

TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA - SER. A I OSA - TOM. 417

ASTRONOMICA - CHEMICA - PHYSICA - MATHEMATICA

**THE STRUCTURAL BASIS FOR
INORGANIC PYROPHOSPHATASE
CATALYSIS AND REGULATION**

by

Heidi Tuominen

TURUN YLIOPISTO
UNIVERSITY OF TURKU
Turku 2011

From the Department of Biochemistry and Food Chemistry
University of Turku
Turku, Finland

Supervised by:

Professor Reijo Lahti, Ph.D.
Department of Biochemistry and Food Chemistry
University of Turku
Turku, Finland

and

Professor Alexander A. Baykov, Ph.D.
A.N. Belozersky Institute of Physico-chemical Biology
Moscow State University
Moscow, Russia

Reviewed by:

Professor Aurelio Serrano, Ph.D.
Institute of Plant Biochemistry and Photosynthesis,
University of Sevilla
Sevilla, Spain

and

Kristian Koski, Ph.D.
Department of Biochemistry and Biocenter Oulu
University of Oulu
Oulu, Finland

Opponent:

Professor Tiina Salminen, Ph.D.
Department of Biosciences
Åbo Akademi University
Turku, Finland

ISBN 978-951-29-4586-3 (PRINT)
ISBN 978-951-29-4587-0 (PDF)
ISSN 0082-7002
Painosalama Oy – Turku, Finland 2011

On niitä, jotka saavat asiat tapahtumaan,
niitä, jotka katsovat, kun asiat tapahtuvat
ja niitä, jotka ihmettelevät mitä mahtoi tapahtua.
- Manfred Kets de Vries

To Vesa, Samu and Jonne

Contents

| | |
|---|-----------|
| List of Original Publications | 5 |
| Abstract | 6 |
| Abbreviations | 7 |
| Abbreviations for Amino Acids | 8 |
| 1 Review of the Literature | 9 |
| 1.1 Inorganic pyrophosphatases..... | 9 |
| 1.1.1 Family I PPases..... | 9 |
| 1.1.2 Family II PPases..... | 12 |
| 1.2 The CBS domain..... | 15 |
| 1.2.1 Proteins with the CBS domain..... | 15 |
| 1.2.2 Natural mutations in the CBS domain..... | 16 |
| 1.2.3 Structures of CBS domain proteins..... | 17 |
| 1.2.4 Ligand binding to the CBS domain..... | 20 |
| 1.2.5 Conformational changes induced by the CBS domain..... | 21 |
| 1.2.6 Function of the CBS domain..... | 23 |
| 1.3 The DRTGG domain..... | 24 |
| 2 Aims of the Study | 25 |
| 3 Methods | 26 |
| 3.1 Cloning, mutagenesis, protein expression and purification (II-IV)..... | 26 |
| 3.2 Characterization of proteins (II-IV)..... | 26 |
| 3.3 Crystallization and structure determination (I and III)..... | 27 |
| 3.4 Homology modeling..... | 29 |
| 4 Results and Discussion | 30 |
| 4.1 Complete catalytic cycle of family I PPases (I)..... | 30 |
| 4.2 Production and initial characterization of CBS-PPases (II and III)..... | 32 |
| 4.3 Overall structure of CBS-PPase (III)..... | 33 |
| 4.4 Ligand binding to CBS domains..... | 34 |
| 4.4.1 Adenylate binding to <i>cp</i> CBS-PPase (III)..... | 34 |
| 4.4.2 Comparison with other CBS domain proteins..... | 35 |
| 4.5 Proposed mechanism of <i>cp</i> CBS-PPase regulation (III)..... | 36 |
| 4.6 The DRTGG domain (III)..... | 38 |
| 4.7 <i>mt</i> CBS-PPase variant analysis (IV)..... | 40 |
| 4.8 Possible physiological role of CBS-PPase regulation..... | 43 |
| 5 Conclusions | 45 |
| Acknowledgements | 47 |
| References | 49 |
| Original Publications | 57 |

List of Original Publications

This thesis is based on the following original publications, which are referred to in the text by the Roman numerals I-IV:

- I Oksanen E., Ahonen A.-K., Tuominen H., Tuominen V., Lahti R., Goldman A. and Heikinheimo P. (2007) A complete structural description of the catalytic cycle of yeast pyrophosphatase. *Biochemistry*, **46**, 1228–1239.
- II Jämsen J.*, Tuominen H.*, Salminen A., Belogurov G. A., Magretova N. N., Baykov A. A. and Lahti R. (2007) A CBS domain-containing pyrophosphatase of *Moorella thermoacetica* is regulated by adenine nucleotides. *Biochem J.* **408**, 327–333.
- III Tuominen H., Salminen A., Oksanen E., Jämsen J., Heikkilä O., Lehtiö L., Magretova N. N., Goldman A., Baykov A. A. and Lahti R. (2010) Crystal structures of the CBS and DRTGG domains of the regulatory region of *Clostridium perfringens* pyrophosphatase complexed with the inhibitor, AMP, and activator, diadenosine tetraphosphate. *J. Mol. Biol.* **398**, 400–413.
- IV Jämsen J., Tuominen H., Baykov A. A. and Lahti R. (2011) Mutational analysis of residues in the regulatory CBS domains of *Moorella thermoacetica* pyrophosphatase corresponding to disease-related residues of human proteins, *Biochem J.* **433**, 497–504.

* These two authors contributed equally to this paper.

The original publications are reproduced in this thesis with permission from the copyright owners.

Abstract

Inorganic pyrophosphatases (PPases) are essential enzymes for every living cell. PPases provide the necessary thermodynamic pull for many biosynthetic reactions by hydrolyzing pyrophosphate. There are two types of PPases: integral membrane-bound and soluble enzymes. The latter type is divided into two non-homologous protein families, I and II. Family I PPases are present in all kingdoms of life, whereas family II PPases are only found in prokaryotes, including archae.

Family I PPases, particularly that from *Saccharomyces cerevisiae*, are among the most extensively characterized phosphoryl transfer enzymes. In the present study, we have solved the structures of wild-type and seven active site variants of *S. cerevisiae* PPase bound to its natural metal cofactor, magnesium ion. These structures have facilitated derivation of the complete enzyme reaction scheme for PPase, fulfilling structures of all the reaction intermediates.

The main focus in this study was on a novel subfamily of family II PPases (CBS-PPase) containing a large insert formed by two CBS domains and a DRTGG domain within the catalytic domain. The CBS domain (named after cystathionine β -synthase in which it was initially identified) usually occurs as tandem pairs with two or four copies in many proteins in all kingdoms of life. The structure formed by a pair of CBS domains is also known as a Bateman domain. CBS domains function as regulatory units, with adenylate ligands as the main effectors. The DRTGG domain (designated based on its most conserved residues) occurs less frequently and only in prokaryotes. Often, the domain co-exists with CBS domains, but its function remains unknown.

The key objective of the current study was to explore the structural rearrangements in the CBS domains induced by regulatory adenylate ligands and their functional consequences. Two CBS-PPases were investigated, one from *Clostridium perfringens* (*cp*CBS-PPase) containing both CBS and DRTGG domains in its regulatory region and the other from *Moorella thermoacetica* (*mt*CBS-PPase) lacking the DRTGG domain. We additionally constructed a separate regulatory region of *cp*CBS-PPase (*cp*CBS). Both full-length enzymes and *cp*CBS formed homodimers. Two structures of the regulatory region of *cp*CBS-PPase complexed with the inhibitor, AMP, and activator, diadenosine tetraphosphate, were solved. The structures were significantly different, providing information on the structural pathway from bound adenylates to the interface between the regulatory and catalytic parts. To our knowledge, these are the first reported structures of a regulated CBS enzyme, which reveal large conformational changes upon regulator binding. The activator-bound structure was more open, consistent with the different thermostabilities of the activator- and inhibitor-bound forms of *cp*CBS-PPase.

The results of the functional studies on wild-type and variant CBS-PPases provide support for inferences made on the basis of structural analyses. Moreover, these findings indicate that CBS-PPase activity is highly sensitive to adenine nucleotide distribution between AMP, ADP and ATP, and hence to the energy level of the cell. CBS-PPase activity is markedly inhibited at low energy levels, allowing PP_i energy to be used for cell survival instead of being converted into heat.

Abbreviations

| | |
|------------------------|---|
| ADP | Adenosine diphosphate |
| AMP | Adenosine monophosphate |
| AMPK | Adenosine monophosphate-activated protein kinase |
| rAMPK γ 1 | Type 1 gamma domain of <i>Rattus norvegicus</i> AMPK |
| mAMPK γ 1 | mammalian (here human or rat) AMPK γ 1 |
| spAMPK | <i>Schizosaccharomyces pombe</i> AMPK |
| AMPPNP | Adenosine 5'-(β,γ -imido)triphosphate |
| AP $_4$ A | Diadenosine 5', 5-P $_1$, P $_4$ -tetraphosphate |
| ATP | Adenosine triphosphate |
| CBS domain | Domain named based on cystathionine β -synthase |
| CBS dimer | Dimer of CBS pair |
| CBS pair | Tight dimer of two CBS domains |
| CBS-PPase | CBS domain containing inorganic pyrophosphatase |
| cpCBS-PPase | <i>Clostridium perfringens</i> CBS-PPase |
| mtCBS-PPase | <i>Moorella thermoacetica</i> CBS-PPase |
| CIC | Chloride-conducting ion channel |
| common family II PPase | Non-regulated PPase family formed by DHH and DHHA2 domains |
| cpCBS | Regulatory part of cpCBS-PPase (residues 66–306) |
| DHH | Domain family, with DHH sequence |
| DHHA2 | DHH-associated domain found in PPases and exopolyphosphatases |
| DRTGG domain | Domain named based on conserved residues |
| ESRF | European synchrotron radiation facility |
| HPrK/P | Histidine containing phosphor carrier protein kinase/ phosphatase |
| IMPDH | Inosine-5'-monophosphate dehydrogenase |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| MgtE | Mg $^{2+}$ transporter family |
| MTA | 5-methyl-5'-thioadenosine |
| MurF | UDP-N-acetylmuramoyl-tripeptide D-Ala-D-Ala ligase |
| NCS | 2-fold non-crystallographic symmetry |
| OpuA | Glycine betaine transporter |
| PDB | Protein data bank |
| P $_i$ | Inorganic phosphate |
| P-loop | Phosphate binding loop |
| PNP | Imidodiphosphate |
| PP $_i$ | Inorganic pyrophosphate |
| PPase | Inorganic pyrophosphatase |
| scPPase | <i>Saccharomyces cerevisiae</i> family I PPase |
| ecPPase | <i>Escherichia coli</i> family I PPase |
| rmsd | Root mean square deviation |
| MRSAD | Molecular replacement with single wavelength anomalous dispersion |
| SAM | S-adenosyl methionine |
| SeMet | Selenomethionine |
| TB | Terrific Broth |
| ZMP | Aminoimidazole 4-carboxamide ribonucleotide |

Abbreviations for Amino Acids

| | | |
|---|-----|---------------|
| A | Ala | Alanine |
| C | Cys | Cysteine |
| D | Asp | Aspartic acid |
| E | Glu | Glutamic acid |
| F | Phe | Phenylalanine |
| G | Gly | Glycine |
| H | His | Histidine |
| I | Ile | Isoleucine |
| K | Lys | Lysine |
| L | Leu | Leucine |
| M | Met | Methionine |
| N | Asn | Asparagine |
| P | Pro | Proline |
| Q | Gln | Glutamine |
| R | Arg | Arginine |
| S | Ser | Serine |
| T | Thr | Threonine |
| V | Val | Valine |
| W | Trp | Tryptophan |
| Y | Tyr | Tyrosine |

1 Review of the Literature

1.1 Inorganic pyrophosphatases

Inorganic pyrophosphatases (PPases, EC 3.6.1.1) catalyze hydrolysis of the pyrophosphate, one of the simplest phosphoryl transfer reactions. Inorganic pyrophosphate (PP_i) is produced in vast amounts from nucleoside triphosphates in vital biosynthetic reactions, such as protein, RNA and DNA synthesis, and hydrolysis of PP_i by PPase provides the necessary thermodynamic pull for these reactions, making this enzyme essential for life [1-5]. PP_i additionally regulates many other cellular processes, including calcification, cell proliferation and iron transport [6]. Consequently, disruption of PP_i metabolism can lead to a variety of pathological conditions [6].

Two types of PPases have been identified: integral membrane-bound and soluble enzymes. Integral membrane-bound ion-translocating inorganic pyrophosphatases pump H^+ [7] or Na^+ ions [8] across biological membranes using energy released upon PP_i hydrolysis. In certain situations, proton pumps can work as PP_i synthases [9]. Soluble PPases are divided into two well-characterized non-homologous families, designated I and II [10, 11], which are described below in more detail.

1.1.1 Family I PPases

Family I PPases, the most common pyrophosphatases found in all kingdoms of life, are among the most extensively characterized phosphoryl transfer enzymes [12, 13]. These enzymes act mainly constitutively without regulation. The only exception found so far is phosphorylation-inhibited plant pollen tube PPase [14]. Within this family, eukaryotic *Saccharomyces cerevisiae* PPase (*scPPase*) and the bacterial enzyme from *Escherichia coli* (*ecPPase*) are the most widely studied.

The X-ray structures of several family I PPases have been solved during the last few decades. *scPPase* was initially crystallized in 1974 [15], and the first apo-*scPPase* structure published in 1981 [16]. Ten years later, the structure of the first *scPPase* product complex was solved [17], but phosphate was misplaced, which was corrected after five years [18, 19]. The first *ecPPase* structures were obtained in 1994 [20, 21]. Subsequently, several *ecPPase* and *scPPase* structures were solved with and without different ligands [22-25], some to very high resolution (*ecPPase* to 1.05 Å [26] and *scPPase* to 1.15 Å [27]), along with those of several variants [26, 28, 29]. The structures of family I PPase from *Thermus thermophilus* [30], *Sulfolobus acidocaldarius* [31], *Pyrococcus horikoshii* [32] and *Mycobacterium tuberculosis* [33] have additionally been determined. In addition, the PPase structures from five other species (*Anaplasma phagocytophilum*, *Ehrlichia chaffeensis*, *Burkholderia pseudomallei*, *Brucella melitensis* and *Rickettsia prowazekii*) (PDB ID: 3LD3, 3LO0, 3GVF, 3FQ3 and 3EMJ, respectively) obtained in the latest structural genomics project

at Seattle Structural Genomics Center for Infectious Disease have been submitted to protein data bank (PDB).

The overall fold of family I PPases is a highly twisted five-stranded β -barrel with some α -helical and β -strand extensions. Family I PPases display a simple cup-like single domain structure where the active site is located at the bottom of the cup (Figure 1). Although shorter bacterial *E. coli* PPase only contains 175 residues and displays 28% sequence identity with the 286-residue eukaryotic *sc*PPase [13], the structural cores of the two enzymes are very similar, with the root mean square deviation of 158 C^α atoms of only 1.45 Å [19]. Active sites are highly conserved. However, the oligomeric structures differ between bacterial and eukaryotic family I PPases, in that the former forms a dimer of trimers while the latter is homodimeric.

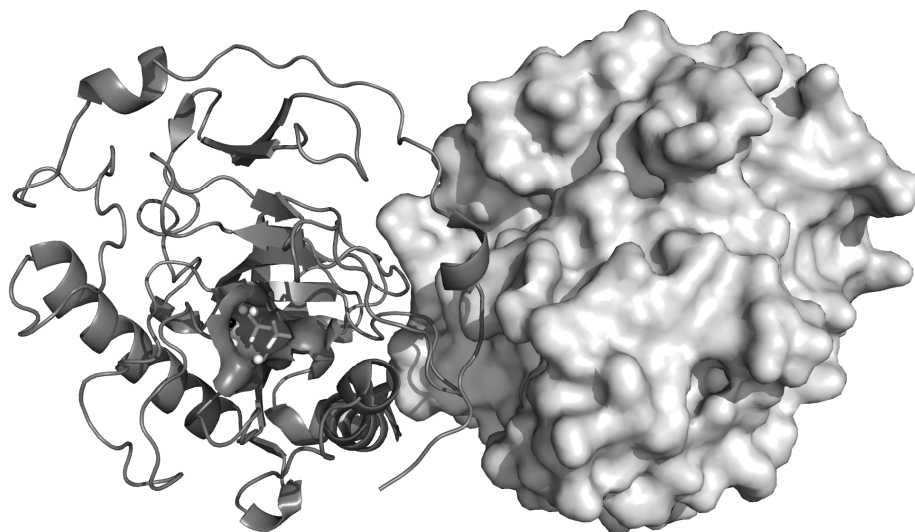
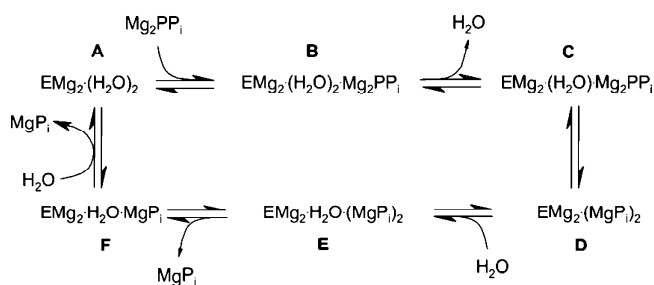


Figure 1. View from top to bottom of the cup (dark gray surface on cartoon) of *sc*PPase (PDB ID: 1E6A). In the fluoride-inhibited structure, F^- (black ball) occupies the position of nucleophilic water, Mn^{2+} ions are presented as white balls, and the substrate as a stick model. One monomer of the dimeric protein is visualized as a cartoon, and the other as a space-filling model.

