**Supplementary Information (SI Appendix)**

**Jin et al. (2016)**

# Structural insights into the LCIB protein family reveals a new group of β-carbonic anhydrases

Shengyang Jin, Jian Sun, Tobias Wunder, Desong Tang, Asaph B. Cousins, Siu Kwan Sze, Oliver Mueller-Cajar, and Yong-Gui Gao

## Supplementary Results

**Residues corresponding to Ser47 in PtLCIB4 are critical to enzymatic activity**

In our main text, we postulated that in the LCIB protein family residues corresponding to Ser47 in PtLCIB4 are critical to enzymatic activity: arginine versus small side chain residues, the former is inactive, whereas the later is active under our experimental condition. Consistent with our hypothesis, sequence and structure alignment of CrLCIB-ΔC, CrLCIC-ΔC, and PtLCIB4 with various known type I β-CAs reveals that they likewise possess small side chain residues (serine and alanine) at the corresponding position of Ser47 in PtLCIB4 (***SI Appendix* Fig. S10a**). The only exceptions are CAs from *Methanobacterium thermoautotrophicum* (MTCA, PDB: 1G5C (1)) and *Streptococcus mutans* (SMCA, PDB: 3LAS), in which a methionine residue is found at this position. MTCA and SMCA exhibit high structural similarity to each other (RMSD 0.96 Å for all atoms) and belong to a distinct subtype of β-CAs, the Cab CAs (1). Both methionine side chains (Met33/39 in MTCA/SMCA) are oriented to the opposite direction to that of Arg121 in CrLCIB-ΔC, hence they do not obstruct the contact between β2 and β2′, which harbor residues significant for central core contacts and consequently for maintaining dimerization (***SI Appendix* Fig. S10b-S10c**). Indeed, the distance between β2 and β2′ is remarkably smaller in SMCA and MTCA (~4 Å) compared to that in CrLCIB-ΔC (~8 Å). The methionine residues are positioned in a mainly hydrophobic environment constituted by both subunits. These structural features clearly indicate that the bulky methionine side chains in SMCA and MTCA do not affect the active site and dimer integrity. On the contrary, their involvement in the formation of a hydrophobic interface enhances the dimerization instead.

## Supplementary Methods

### 1. Cloning

The CrLCIB (XM_001698292.1), CrLCIC (AB168094.1), CsLCIB (WP_012627907.1) and the FjLCIB (WP_012026570.1) genes were codon optimized for expression in *E. coli* and synthesized by GenScript. The PtLCIB3 (XP_002177507.1), PtLCIB4 (BAV00141.1), and PtCA1 (AAL07493.1) gene was amplified from *P. tricornutum* cDNA. The gene fragments encoding CrLCIB-ΔC (residues 53-346) and CrLCIC-ΔC (residue 40-344) were cloned into pDBHT-2-TEV vector. Full length PtLCIB4, CsLCIB, FjLCIB and a truncated form of PtLCIB3 (residue 57-517) were cloned into pETDuet-1 vector (Novagen). The gene fragment encoding PtCA1 (N-terminal truncated, residue 47-282) was cloned into pHUE vector, permitting precise cleavage of the N-terminal $His_6$-Ubiquitin fusion protein (2). The mature form of CrLCIB (residue 53-448) and CrLCIC (residue 40-404) were inserted into pGEX-6P-1 vector. CrLCIB-LCIC co-expression construct was made by cloning CrLCIB (residue 53-448) fused with a N-terminal GST protein into the first multiple cloning site (MSC) of the pETDuet-1 vector, and CrLCIC (residue 40-404) into the second MSC. All mutants of PtLCIB4 and PtLCIB3 were generated by site-directed mutagenic PCR with KAPA HIFI DNA polymerase (KAPA

Biosystems), following the manufacturer's protocol. All the constructs were verified by DNA sequencing.

## 2. Protein expression and purification

For all recombinant proteins, *E. coli* strain BL21(DE3) transformed with the expression plasmids harboring target genes was grown to stationary phase in the presence of antibiotics, and the overexpression was induced overnight with 0.5 mM IPTG at 16 °C. The overexpressed CrLCIB-ΔC and CrLCIC-ΔC proteins were purified from cell lysate by Ni-NTA affinity chromatography followed by $His_6$-tag cleavage by TEV protease. The proteins were further purified by ion exchange chromatography and finally gel filtration using a HiLoad 16/600 Superdex200 column (GE Healthcare) in 20 mM Tris pH8.0, 150 mM NaCl and 5 mM β-mercaptoethanol (β-ME), or without β-ME when used for CA activity assays. The wild-type and mutants of PtLCIB3 and PtLCIB4, PtCA1, CsLCIB, and FjLCIB proteins were purified by Ni-NTA affinity chromatography (with $His_6$-ubituitin tag cleavage by USP2 protease for PtCA1 only) and subsequently gel filtration with HiLoad 16/600 Superdex200 column (GE Healthcare) in 20 mM Tris pH8.0, 150 mM NaCl. CrLCIB and CrLCIC are purified by glutathione-linked sepharose affinity chromatography followed by GST tag cleavage by PreScission protease. The proteins were further purified by ion exchange chromatography and then gel filtration with HiLoad 16/600 Superdex200 column (GE Healthcare) in 20 mM Tris pH8.0, 150 mM NaCl. The CrLCIB-LCIC complex was purified by Ni-NTA affinity chromatography followed by His-GST tag cleavage by PreScission protease. The protein complex was further purified by ion exchange chromatography and finally gel filtration with HiLoad 16/600 Superdex200 column (GE Healthcare) in 20 mM Tris pH8.0, 150 mM NaCl.

For the enrichment of the CrLCIB-LCIC complex from *C. reinhardtii* extracts, cells grown in Sueoka's high salt medium (3) under airlift conditions were harvested and resuspended in lysis buffer (100 mM Tris-HCl, 100 mM NaCl, 0.5 mM EDTA pH 8.0, 5m M DTT) supplemented with phenylmethylsulfonyl fluoride (PMSF) and PierceTM Protease Inhibitor Tabletes, EDTA-free (Thermo Scientific). After disruption of the cells by sonication and centrifugation with 48000g for 30 min at 4 °C, the supernatant was diluted 1:1 with MilliQ water (Merck Millipore) and applied to anion exchange chromatography (Source 30Q, GE Healthcare) using buffer A (20 mM Tris pH8.0, 50 mM NaCl, 2.5 mM DTT) and buffer B (20 mM Tris pH8.0, 1 M NaCl, 2.5 mM DTT). Fractions eluting between 145 – 335 mM NaCl were combined, diluted to 50 mM NaCl and again separated by anion exchange chromatography (source 15Q using Buffer A and Buffer B). Fractions showing the bands of interest (210 – 260 mM NaCl) were pooled and applied to a gel filtration column (Superdex200 Increase 10/300, GE Healthcare) equilibrated with 20 mM Tris pH8.0, 50 mM NaCl, and 1 mM DTT. High molecular fractions highly enriched in the CrLCIB-LCIC complex were pooled (elution volume 9.5 – 11.75 ml) and subjected to carbonic anhydrase activity assays and mass spectroscopy analysis.
Analytical gel filtration analysis of all proteins were carried out using a Superdex200 Increase 3.2/300 column in 20 mM Tris pH8.0, 50 mM NaCl.

