harvesting proteins. Proc Natl Acad Sci USA 93: 15244-15248

Meskauskiene R, Apel K (2002) Interaction of FLU, a negative regulator of tetrapyrrole biosynthesis with the glitamyl-tRNA reductase requires the tetratricopeptide repeat domain of FLU. FEBS lett. 532: 27-30

Moser J, Schubert WD, Beier V, Bringemeier I, Jahn D, Heinz DW (2001) V-shaped structure of glutamyl-tRNA reductase, the first enzyme of tRNA-dependent tetrapyrrole biosynthesis. EMBO J 20: 6583-6590

Oster U, Tanaka R, Tanaka A, Rudiger W (2000) How chlorophyll *b* is made: Cloning and

functional expression of the gene encoding the key enzyme for chlorophyll *b* biosynthesis (CAO) from *Arabidopsis thaliana*. Plant J 21: 305-310

Rochaix JD (1980) Restriction fragments from *Chlamydomonas* chloroplast DNA. Methods Enzymol. 65: 785-795.

Satoh S, Ikeuchi M, Mimuro M, Tanaka A (2001) Chlorophyll *b* expressed in cyanobacteria functions as a light-harvesting antenna proteins in photosystem I through flexibility of the proteins. J Biol Chem 276: 4293-4297

Strand Å, Asami T, Alonso J, Ecker JR, Chory J (2003) Choloroplast to nucleus communication triggered by accumulation of Mg-potoporphyrinIX. Nature 421: 79-83

Tanaka A, Tanaka Y, Tsuji H (1994) Preferential accumulation of apoproteins of the light-harvesting chlorophyll *a/b*-protein complex in greening barley leaves treated with 5-aminolevulinic acid. Planta 192: 92-97

Tanaka A, Ito H, Tanaka R, Tanaka NK, Yoshida K (1998) Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. Proc. Natl Acad. Sci USA 95: 12719-12723

Tanaka R, Koshino Y, Sawa S, Ishiguro S, Okada K, Tanaka A (2001) Overexpression of chlorophyllide *a* oxygenas (CAO) enlarges the antenna size of photosystem II in *Arabidopsis thaliana*. Plant J 26: 365-373

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature 408: 796-815

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24: 4876-4882.

Ting CS, Rocap G, King J, Chisholm SW (2002) Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. Trends Microbiol 10: 134-142

Tomitani A, Okada K, Miyashita H, Matthijs HC, Ohno T, Tanaka A (1999) Chlorophyll *b* and phycobilins in the common ancestor of cyanobacteria and chloroplasts. Nature 400: 159-162

Vavilin D, Xu H, Lin S, Vermaas W (2003) Energy and electron transfer in photosystem II of a chlorophyll *b*-containing *Synechosystis* sp. PCC 6803 mutant. Biochemistry 42: 1731-1746

Williams JG (1988) Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methos in *Synechosystis* 6803. Methods Enzymol 167: 766-779

Xu H, Vavilin D, Vermaas W (2001) Chlorophyll *b* can serve as the major pigment in functional photosystem II complexes of cyanobacteria. Proc Natl Acad Sci USA 98: 14168-14173

Legends for figure

Fig. 1 HPLC analysis of chlorophylls from transformant mutants. *Synechocystis* sp. PCC6803 was transformed with *PhCAO* (PhCAO), and chlorophylls of transformants were extracted with 100% acetone and subjected to HPLC as described in Materials and methods. Wild types of *C. reinhardtii* and *Synechocystis* sp. PCC6803 (WT) were used as controls. Arrows indicate the peaks of chlorophyll *b*.

Fig. 2a, b Alignment of the amino acid sequence of CAO of *P. hollandica* with those of other organisms. **a** Conserved residues are indicated by a black background. Dashes indicate absent residues. The C-domain has a Rieske center and a non-heme binding site. Arrows indicate the predicted cleavage site of the transit peptide of *AtCAO*. Transit peptides of *OsCAO* could not be predicted by TargetP, and we assumed that the cleavage site is the same as that for *OsCAO*. *AtCAO*: *A. thaliana CAO*, *OsCAO*: *O. sativa CAO*, *PhCAO*: *P. hollandica CAO*. **b** This figure shows the domain structures of *AtCAO* and *PhCAO*. Chlorophyte CAO has 3 domains except for transit peptides. Prochlorophyte CAO consists of a C-domain with a

C-terminal extension of 32 amino acids.