PPases are metal-dependent enzymes that use divalent activation cations (Mg^{2+} , Zn^{2+} , Mn^{2+} and Co^{2+}). Ca^{2+} is a competitive inhibitor [34]. Mg^{2+} , which binds with micromolar affinity, confers high substrate specificity for inorganic pyrophosphate [35], the only physiologically relevant substrate. PPase catalysis requires three or four metal ions per substrate molecule [36], depending on the pH and PP_i concentration. In yeast cells, the PP_i concentration and pH favor the three-metal ion mechanism [36], whereas at the low pH used for crystallization, the four-metal mechanism predominates. Hydrolysis of the metal-complexed pyrophosphate is accelerated by a factor of 10^{10} , compared to the uncatalyzed reaction [37].

The catalytic reaction of these enzymes is well-characterized. The effects of all eighteen charged and tyrosine residues in identical *ec*PPase and *sc*PPase active sites have been studied via conservative site-directed mutagenesis [38]. Kinetic analyses have disclosed six separate reaction steps, as shown in Scheme 1. The enzyme is activated by two metal ions (Scheme 1, A), and one or two additional metal ions interact with the substrate (Scheme 1, B). The metal positions in the active site are termed M1–M4 in the proposed order of metal binding, where M1 and M2 are the activating metal ions, and M3 and M4 bind with substrate [19]. The enzyme undergoes isomerization (Scheme 1, B \rightarrow C), as evident from both enzymological [39, 40] and structural data [27]. The generation of the reactive nucleophile, O_{Nu} , between M1 and M2 is aided by a low-barrier hydrogen bond between D117 and the nucleophile [27]. The protonation state of the nucleophile has not been directly observed, and is therefore denoted O_{Nu} . After the hydrolysis step (Scheme 1, D), the two phosphate–metal complexes are individually released (Scheme 1, E and F), and the enzyme returns to the two metal ion form (Scheme 1, A).



Scheme 1. Six-state mechanism of *sc*PPase. (paper I)

The rate-determining step for the forward reaction with Mg^{2+} as the cofactor is PP_i hydrolysis (Scheme 1, C \rightarrow D), whereas in the reverse direction leading to net synthesis, the rate-determining step is the preceding isomerization (Scheme 1, C \rightarrow B) [41]. Isomerization decreases the pK_a of the nucleophile to 5.85 [40], which is structurally explained by the two-water bridge collapsing into a single nucleophilic water between the M1 and M2 metal ions.

In the earlier *sc*PPase structures, the complexed metal ions were the less active cofactors, Mn^{2+} [19, 27] or Co^{2+} (PDB ID: 1M38 [42]), both of which bind phosphate with higher affinity [43]. The identity of the metal cofactor determines the order of phosphate release [44, 45], so that the leaving group phosphate, P2, is released before the electrophilic phosphate, P1, after PP_i hydrolysis in the presence of Mg^{2+} . This order is reversed with Mn^{2+} as the cofactor [46].

Intermediates A and C–E in the proposed kinetic scheme (Scheme 1) have been described earlier [27]. The *sc*PPase· Mn_2 structure at 2.2 Å resolution (PDB ID: 1WGI [19]) represents intermediate A, where the metal sites, M1 and M2, are filled. The fluoride-inhibited product complex at 1.9 Å (PDB ID: 1E6A [27]) representing intermediate C contains four metal ions, PP_i , and a fluoride ion replacing the nucleophilic O_{Nu} (Figure 1). The product complex at 1.15 Å resolution (PDB ID: 1E9G

[27]) contains two conformations of the hydrolyzed product, which correspond to intermediates D and E, where the latter is a hydrated rearrangement of D. Structures of the intermediates B and F are yet to be solved.

1.1.2 Family II PPases

Family II PPases, also designated type C PPases [47, 48], were discovered more recently in 1998 [10, 11]. These enzymes exist almost exclusively in Bacteria, mainly in Firmicutes (Table 1), including several human pathogens. Most of family II PPases (common family II PPases) are composed of two well-defined domains, N-terminal DHH and C-terminal DHHA2, and belong to the DHH (Asp-His-His) family of phosphohydrolases (Figure 1 in original publication III) [49]. One-quarter of family II PPase enzymes contain a 250-residue insert within the N-terminal domain. The insert (regulatory region) consists of a DRTGG domain between two CBS domains (both type of domains are described more details later). These “longer” PPases are denoted CBS-PPases to distinguish them from common family II PPases. The DRTGG domain is missing in *Moorella thermoacetica*, *Syntrophomonas wolfei* and *Syntrophothermus lipocalidus* CBS-PPases (Table 1).

Several structures of common family II PPases are available. Initially, the structures were solved in 2001 from *Streptococcus mutans* (PDB ID: 1I74) [50], *Streptococcus gordonii* (PDB ID: 1K20) [48] and *Bacillus subtilis* (PDB ID: 1K23) [48] enzymes with variable domain interface conformations and different Mn^{2+} and sulfate combinations. Subsequently, product analog complexes were solved, including *S. gordonii* PPase complexed with sulfate and Zn^{2+} (PDB ID: 1WPP) [51] and metal-free *B. subtilis* PPase complexed with high-affinity sulfate (PDB ID: 1WPN) [51]. Substrate analog complexes include imidodiphosphate (PNP) complexes of *Streptococcus agalactiae* PPase with two Mn^{2+} and a Mg^{2+} (PDB ID: 2ENX) [52], fluoride-inhibited *B. subtilis* PPase with Mg^{2+} (Figure 2) [53] and *B. subtilis* H98Q variant PPase with two Mn^{2+}/Fe^{2+} and two Mg^{2+} ions (PDB ID: 2IW4) [53].

Family II PPases are homodimeric (Figure 2), similar to the eukaryotic family I PPase. A flexible linker connects the N- and C-terminal domains, with the active site located at the domain interface [48, 50]. The C-terminal domain of family II PPase contains the high-affinity substrate-binding site, whereas the catalytic site that binds the nucleophile-coordinating metal cations is located in the N-terminal domain. Closure of the C-terminal domain onto the N-terminal region creates a catalytically competent conformation by bringing the electrophilic phosphate of substrate into the catalytic site [51]. Substrate binding to the C-terminal domain in the open conformation causes domain closure [51].

Table 1. Inorganic pyrophosphatases in Firmicutes based on fully sequenced genomes in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [56].

| | | | | | | | | | |
|--|----|----------------------------------|----|-----------------------------------|----|---|-------|----|--|
| Bacillales | | | | | | | | | |
| <i>Alicyclobacillus acidocaldarius</i> | II | <i>Lysinibacillus sphaericus</i> | II | <i>L. brevis</i> | II | <i>Acidaminococcus fermentans</i> | II | | |
| <i>Anoxybacillus flavithermus</i> | I | <i>Macrococcus caseolyticus</i> | II | <i>L. casei</i> | II | <i>Anaeroecellum thermophilum</i> | + | | |
| <i>Bacillus subtilis</i> | II | <i>Oceanobacillus iheyensis</i> | II | <i>L. gasseri</i> | II | <i>Anaerococcus prevotii</i> | II | + | |
| <i>B. halodurans</i> | I | <i>Paenibacillus</i> sp. JDR-2 | II | <i>L. reuteri</i> | II | <i>Ammonifex degensii</i> | II | | |
| <i>B. anthracis</i> | II | <i>Staphylococcus aureus</i> | II | <i>L. helveticus</i> | II | <i>Caldicellulosiruptor saccharolyticus</i> | + | | |
| <i>B. cereus</i> | II | <i>S. epidermidis</i> | II | <i>L. fermentum</i> | II | <i>Candidatus Desulforudis audaxviator</i> | I,II+ | | |
| <i>B. cytotoxicus</i> | II | <i>S. haemolyticus</i> | II | <i>L. rhamnosus</i> | II | <i>Carboxydothermus hydrogenoformans</i> | + | | |
| <i>B. thuringiensis</i> | II | <i>S. saprophyticus</i> | II | <i>L. crispatus</i> | II | <i>Clostridiales</i> genomsp. | II | | |
| <i>B. weihenstephanensis</i> | II | <i>S. carnosus</i> | II | <i>Pedococcus pentosaceus</i> | II | <i>Coprothermobacter proteolyticus</i> | II | + | |
| <i>B. licheniformis</i> | II | <i>S. lugdunensis</i> | II | <i>Enterococcus faecalis</i> | II | <i>Desulfitobacterium hafriense</i> | II | + | |
| <i>B. clausii</i> | II | <u>Lactobacillales</u> | II | <i>Oenococcus oeni</i> | II | <i>Desulfotomaculum reducens</i> | II | + | |
| <i>B. amyloliquefaciens</i> | II | <i>Lactococcus lactis</i> | II | <i>Leuconostoc mesenteroides</i> | II | <i>D. acetoxidans</i> | + | | |
| <i>B. pumilus</i> | II | <i>Streptococcus pyogenes</i> | II | <i>L. citreum</i> | II | <i>Eubacterium eligens</i> | II | + | |
| <i>B. pseudofirmus</i> | II | <i>S. pneumoniae</i> | II | <i>L. kimchii</i> | II | <i>E. rectale</i> | II | + | |
| <i>B. megaterium</i> | II | <i>S. agalactiae</i> | II | <u>Clostridia</u> | II | <i>Finexgoldia magna</i> | II | + | |
| <i>B. selenitireducens</i> | II | <i>S. mutans</i> | II | <i>Clostridium acetobutylicum</i> | II | <i>Halothermothrix orenii</i> | II | + | |
| <i>B. tustiae</i> | I | <i>S. thermophilus</i> | II | <i>C. perfringens</i> | II | <i>Hellobacterium modesticaldum</i> | II | + | |
| <i>Brevibacillus brevis</i> | II | <i>S. sanguinis</i> | II | <i>C. tetani</i> | II | <i>Natranaerobius thermophilus</i> | II | + | |
| <i>Exiguobacterium sibiricum</i> | II | <i>S. suis</i> | II | <i>C. novyi</i> | II | <i>Pelotomaculum thermopropionicum</i> | + | | |
| <i>Exiguobacterium</i> sp. AT1b | II | <i>S. gordonii</i> | II | <i>C. thermocellum</i> | II | <i>Symbiobacterium thermophilum</i> | I | + | |
| <i>Geobacillus kaustophilus</i> | II | <i>S. equi</i> | II | <i>C. difficile</i> | II | <i>Syntrophomonas wolfei</i> | II* | + | |
| <i>G. thermodenitrificans</i> | I | <i>S. uberis</i> | II | <i>C. botulinum</i> | II | <i>Syntrophothermus lipocalidus</i> | II* | + | |
| <i>Geobacillus</i> sp. WCH70 | I | <i>S. dysgalactiae</i> | II | <i>C. beijerinckii</i> | II | <i>Thermincola</i> sp. JR | II | + | |
| <i>Geobacillus</i> sp. Y412MC10 | II | <i>S. gallolyticus</i> | II | <i>C. kilyveri</i> | II | <i>Thermoanaerobacter tengcongensis</i> | II | + | |
| <i>Geobacillus</i> sp. Y412MC61 | I | <i>S. mittis</i> | II | <i>C. phytofermentans</i> | II | <i>T. sp. X514</i> | II | + | |
| <i>Geobacillus</i> sp. C56-T3 | I | <i>Lactobacillus plantarum</i> | II | <i>C. cellulolyticum</i> | II | <i>T. pseudethanolicus</i> | II | + | |
| <i>Listeria monocytogenes</i> | II | <i>L. johnsonii</i> | II | <i>Alkaliphilus</i> | II | <i>T. italicus</i> | II | + | |
| <i>L. innocua</i> | II | <i>L. acidophilus</i> | II | <i>A. metalliredigens</i> | II | <i>T. mathranii</i> | II | + | |
| <i>L. welshimeri</i> | II | <i>L. sakei</i> | II | <i>A. oremlandii</i> | II | <i>Veillonella parvula</i> | II | | |
| <i>L. seeligeri</i> | II | <i>L. salivarius</i> | II | <u>Others</u> | II | <u>Moorella</u> | II* | ++ | |
| | II | <i>L. delbrueckii</i> | II | <i>Acetohalobium arabaticum</i> | + | <i>Moorella thermoacetica</i> | II* | ++ | |

I, family I PPase; II, common family II PPase; II*, CBS domain-containing subfamily of family II PPase; II*, CBS-PPase missing DRITGG domain; +, membrane-bound PPase

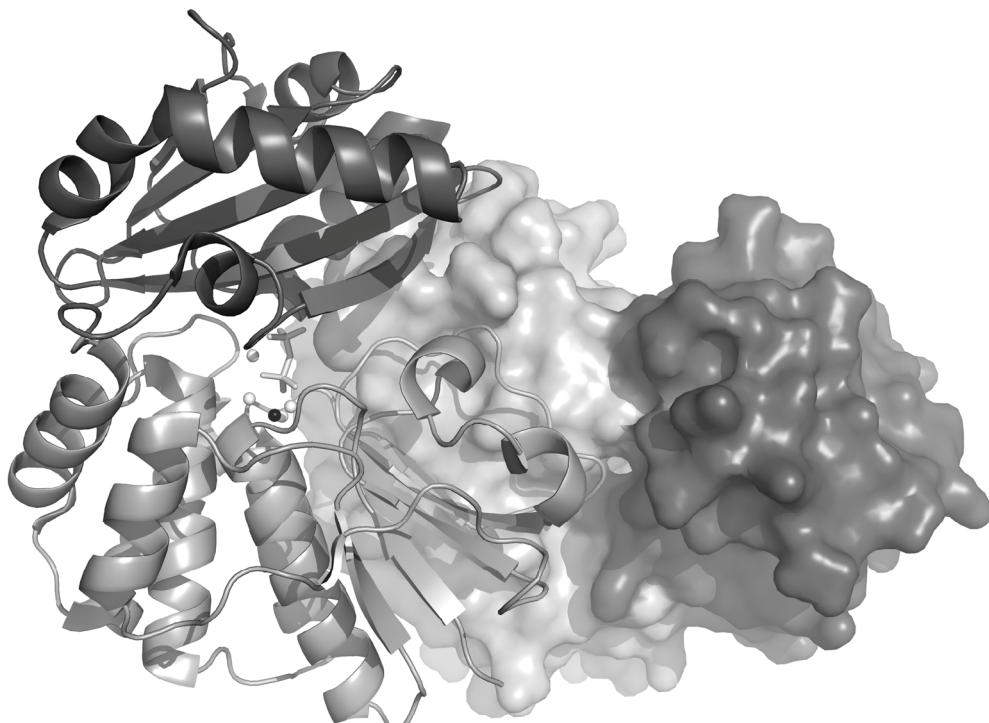


Figure 2. Dimeric common family II PPase formed by N-terminal DHH (light gray) and C-terminal DHHA2 (dark gray) domains (PDB ID: 2HAW); otherwise coloring is similar to that in Figure 1. The active site is located between two flexible domains and occupied with F^- , Mg^{2+} and the substrate analogy PNP.

Both soluble PPase families catalyze the same PP_i hydrolysis reaction, and have similar active sites [48, 50] but significantly different functional properties. Unlike family I PPases, family II enzymes are most effectively activated by Mn^{2+} or Co^{2+} , which bind with nanomolar affinity [54]. Family II PPases with these metal ions as cofactors are about 10-fold more active than family I PPases with Mg^{2+} (k_{cat} of 1700-3300 s^{-1} versus 110-330 s^{-1}) [47, 54, 55]. In addition, divalent cations, in particular, Mn^{2+} , promote enzyme dimerization, where the dimer is the active form of family II PPase [54].

The crystal structures revealed that the preference for Mn^{2+} over Mg^{2+} in family II PPases is at least partly attributable to the active site histidines that bind the metal ions and bidentate carboxylate coordination of the metal ion in the binding site [50]. Substrate binding to family II PPase appears to require a change in the coordination number of the high-affinity metal site from five to six [51]. Varied five/six-coordinated geometry is typical for transition metals, such as Mn^{2+} or Co^{2+} , but not Mg^{2+} , which is almost always six-coordinated [57]. Consequently Mn^{2+} and Co^{2+} fit the catalytic machinery better than Mg^{2+} . Furthermore, Zn^{2+} , which does not form the six-coordination complex easily [57], is a very poor activator of family II PPase [51, 55].

Most highly active common family II PPases appear to be constitutively active, but pathogenic *Streptococcus agalactiae* and *Streptococcus pneumonia* common family II PPases are regulated via kinase-dependent phosphorylation [58, 59]. Disruption of kinase or phosphatase function affects cell growth, segregation and virulence in *S. agalactiae* [58]. Other homologous pathogenic bacterial species, such as Streptococci, Staphylococci and Enterococci, may be similarly regulated [58]. In *S. agalactiae* family II PPase, four potential phosphorylation sites have been identified in the dimer interface area [52], but further studies are needed for distinguish the actual phosphorylation site.