## 3. Crystallization

Following size exclusion the pure proteins were concentrated to 10 mg/ml for CrLCIB-ΔC and CrLCIC-ΔC, and 15 mg/ml for PtLCIB4, PtLCIB4 H88A, and PtLCIB4 S47R, respectively. Crystals were produced using the hanging-drop vapor-diffusion method. Native CrLCIB-ΔC was crystallized in 100 mM Bis-Tris pH 5.5, and 25% w/v PEG3350. Native CrLCIC-ΔC was crystallized in 200 mM lithium sulfate monohydrate, 100 mM Bis-Tris pH 5.5, and 25% w/v PEG3350. Native PtLCIB4 and the PtLCIB4 H88A mutant were crystallized in 200 mM sodium acetate trihydrate, 100 mM Tris pH8.0, and 27% w/v PEG4000. The native PtLCIB4 S47R mutant was crystallized in 200 mM sodium chloride, 100 mM Bis-Tris pH6.5 and 25% w/v PEG3350. Pt-derived CrLCIC-ΔC crystals were obtained by soaking native CrLCIC-ΔC crystals into 10 mM $K_2Pt(NO_2)_4$ for 18 min. All crystals were cryo-protected in their corresponding well solutions supplemented with 15-25% v/v glycerol and frozen in liquid nitrogen prior to data collection.

## 4. Data collection and structure determination

The datasets for native CrLCIB-ΔC and PtLCIB4 H88A were collected at beamline PX1 of the Swiss Light Source (Villigen, Switzerland). The dataset of native CrLCIC-ΔC crystal was collected at beamline BL13B1 of the National Synchrotron Radiation Research Center (Taiwan, R.O.C.). The dataset of PtLCIB4 and PtLCIB4 S47R were collected at beamline MXII of the Australian Synchrotron (Melbourne, Australia). The phasing dataset for Pt-derived CrLCIC-ΔC was collected at beamline 5.0.2 of the Advanced Light Source (Berkeley, CA).

CrLCIB-ΔC and PtLCIB4 H88A native dataset was processed by XDS (4). Other datasets were processed by MOSFLM (5). Merging and scaling were done by SCALA (6). Phase determination of CrLCIC-ΔC dataset was performed by the SAD method. A single Pt site was found by SHELXD (7) , the initial experimental phase was calculated using PHENIX AUTOSOL (8) and the model was built using PHENIX AUTOBUILD (9). The figure of merit of the phasing data set was 0.72. The structures of CrLCIB-ΔC, PtLCIB4 and its mutants were solved by molecular replacement with PHASER (10) using the CrLCIC-ΔC structure as a search model. All models were built using PHENIX AUTOBUILD (9) and COOT (11). Refinement was performed using COOT and PHENIX.REFINE (12). The statistics of data collection and refinement are summarized in ***SI Appendix*** **Table S1**. All the structural figures were rendered in PyMol (Schrödinger, LLC).

## 5. Carbonic anhydrase assay

Carbonic anhydrase activity was measured using a membrane-inlet mass spectrometer (MIMS) to measure the rates of $^{18}O_2$ exchange from labeled $^{13}C^{18}O_2$ to $H_2{}^{16}O$ at 25 °C with a total carbon concentration of 0.5 mM (13,14). The hydration rates were calculated from the enhancement in the rate of $^{18}O$ loss over the uncatalyzed rate, and the non-enzymatic first-order rate constant was applied at pH

7.4 (14). The activity of CA is expressed as first-order rate constant for the hydration of $CO_2$ ($k_{CA}$; μmol mg protein$^{-1}$ s$^{-1}$ μM $CO_2^{-1}$) at 25 $^{o}$C. For each sample three to six replicate measurements were performed.

Determination of enzymatic CA activity was also carried out using the potentiometric method described by Wilbur and Anderson (15). Briefly, 1.5 ml of 20 mM Tris/sulfuric acid buffer pH 8.3 (adjusted at room temperature) was chilled on ice, containing putative CAs in the range of 0 – 0.25 mg. After starting the reaction by adding 1 ml ice-cold CO2-saturated water, the time (T) required for the pH to drop from 8.3 to 6.3 was recorded. Wilbur-Anderson units (WAU) were calculated as a measure for CA activity using the following equation:

WAU per mg = $(T_0 – T)/( T \times MASS \times 2)$

with

$T_0$ = average time (sec) for pH drop of protein free reaction, adjusted between 70-100 sec.

T = average time (sec) for pH drop of protein (CA) containing reaction.

MASS = protein (mg) added to each reaction.

## 6. Post-translational modification (PTM) analysis by mass spectrometry

The gel band was analyzed for post translational modifications as previously described (17). Briefly, the targeted gel band was excised. Protein in gel pore was reduced by DTT, alkylated with IAA and then digested by trypsin overnight. The tryptic peptides were analyzed using a Dionex Ultimate 3000 RSLC nanoLC coupled to Q-Exactive LC-MS/MS system (Thermo Electron, San Jose, CA). The LC-MS/MS data were then analyzed by Mascot (Matrix Sciences, London, UK) and PEAKS (Waterloo, ON, Canada) protein database search software. In order to obtain high confidence PTM sites, only sites identified by both programmes were used for further interpretation. In Mascot search, raw data files were converted into the mascot generic file format using ProteomeDiscoverer version 1.4 (Thermo Electron, San Jose, CA) with the MS2 spectrum processor for de-isotoping the MS/MS spectra as previously described (17). The UniProt *C. reinhardtii* database (15,176 sequence, 6,968,649 residues downloaded on 9 September 2016) was used for database search. The analysis was performed using an in-house Mascot server (version 2.4.1, Matrix Science, London, UK) with MS tolerance of 10 ppm and MS/MS tolerance of 0.02 Da. Two missed trypsin cleavage sites per peptide were tolerated. Carbamidomethylation (C) was set as a fixed modification, while oxidation (M), deamidation (NQ), Glycerophospho (ST), Phosphorylation (ST) and Phosphorylation (Y) were variable modifications. The annotated MS/MS spectra of the Mascot detected PTMs were exported as Mascot Peptide View. In another independent analysis, LC-MS/MS raw data were searched by PEAKS studio (version 7.0, Bioinformatics Solutions, Waterloo, Canada) using the PEAKS PTM function to confirm the result. 10 ppm MS and 0.02 ppm MS/MS tolerances were used for all data analysis.