Fig. 3 HPLC analysis of chlorophyll from transformant mutants. *Synecocystis* sp. PCC6803 was transformed by the ABC-domain (ABC), the BC-domain (BC) or the C-domain (C) of *AtCAO*. Chlorophylls of transformants were extracted with 100% acetone and subjected to HPLC. Wild types of *C. reinhardtii* and *Synechocystis* sp. PCC6803 (WT) were used as controls. Arrows indicate the peaks of chlorophyll *b*.

Fig. 4 N-terminal extensions of enzymes for chlorophyll biosyntheses. Amino acid sequences of the enzymes for chlorophyll biosynthesis of *Synechocystis* sp. PCC6803 were aligned with those of *A. thaliana*. Zero indicates the N-terminus of cyanobacteria enzymes. *hemA*: Glutamyl-tRNA reductase, *hemL*: Glutamate-1-semialdehyde aminotoransferase, *hemB*; ALA dehydratase, *hemC*: Porphobilinogen deaminase, *hemE*: Uroporphyrinogen III decarboxylase, *hemF*: Coproporphyrinogen III oxidase, *chlD*: Magnesium chelatase, *chlI*: Magnesium chelatase, chlH: Magnesium chelatase, *chlM*: Mg protoporphyrin IX methyl transferase, *porA*: Protochlorophyllide oxidoreductase, *chlP*: Chlorophyll synthetase, *chlG*: Geranylgeranyl hydrogenase, *CAO*: Chlorophyll *a* oxygenase.

Fig. 5 Alignment of the amino acid sequence of *AtCAO* with that of the CAO homologs. The black background indicates conserved residues. Dashes indicate absent residues. Amino acid sequences without transit peptides were compared. These sequences have a Rieske center and a non-heme binding site.