1.2 The CBS domain

The CBS domain (Pfam: PF00571 [60]) is a small intracellular region usually present in proteins as a tandem pair with two (or less frequently, four) copies. Two CBS domains form a CBS pair or Bateman domain, designated based on Alexander Bateman, who was the first to identify this structure while studying genome internal duplications in *Methanocaldococcus jannaschii* in 1997 [61]. He also named this cystathionine β -synthase-like domain 'CBS domain' [62].

1.2.1 Proteins with the CBS domain

According to the Pfam database [60], CBS domains have been identified in nearly 30 000 protein sequences of numerous cytosolic and membrane-associated enzymes and channels from all kind of species [63]. Typically, orthologous proteins in some species do not contain the CBS pair whereas others have it, indicating that this structure is not essential for protein function [64-66]. The most extensive research to date has been performed on the CBS domain-linked human hereditary disease-causing proteins, AMP-activated protein kinase (AMPK), chloride-conducting ion channel (ClC), cystathionine β -synthase and inosine-5'-monophosphate dehydrogenase (IMPDH) [63]. The group of Martínez-Cruz is currently focusing on CBS domain-containing proteins from *M. jannaschii*. This hyperthermophilic organism contains 15 proteins with the CBS domain [67, 68]. More recent studies have investigated a Mg²⁺ transporter (MgtE) and glycine betaine ABC transporter (OpuA) [69, 70]. Several CBS domain-only protein structures with two or four CBS domain repeats have been solved, but their partners and physiological functions remain unknown [71-75]. The coordinates for many CBS domain structures have recently been deposited to the Protein Data Bank (PDB ID: 1PBJ, 1YAV, 1VR9, 2O16, 2P9M, 2EMQ, 2YZI, 2YZQ, 3DDJ, 3FHM, 3K6E, 3LQN, 3FNA, 3LFR, 3LHH, 3LV9, 3FV6, 3FVR, 3FVS and 3OCO), but no descriptions have been published.

CBS domains are widely distributed among proteins with distinct functions. The most extensively studied CBS domain-containing proteins are presented below. Cystathionine β -synthase is a homotetrameric protein with the regulatory heme domain located at the N-terminus, catalytic part in the middle, and CBS pair at the C-terminus [76]. AMPK is a heterotrimeric protein consisting of α , β and γ subunits, whereby the γ subunit contains a tetra repeat of CBS domains [77]. Mammals have three paralogous γ

subunits (AMPK γ 1-3), of which AMPK γ 1 is the shortest and contains only CBS domains.

CICs are homodimeric integral membrane proteins with a CBS domain pair in the cytoplasmic C-terminal region. CIC-0 from the electric organ of the electric ray, *Torpedo marmorata*, was initially identified and cloned [78], followed by nine paralogous mammalian CICs (CIC-1 to CIC-7, CIC-Ka and CIC-Kb), which vary in terms of length of the cytoplasmic C-terminal region (residues 155 to 398). Mammalian CIC channels exhibit differential tissue localization [66]. CIC-0, CIC-1, CIC-2, CIC-Ka, and CIC-Kb share about 50–60% sequence identity. These proteins are located on the cell membrane and denoted 'muscle-type CICs'. CIC-3 to CIC-5 form a second group, which are localized on the membranes of intracellular vesicles. The third subfamily comprises CIC-6 and CIC-7, which are broadly expressed in various tissues.

Two paralogous IMPDHs exist in humans (hIMPDH1 and hIMPDH2). IMPDH is a homotetrameric protein in which the CBS pair is inserted within the catalytic sequence but located as a separate domain at the protein surface [79]. The typical homodimeric bacterial MgtE Mg^{2+} transporter consists of a cytosolic N-terminal part including the N-domain and a CBS pair, as well as a transmembrane C-terminal region [69]. OpuA is composed of two identical nucleotide-binding domains fused to the CBS pair and two identical substrate-binding domains fused to the transmembrane domain [80].

1.2.2 Natural mutations in the CBS domain

Point or deletion mutations in the CBS domains of human proteins (Figure 3) cause various hereditary diseases. Several natural mutations found in the CBS domain of human cystathionine β -synthase [81-83] trigger homocystinuria either through loss of enzyme activity (V456P mutation) [83] or regulatory function (D444N mutation) [81]. Amino acid substitutions (R224P or D226N) in the CBS domain of hIMPDH1 cause retinitis pigmentosa, a hereditary disease, which can lead to blindness [84, 85]. Mutations in CICs are linked to several hereditary diseases, some of which are located in the CBS domains. Point mutations in CIC-K β cause Bartter syndrome [86], substitutions in CIC-7 lead to osteopetrosis [87-89] and several different-length C-terminal truncations in CIC-5 (shortest severe deletion after R648) trigger Dent disease [90, 91]. Point mutations in the CBS domain of AMPK are associated with two hereditary diseases, Wolff-Parkinson-White syndrome [92] and familial hypertrophic cardiomyopathy [93]. Furthermore, the R250Q substitution in Hampshire pig AMPK γ 3 (pAMPK γ 3) enhances the glycogen content in skeletal muscle about 70%, leading to deterioration of pork meat quality [94, 95]. Similar effect is observed in humans where equivalent mutation R225W of AMPK γ 3 increased about 90% glycogen content in muscle cells of obese people [96].

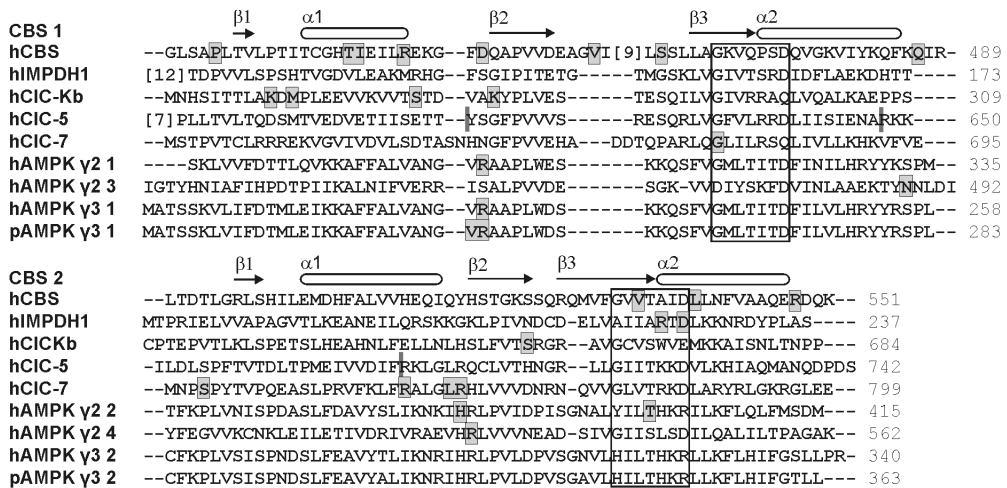


Figure 3. Sequence alignment of CBS domains 1 and 2 of proteins with disease-causing mutations. Disease-associated residues are shaded in gray. The C-terminal deletion is presented with a gray bar. CBS protein designation (the h prefix indicates human protein): hCBS, human cystathionine β -synthase; pAMPK γ 3, pig AMPK γ 3. For AMPKs containing two pairs of CBS domains per subunit, CBS domain numbers are presented after protein names. Sequence extensions are indicated as the number of residues in parentheses. Sequence numbering is shown on the right. Core CBS domain secondary structures based on hCLC-5 are presented above alignments. Ribose phosphate binding motifs are highlighted in boxes.

1.2.3 Structures of CBS domain proteins

CBS domains about 60 residues in length from various proteins and organisms have low sequence conservation within protein families, but display conserved three-dimensional structures [97]. CBS domains fold into an α/β structure with internal symmetry [71]. These form a typical core β - α - β - β - α fold [62], along with an additional small helix and often, a short β -sheet at the N-terminal region of the domain (Figure 4a) located adjacent to the neighboring CBS domain (Figure 4b) [73]. The CBS pair is a tight structure, since these domains complement each other, even when expressed separately [98] and from different proteins [99].

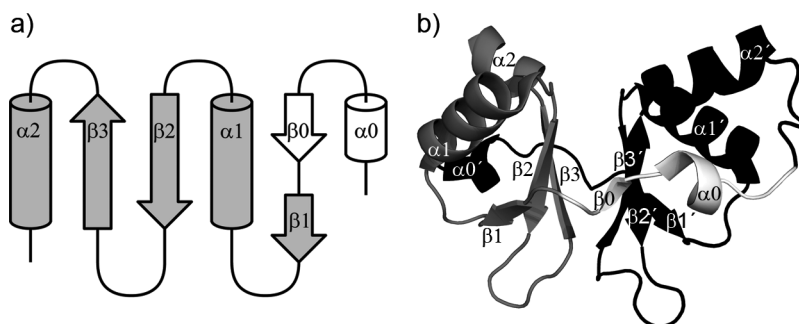


Figure 4. Topology of the CBS domain. (a) CBS domain core is presented in gray and the common CBS linker in white. The figure has been generated with TopDraw [100]. (b) A CBS domain is colored and labeled as (a) and pairing CBS domain with black. Labels for pairing CBS domain include prime. The β -Strands (β 1-3) form a β -sheet and β 0 is part of the pairing CBS domain β -sheet beside on β 3' (PDB ID: 2RIF).

Several X-ray structures of CBS domains have been solved. The first structure was that of intact IMPDH from *Streptococcus pyogenes*, which was determined before identification of the CBS domain ligands [79]. Catalytic domains of the homotetrameric enzyme form intermolecular interactions and CBS domains form a pair that is a separate flexible structure on the protein surface. Thereafter, several structures of separate regulatory CBS domains or functional parts of proteins have been determined. For instance, the structure of human cystathionine β -synthase was solved without CBS domains, since the full-length protein has a tendency to aggregate [101]. The protein is dimeric in this structure, and it is speculated that its CBS pair is responsible for tetramerization of full-length protein in solution [102].

The structure of full-length CIC from *E. coli* has been solved. Interestingly, this transporter does not contain CBS domains in the cytoplasmic region, unlike human CICs [103]. In addition, three structures of CIC CBS pairs have been determined, specifically, CIC-0 [104], CIC-5 [105] and CIC-K α [106]. The CBS pair of CIC-5 dimerizes via β 1-strand interactions to form a V-shaped oligomeric structure, which brings two CBS2 domains into close proximity to each other, with few interactions between the remote CBS1 counterparts (Figure 5a). This observation was confirmed in solution studies and is thus not a crystallization artifact [105]. CIC-K α dimerizes similar to CIC-5, and while CIC-0 crystal packing is different, mutagenesis studies have shown that dimerization in solution is similar to that of the other two CICs [105].

AMPKs are among the most intensively studied CBS domain-containing proteins with bound adenylate ligands. Several AMPK structures have been solved, but none determined for the full-length heterotrimeric protein. Separate domain structures have been established for α [107] and β subunits [108], as well as a fragment of the γ 1 subunit comprising the CBS3 and CBS4 domains [109, 110]. Moreover, core structures including the γ 1 subunit and closely related parts of α and β subunits have been solved [111-113]. The CBS pairs of the AMPK γ 1 subunit dimerize differently, compared with CICs, and form an elliptical disk-like structure (Figure 5b) about 60 Å in diameter and 30 Å thick, which contacts helices α 1 and α 2 of CBS domains [112]. In

both CBS3-4 pair structures, the CBS3 domain interacts with the CBS4 domain from the other monomer, and vice versa (head-to-tail assembly) [109, 110]. Full-length $\gamma 1$ subunit with a CBS domain tetra repeat forms a similar disk-like dimer, but displays head-to-head assembly in which CBS1 interacts with CBS3 and CBS2 with CBS4, which appears to be a biological unit in solution [111-113].

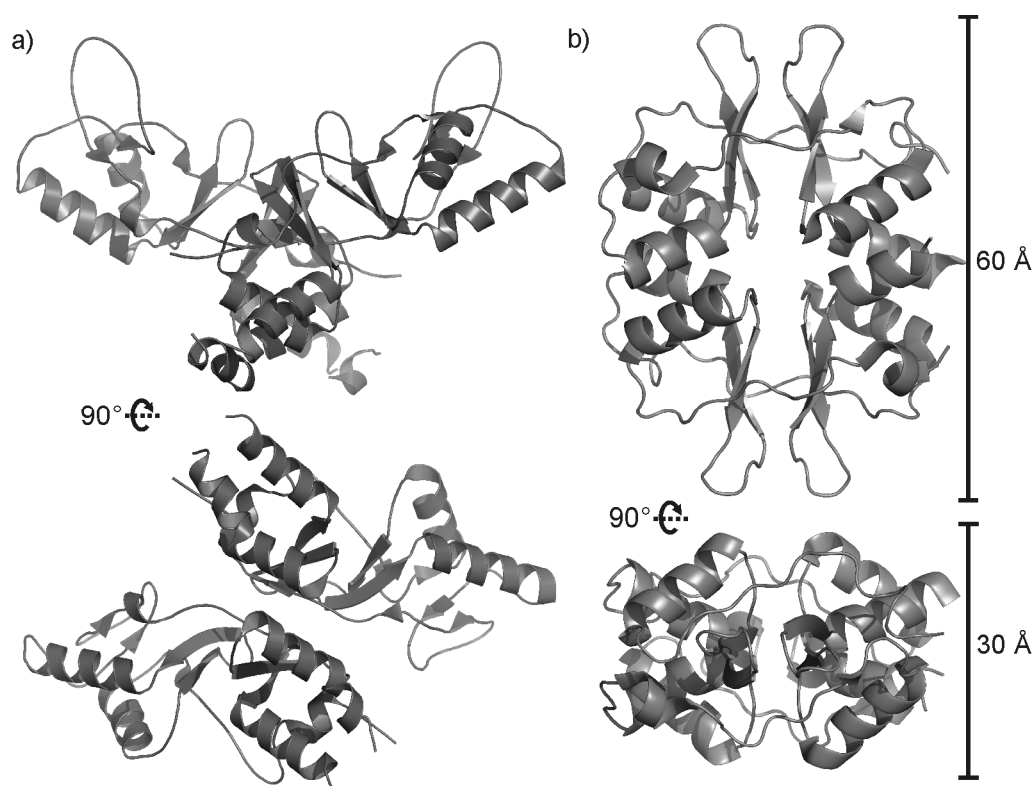


Figure 5. Dimerization patterns of the CBS pair in CIC-5 (PDB ID: 2JA3) (a) and tetra-repeat AMPK $\gamma 1$ (PDB ID: 2OXX) (b) are different. In the CIC dimer, CBS2 domains are adjacent to each other, whereas in AMPK and most other proteins, CBS pairs dimerize by forming a disk-like assembly.

Structures of several CBS domain-only proteins have been solved, but their functions and possible partner proteins remain unknown at present. The first structure was obtained for *Thermotoga maritima* TM0935 protein (PDB ID: 1O50) [71]. An *in situ* proteolysis project yielded one *Nitrosomonas europaea* CBS domain-only structure (NE2398) (PDB ID: 2RC3) in which 15 C-terminal residues were deleted [72]. The first archaea structure was obtained from hyperthermophilic *Sulfolobus tokodaii* (ST2348, PDB ID: 2EF7) [74]. CBS domain-only protein from *Mycobacterium tuberculosis* was designated 'hypoxic response protein 1' (HRP1, Rv2626c, PDB ID: 1XKF and 1Y5H), based on the observed 12- to 107-fold upregulation during hypoxia [75]. Bacterial *Pyrobaculum aerophilum* PAE2072 (PDB ID: 2RIH and 2RIF) [73] and *M. jannaschii* MJ1225 [68] have additionally been identified as CBS domain-only proteins. The common feature in all these proteins, with one exception (PDB ID: 1XKF), is the elliptical disk-like assembly, similar to the

AMPK γ subunit, which appears to be a typical assembly for CBS domains. Differences are evident in dimerization. Although all proteins, except TM0935, display head-to-head assembly, the dimer interface involves both CBS domains in TM0935, NE2398, PAE2072 and MJ1225 structures, whereas only CBS1 domains are involved in dimerization of HRP1 (PDB ID: 1Y5H) and CBS2 domains in ST2348. The biological relevance of these different contacts is yet to be established.

In addition to bacterial IMPDH, proteins with known full-length structures include CBS-Zn ribbon-like protein from *Thermoplasma acidophilum* (TA0289) [97] and the membrane Mg^{2+} transporter, MgtE, from *Thermus thermophilus* [69]. The CBS pair of TA0289 appears to play a role in dimerization, whereas the CBS pair of MgtE works as a regulator. Both proteins display the dimeric disk-like assembly of CBS domains.