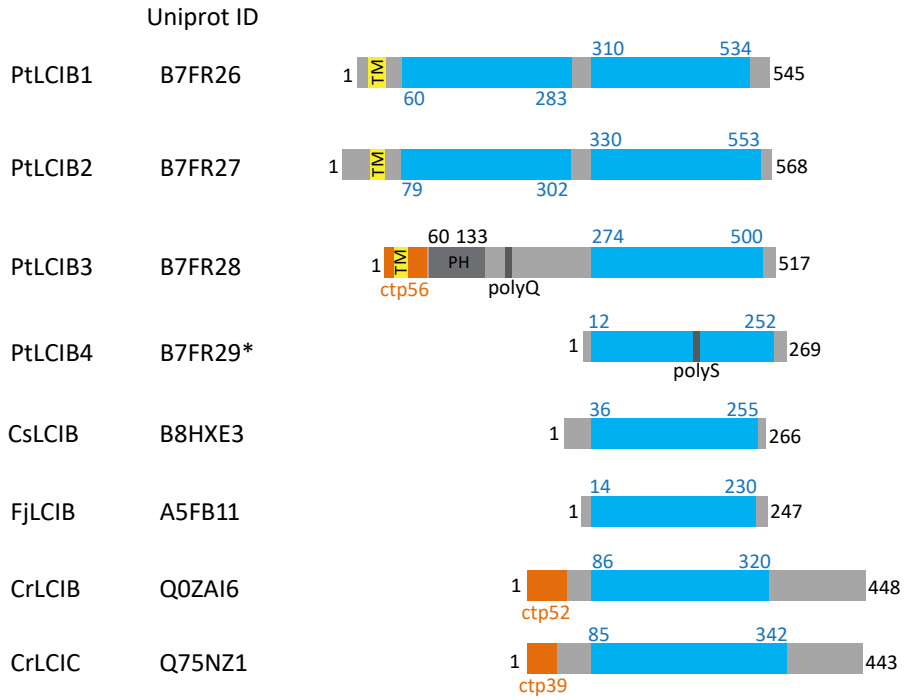
## Supplementary References


1. Strop P, Smith KS, Iverson TM, Ferry JG, & Rees DC (2001) Crystal structure of the "cab"-type beta class carbonic anhydrase from the archaeon Methanobacterium thermoautotrophicum. *The Journal of biological chemistry* 276(13):10299-10305.
2. Catanzariti AM, Soboleva TA, Jans DA, Board PG, & Baker RT (2004) An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein science : a publication of the Protein Society* 13(5):1331-1339.
3. Sueoka N (1960) Mitotic Replication of Deoxyribonucleic Acid in Chlamydomonas Reinhardi. *Proceedings of the National Academy of Sciences of the United States of America* 46(1):83-91.
4. Kabsch W (2010) Xds. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2):125-132.
5. Battye TG, Kontogiannis L, Johnson O, Powell HR, & Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography* 67(Pt 4):271-281.
6. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography* 50(Pt 5):760-763.
7. Sheldrick GM (2008) A short history of SHELX. *Acta crystallographica. Section A, Foundations of crystallography* 64(Pt 1):112-122.
8. Terwilliger TC, *et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta crystallographica. Section D, Biological crystallography* 65(Pt 6):582-601.
9. Terwilliger TC, *et al.* (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta crystallographica. Section D, Biological crystallography* 64(Pt 1):61-69.
10. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, & Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta crystallographica.*
11. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 4):486-501.
12. Afonine PV, *et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography* 68(Pt 4):352-367.
13. Badger MR & Price GD (1989) Carbonic anhydrase activity associated with the cyanobacterium Synechococcus PCC7942. *Plant physiology* 89(1):51-60.
14. Boyd RA, Gandin A, & Cousins AB (2015) Temperature response of C4 photosynthesis: Biochemical analysis of Rubisco, Phosphoenolpyruvate Carboxylase and Carbonic Anhydrase in *Setaria viridis*. *Plant physiology* 169:1850-1861.
15. Jenkins CL, Furbank RT, & Hatch MD (1989) Mechanism of $C_4$ photosynthesis A model describing the inorganic carbon pool in bundle sheath cells. *Plant physiology* 91(4):1372-1381.

16. Wilbur KM & Anderson NG (1948) Electrometric and colorimetric determination of carbonic anhydrase. *The Journal of biological chemistry* 176(1):147-154.
17. Hao P, Ren Y, Tam JP, & Sze SK (2013) Correction of errors in tandem mass spectrum extraction enhances phosphopeptide identification. *Journal of proteome research* 12(12):5548-5557.
18. Tamura K, Stecher G, Peterson D, Filipski A, & Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* 30(12):2725-2729.
19. Yamano T, *et al.* (2010) Light and low-CO2-dependent LCIB-LCIC complex localization in the chloroplast supports the carbon-concentrating mechanism in Chlamydomonas reinhardtii.

**Supplementary Figures**

Figure S1




* The PtLCIB4 sequence used in this study does not exactly match the sequence under B7FR29 but is identical with EST sequence emb|CU695874.1|.

**Figure S1. Schematic representation of LCIB homologues.** Schematic representation of the domain structure of the four *P. tricornutum* LCIB isoforms and other LCIB homologues addressed in this study. Blue bars represent domains homologous to the *P. tricornutum* LCIB-module defined as residues 12-252 of PtLCIB4. TM, predicted transmembrane helix; ctp, predicted chloroplast transit peptide; PH, pleckstrin homology domain; and poly Q/S, glutamine/ serine rich motifs.