Fig.1

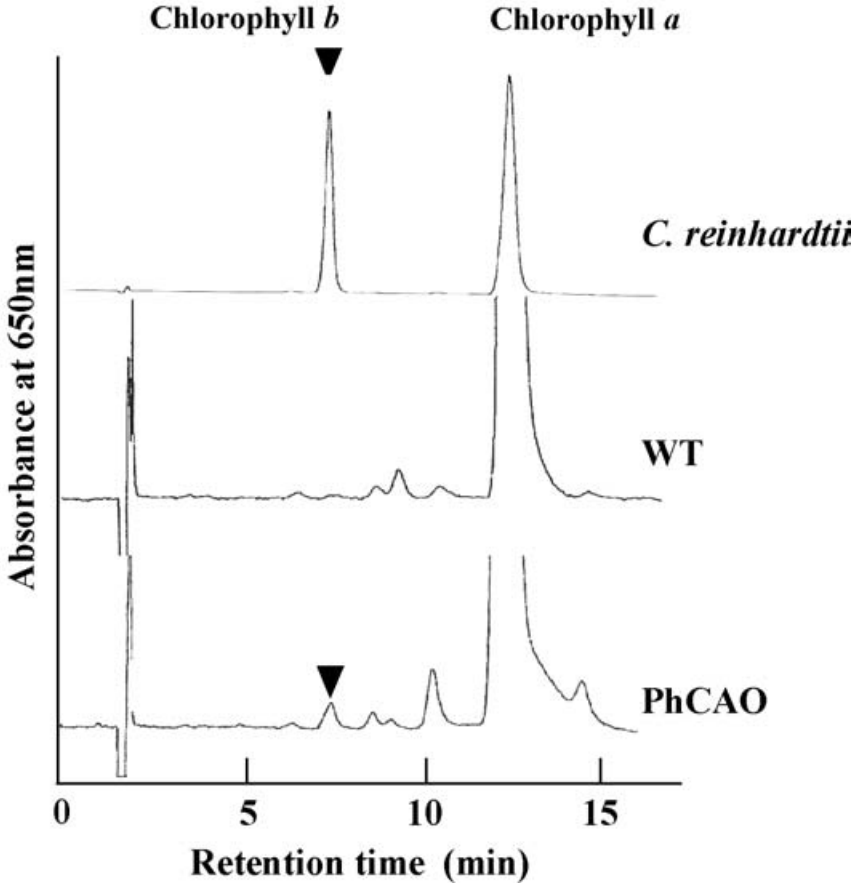
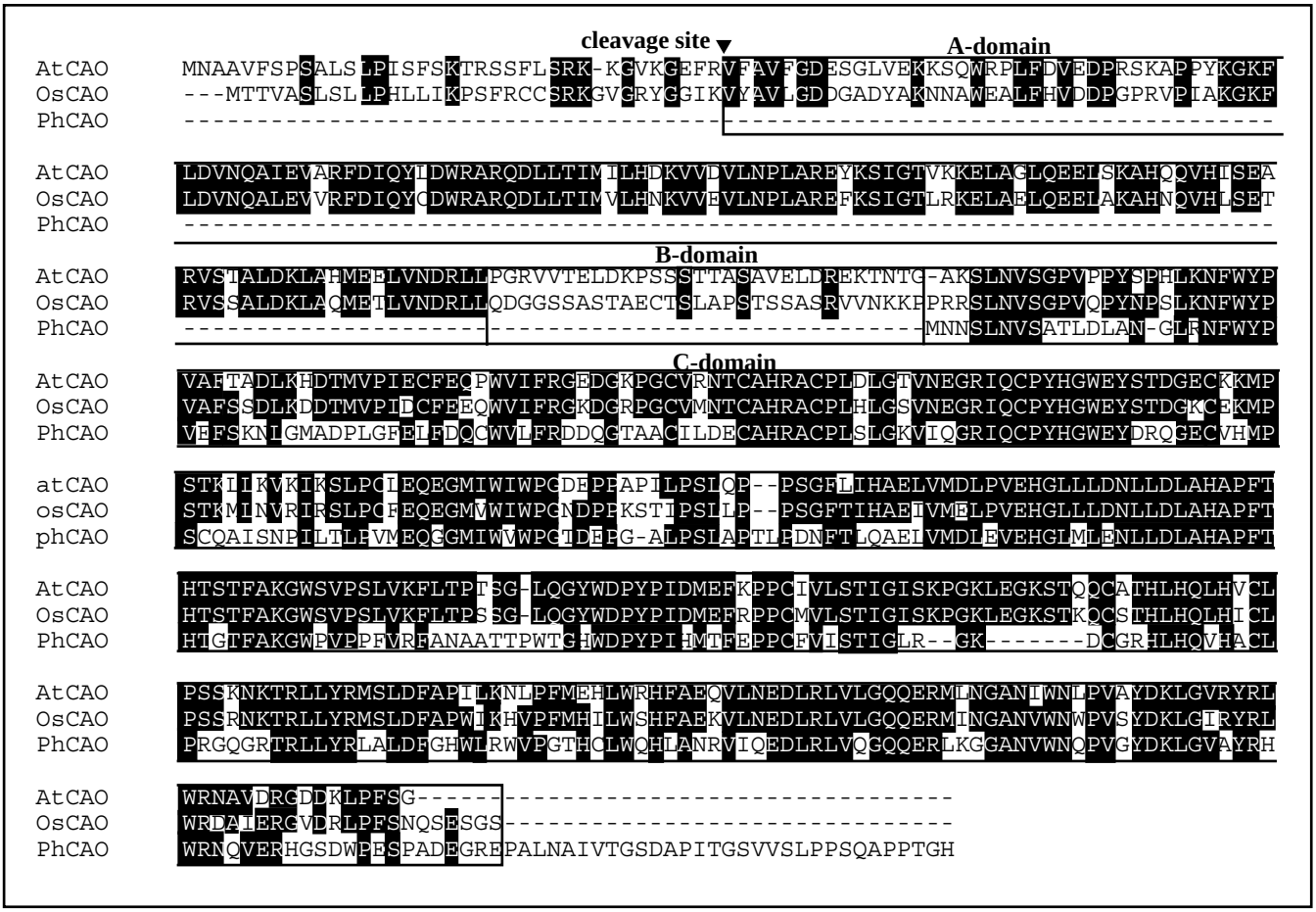

Chlorophyll b
Chlorophyll a
C. reinhardtii
WT
PhCAO
Absorbance at 650nm
0
5
10
15
Retention time (min)