1.2.4 Ligand binding to the CBS domain

The first CBS domain ligand was identified when cystathionine β -synthase S-adenosyl methionine (SAM) activation was prevented upon deletion of the C-terminal CBS domains [102]. Subsequently, Hardie and Hawley [114] discovered that AMPK is activated by AMP and inhibited by ATP. In 2004, Scott and co-workers [115] demonstrated for the first time adenylate binding to separate CBS domains of human hereditary disease-associated proteins. The group showed that tetra repeats (CBS1-4) of AMPK γ 1-3 interact with both AMP and ATP, but bind AMP more strongly. Moreover, the CBS pair of IMPDH2 bound ATP more strongly than AMP. Weak ATP binding was also observed with the CBS pair of CIC-2, and SAM bound to the CBS pair of cystathionine β -synthase. This was a breakthrough study in identifying CBS domain ligands.

Ligand binding studies facilitated the determination of several adenylate-bound structures [67, 68, 72, 73, 105, 109, 111-113] (PDB ID: 3DDJ, 3FHM, 3FNA, 2YZQ, 2FWR, 3FWS and 3LFR). The CBS domain ligands identified to date include AMP, ADP, ATP, NAD^+ , NADH, SAM and 5'-deoxy-5'-methylthioadenosine (MTA). Furthermore, some structures contain bound phosphate (PDB ID: 3K6E and 2O16), sulfate (PDB ID: 1YAV and 2RIH [73]) or a ligand analog (e.g., the ATP analog, adenosine 5'-(β,γ -imido)triphosphate (AMPPNP) or the AMP analog, aminoimidazole 4-carboxamide ribonucleotide (ZMP)).

Ligands bind to the cleft within the CBS pair. Each CBS pair has two potential ligand binding sites on both sides, which contain some conserved elements, for instance, a ribose phosphate-binding motif (CBS motif, Figure 3), Ghx(T/S)x(T/S)D (where h is a hydrophobic residue and x is any residue) [109], responsible for AMP binding. The hydrophobic residue (h) interacts with the adenine ring, both Thr/Ser residues bind phosphate, and aspartate binds ribose hydroxyls. The backbone atoms of the hydrophobic residue contact adenine N1 (-N) and N6 (-O). The pairing CBS domain provides the other hydrophobic residue to interact with the adenine ring. There is notable variation in the binding residues, which may explain the diversity of CBS domain ligands.

Typically, a CBS pair binds only one adenylate ligand on one side, as observed in the first adenylate-bound structures of the CBS pair from CIC-5 (Figure 6a) [105]. ADP and ATP bind to the same site, with the exception of AMPK γ , in which the CBS domain tetra repeat binds to all possible sites. *Rattus norvegicus* AMPK γ 1 (rAMPK γ 1) has one strong AMP binding site in the CBS4 domain, which maintains interactions with the ligand through the protein purification process. Bound AMP cannot be replaced by soaking with other adenylate nucleotides [112]. The same site binds the drug molecule, ZMP [109, 113]. Mammalian (human and rat) AMPK γ 1 (mAMPK γ 1) CBS1 and CBS3 motifs bind at least AMP and ATP, depending on ligand availability and concentration [109]. While ligand binding to the CBS2 motif in mAMPK γ 1 has not been demonstrated, it is an ADP-specific site in *Schizosaccharomyces pombe* AMPK (*sp*AMPK).

1.2.5 Conformational changes induced by the CBS domain

Several structures of different adenylate ligands bound to same site have been published, but do not reveal how the regulation signal is transmitted to the functional part of the protein. For example, only minor differences are observed between CIC-5 structures with different ligands (Figure 6a) [105]. No structures have been solved for the same protein with and without ligand. For the *Pyrobaculum aerophilum* CBS-domain-containing protein, PAE2072, one structure has four AMPs, while the other contains four sulfates (Figure 6b) [73]. However, no significant conformational changes are evident between the two complexes.

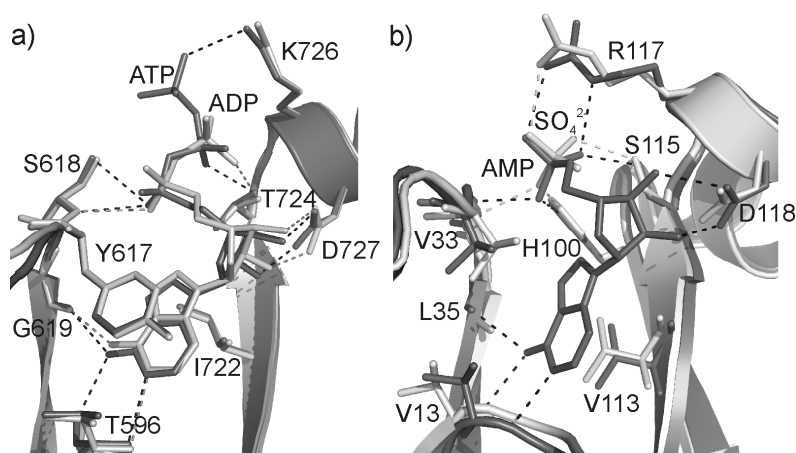


Figure 6. (a) ADP (light gray) and ATP (dark gray) bind similarly to the CBS pair in CIC-5 (PDB ID: 2JA3 and 2J9L, respectively), with no evident conformational changes between the two complexes. (b) Minor movements are observed in the CBS domain-only protein PAE2072 from *Pyrobaculum aerophilum* when sulfate is bound instead of AMP (PDB ID: 2RIH and 2RIF, respectively). Binding residues are labeled, and presented as a ball-and-stick model.

Significant conformational differences are observed in only the CBS pair of *M. jannaschii* MJ0100 protein structures bound to different adenylates. The function of this protein is still unknown. Interestingly, crystallization of the CBS pair of MJ0100

with no added ligands in different forms revealed structures containing different combinations of SAM and MTA, which induced distinct conformations. Based on the structures, it was hypothesized that SAM binds initially in such a way that the methionyl group of the ligand remains disordered and the dimer interface is 23° open (Figure 7). Subsequently, SAM adopts a more compact conformation, allowing a second SAM molecule to bind (interface 12° open). In the third structure, both SAM molecules are ordered (interface 9° open), facilitating the final binding of one or two MTA molecules and dimer closure (Figure 7). These binding steps do not significantly affect monomer assembly; maximal rmsd (root mean square deviation) of C^α with and without ligand is 0.88 Å [67].

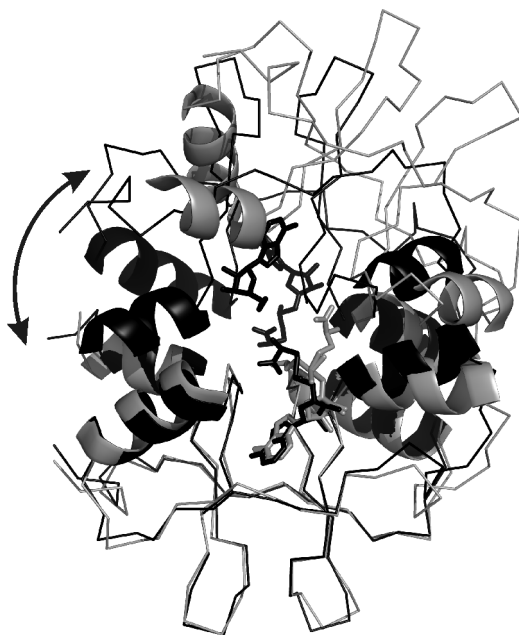


Figure 7. MJ0100 protein [67] in open (PDB ID: 3KPB, gray) and closed (PDB ID: 3KPD, black) conformations. In the conformation opened by 23°, one SAM molecule is bound, whereas the closed conformation contains three ligands. Dimer interface helices are highlighted as cartoons. The arrow indicates CBS2 interface opening.

Conformational changes were also observed in two other CBS domain-containing proteins in the absence of adenylate ligands. MgtE structures display conformational changes due to fluctuating Mg²⁺ concentrations, where the CBS2 domain interface opens/closes by 20° [69]. Movement was additionally observed in the other domain, which was rotated out by 120°, but it is unclear how this change mediates protein function. Fluorescence resonance energy transfer measurements disclosed a significant movement in CIC-0 [116], although the regulator is yet to be identified.

1.2.6 Function of the CBS domain

Functional studies on CBS domain-containing proteins have disclosed that they typically maintain their structure and function upon deletion of the CBS regions. Deletion of the CBS pair in IMPDH [117] and cystathionine β -synthase [76] did not affect catalytic activity. CBS domains only appear essential for function or cellular localization in CICs [118, 119].

Several functions have been proposed for CBS domains. One obvious role is protein dimerization, which is evident in crystal structures. For instance, cystathionine β -synthase is tetrameric in solution [76, 102], whereas the solved structure without CBS domains is dimeric [101], suggesting that CBS domains are fully responsible for tetramerization.

A second possible function of the CBS pair is autoinhibition. Cystathionine β -synthase [76] can be activated upon binding to SAM or deleting the C-terminal CBS domain-containing region [102]. Furthermore, the point mutations I435T, P422L and S466L in cystathionine β -synthase weaken autoinhibition, similar to R152Q and R171Q in rAMPK γ 1 [120] and R225Q in pAMPK γ 3 [95]. In contrast, hCIC-1 ligand binding site variants (T636A, P638A, H847A and L848A) suppress protein function, similar to that observed upon ATP binding to wild-type [121].

Sensing of adenylate ligands, i.e., energy level of cells, is a third and probably most important function of CBS domains. AMP-activating AMPK is most extensively characterized in this respect, and works as a whole human body energy balance sensor. When energy is consumed, the ATP level decreases and both AMP and ADP levels increase. Since adenylate kinase maintains a constant ATP:ADP ratio by catalyzing the reaction: $2\text{ADP} \rightleftharpoons \text{ATP} + \text{AMP}$, the AMP concentration displays the largest variations during cell stress, and is therefore a more effective sensor molecule than ADP [114]. However, AMP is not used as an effector in all AMPKs, since lower eukaryotic *S. cerevisiae* AMPK (*scAMPK*) does not respond to AMP, but to lowered glucose levels via phosphorylation. Alkaline stress and increased ionic strength activate *scAMPK* [122].

Adenylate sensing by the CIC family has been also studied intensively. While CIC-4 is activated by ATP [123], CIC-1 [105, 121] and the plant CIC family member *Arabidopsis thaliana* nitrate/proton antiporter (AtCICa) [124] are inhibited by ATP. The role of the CBS domain has been confirmed with ATP binding studies on CIC-2 [115], ligand-bound CIC-5 structures [105] and mutagenesis studies on ligand binding in AtCICa [124].

Based on the observed *scAMPK* activation at increasing ion strength, it is proposed that ion sensing is a fourth function of the CBS domain. This function has been comprehensively investigated with the ABC transporter, OpuA. *Lactococcus lactis* OpuA is deactivated at low ionic strength and activated above a threshold ionic strength, and deletion of the CBS2 domain makes protein almost insensitive to ionic strength [125]. Mahmood and co-workers [70] showed that positively charged residues

in a loop between $\beta 2$ and $\beta 3$ (Figure 4) are essential for ionic strength sensing. MgtE is another ion-sensing CBS domain protein, responding to the Mg^{2+} level. The structure of *Thermus thermophilus* MgtE contains several Mg^{2+} ions, with one bound to the adenylate binding site of CBS domain whereby the CBS2 dimer interface is opened by 20° and the other domain rotated by 120° in the absence of Mg^{2+} [69].

DNA and RNA binding is the fifth possible function of the CBS domain, in view of the finding that bacterial IMPDH binds 100 base-long single-stranded DNA and RNA in the nanomolar range *in vivo* [126]. However, there is no evidence of nucleic acid binding to the CBS domain.

1.3 The DRTGG domain

The DRTGG domain (Pfam: PF07085), designated based on its most conserved residues, consists of about 120 amino acids, and is often associated with CBS domains. Bateman speculated that this domain is very distantly related to the CBS domain pair. While there are no significant sequence similarities between these regions, its length and association with CBS domains supports this theory (Pfam database [60]).

The protein data bank contains one separate DRTGG domain structure of a hypothetical protein from *Archaeoglobus fulgidus* (PDB ID: 2IOJ), but no data on this structure are currently available. Only two studies have focused on the DRTGG domain, both with phosphotransacetylase (Pta), which contains the DTRGG domain in a 350-residue N-terminal extension before the catalytic region. The results of one study showed that the DRTGG domain mutation, R252H, in the N-terminal region of *Salmonella enterica* Pta prevents NADH inhibition and enhances pyruvate activation, compared with wild-type, indicating that this region plays a role in NADH and pyruvate regulation. No evidence is available on where the two effectors bind in the phosphotransacetylase N-terminal region [127]. Yet another study has shown that DTRGG domain truncations in the *Escherichia coli* Pta N-terminal part have no effects on regulation and catalytic activity [128]. Therefore, the function of the DRTGG domain remains largely unknown at present.

2 Aims of the Study

This work was part of a larger project focusing on the catalytic mechanism and regulation of soluble PPases. The specific aims of my Ph.D. thesis were as follows:

1. Crystallization of wild-type and variant *S. cerevisiae* PPases for determining the structures of missing intermediates in the reaction mechanism with its natural metal cofactor, magnesium (paper I).
2. Production, crystallizing and solving structures of the regulatory region of *Clostridium perfringens* CBS-PPase with activating and inhibitory ligands to elucidate its mechanism of regulation by nucleotides (paper III).
3. Cloning, expression, purification and mutagenesis of *Moorella thermoacetica* CBS-PPase for functional studies (papers II and IV).

3 Methods

Detailed descriptions of the materials and methods employed are presented in the original publications I-IV.

3.1 Cloning, mutagenesis, protein expression and purification (II-IV)

Genomic DNA samples extracted from *C. perfringens* strain 13 and *M. thermoacetica* strain ATCC 35608 were obtained from German Collection of Microorganisms and Cell Cultures. Genes encoding *C. perfringens* CBS-PPase (*cpCBS*-PPase, GenBank NP_562971, Swiss-Prot Q8XIQ9) and *M. thermoacetica* CBS-PPase (*mtCBS*-PPase, GenBank NC_007644, Swiss-Prot Q2RFU3), as well as the regulatory region of *cpCBS*-PPase (residues 66-306, *cpCBS*) were amplified with PCR. *NdeI*-*XhoI* fragments were ligated into pET36+ plasmid transformed into the *E. coli* expression strain BL21-CodonPlus(DE3)-RIL (Stratagene). The *NdeI*-*XhoI* fragment of *cpCBS* was additionally ligated into the pET15b+ plasmid. For mutagenesis purposes, the *XbaI*-*XhoI* fragment of *mtCBS*-PPase was moved to pBLUESCRIPT SK⁻ plasmid, and point mutations generated with the QuikChange kit (Stratagene).

All CBS-PPase proteins were expressed in Terrific Broth (TB) medium, except the selenomethionyl derivative (SeMet), which was produced in B834(DE3) *E. coli* strain (Novagen) grown in commercial SelenoMet Medium Base with Nutrient Mix (Molecular Dimensions Ltd). Protein expression in TB medium was induced with 0.4 mM isopropyl β -thiogalactopyranoside at A_{600} of 1.2, and cultures incubated for a further 3 h. Full-length proteins were purified with ionic exchange and size exclusion chromatography. To achieve higher purity for crystallization purposes, a histidine-affinity tag was added. *cpCBS* was expressed with both the C-terminal His₈ tag and N-terminal thrombin cleavable His₆-tag. His-tagged proteins were purified using immobilized-metal affinity and size exclusion chromatography. Thrombin digestion was performed overnight at room temperature using the affinity column. Concentrated proteins were stored at -70°C. Protein purity was assessed with SDS-PAGE and homogeneity with native PAGE. Protein concentrations were estimated from their A_{280} values.

3.2 Characterization of proteins (II-IV)

(II-IV) After optimization of reaction conditions, enzyme activity measurements were mainly performed at 25°C in 25-40 ml of 100 mM MOPS/KOH buffer (pH 7.2), including 0.1 mM CoCl₂, 5 mM MgCl₂ and 0.16 mM PP_i for *mtCBS*-PPase or 0.1 mM CoCl₂, 20 mM MgCl₂ and 0.23 mM PP_i for *cpCBS*-PPase. Reactions were initiated either by adding enzyme or PP_i. P_i formation was monitored using a continuous P_i analyzer. Due to low *mtCBS*-PPase activity, a more sensitive method was used, and reactions were followed for longer time-periods of up to 10 minutes.

(II & III) The oligomeric states of proteins with and without ligands and at different protein concentrations were determined via analytical ultracentrifugation on a Spinco E instrument (Beckman Instruments). Glutaraldehyde cross-linking experiments were conducted in parallel to support sedimentation analysis data.

(III) Thermostability measurements were performed using a differential scanning fluorometer (ThermoFluor) with the MiniOpticon Real-Time PCR System (BioRad). The assay mixture was placed in 96-well plates containing nucleotides at variable concentrations, full-length *cp*CBS-PPase, HEPES/KOH, pH 7.5, and SYPRO Orange (Invitrogen). Plates were heated from 30 to 99.9°C, and fluorescence monitored at 523–543 nm.

3.3 Crystallization and structure determination (I and III)

Wild-type and variant *sc*PPase proteins were stored at -20°C in buffered 50% glycerol [38]. Glycerol was removed with buffer exchange before crystallization. CBS-PPases were stored at -70°C without cryoprotectant. All proteins were crystallized with the vapor diffusion method, *sc*PPases in sitting drops at 4°C and CBS-PPases in hanging drops at room temperature. The *sc*PPases contained 5 mM MgCl₂ and 1 mM P₁ as ligands in crystallization, whereas *cp*CBSs contained 0.15 mM AMP or 0.25 mM diadenosine tetraphosphate (AP₄A) as regulators.