Figure S2

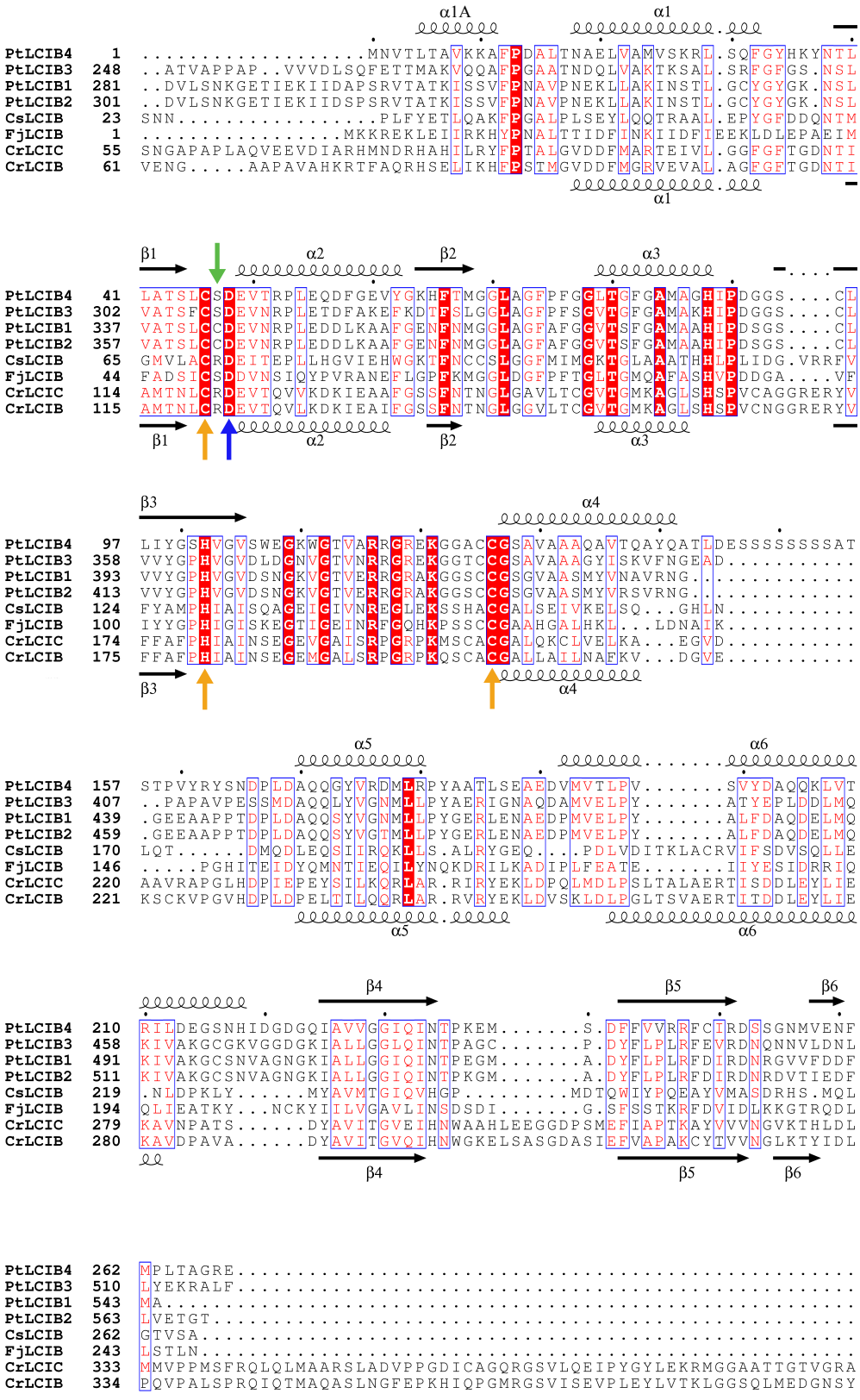



**Figure S2. Sequence alignment of LCIB homologues.** Secondary structure elements of PtLCIB4 and CrLCIB are annotated on the top and at the bottom of the alignment, respectively. Only the segments of sequences aligning with PtLCIB4 are shown. Both PtLCIB1 and PtLCIB2 contain two highly conserved (> 68% sequence identity) LCIB-like domains and the C-terminal ones were used for the alignment. Note that the critical residues at the active site, including Cys-His-Cys (indicated by orange arrows) for zinc binding and the catalytic Asp residue (indicated by a blue arrow), are conserved. The position of Arg121/Ser47 in CrLCIB/PtLCIB4 is indicated by a green arrow.

Figure S3

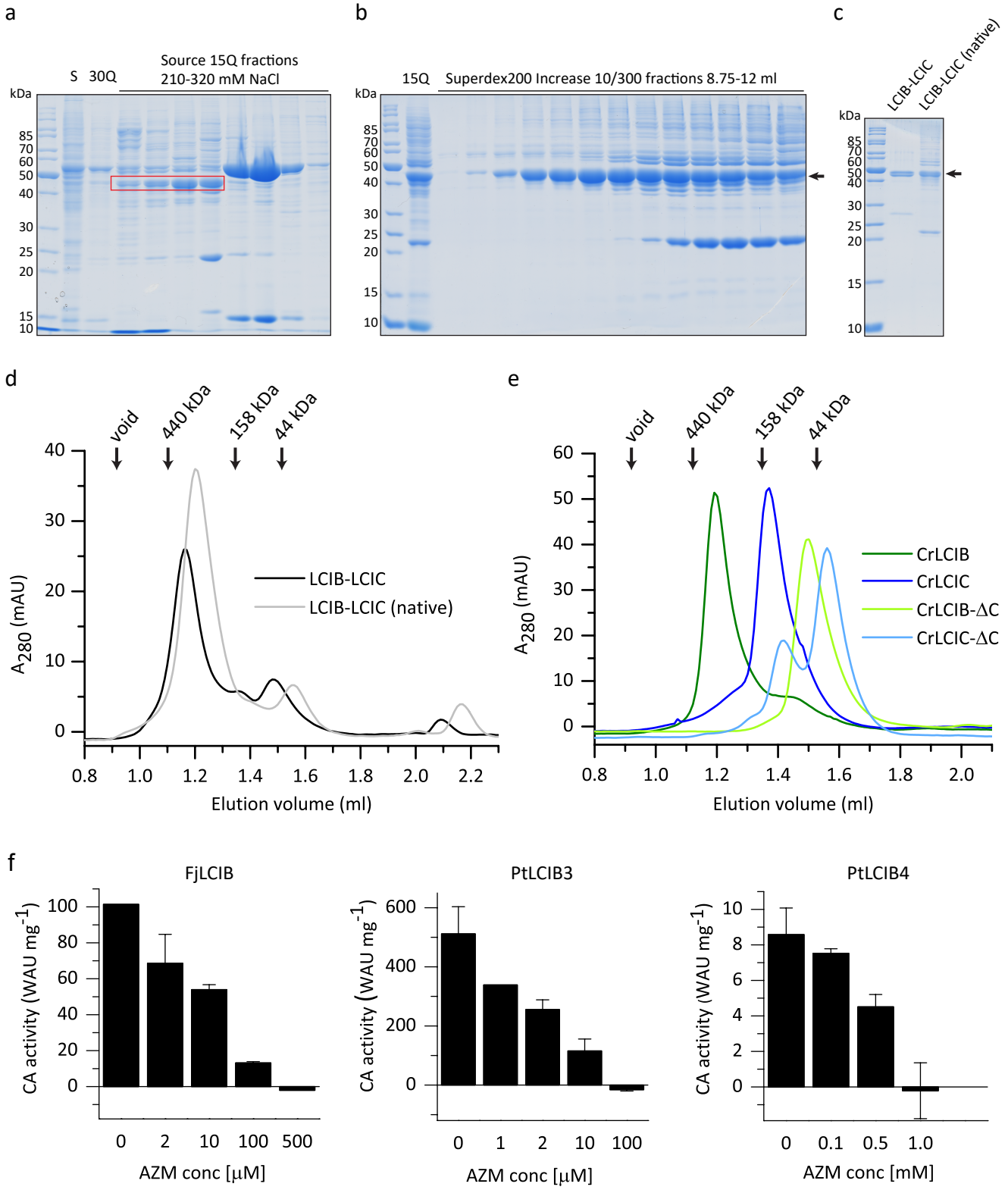



**Figure S3. Native CrLCIB-LCIC complex purification, CrLCIB/C-ΔC gel filtration and AZM inhibition of CA active LCIB homologues.** Enrichment of CrLCIB-LCIC complex from *C. reinhardtii* extract. **a,** Coomassie-stained 12% SDS-PAGE analysis of the total soluble fraction (S), combined Source 30Q fractions (30Q, elutions with 145 – 335 mM NaCl), and Source 15Q fractions eluting with 210 – 320 mM NaCl. Bands representing CrLCIB-LCIC complex are boxed in red (elutions with 210 – 260 mM NaCl). **b,** Coomassie-stained 12% SDS-PAGE analysis of combined fractions after Source 15Q (15Q, elutions with 210 – 260 mM NaCl) and Superdex 200 Increase 10/300 elutions (8.75 – 12 ml). **c,** Coomassie-stained 15% SDS-PAGE analysis of the CrLCIB-LCIC purified from *E. coli* (LCIB-LCIC) and from *C. reinhardtii* (LCIB-LCIC (native), combined gel filtration elutions 9.5 – 11.75 ml). **d,** Analytical gel filtration of CrLCIB-LCIC purified from *E. coli* (LCIB-LCIC) and from *C. reinhardtii* (LCIB-LCIC (native)). **e,** Analytical gel filtration of CrLCIB/C and CrLCIB/C-ΔC. **f,** Wilbur-Anderson CA activity assays of the recombinant FjLCIB, PtLCIB3, and PtLCIB4 proteins in the presence of 0 – 1000 μM AZM (acetazolamide).