Fig.2

a


cleavage site
A-domain
B-domain
C-domain
AtCAO
OsCAO
PhCAO
atCAO
osCAO
phCAO

b

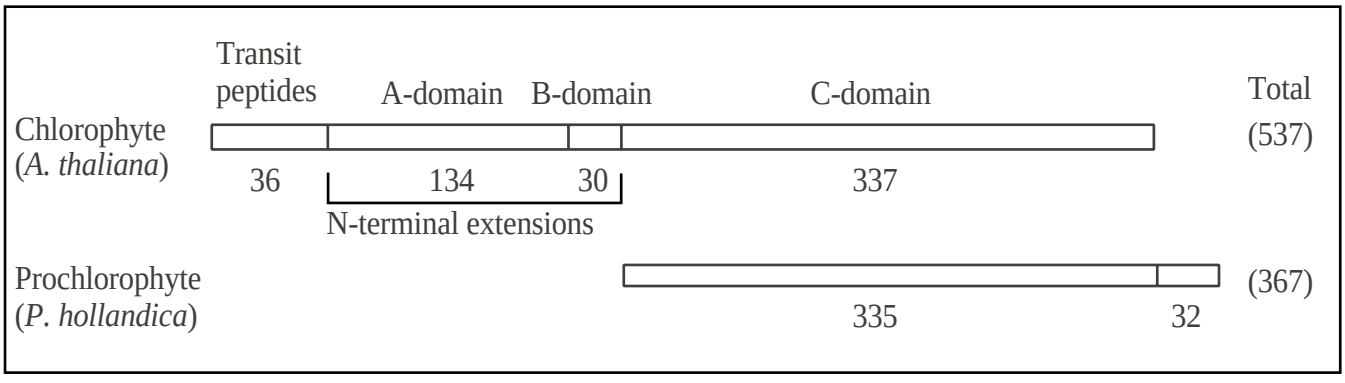

Transit peptides
A-domain
B-domain
C-domain
Total
Chlorophyte (A. thaliana)
(537)
36
134
30
337
N-terminal extensions
Prochlorophyte (P. hollandica)
(367)
335
32