Wild-type *sc*PPase data were collected at synchrotron beam BM01 at ESRF (Grenoble, France) and all variant data collected with a Rigaku or Nonius generator and R-AXIS IV detector. All *sc*PPases, including wild-type and the seven variants, crystallized to the P2₁ space group with almost identical unit cell parameters ($a = 51.58\text{--}51.92 \text{ \AA}$, $b = 93.00\text{--}93.55 \text{ \AA}$, $c = 69.17\text{--}70.11 \text{ \AA}$ and $\beta = 99.33\text{--}100.01^\circ$). Resolution was at least to 1.9 Å and optimally to 1.5 Å. Data were processed with Denzo [129]. Phases were from molecular replacement by using MOLREP [143] where the model structure was Mg²⁺-bound *sc*PPase (PDB ID: 1HUK) [Swaminathan and others, unpublished]. Density fitting was performed with ARP/wARP [130], Refmac5 [131] used for refinement, and manual model building performed with O [132] and Coot [133]. Procheck was applied for structure validation [134].

All *cp*CBS data were collected at ESRF. Both types of *cp*CBS complexes crystallized to space group P2₁2₁2₁ with similar unit cell parameters (AMP-bound, $a = 58.96 \text{ \AA}$, $b = 79.61 \text{ \AA}$, $c = 116.36 \text{ \AA}$; AP₄A-bound, $a = 63.85 \text{ \AA}$, $b = 71.76 \text{ \AA}$, $c = 116.33 \text{ \AA}$). The complexes displayed slight differences in crystal packing (Figure 8). We managed to crystallize His₈-tagged *cp*CBS (with and without SeMet labeling) with AMP and His₆ tag-removed protein with AP₄A. Data were processed with XDS [135]. Both data sets were obtained to 2.3 Å resolution. Phasing appeared difficult due to pseudosymmetry. During processing, it was observed that every second spot was missing from axes a and c and therefore they had clear 2₁ screw axis, but b was not clear. According to Truncate [136] analysis, AMP-bound data appeared pseudosymmetric, based on the peak in the Patterson map (Table 2), which appeared to

be due to the 2-fold non-crystallographic symmetry (NCS) axis parallel to the b -axis. For AP₄A, the NCS axis was at an angle of 18.5° in relation to the b -axis (Figure 8). The L-test [137] (Table 2) in Xtriage [138] gave no indication of twinning.

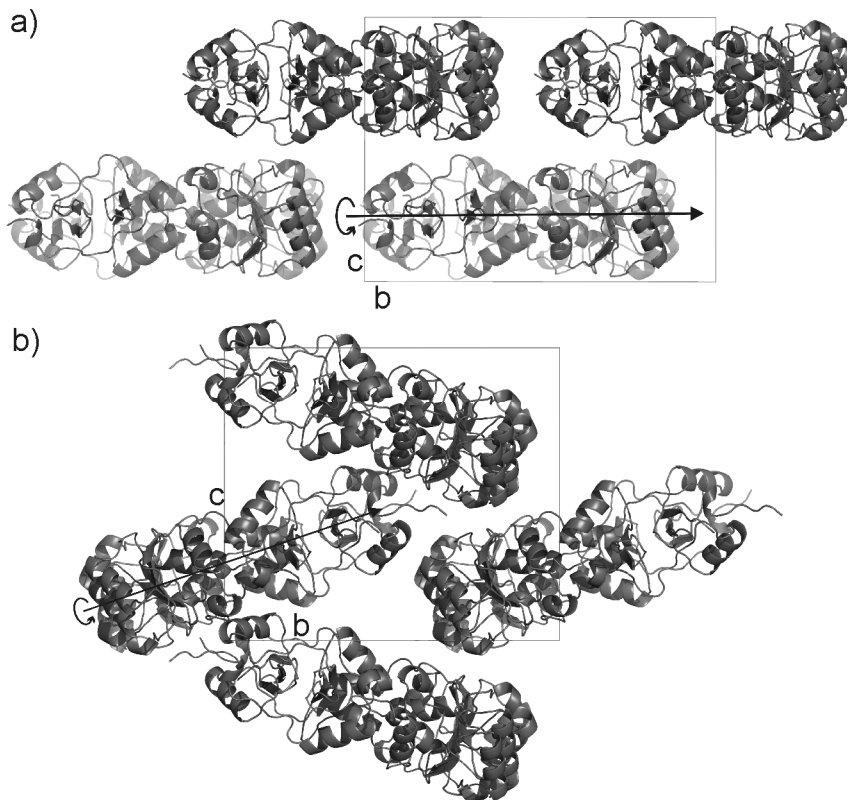


Figure 8. Different crystal packing structures of *cpCBS*. (a) Data for the AMP-bound structure are pseudosymmetric due to the parallel two-fold NCS axis (arrow) relative to unit cell b -axis. (b) In the AP₄A-bound crystal structure, the NCS axis is rotated by 18.5°, relative to the b -axis.

Table 2. Patterson and L-test analysis

| | AMP | SeMet AMP | AP ₄ A |
|-------------------------|---------------------|---------------------|-------------------|
| Patterson peak | 0.500, 0.500, 0.224 | 0.500, 0.500, 0.222 | |
| % of origin | 40 | 45 | |
| Mean $ L $ ¹ | 0.457 | 0.469 | 0.483 |
| Mean L^2 ² | 0.288 | 0.300 | 0.314 |
| Twin fraction | 0.04 | 0.03 | 0.01 |

¹ for twinned 0.375, untwinned 0.500

² for twinned 0.200, untwinned 0.333

Phasing of pseudosymmetric data was difficult manually. After testing different methods, we used a combination of XDS and Auto-Rickshaw pipeline [139] to solve the phases of pseudosymmetric data. Used method was the MRSAD [140] where single wavelength anomalous dispersion method is combined with the molecular

replacement. ShelxD [141] was effective in locating 8 selenium atoms out of a total of 18. The figure of merit was 0.589 after refinement with MLPHARE [142], and density modification with DM [143] and ARP/wARP [130] allowed building of only 130 residues into the electron density map. This model and phases were improved by feeding the resulting model back into Auto-Rickshaw using the MRSAD method together with the sequence. In combination with manual model building, this helped ARP/wARP to build 90% of the residues. The AP₄A structure was solved by molecular replacement using MOLREP [144] using the AMP structure as a template. Structures were refined with REFMAC5 [131] through the CCP4 interface [145]. Manual model building was performed with Coot [133]. Coot and MolProbity [146] were used for structure validation. TLS refinement [147] lowered *R*-factors and helped to maintain reasonable geometry.

3.4 Homology modeling

For modeling of full-length *cp*CBS-PPase, we used structures of common family II PPases (PDB ID: 1K20, 1WPP, 2EBO, and 2HAW) and *cp*CBS complexes with AMP and AP₄A as templates. Alignments were performed with Bodil [148]. 2D structures for the linker areas 65–73 and 296–307 were predicted with PSIPRED [149, 150], and homology modeling performed with Modeller 9v6 [151]. PSIPRED predicted that the linker region between positions 66 and 72 forms an α -helix. The linker region (positions 296–307) was disordered in our structures, and prediction did not provide a reliable structure. However, we modeled residues 296–305 as an α -helix, based on other CBS domain structures.

4 Results and Discussion

4.1 Complete catalytic cycle of family I PPases (I)

In paper I, wild-type and seven variant *sc*PPases (Table 3) were crystallized with phosphate and the physiological metal cation, Mg^{2+} , whereas previous structures contained Mn^{2+} or Co^{2+} , which allowed efficient structure determination due to anomalous signals. Interestingly, all Mg^{2+} -bound crystals displayed one monomer in an open form (A) and the other in a closed conformation (B) due to crystal packing. The loop of residues 103–115 was ordered in the closed form, but disordered in the open form. Seven conservative point mutants of residues around the electrophilic phosphorus, P2, and at the bottom of the active site were generated, including E48D, Y93F, D115E, D117E, D120E, D120N and D152E (Figure 2 in original publication I). The active site conformations depended on the PPase variant, and were structurally equivalent to various intermediates along the reaction pathway.

Most of the structures were very similar to that of wild-type at the backbone level, where overall rmsd between the structures was from 0.1 Å to 0.28 Å for 218–263 superimposed C^α atoms. A monomers have more differences than B monomers, which adopt the same conformation as wild-type, except the closed monomer of the D152E mutant in which the active site is more disordered. Structures were determined to 1.5–1.9 Å resolution, and the electron density maps allowed Mg^{2+} ions to be identified based on their geometry. Each active site contains one to four Mg^{2+} ions and zero to two phosphate ions (Table 3) depending on the mutation, and possibly, exact soaking time. In two open conformation structures (wt and D152E), the leaving phosphate (P_{exit}) exists on the “edge of cup” (Figure 1). One extra Mg^{2+} ($Mg9$) is located nearby in the closed form of D152E. These structures may be assigned to several intermediates of the different reaction steps.

Table 3. Active site content of Mg^{2+} -complexed *sc*PPases. (modified from paper I)

| | monomer A active site ligands | monomer B active site ligands |
|------------------------|-------------------------------|-------------------------------|
| <i>wt</i> ^a | Mg1, Mg2, P_{exit} | Mg1, Mg2, Mg3, P1 |
| E48D | Mg1, Mg3, P1 | Mg1, Mg3, P1 |
| Y93F | Mg1, Mg2 | Mg1 |
| D115E | Mg1, Mg2, Mg3, Mg4, P1, P2 | Mg1, Mg2, Mg3, Mg4, P1, P2 |
| D117E | Mg1, Mg2 | Mg1, Mg2, Mg4, P2 |
| D120E | Mg1, Mg3, P1 | Mg1, Mg3, $P1^b$ |
| D120N | Mg1, Mg3, P1 | Mg1, P1 |
| D152E | Mg2, P_{exit} | Mg2, Mg9 |

^a wild-type, ^b Likely, two conformations

PPases catalyze the hydrolysis of pyrophosphate, and require three to four divalent metal cations, depending on pH. The magnesium ion confers highest catalytic activity. Together with previously reported structures, the solved *sc*PPase structures provide the first complete structural description of the catalytic cycle. The mutations affected metal binding and the hydrogen bonding network in the active site, providing an insight into

the roles of specific residues. Analysis of the wild-type and four variant proteins provided information on the different reaction intermediates. Based on the new structures, the reaction mechanism scheme was revised (Figure 9a) to include two alternative routes for product release and nucleophile regeneration.

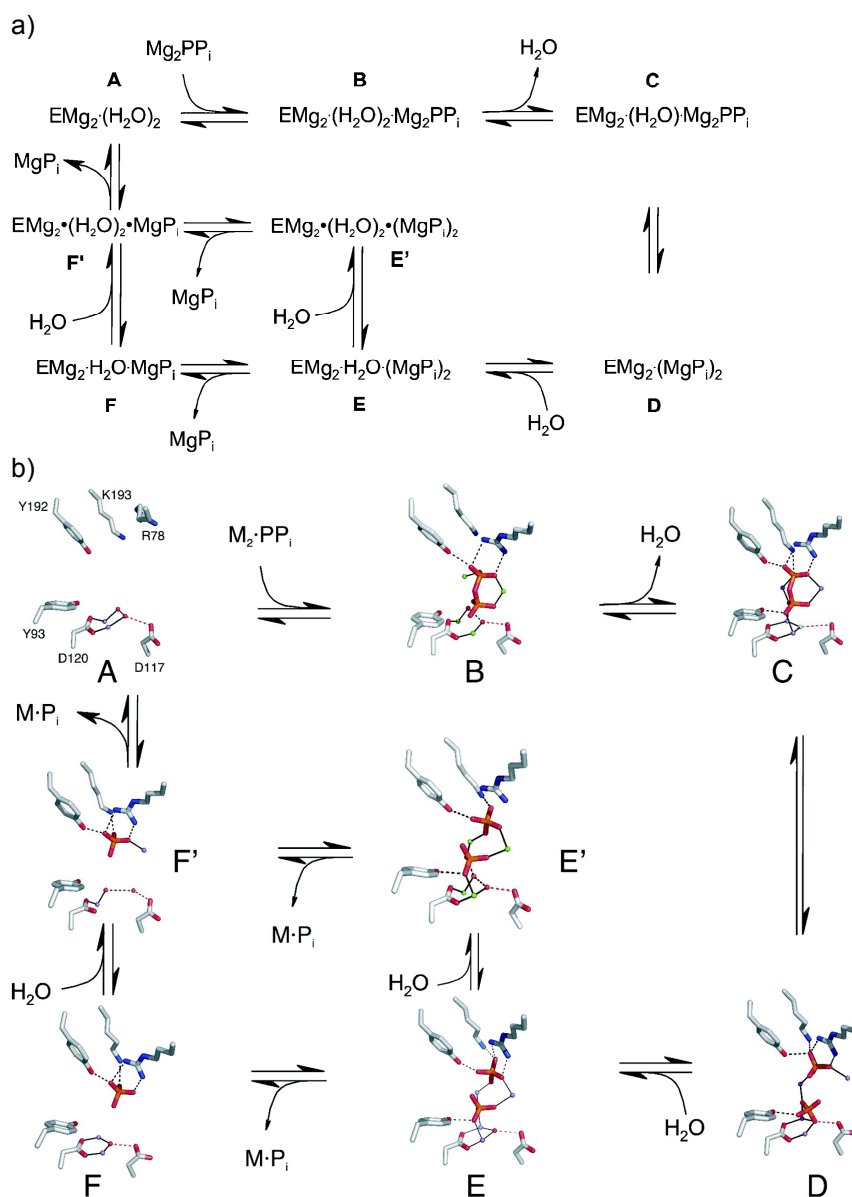


Figure 9. (a) Revised reaction scheme and (b) active site structures. Residues R78, Y93, D117, D120, Y192, and K193 are labeled in intermediate A. Metals, phosphates, and the two water molecules bridging M1 and M2 are further shown. PDB IDs for intermediates: A, 1WGI [19]; B, D115E_A with PP_i modeled; C, 1E6A [29]; D, 1E9G [27] conformation B(1); E, 1E9G conformation A(1); E', D115E_A; F, wt_B; F', E48D_A (paper I).

Several structures are involved in step A, including monomer A of wild-type, Y93F and D117E (Figure 9b) encompassing both Mg1 and Mg2 with two water molecules in between. Similar structure and composition were previously observed with Mn²⁺ ions (PDB ID: 1WGI [19]). The structure of the substrate-bound enzyme before nucleophile attack (intermediate B) was not solved, but modeled for D115E_A by inserting PP_i in the place of two P_is. The situation after reorganization (state C) is presented for the D117E_B structure, even though P1 and Mg3 are missing due to mutation. Furthermore, the active site is fully occupied in the fluoride-inhibited Mn²⁺ structure (PDB ID: 1E6A [27]). State D is represented by the high-resolution Mn²⁺ structure (PDB ID: 1E9G [27]). All four product-releasing intermediates are represented as Mg-bound variant structures, specifically, state E in the D115E_B structure where one water molecule binds and weakens P1 interactions, state E' in D115E_A (also PDB ID: 117E [29] and 1E9G [27]) where another water molecule is added, and alternative state F in E48D_B as well as wt_B where the Mg3-P1 complex is released before the addition of a second water molecule (step leading to F') in E48D_A. Determination of these structures completes the X-ray structural studies by providing all kinetically recognized reaction intermediates, including two alternative product release and nucleophile regeneration pathways that cannot be distinguished based on kinetic data and possibly operate in parallel.

4.2 Production and initial characterization of CBS-PPases (II and III)

We have cloned, expressed, purified and characterized full-length CBS-PPase from *Clostridium perfringens* and the enzyme from *Moorella thermoacetica*, one of three known CBS-PPases without the DRTGG domain. Several other CBS-PPases cloned were less amenable to purification than *cp*CBS-PPase. The solubility of this protein was approximately 15 mg/ml. The protein did not move as a sharp band on native PAGE (data not shown). Shorter *mt*CBS-PPase was easier to handle, as it could be concentrated to 100 mg/ml, and migrated as a sharp band on native PAGE, although the purity assessed using SDS-PAGE was not as good as that for single-banded *cp*CBS-PPase (data not shown).

The *cp*CBS-PPase activity was very low in the presence of Mn²⁺ and Mg²⁺. Trials with different metals showed that Co²⁺ and Ni²⁺ (unpublished result) are the best cofactors for both CBS-PPases. Magnesium was also needed for optimal catalytic activity, and thus, subsequent studies employed both Co²⁺ and Mg²⁺. This Co²⁺ / Ni²⁺ preference was an interesting finding, as both these metals are essential trace elements for *Moorella thermoacetica* [152].

The catalytic constant, k_{cat} , appeared quite low for CBS-PPases, specifically, 1.7 s⁻¹ for *mt*CBS-PPase and 20 s⁻¹ for *cp*CBS-PPase, which is over 100 times lower than that for common family II PPases. CBS domains may thus play an autoinhibitory role in CBS-PPases.