Figure S4

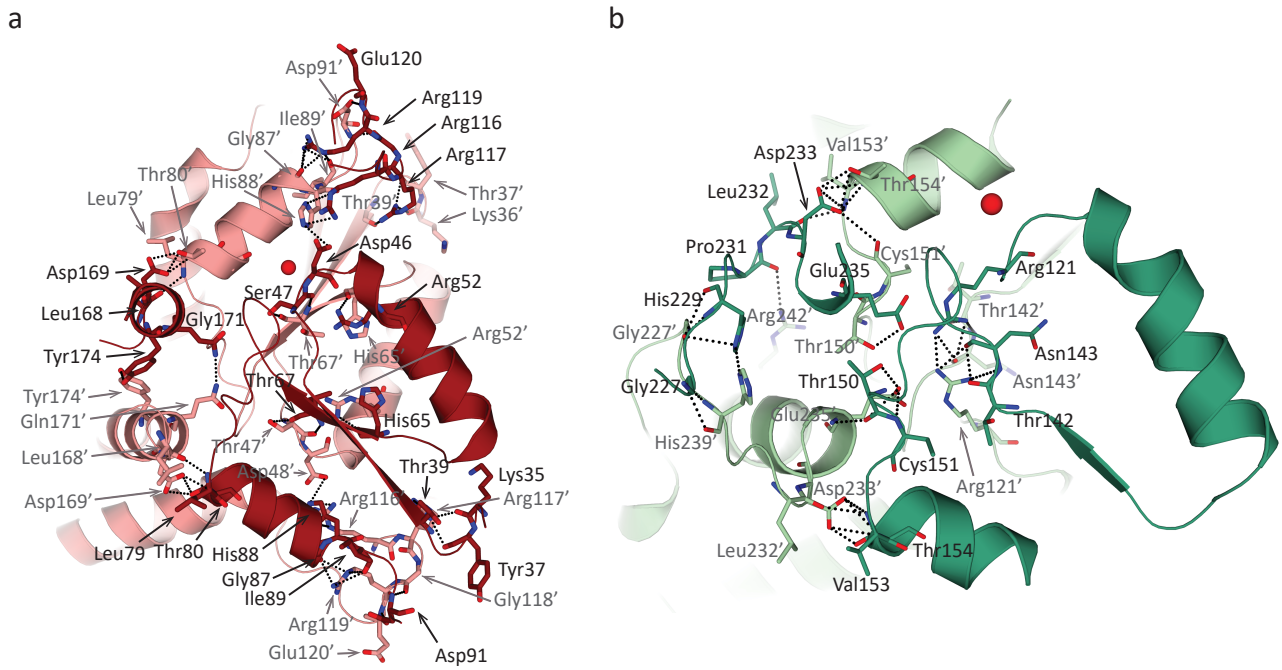



**Figure S4. Detailed interactions at the dimeric interface of CrLCIB-ΔC and PtLCIB4.** Zn is represented as red sphere. **a,** The two subunits of the PtLCIB4 dimer are shown in dark and pale red, respectively. Residues involved in dimeric interaction are labeled. **b,** The two subunits of CrLCIB-ΔC dimer are shown in dark and pale green, respectively. Residues involved in dimeric interaction are labeled.

Figure S5

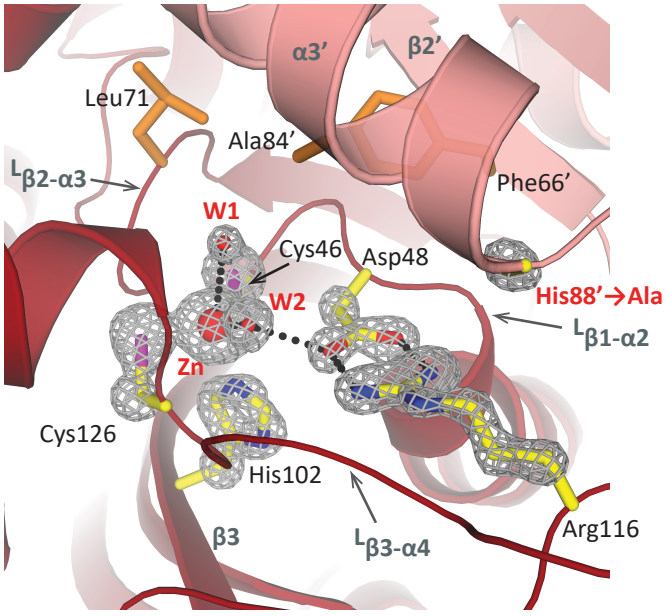



**Figure S5. The active site of the PtLCIB4 H88A mutant.** The two subunits of the PtLCIB4 H88A dimer are colored dark and pale red. The polar and hydrophobic residues in the active site are colored yellow and orange, respectively. *Fo - Fc* electron density map (grey mesh) of active site residues is calculated with omission of the enveloped atoms.

Figure S6

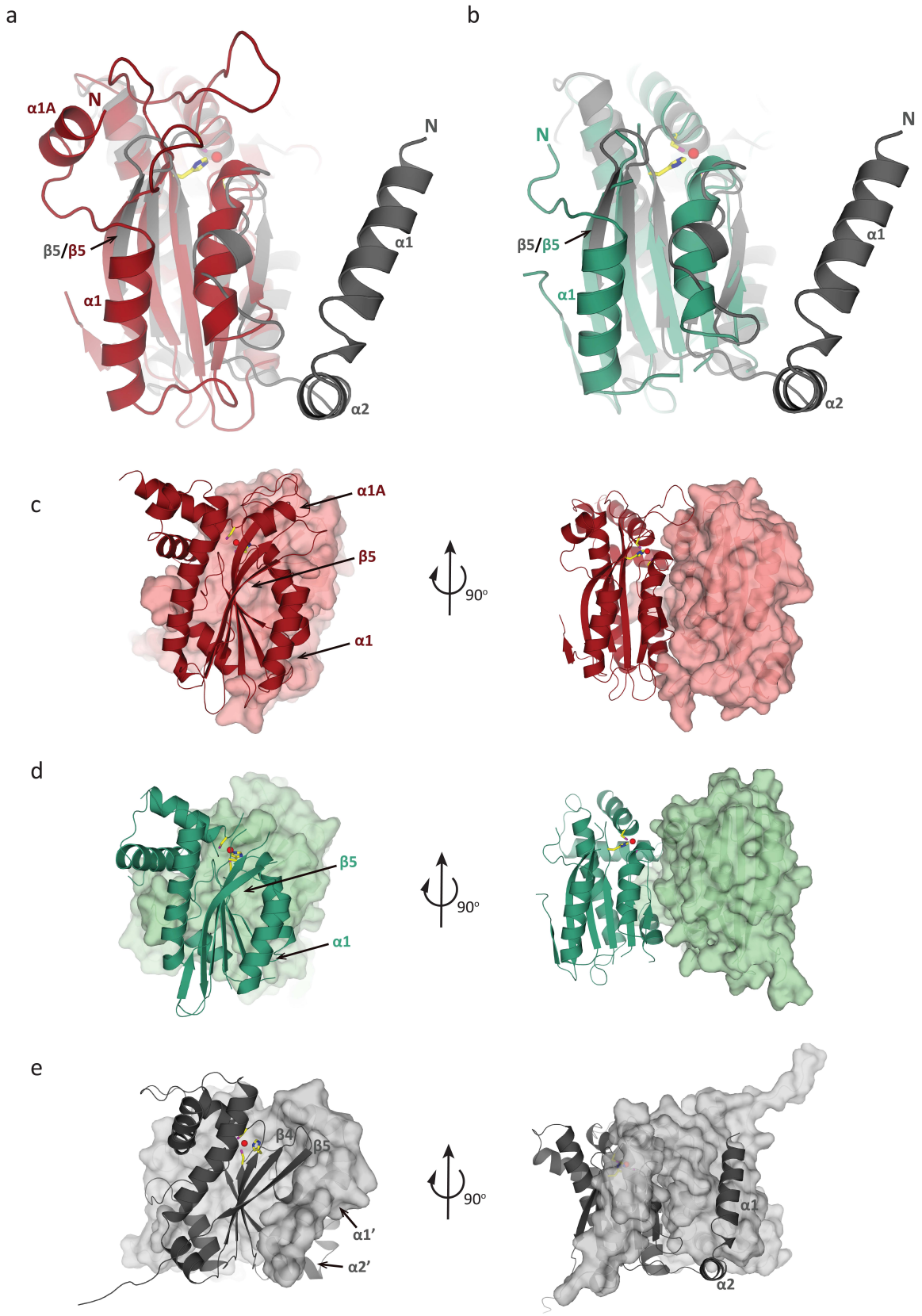



**Figure S6. Structural comparison of the N-terminal extension of CrLCIB-ΔC, PtLCIB4, and PSCA.** The N-terminus is indicated by "N" and the relevant secondary structure elements are labeled. **a,** Superposition of the PtLCIB4 (red) and PSCA (grey) structures. **b,** Superposition of the CrLCIB-ΔC (green) and PSCA (grey) structures. The dimerization arrangement of PtLCIB4 (**c**), CrLCIB-ΔC (**d**), and PSCA (**e**) are shown with one of the subunits depicted in cartoon and the other as surface.