Fig.3

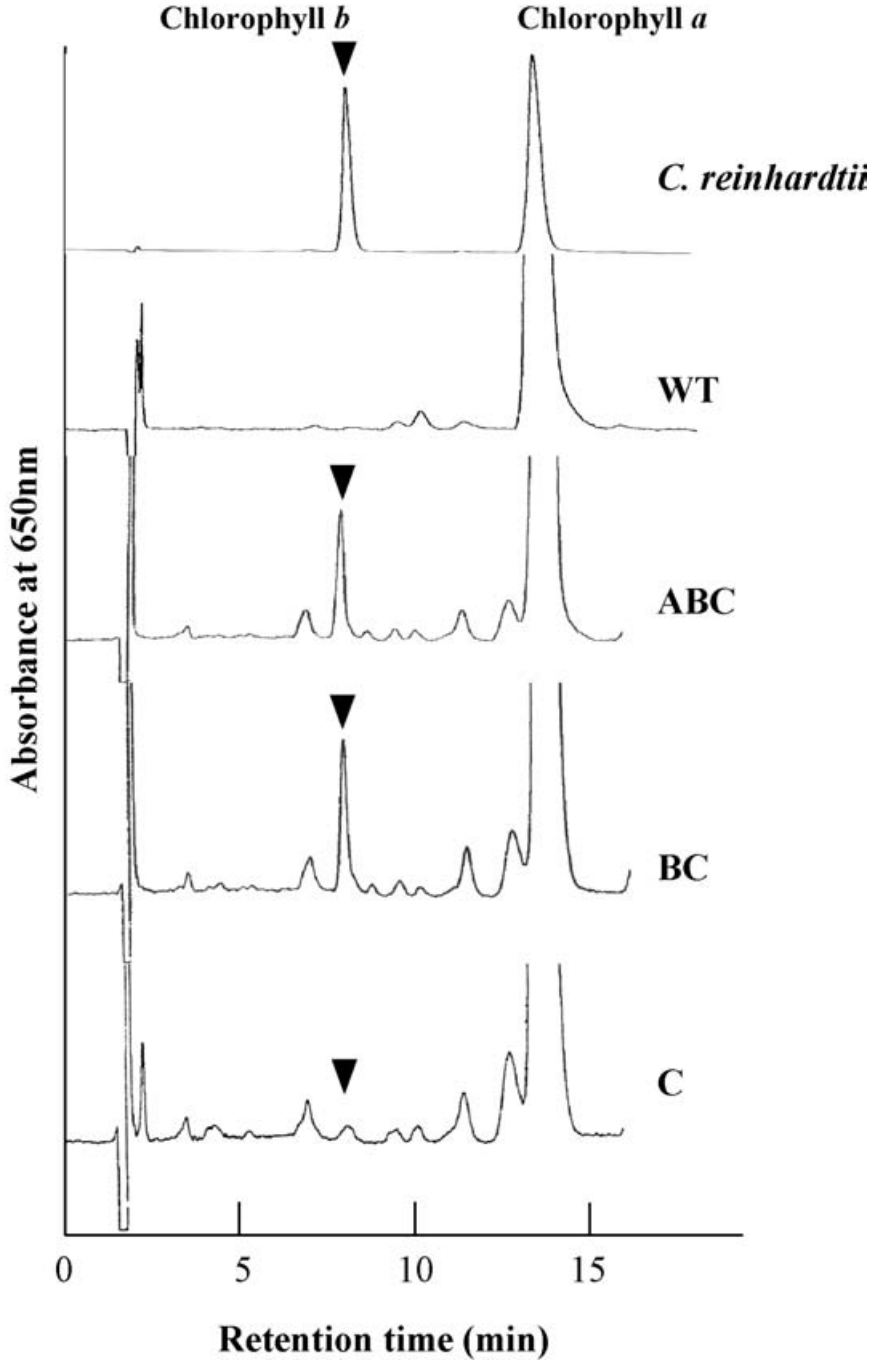

Chlorophyll b
Chlorophyll a
C. reinhardtii
WT
ABC
BC
C
Absorbance at 650nm
0
5
10
15
Retention time (min)