Furthermore, CBS-PPases were found to be regulated by nucleotides. Nucleotide specificity was estimated using *mt*CBS-PPase. Adenosine compounds induced the most significant effects. ATP activated, while ADP and AMP inhibited CBS-PPases (II). The nucleotides were effective at nanomolar (ADP) or micromolar (AMP and ATP) concentrations, representing 100–10,000-fold stronger binding than that in other CBS domain-containing proteins [105, 115, 124]. The two CBS-PPases investigated had different nucleotide specificities. AMP was the strongest inhibitor, and diadenosine tetraphosphate (AP₄A) activated *cp*CBS-PPase 2.5-fold (III).

4.3 Overall structure of CBS-PPase (III)

Extensive crystallization trials were performed with both full-length CBS-PPases. Small, straight needles were obtained once with *cp*CBS-PPase, but these were too small for structural studies and could not be reproduced. The best *mt*CBS-PPase crystals obtained diffracted to 4.5 Å, but optimization did not lead to improvements in crystal quality. Crystallization of these proteins, which contain several flexible domains and potential stabilizing ligands, appeared a significant challenge.

Accordingly, a truncated version of *cp*CBS-PPase was constructed by deleting the catalytic region, keeping in mind that homologous structures are already available. We were thus able to crystallize and solve the structure of the regulatory part CBS1-DRTGG-CBS2 (*cp*CBS). *cp*CBS-His₈ was crystallized with AMP, and *cp*CBS with the His-tag removed (via thrombin cleavage) with AP₄A. For phasing purposes, the *cp*CBS-His₈-AMP complex was crystallized with SeMet protein.

The CBS domain fold in *cp*CBS structures was similar to those of other CBS domain structures whereby the core topology forms a β - α - β - β - α assembly (Figure 4, 10), but the last helix of the CBS2 domain was disordered. A unique feature of our structures is the presence of the DRTGG domain insert between the CBS domains, in contrast to the successive CBS domains in all other structures.

The CBS1 and DRTGG domains are mainly well-ordered in the *cp*CBS structures, while the CBS2 domain appears more flexible, in view of its higher *B*-factor values. The two monomers in the asymmetric unit are similar in both structures, where rmsd per C^α is below 0.6 Å. The only significant difference is in the CBS1-DRTGG linker, whereby monomers adopt different conformations in the activator structure due to crystal contacts and the same areas are disordered in the inhibitor structure. Both structures display one special feature. Tyr124, located at the NCS axis, adopts two conformations within a very confined space by forming hydrogen bonds either with the carbonyl oxygen of Arg240 or, in unusual manner, with Met125 [153]. While the refinement was somewhat complicated due to translational pseudosymmetry in the inhibitor complex, the ligand-binding site is well-ordered in both structures.

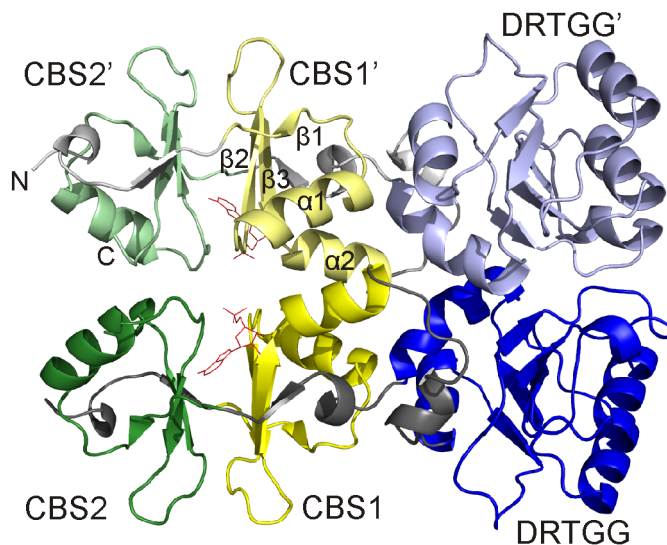


Figure 10. The overall structure of dimeric *cpCBS-His₈* complexed with AMP. The CBS1 core is presented in yellow, the DRTGG domain is in blue, and the CBS2 domain is in green, while linker areas are shown in gray. The second subunit is presented in similar, but lighter colors. AMP is depicted in red. The CBS domain core topology, $\beta 1-\alpha 1-\beta 2-\beta 3-\alpha 2$, is indicated in CBS1'. (paper III)

cpCBS forms a homodimer (Figure 10) in which all three domains interact with the adjacent identical domains of the other monomer. The dimeric structure of soluble *cpCBS* has been confirmed with sedimentation and cross-linking experiments (III). Full-length *mtCBS-PPase* (II) is also homodimeric, similar to common family II PPases [154]. CBS-PPases appear to be tighter dimers, since dissociation is not observed at low enzyme concentrations, in contrast to common family II PPases [54].

4.4 Ligand binding to CBS domains

4.4.1 Adenylate binding to *cpCBS-PPase* (III)

The *cpCBS* crystals grown in the presence of AMP contained two AMP molecules bound per dimer (i.e., one AMP per CBS pair). The nucleotide was clearly visible in the difference map between the CBS1 and CBS2 domains on the side of CBS1 (Figure 11a). The adenine moiety forms hydrophobic interactions with Met114 of CBS1 and Tyr278 of CBS2, as well as hydrogen bonds with the main-chain atoms of CBS2 Asn280 and Val258 in the linker region before the CBS2 domain core. Ribose hydroxyls form hydrogen bonds with Asn119 and Thr253 in linker before the CBS2 core. Phosphate is coordinated to Ser116 and Ser118 hydroxyls and Lys100' (the primed number refers to the symmetry-related residue from the other subunit of the dimer). A water molecule bound to Ser101 and Asn280 forms hydrogen bonds with both adenine N7 and phosphate.

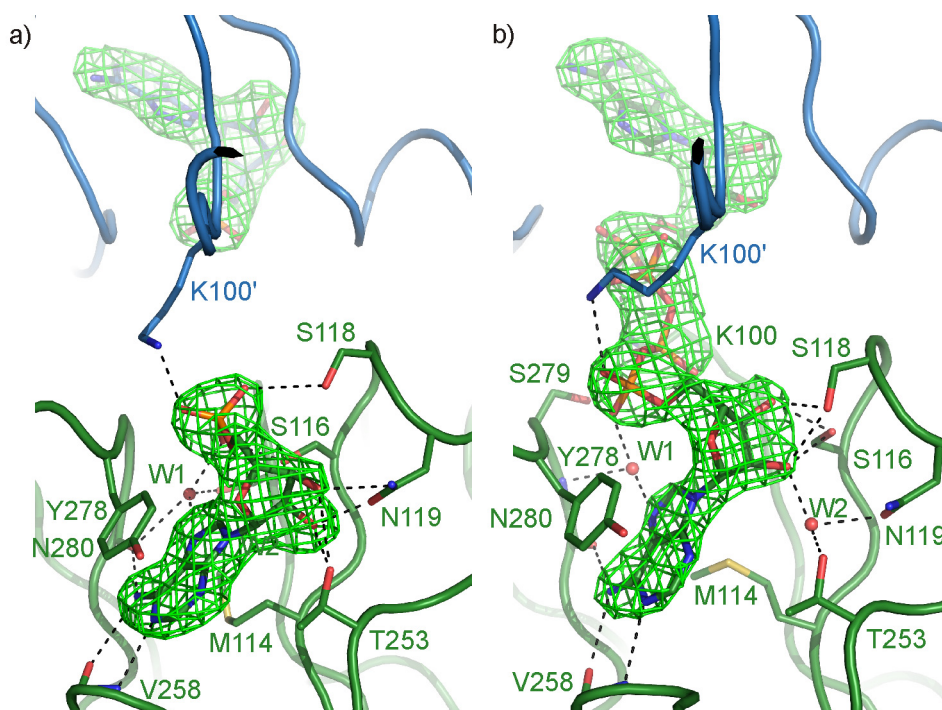


Figure 11. Positive difference maps (contoured at 3.0σ) of bound AMP (a) and AP₄A (b). The lower subunits in (a) and (b) are shown in detail, with residue numbers. Monomer A is shown in green and monomer B in blue. Red balls refer to water molecules. CBS1 motif residues are from Met114 to Asn 119. (paper III)

In *cp*CBS crystals grown in the presence of AP₄A, only one ligand molecule is present (Figure 11b), which occupies two binding sites, which are same as for AMP. While AP₄A adenine interactions are similar to those in the AMP-bound structure, ribose and phosphate interactions are clearly different. Asn119 and Thr253, which form hydrogen bonds with ribose in the AMP structure, interact with the activator only via a water molecule. Instead, ribose is hydrogen-bonded to Ser116 and Ser118 in the *cp*CBS–AP₄A complex. Two phosphates of AP₄A bind to Ser279 and Lys100'. In contrast, the contacts between AMP and the CBS2 domain are similar to those in the AP₄A-bound structure.

4.4.2 Comparison with other CBS domain proteins

Binding of AMP by the CBS domain pair is similar, but not identical to that reported previously for other CBS proteins [73, 109, 111-113]. The adenine ring is invariably bound by two hydrophobic residues, which are aligned with Met114 and Tyr278 of *cp*CBS, and via two main chain interactions (Figure 11). Asn119 in CBS1 motif interacting with the ribose moiety in *cp*CBS is Asp in all AMP-bound structures. The second ribose-binding residue, Thr in *cp*CBS and Ser in the CBS domain-containing protein from *Sulfolobus solfataricus* (PDB ID: 3DDJ), is Arg in AMP-activated protein kinase [112], Asp in possible d-arabinose 5-phosphate isomerase yrbH from *E. coli*

(PDB ID: 3FNA), and Asn in the CBS domain protein PAE2072 from *Pyrobaculum aerophilum* (PDB ID: 2RIF). The highest variability is observed in the phosphate-binding residues. Ser116 and Ser118 are nearly always aligned with Ser or Thr in other proteins (Ser118 is replaced with Arg in the PAE2072 protein from *P. aerophilum*) and the residue corresponding to Lys100 is typically replaced with non-aligning Lys, His or Ser in other structures.

4.5 Proposed mechanism of *cp*CBS-PPase regulation (III)

The overall structures of the two *cp*CBS complexes superimpose with an rmsd of 1.32 Å for all equivalent residues (460 C^α atoms). CBS2 domain movement is responsible for this deviation. The CBS2 domains, including the linker before the CBS1 core (99 C^α atoms), superimpose with rmsd of 2.32 Å, whereas the rmsd value is only 0.71 Å for the rest of the structure (362 C^α atoms). A clear conformational transition is evident from the inhibitor- to activator-bound form, where different orientations of the adenosine rings of the activator lead to displacement of Tyr278 and the whole RY²⁷⁸SN loop, and consequently, a more open CBS2 domain dimer interface (Figure 12).

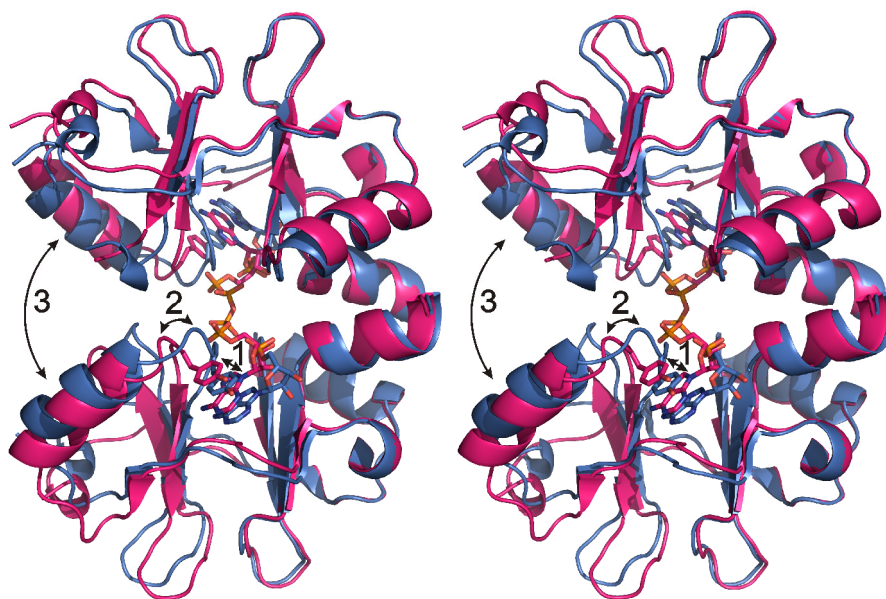


Figure 12. Stereo view of the superimposed AMP- (blue) and AP₄A-bound (red) CBS pair structures of *cp*CBS representing movement of Tyr278 along aromatic interactions to the adenine ring of ligand (1), conformational changes of the RY²⁷⁸SN loop (2), and opening of the CBS2 dimer interface in the AP₄A complex (3). (paper III)

Ligand-induced conformational changes in CBS domains were confirmed by comparing the effects of the inhibitor and activator on the thermostability of full-length *cp*CBS-PPase (Figure 13). The melting point of AMP-bound enzyme was 7°C higher and that of the AP₄A-bound form was 10°C lower, compared with the ligand-free enzyme.

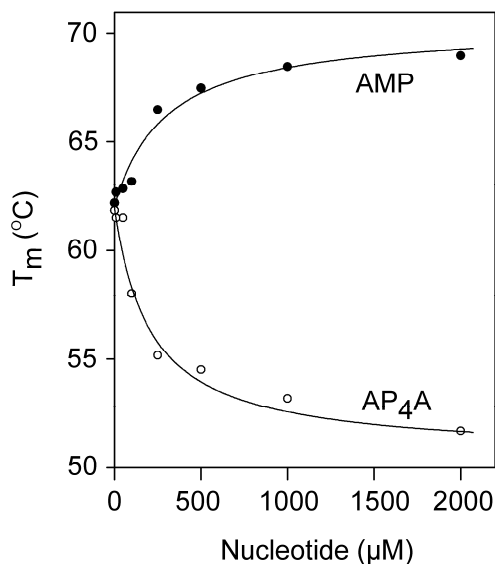


Figure 13. ThermoFluor measurement of *cpCBS*-PPase thermostability in the presence of AMP and AP_4A .

The available structures of several ligand complexes of MJ0100 demonstrate remarkable movement [67]. A comparison of ligand-induced *cpCBS* structure opening with that of MJ0100 revealed the most marked movement in the CBS2 domain interface in both proteins. Interestingly, similar displacement of a phenylalanine residue from the dimer interface to its own CBS domain interior was observed in activated *cpCBS*-PPase and 12°- and 23°-opened MJ0100.

To obtain an insight into the pathway of the regulation signal to the enzyme active site, full-length *cpCBS*-PPase was modeled based on known common family II PPases and *cpCBS* insert structures (Figure 14). According to this model, the enzyme active sites are located 25 Å from the nucleotide-binding sites. Two possible mechanisms are envisaged. One assumes that the conformational change in the CBS domains allows more freedom of movement for the DHHA2 domain. DHHA2 domains are in close proximity to the moving CBS2 domain interface, and are expected to move during substrate binding and product release [48, 50]. Notably, the active site is formed by residues from DHH and DHHA2 domains connected by a flexible linker sequence. We speculate that in the AMP-inhibited state, the closed CBS2 domain interface prevents this movement and DHHA2 domains move freely in the open activated state.

The other mechanism assumes that the insert causes disorder in the active site via direct effects on important residues. The basis for this mechanism is the finding that short linkers between DHH and CBS domains (residues 66–70 and 305–308) are adjacent to the DHH domain chain segment (residues 315–331), which connects two highly conserved active site motifs, DHNE^{314} and DHH^{334} . A combination of these two mechanisms is additionally possible.