Figure S7

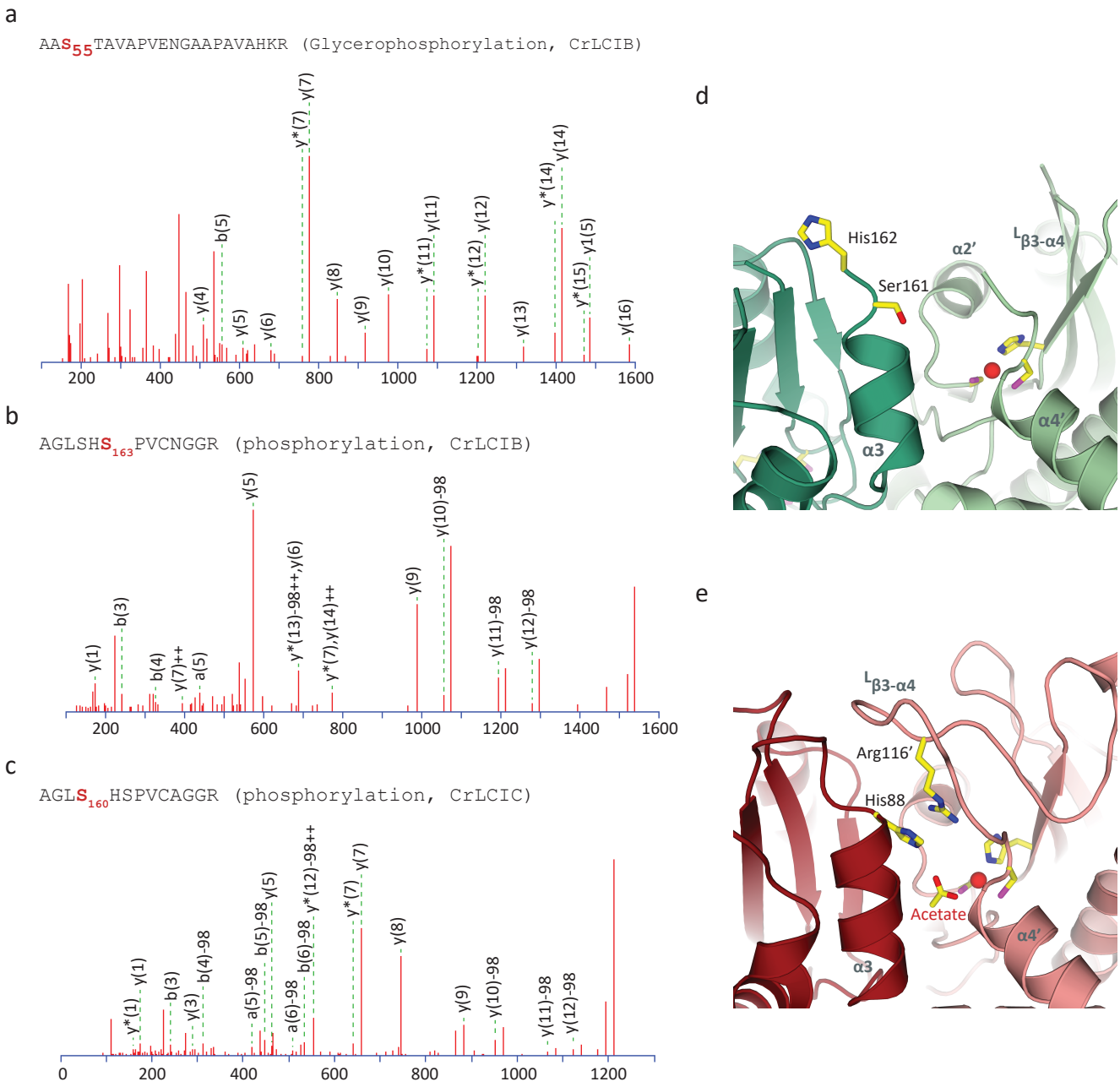



**Figure S7. LC-MS/MS based post-translational modification analysis of the native CrLCIB-LCIC complex and modelling of CrLCIB-ΔC structure.** The phosphorylated peptides were identified with high confidence. The annotated MS/MS spectra showed the assignment of the peaks to b- and y-ions fragments, and the peptides were identified by both Mascot and PEAKS software with significant scores. The Mascot annotated spectra generated for one N-terminal glycerophosphorylation in CrLCIB (**a**), one serine phosphorylation in CrLCIB (**b**) and in CrLCIC (**c**) are shown with the identified peptide and modification site. **d,** Modelled CrLCIB-ΔC dimer is shown in dark and pale green. His162 (corresponds to the catalytically important His88 in PtLCIB4) and Ser161 are shown in yellow sticks. The Zinc ion is shown as red sphere with the coordinating residues shown in yellow sticks. The C-terminal end of Lβ3-α4 is indicated while the rest is disordered, including the phosphorylated Ser163. A similar view of the PtLCIB4 dimer with the His88 residue positioned in a catalytically active conformation (toward the acetate) is also shown for comparison (**e**).

Figure S8

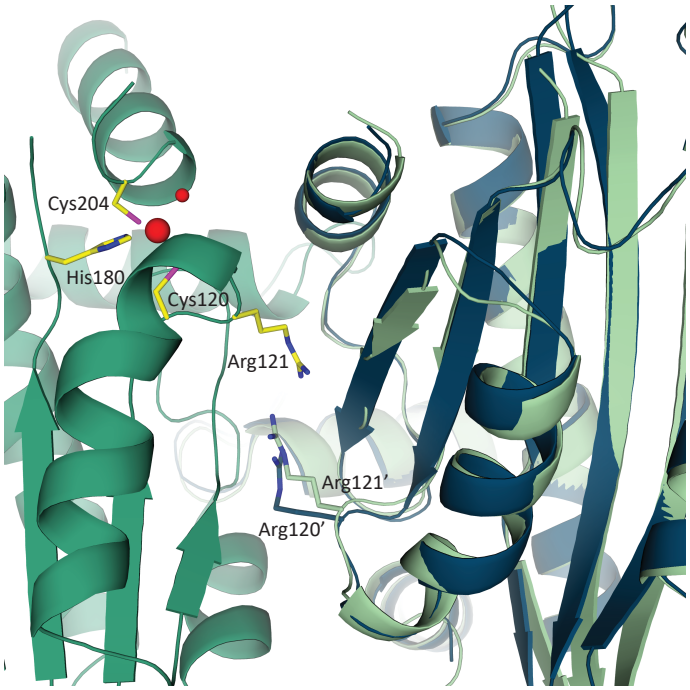



**Figure S8. A proposed model of the CrLCIB-LCIC heterodimer.** The two subunits of the CrLCIB-ΔC homodimer are colored green and pale green. The CrLCIC-ΔC subunit (blue) is aligned onto one subunit (pale green) of CrLCIB-ΔC dimer, with a RMSD of 0.6 Å for all atoms. The residues at the active site of CrLCIB-ΔC are shown as yellow sticks. Zn and water are shown in larger and smaller red spheres, respectively. The Arg121/120 of CrLCIB/C are depicted and labeled.