Fig.4

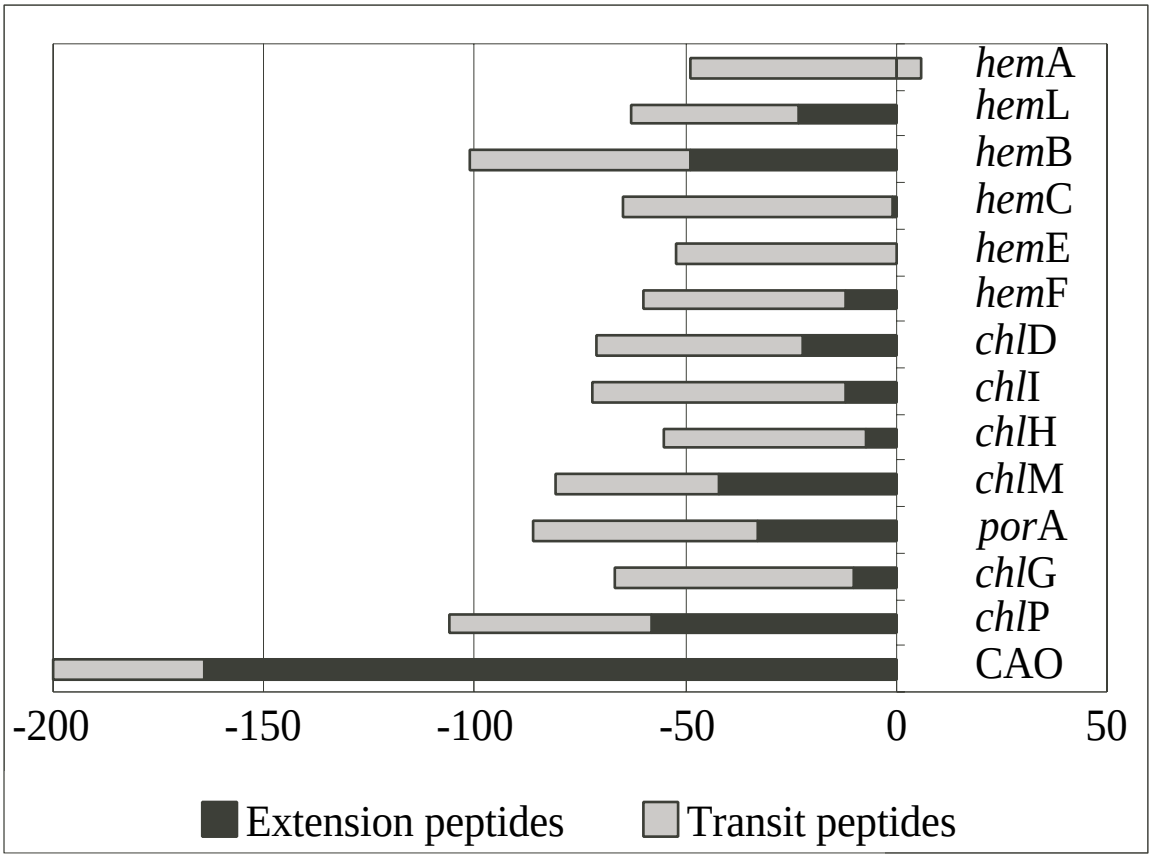
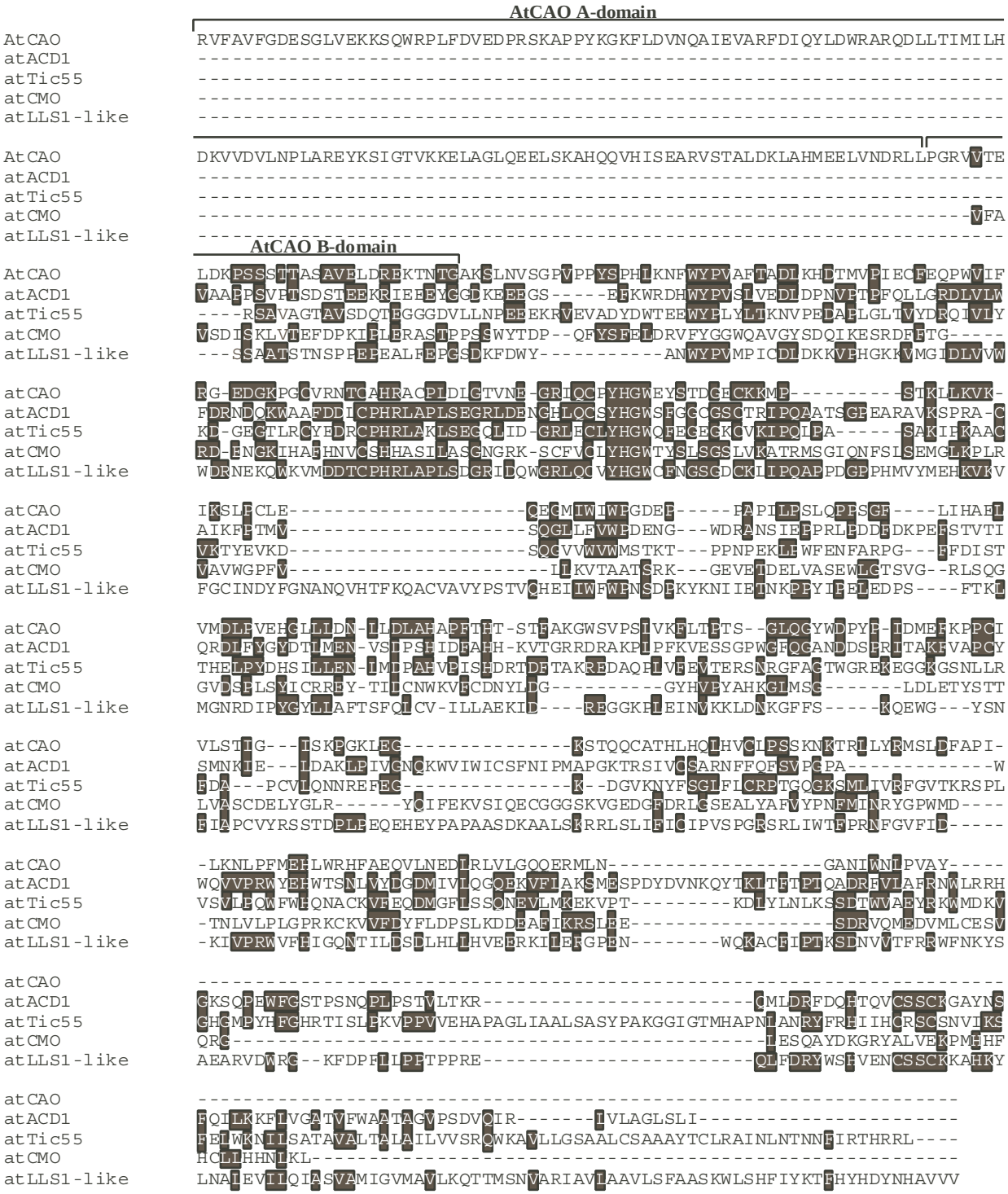

hemA
hemL
hemB
hemC
hemE
hemF
chlD
chlI
chlH
chlM
porA
chlG
chlP
CAO
-200
-150
-100
-50
0
50
Extension peptides
Transit peptides

Fig.5


AtCAO A-domain
AtCAO B-domain
AtCAO
atACD1
atTic55
atCMO
atLLS1-like