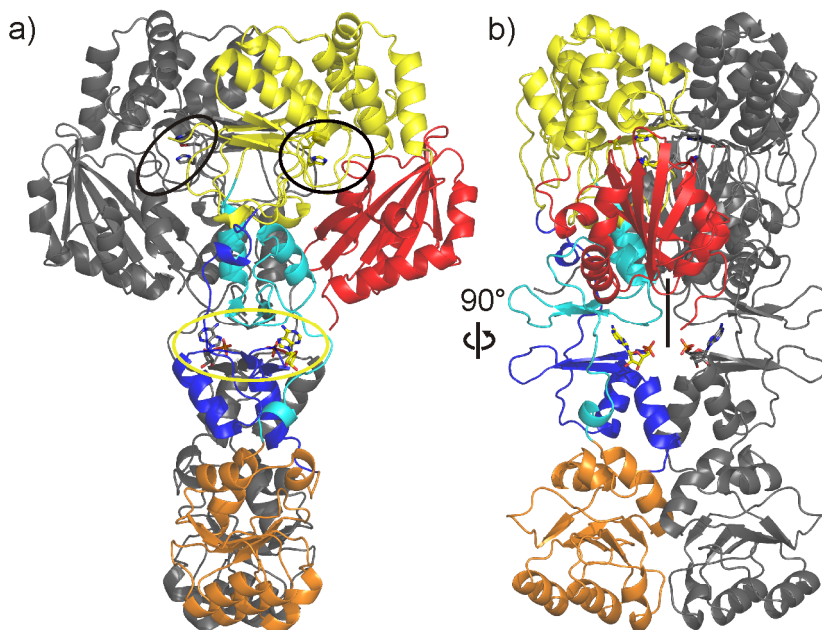


Figure 14. (a and b) Two views of the modeled *cpCBS*-PPase domains complexed with AMP. In one subunit, the main PPase domains are shown in yellow (DHH) and red (DHHA2), CBS1 and CBS2 domains presented in cyan and blue, respectively, and the DRTGG domain in orange. The other subunit is depicted in gray. Active sites are within black circles and regulator binding sites within the yellow circle. In (b), the molecule is rotated by 90°. The DHHA2 domain is located on the flexible CBS2 domain interface (black bar). (paper III)

4.6 The DRTGG domain (III)

Our group initially described the DRTGG domain, which is present in many proteins. This domain is located between two CBS domains in CBS-PPases, and has an α/β structure. Domain topology is $\alpha 1-\beta 1-\beta 2-\alpha 2-\beta 3-\alpha 3-\beta 4-\alpha 4-\beta 5-\alpha 5$, where the β -strand direction in the sheet is depicted with arrows in Figure 15. The conserved amino acids, $D^{189}R^{190}$ and $T^{208}G^{209}G^{210}$ (numbered based on full-length *cpCBS*-PPase, Figure 16), are in close proximity on the domain surface, near the dimer interface, and located in loops after strands 3 and 4, respectively. The dimer contacts of the DRTGG domains are mainly formed by hydrophobic interactions of helices 5. The DRTGG domains are almost identical in both *cpCBS* structures (rmsd is 0.362 for all 220 C_{α} atoms in dimer), and contain no bound ligand.

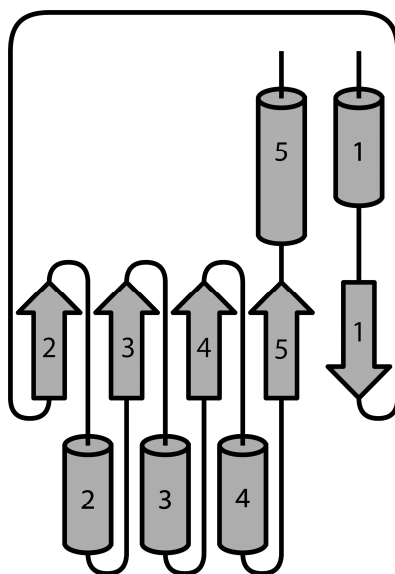


Figure 15. Topology of the DRTGG domain. (paper III)

To date, the DRTGG domain structure of only one more protein, putative phosphotransacetylase (PtaN) from *A. fulgidus* (PDB ID: 2IOJ), has been determined. Its topology and overall structure are identical, with an rmsd of 1.88 for 90 C α atoms. The greatest difference is in quaternary structure. Based on PISA [155], the crystal contacts of 2IOJ are not preserved in solution and the protein appears monomeric. A R252H mutation in β 3 prevents NADH inhibition and enhances pyruvate activation of *S. enterica* phosphotransacetylase [127], but no binding of these effectors to the DRTGG domain has been demonstrated [128].

The structures of two distant relatives of the DRTGG domain are available. The DRTGG domain belongs to the MurF_HprK_N clan (Pfam: PF07085), which also contains N-terminal domains of MurF (UDP-N-acetylmuramoyl-tripeptide D-Ala-D-Ala ligase) and HPrK/P (histidine containing phosphor carrier protein kinase/phosphatase). The MurF N-terminal domain has the same β -sheet topology as the DRTGG domain and binds inhibitors in two structures of MurF from *Streptococcus pneumoniae* (PDB ID: 2AM1 and 2AM2) [156]. MurE from *Mycobacterium tuberculosis* (PDB ID: 2WTZ) is analogous to MurF, and binds the UDP moiety of substrate in the same place [157]. An analogous area in *cpCBS*, located in the loop adjacent to the poorly ordered DR loop containing conserved A169 (Figure 16) may thus represent a potential binding site for an unrecognized ligand. Furthermore, the *Staphylococcus xylosus* HPrK/P structure (PDB ID: 1KO7) [158] contains two phosphate molecules bound at the dimer interface of Hpr_kinase_N domains, suggesting that a similar cavity found in *cpCBS* may also be a binding site for some ligand.

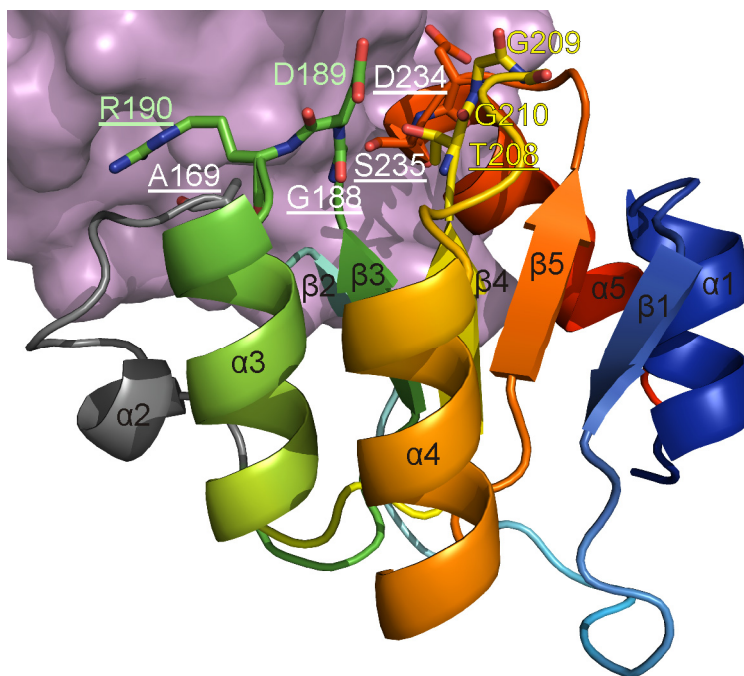


Figure 16. The DRTGG domain of *cpCBS* is colored from blue (N-terminus) to red (C-terminus). The other DRTGG monomer is depicted as a violet surface. The poorly ordered loop-helix 2 region is gray. Conserved residues close to the subunit interface are presented as a ball-and-stick model. The residues belonging to the conserved DR and TGG motifs are presented in green and yellow, respectively. The most conserved residues are underlined. (paper III)

4.7 *mtCBS*-PPase variant analysis (IV)

The effects of ligand binding were determined by mutating eight residues in potential regulatory ligand binding areas of *mtCBS*-PPase. The majority of residues were substituted with Ala or Gly, and Tyr169 was conservatively replaced with Phe. The residues for substitution were selected based on the mutations in the CBS domain-containing proteins that cause human hereditary diseases; some of the most conserved and adjacent bulky residues were also mutated (Table 5). The disease-causing mutations (Figure 3), D444N in cystathionine β -synthase and R302Q in AMPK γ 2, correspond to Lys100 of *mtCBS*-PPase. The L766P and R767Q/W mutations in CIC-7 correspond to Tyr169 and Arg170 of *mtCBS*-PPase, respectively. R224P and D226N in IMPDH 1 correspond to *mtCBS*-PPase Arg187 and His189. Furthermore, R225W (corresponding to *mtCBS*-PPase Lys100) in human AMPK γ 3 and V249I and R250G (corresponding to *mtCBS*-PPase Val99 and Lys100, respectively) in pAMPK γ 3 trigger changes in the glycogen content of cells. Val99, Arg168, Tyr169, Arg170, and Arg187 (*mtCBS*-PPase numbering) are highly conserved in CBS-PPases [159]. Furthermore, the *cpCBS* structure suggests that Tyr169 stacks with the adenine ring and Lys100 binds phosphate. Residues at positions 168–170 form part of the RYRN sequence corresponding to the RYSN loop of *cpCBS*-PPase, which undergoes the most

significant conformational changes between the activator- and inhibitor-bound forms of the enzyme. Val99 of CBS2 is located at the same relative position as CBS1 Tyr169. Residues 187–189 of CBS2 include His189 corresponding to ribose-binding residues in the nucleotide-bound *sp*AMPK [111] and ClC-5 [105] structures.

Table 5. Conservation of mutated residues based on 116 CBS-PPase sequences (one sequence per species selected from UniProt [159]).

| Mutated residue* | identity | similarity | variation within similar |
|------------------|----------|------------|--------------------------|
| V99/99 | 25% | 78% | I/L/V/M/F |
| K100/100 | 33% | 48% | K/R |
| R168/277 | 64% | 76% | R/K |
| Y169/278 | 33% | 86% | Y/F/H/V/L/I |
| R170/279 | 84% | 84% | R/K |
| R187/296 | 86% | 91% | R/K |
| Y188/297 | 22% | 24% | Y/F |
| H189/298 | 37% | | |

* Numbering based on *mt*CBS-PPase and *cp*CBS-PPase, respectively.

Based on the effects of mutations on k_{cat} , variant enzymes were divided into three groups: 1) similar catalytic activity as wild-type (K100G), 2) more active (V99A, R168A, Y169A, Y169F, Y188A, and H189A), and 3) less active (R170A and R187G) enzymes.

The majority of variant enzymes were inhibited by AMP, analogous to wild-type. Conversely, the R168A and Y169A variants were activated by AMP. The residual activity of the enzyme-AMP complex was increased in the Y188A and H189A variants. The AMP-binding affinity ($K_{i,app}$) was lowest for V99A, slightly decreased in K100G and Y188A, remarkably decreased in R168A and Y169A, not significantly affected in the Y169F and R170A variants, and notably elevated in the R187G and H189A mutant proteins (Table 6). In most cases, the effects on ADP affinity were parallel to those of AMP. Direct binding measurements with the membrane filtration assay disclosed that the stoichiometry of AMP binding increased from 1 to 2 ligands per monomer in the Y169F, R170A, R187G, and Y188A variants, implying occupancy of both CBS domains in the pair. Based on the *mt*CBS-PPase model, it is proposed that there is no space for two adenylate ligands in the wild-type CBS domain pair, but space may arise upon converting a bulky residue to the smaller Ala or Gly.

Most substitutions appeared to enhanced enzyme activity by weakening autoinhibition, providing further support for the autoinhibitory role of CBS domains. Furthermore, we observed reversal of the AMP effect from inhibition to activation in the R168A and Y169A substituents. To our knowledge, no studies to date have reported changes in the direction of a nucleotide effect as a result of amino acid substitution in CBS domain-containing proteins.

Table 6. Catalytic and nucleotide-binding parameters for wild-type and variant *mtCBS-PPases* (IV). Values showing large deviations from wild-type enzyme are presented in bold. Parameter values for catalytic activity and nucleotide inhibition refer to substrate-activated enzyme.

| Enzyme variant | Catalytic activity | | | AMP inhibition/activation | | ADP inhibition | | AMP binding | |
|----------------|--------------------------------------|-------------------------|--------------------------------------|---------------------------|--------------------------------------|-----------------------|-------------------------|---------------------------------|--|
| | k_{cat} (s^{-1}) | K_m (μM) | $K_{i,\text{app}}$ (μM) | Residual activity (%) | $K_{i,\text{app}}$ (μM) | Residual activity (%) | K_d (μM) | Binding stoichiometry (mol/mol) | |
| WT | 2.2±0.2 | 8±3 | 0.6±0.1 | 6±2 | 1.0±0.4 | 14±5 | 13±1 | 0.94±0.01 | |
| V99A | 7.3±0.5 | 3.7±1.6 | > 1000 | n.d. | > 1000 | n.d. | > 2000 | n.d. | |
| K100G | 1.9±0.1 | 13±4 | 1.7±0.2 | 4±2 | 0.13±0.03 | 8±3 | 340±90 | 1.1±0.2 | |
| R168A | 2.8±0.3 | 11±5 | 11±5 | 280±30 | 0.8±0.2 | 12±4 | 35±8 | 1.06±0.07 | |
| Y169A | 5.0±0.7 | 9.4±2.8 | 8±3 | 122±2 | > 1000 | n.d. | 1000±200 | 0.7±0.3 | |
| Y169F | 7.1±0.8 | 12±6 | 0.39±0.06 | <2 | 0.017±0.001 | <2 | 50±10 | 1.8±0.1 | |
| R170A | 0.9±0.1 | 7.9±2.6 | 0.9±0.1 | 2±3 | 57±36 | <20 | 17±1 | 2.0±0.1 | |
| R187G | 1.0±0.1 | 5.7±2.4 | 0.051±0.007 | <4 | 0.009±0.001 | <2 | 27±7 | 1.6±0.1 | |
| Y188A | 5.8±0.9 | 15±9 | 2.0±0.7 | 37±4 | 1.8±0.8 | 5±1 | 150±50 | 2.1±0.2 | |
| H189A | 3.8±0.2 | 17±4 | 0.0015±0.0006 | 44±3 | 0.0011±0.0004 | 5±3 | 1400±100 | 1.3±0.1 | |

The effects of substitutions on nucleotide binding affinity can be explained in terms of changes in nucleotide contacts (group 1 residues), whereas effects on activity may be attributable to alterations in the closure/opening of the CBS domain interface (group 2 residues). Group 1 residues include Lys100, Tyr169 and Arg170. Group 2 residues, which do not interact with bound nucleotide in the *cp*CBS-PPase structure [8], include Arg168, Arg187 and Tyr188 located at the domain interface. R168 is present within the moving RYRN loop that probably plays a pivotal role in transmitting the regulatory signal to catalytic domains. Highly conserved R187 (Table 5) interacts with the RYRN loop and has different conformations in the two *cp*CBS structures. The Y188A substitution weakens dimer interface interactions, thereby favoring an open, activated conformation of the CBS domain pair.

4.8 Possible physiological role of CBS-PPase regulation

Constitutively active soluble PPases convert chemical energy to heat by hydrolyzing high-energy phosphate compounds. This is affordable during fast logarithmic growth when considerable energy is available and PP_i hydrolysis favors biosynthetic reactions. At low energy and other stress conditions, this results in wasting energy unless PPase activity is suppressed. There is some evidence of family I and family II PPase inhibition upon phosphorylation [14, 58], but it is not known if phosphorylation occurs more widely within these families. The CBS domain pair is another type of down and up regulator of PPases. CBS-PPases respond to cell energy levels via down-regulation under low energy conditions (II and III). Notably, *mt*CBS-PPase is activated at high energy levels. Saved energy therefore becomes available for energy-consuming events in cells.

According to the complete genome sequences in the KEGG database [56], almost all organisms have at least one inorganic pyrophosphatase. All eukaryotic organisms only contain soluble family I PPases, except plants and protists, which additionally contain membrane-bound PPase. The only known exception is the eukaryotic protozoan parasite *Giardia lamblia*, which contains two CBS-PPases. Family I PPase is most common also in Archaea and Prokaryotes. Separate bacterial families usually have one type of soluble PPase, and occasionally, a membrane-bound PPase. There are, however, limited cases of different PPase composition in species within some genera. This finding is suggestive of horizontal gene transfer at some stage of evolution.

A more comprehensive analysis of Bacteria indicates that many bacterial families, such as Beta-, Gamma- and Epsilonproteobacteria, mainly contain family I PPases, whereas others, like Firmicutes, contain family II PPases. The exception is the *Vibrio* genus, which contains both family I and II PPases [160]. CBS-PPases are thus found mainly in Firmicutes, and sporadically in other Bacteria and Archaea. One strict rule is that membrane-bound PPase never coexists with common family II PPase, but only with CBS-PPase or family I PPase. This finding suggests that membrane-bound PPase is a potential PP_i user in cell stress, which is supported by the finding that overexpression of the membrane-bound enzyme in plants increases their drought and salt tolerance [161, 162]. Indirect support also comes from the observation that

Fusobacterium nucleatum has both membrane-bound PPase and CBS-PPase, whereas all other *Fusobacterium* species (*Leptotrichia buccalis*, *Sebaldella termitidis* and *Streptobacillus moniliformis*) that have lost membrane PPase contain common family II PPase (based on KEGG [56]).

5 Conclusions

Structural studies on *sc*PPase active site variants with the natural metal cofactor, Mg^{2+} , have facilitated establishing of the complete reaction scheme in terms of the three-dimensional structures of all eight reaction intermediates. Our data support two parallel routes for product release. With this complete structural scheme, the soluble family I PPase has become one of the most well-characterized phosphoryl transfer enzymes.

This study mainly focused on a recently discovered subfamily of family II PPase, CBS-PPase. Our results provide support for the theory that the CBS domain insert acts as an autoinhibitor of the enzyme. Furthermore, *mt*CBS-PPase is inhibited by AMP and ADP and activated by ATP, i.e., it works as an energy state sensor, thereby regulating biosynthetic reactions via changing the concentration of the inhibitor, PP_i . Effective adenine nucleotide concentrations are nanomolar or micromolar, which are significantly stronger than that previously reported for CBS domain proteins.

We have presented direct evidence that *cp*CBS-PPase binds the unique activator, AP_4A , which is a novel ligand for CBS domains. AP_4A is accumulated during heat shock and oxidative stress. Hypothetically, the enzyme may play a role in gas gangrene caused by *C. perfringens*, and thus act as a potential target for drug development.