Figure S9

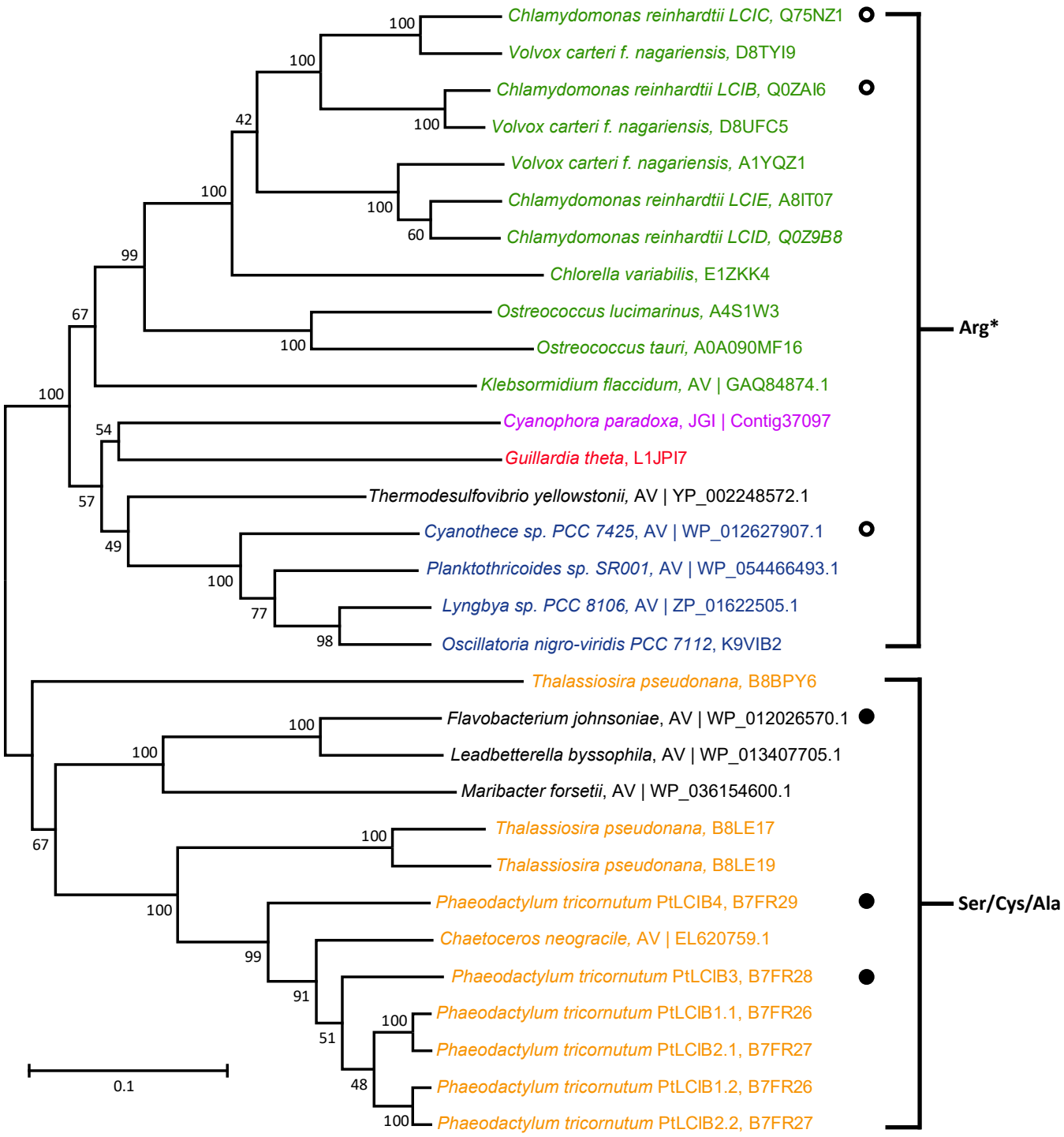



**Figure S9. Phylogenetic tree of selected LCIB homologues.** A partial sequence alignment of LCIB homologue sequences corresponding to residues (12-252) of PtLCIB4 was performed. The alignment was then used to produce a phylogenetic tree by Neighbor-Joining in the MEGA6 (18) software suite using the p-distance model and pairwise gap deletion. Bootstrap values are the percentage from 1000 replicates where a given node was present. For PtLCIB1/2 which contain two LCIB-homologous modules, each module was aligned separately and annotated .1 or .2. Among the tested LCIB homologues, the CA active ones are indicated by a solid sphere and the CA inactive ones by a hollow sphere. The Uniprot accession number (no prefix), NCBI accession version identifier (AV) or Joint Genome Institute ID (JGI) is shown for each protein after the species name. Green: green algae; Orange: diatoms; Red: flagellate cryptomonad algae; Blue: cyanobacteria; and Purple: other algae. Black: other bacteria. "*": All sequences bracketed have arginine at the discussed position except for LCIE, which contains a histidine instead. LCIE also lacks one of the cysteines that putatively coordinates the zinc ion. At the position corresponding to Ser47 in PtLCIB4, Ser, Cys or Ala was found in the other LCIB family members within the bracket indicated by Ser/Cys/Ala. The tree is extended from Yamano et al (19).

Figure S10

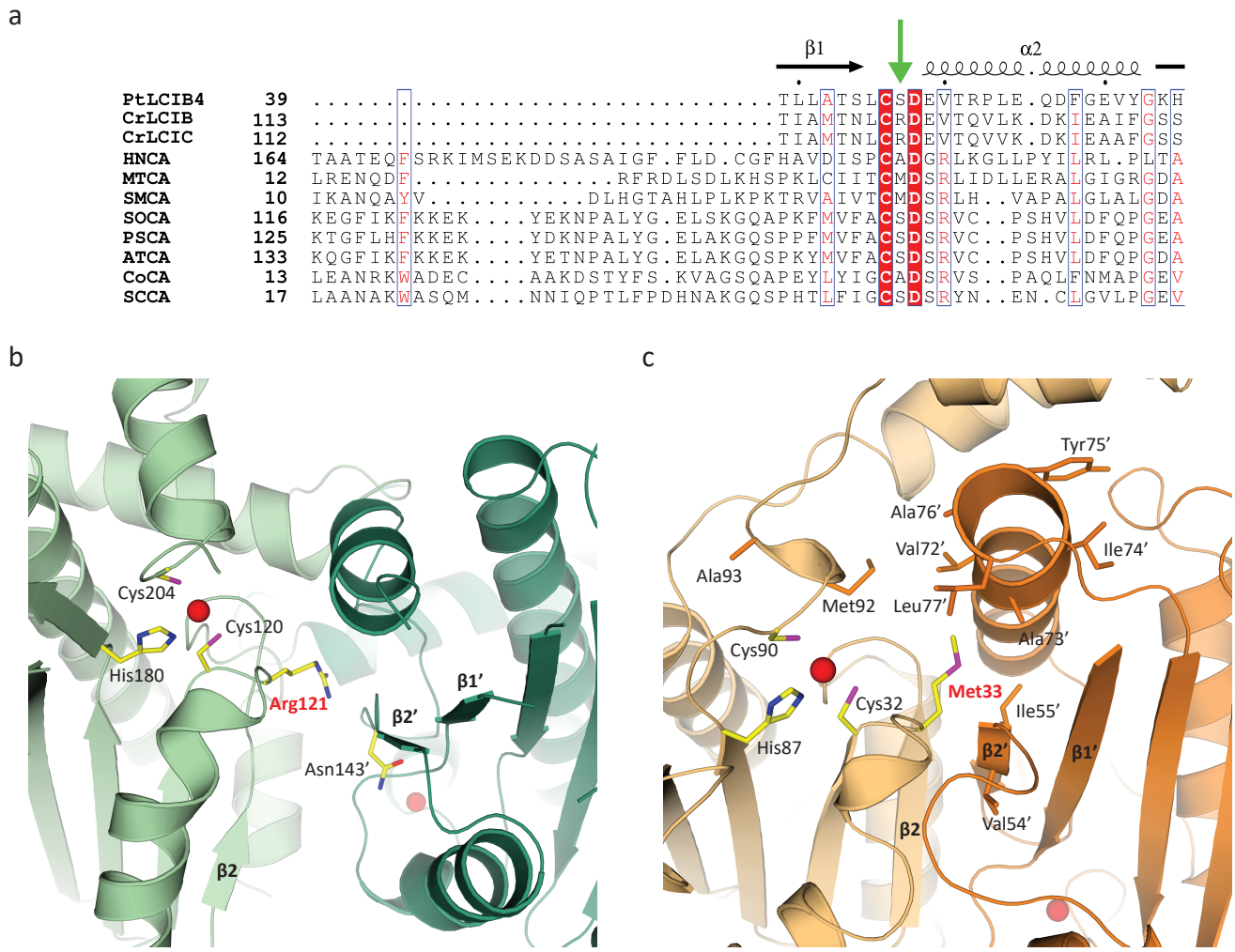



**Figure S10. Sequence alignment and structure comparison of CrLCIB-ΔC, CrLCIC-ΔC and PtLCIB4 with known β-CAs. a,** Partial sequence alignment of PtLCIB4, CrLCIB-ΔC and CrLCIC-ΔC with known β-CAs are conducted by Clustal Omega and rendered in ESPript 3.0. The variation of Ser47 in PtLCIB4 across species is indicated by a green arrow. HNCA, CA from *Halothiobacillus neopolitanus*. MTCA, CA from *Methanobacterium thermoautotrophicum*. SMCA, CA from *Porphyridium purpureum*. SOCA, CA from *Spinacia oleracea*. ATCA, *Arabidopsis thaliana*. CoCA, CA from *Coccomyxa*. SCCA, *Saccharomyces cerevisiae*. The two subunits of PtLCIB4 dimer (**b**), MTCA dimer (**c**) are colored green/pale green and orange/pale yellow, respectively. Zn is shown as red sphere. The residues near the active site are shown as yellow sticks. Hydrophobic residues of MTCA are shown in orange sticks. Despite both being a large side chain, the positioning of Arg121 and Met33 relative to the dimer interface is apparently different.

# Supplementary Table

**Table S1 Data collection and refinement statistics**

| Crystals | CrLCIC Pt-derivative | CrLCIB-ΔC | CrLCIC-ΔC | PtLCIB4 | PtLCIB4 H88A | PtLCIB4 S47R |
|---|---|---|---|---|---|---|
| **PDB code** | | 5K5W | 5B5X | 5B5Y | 5B5Z | 5B60 |
| **Data collection** | | | | | | |
| Wavelength (Å) | 1.000 | 1.000 | 1.000 | 0.954 | 1.000 | 0.954 |
| Space group | I 2 2 2 | $P2_12_12_1$ | I 2 2 2 | P1 | P1 | C 1 2 1 |
| Cell Dimensions | | | | | | |
| a (Å) | 56.2 | 43.8 | 56.0 | 47.9 | 48.08 | 68.55 |
| b (Å) | 87.8 | 80.2 | 88.0 | 48.0 | 48.12 | 66.11 |
| c (Å) | 100.8 | 143.8 | 100.8 | 60.0 | 59.92 | 69.06 |
| α (º) | 90.0 | 90.0 | 90.0 | 88.0 | 76.78 | 90.0 |
| β (º) | 90.0 | 90.0 | 90.0 | 102.79 | 87.96 | 101.0 |
| γ (º) | 90.0 | 90.0 | 90.0 | 93.31 | 86.05 | 90.0 |
| Resolution (Å) | 49.06-2.05 (2.16-2.05)[a] | 71.88-2.59 (2.75-2.59) | 66.31-2.50 (2.54-2.50) | 58.48-1.75 (1.84-1.75) | 58.32-1.60 (1.69-1.60) | 67.79-2.20 (2.32-2.20) |
| $R_{sym}$ (%) | 6.9 (30.2) | 10.2 (72.0) | 9.7 (48.8) | 9.5 (50.1) | 3.8 (11.0) | 10.5 (83.8) |
| <I/σ> | 15.2 (4.9) | 10.45 (1.7) | 27.6 (3.0) | 9.9 (3.7) | 18.6 (8.4) | 11.0 (2.7) |
| Redundancy | 5.8 (5.1) | 3.3 (3.4) | 5.8 (4.5) | 3.7 (3.8) | 3.5 (3.5) | 8.8 (9.0) |
| Unique reflection | 16092 | 16427 | 8828 | 52517 | 68811 | 15494 |
| Completeness (%) | 99.2 (98.1) | 97.3 (95.0) | 98.3 (90.0) | 93.5 (93.5) | 95.1 (94.0) | 99.5 (99.6) |
| **Refinement** | | | | | | |
| Resolution (Å) | | 41.93-2.59 | 34.47-2.51 | 36.66-1.75 | 47.96-1.60 | 47.16-2.20 |
| No. reflections | | 15992 | 8636 | 49104 | 65411 | 15400 |
| $R_{work}/R_{free}$ (%) | | 26.3/27.5 | 21.6/28.1 | 17.5/20.2 | 18.4/21.3 | 25.0/25.4 |
| r.m.s.d bonds/angle | | 0.01/1.85 | 0.01/1.27 | 0.01/1.15 | 0.01/1.14 | 0.01/1.27 |
| Number of atoms | | | | | | |
| Protein | | 3231 | 1714 | 3592 | 3594 | 1603 |
| Ligand | | 0 | 0 | 8 | 0 | 0 |
| Ion | | 2 | 6 | 2 | 2 | 2 |
| Water | | 64 | 14 | 270 | 352 | 24 |
| B factors | | | | | | |
| Protein | | 44.2 | 49.5 | 19.9 | 24.3 | 52.2 |
| Ligand/ion | | 34.9 | 49.7 | 19.2 | 14.6 | 55.2 |
| Water | | 40.4 | 46.9 | 26.6 | 31.6 | 48.9 |
| r.m.s. deviations | | | | | | |
| Bond length (Å) | | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| Bond angles (º) | | 1.85 | 1.27 | 1.15 | 1.14 | 1.27 |
| Ramachandran | | | | | | |
| Favored (%) | | 91.34 | 96.31 | 99.36 | 98.73 | 94.37 |
| Allowed (%) | | 7.18 | 3.23 | 0.64 | 0.85 | 4.69 |
| Outlier (%) | | 1.49 | 0.46 | 0.00 | 0.42 | 0.94 |

[a] Values in parentheses are for the highest-resolution shell.