The structural mechanism of CBS-PPase activation by AP_4A and inhibition by AMP was addressed by solving the structures of the regulatory part of *cp*CBS-PPase complexed with these nucleotides. We have shown that CBS-PPase forms a homodimer in which the CBS domain pairs and DRTGG domains dimerize face-to-face, similar to the DHH domains in common family II PPase. Comparison of these structures has allowed the discovery of the most significant conformational changes associated with nucleotide binding to CBS domains for the first time. The major difference between the activator and inhibitor complexes is that the latter has a much less open CBS domain interface due to displacement of a RYRN loop and a Tyr residue.

A homology model of full-length *cp*CBS-PPase was generated to provide an insight into the pathway of regulatory signal transmission from CBS domains to the catalytic part, which are located 25 Å apart. Two probable regulation mechanisms are suggested. One mechanism proposes that inhibitor-bound CBS domains arrest catalytic DHHA2 domain movement, while the other suggests a direct effect on active site residues by the regulatory part inserted within the other catalytic DHH domain. The two mechanisms may also act in parallel.

Mutational and functional studies on *mt*CBS-PPase have led to the identification of several residues responsible for nucleotide binding to CBS domains, binding specificity and direction of subsequent effects on activity (activation or inhibition).

Our group initially described the structure of the DRTGG domain, which is also found in many other proteins. In *cp*CBS-PPase, DRTGG is involved in protein

dimerization. Further studies are required to elucidate its possible role in activity regulation.

Genome sequence analyses have disclosed that membrane-bound PPases never occur together with highly active common family II PPases, but are common partners of CBS-PPases. This combination would permit CBS-PPase inhibition under stress conditions, allowing membrane-bound PPases to use PP_1 energy to drive proton and sodium pumping.

Acknowledgements

This work was carried out during the years 2004 and 2011 at the Department of Biochemistry and Food Chemistry, University of Turku. Professors Timo Lövgren and Heikki Kallio, the former and present head of the department are acknowledged for providing the working facilities.

I am grateful to my supervisors Professor Reijo Lahti and Professor Alexander Baykov for their help, guidance and support during my thesis project. I acknowledge the Professor of Biochemistry, Jyrki Heino, for the good working environment. Professor Aurelio Serrano and Ph.D. Kristian Koski are acknowledged for kindly reviewing this thesis.

I owe my thanks to Master Thesis worker Outi Heikkilä (née Lipiäinen) who was initial investigator along with her supervisor Ph.D. Anu Salminen on CBS-PPases whose steps were easy to follow. I thank Anu of her kindness, understanding, help and recommendation to joint to PPase group and encouragement to continue working in group.

Master students Tiina Myyryläinen, Miina Lehestö, Maria Salmela, Pekka Patrikainen, Annamari Paino and Pirjo Tissari Minna Peippo gave variety for daily working during their four week practising and summer work. I am grateful to got the chance to supervise them. They all were effective and pushed my work forward and motivated me to go further. I also acknowledge Master Thesis workers Juho Vuononvirta and Mari Jääskeläinen, it was interesting to work with them and both improved my skills as a supervisor.

Thanks to CBS-PPase colleague M.Sc. Joonas Jämsen, it was interesting and educational to work with you. Warm thanks go to current and former group members Ph.D. Marko Tammenkoski, Ph.D. Anssi Malinen, Ph.D. Georgiy (Gosha) Belogurov, M.Sc. Heidi Luoto and all Master students. I thank departments technical staff Jani Sointusalo, Anu Hirvensalo, Pirkko Haavisto, Timo Valta, Linnea Samila, Venla Tyystjärvi, Noora Häkkinen and secretary Satu Jasu who all have done their own work well and that way have enabled me to work effectively within research and teaching.

I thank Adrian Goldman of years in his group where I learnt basics of X-ray crystallography. I feel that still during my thesis work my second homelike working place was at Institute of Biotechnology in University of Helsinki in “territory” of Professor Adrian Goldman and Professor Pirkko Heikinheimo. People there have always been so helpful, friendly and working on same field as I and therefore it was easy to discuss with them. After several working days I have had feeling that I have been in Helsinki even though I have been working in Turku while I have got advice mainly via e-mails. Special thanks of that go to Ph.D. Lari Lehtiö who helped me a lot with Linux operating system and later on analysing final structures, and to Ph.D. Esko Oksanen for invaluable help on maintaining Unix in Macintosh, on data collection and on all other things during ESRF visits in Grenoble. Thanks to Seija Mäki with help of

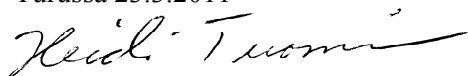
robotic crystallization, Ph.D. Tommi Kajander and others who have helped me and tested my crystals in Grenoble. I thank M.Sc. Heidi Repo for help with ThermoFluor measurements. Thanks to Professor Pirkko Heikinheimo, M.Sc. Marika Mäkinen and M.Sc. Elina Kuokkanen who all have carried my heavy dewar in train from Turku to Viikki. I kindly thank Ph.D. Tassos Papageorgiou and Professor Tiina Salminen who loaned their dewars for carrying crystals before we got our own one. Very important step was also participating to EMBO'07 – A course on Exploiting Anomalous Scattering in Macromolecular Structure Determination where my data was as a test data on phasing with autoRickshaw pipeline. There with help of Ph.D. Santosh Panjikar, thanks to him, phases of my pseudosymmetric data were solved.

Big thanks go to Professor Mark Johnson, director of ISB graduate school (National Doctoral Programme in Informational and Structural Biology), and its former and current secretaries Kaija Söderlund and Fredrik Karlsson. Moreover, I thank all organisers of all very nice winter meetings in Northern Finland and spring meetings mainly in archipelago, which have given many new contacts, much exercise and enhanced presentation skills. Furthermore, ISB gave chance to go to several other meetings by giving travel supports. I thank from my heart all my good friends and special thanks to Ph.D. Heli Elovaara, Ph.D. Riikka Ihalin, Lic.Med. Päivi Slama and Lic.Med. Päivi Helmiö, who all trusted in me, motivated and created confidence that I can do this.

Kiitos vanhemmilleni, Terttu ja Mauri Heiskaselle, jotka aina ovat valmiita huolehtimaan lapsista ja eläimistä minun matkustellessa ja juhliessa töiden merkeissä. Kiitokset perheelleni, puolisololleni Vesalle, esikoisellemme Samulle ja kuopuksellemme Jonnelle, jotka ovat antaneet väriä elämään ja vastapainoa tutkimustyölle. Kiitokset myös muille lähisukulaisille avusta arjessa ja tekemällä juhlapäivistä juhlia. Kiitokset kuuluvat myös karvaisille kavereille, aamuherättäjälle kehräävälle Somali kissa Oliverille sekä liikkeessä ja valppaana pitävälle Samojedinkoira Ellille.

This work was financially supported by The Academy of Finland, Ministry of Education and Finnish Cultural Foundation.

Turussa 23.3.2011



References

- 1 Kornberg, A. (1962) *Horizons in Biochemistry*. Academic Press, New York
- 2 Chen, J., Brevet, A., Fromant, M., Lévêque, F., Schmitter, J., Blanquet, S. and Plateau, P. (1990) Pyrophosphatase is essential for growth of *Escherichia coli*. *J Bacteriol.* **172**, 5686-5689
- 3 Lundin, M., Baltscheffsky, H. and Ronne, H. (1991) Yeast PPA2 gene encodes a mitochondrial inorganic pyrophosphatase that is essential for mitochondrial function. *J Biol Chem.* **266**, 12168-12172
- 4 Sonnewald, U. (1992) Expression of *E. coli* inorganic pyrophosphatase in transgenic plants alters photoassimilate partitioning. *Plant J.* **2**, 571-581
- 5 Ogasawara, N. (2000) Systematic function analysis of *Bacillus subtilis* genes. *Res Microbiol.* **151**, 129-134
- 6 Heinonen, J. (2001) *Biological role of inorganic pyrophosphate*. Kluwer academic publishers, Massachusetts
- 7 Moyle, J., Mitchell, R. and Mitchell, P. (1972) Proton-translocating pyrophosphatase of *Rhodospirillum rubrum*. *FEBS Lett.* **23**, 233-236
- 8 Malinen, A., Belogurov, G., Baykov, A. and Lahti, R. (2007) Na⁺-pyrophosphatase: a novel primary sodium pump. *Biochemistry.* **46**, 8872-8878
- 9 Baltscheffsky, H., Von Stedingk, L., Heldt, H. and Klingenberg, M. (1966) Inorganic pyrophosphate: formation in bacterial photophosphorylation. *Science.* **153**, 1120-1122
- 10 Shintani, T., Uchiumi, T., Yonezawa, T., Salminen, A., Baykov, A., Lahti, R. and Hachimori, A. (1998) Cloning and expression of a unique inorganic pyrophosphatase from *Bacillus subtilis*: evidence for a new family of enzymes. *FEBS Lett.* **439**, 263-266
- 11 Young, T., Kuhn, N., Wadson, A., Ward, S., Burges, D. and Cooke, G. (1998) *Bacillus subtilis* ORF yybQ encodes a manganese-dependent inorganic pyrophosphatase with distinctive properties: the first of a new class of soluble pyrophosphatase? *Microbiology.* **144** (Pt 9), 2563-2571
- 12 Cooperman, B., Baykov, A. and Lahti, R. (1992) Evolutionary conservation of the active site of soluble inorganic pyrophosphatase. *Trends Biochem Sci.* **17**, 262-266
- 13 Baykov, A., Cooperman, B., Goldman, A. and Lahti, R. (1999) Cytoplasmic inorganic pyrophosphatase. *Prog Mol Subcell Biol.* **23**, 127-150
- 14 de Graaf, B., Rudd, J., Wheeler, M., Perry, R., Bell, E., Osman, K., Franklin, F. and Franklin-Tong, V. (2006) Self-incompatibility in Papaver targets soluble inorganic pyrophosphatases in pollen. *Nature.* **444**, 490-493
- 15 Bunick, G., McKenna, G., Colton, R. and Voet, D. (1974) The x-ray structure of yeast inorganic pyrophosphatase. Crystal properties. *J Biol Chem.* **249**, 4647-4649
- 16 Harutyunyan, E., Terzyan, S. S., Voronova, A. A., Kuranova, I., Smirnova, E. A., Vainshtein, B. K., Höfne, W. and Hansen, G. (1981) An X-ray study of yeast inorganic pyrophosphatase at 3 Å resolution. *Dokl Akad Nauk SSSR.* **258**, 1481-1485
- 17 Chirgadze, N., Kuranova, I., Nevskaya, N., Teplyakov, A., Wilson, K., Strokopytov, B., Harutyunyan, E. and Hohne, W. (1991) Crystal structure of MnPi complex of yeast inorganic pyrophosphatase at 2.35 Å resolution. *Crystallography Reports.* **36**, 128-132
- 18 Harutyunyan, E., Kuranova, I., Vainshtein, B., Höhne, W., Lamzin, V., Dauter, Z., Teplyakov, A. and Wilson, K. (1996) X-ray structure of yeast inorganic pyrophosphatase complexed with manganese and phosphate. *Eur J Biochem.* **239**, 220-228
- 19 Heikinheimo, P., Lehtonen, J., Baykov, A., Lahti, R., Cooperman, B. and Goldman, A. (1996) The structural basis for pyrophosphatase catalysis. *Structure.* **4**, 1491-1508
- 20 Kankare, J., Neal, G., Salminen, T., Glumoff, T., Cooperman, B., Lahti, R. and Goldman, A. (1994) The structure of *E. coli* soluble inorganic pyrophosphatase at 2.7 Å resolution. *Protein Eng.* **7**, 823-830
- 21 Oganessyan VY, Kurilova, S., Vorobyeva, N., Nazarova, T., Popov, A., Lebedev, A., Avaeva, S. and Harutyunyan, E. (1994) X-ray crystallographic studies of recombinant inorganic pyrophosphatase from *Escherichia coli*. *FEBS Lett.* **348**, 301-304
- 22 Kankare, J., Salminen, T., Lahti, R., Cooperman, B., Baykov, A. and Goldman, A. (1996) Crystallographic identification of metal-binding sites in *Escherichia coli* inorganic pyrophosphatase. *Biochemistry.* **35**, 4670-4677
- 23 Kankare, J., Salminen, T., Lahti, R., Cooperman, B., Baykov, A. and Goldman, A. (1996) Structure of *Escherichia coli* inorganic pyrophosphatase at 2.2 Å resolution. *Acta Crystallogr D Biol Crystallogr.* **52**, 551-563
- 24 Avaeva, S., Kurilova, S., Nazarova, T., Rodina, E., Vorobyeva, N., Sklyankina, V., Grigorjeva, O., Harutyunyan, E., Oganessyan, V., Wilson,

- K., Dauter, Z., Huber, R. and Mather, T. (1997) Crystal structure of *Escherichia coli* inorganic pyrophosphatase complexed with SO_4^{2-} . Ligand-induced molecular asymmetry. *FEBS Lett.* **410**, 502-508
- 25 Harutyunyan, E., Oganessyan, V., Oganessyan, N., Avaeva, S., Nazarova, T., Vorobyeva, N., Kurilova, S., Huber, R. and Mather, T. (1997) Crystal structure of holo inorganic pyrophosphatase from *Escherichia coli* at 1.9 Å resolution. Mechanism of hydrolysis. *Biochemistry*. **36**, 7754-7760
- 26 Samygina, V., Moiseev, V., Rodina, E., Vorobyeva, N., Popov, A., Kurilova, S., Nazarova, T., Avaeva, S. and Bartunik, H. (2007) Reversible inhibition of *Escherichia coli* inorganic pyrophosphatase by fluoride: trapped catalytic intermediates in cryo-crystallographic studies. *J Mol Biol.* **366**, 1305-1317
- 27 Heikinheimo, P., Tuominen, V., Ahonen, A., Teplyakov, A., Cooperman, B., Baykov, A., Lahti, R. and Goldman, A. (2001) Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase. *Proc Natl Acad Sci U S A.* **98**, 3121-3126
- 28 Avaeva, S., Rodina, E., Vorobyeva, N., Kurilova, S., Nazarova, T., Sklyankina, V., Oganessyan, V., Samygina, V. and Harutyunyan, E. (1998) Three-dimensional structures of mutant forms of *E. coli* inorganic pyrophosphatase with Asp→Asn single substitution in positions 42, 65, 70, and 97. *Biochemistry (Mosc).* **63**, 671-684
- 29 Tuominen, V., Heikinheimo, P., Kajander, T., Torkkel, T., Hyytiä, T., Käpylä, J., Lahti, R., Cooperman, B. and Goldman, A. (1998) The R78K and D117E active-site variants of *Saccharomyces cerevisiae* soluble inorganic pyrophosphatase: structural studies and mechanistic implications. *J Mol Biol.* **284**, 1565-1580
- 30 Teplyakov, A., Obmolova, G., Wilson, K., Ishii, K., Kaji, H., Samejima, T. and Kuranova, I. (1994) Crystal structure of inorganic pyrophosphatase from *Thermus thermophilus*. *Protein Sci.* **3**, 1098-1107
- 31 Leppänen, V., Nummelin, H., Hansen, T., Lahti, R., Schäfer, G. and Goldman, A. (1999) *Sulfolobus acidocaldarius* inorganic pyrophosphatase: structure, thermostability, and effect of metal ion in an archaeal pyrophosphatase. *Protein Sci.* **8**, 1218-1231
- 32 Liu, B., Bartlam, M., Gao, R., Zhou, W., Pang, H., Liu, Y., Feng, Y. and Rao, Z. (2004) Crystal structure of the hyperthermophilic inorganic pyrophosphatase from the archaeon *Pyrococcus horikoshii*. *Biophys J.* **86**, 420-427
- 33 Tammenkoski, M., Benini, S., Magretova, N., Baykov, A. and Lahti, R. (2005) An unusual, His-dependent family I pyrophosphatase from *Mycobacterium tuberculosis*. *J Biol Chem.* **280**, 41819-41826
- 34 Moe, O. and Butler, L. (1972) Yeast inorganic pyrophosphatase. 3. Kinetics of Ca^{2+} inhibition. *J Biol Chem.* **247**, 7315-7319
- 35 Sperow, J., Moe, O., Ridlington, J. and Butler, L. (1972) Yeast inorganic pyrophosphatase VI. studies of specificity and mechanism *J Biol Chem.* **248**, 2062-2065
- 36 Baykov, A. and Shestakov, A. (1992) Two pathways of pyrophosphate hydrolysis and synthesis by yeast inorganic pyrophosphatase. *Eur J Biochem.* **206**, 463-470
- 37 Cooperman, B. (1982) The mechanism of action of yeast inorganic pyrophosphatase. *Methods Enzymol.* **87**, 526-548
- 38 Heikinheimo, P., Pohjanjoki, P., Helminen, A., Tasanen, M., Cooperman, B., Goldman, A., Baykov, A. and Lahti, R. (1996) A site-directed mutagenesis study of *Saccharomyces cerevisiae* pyrophosphatase. Functional conservation of the active site of soluble inorganic pyrophosphatases. *Eur J Biochem.* **239**, 138-143
- 39 Pohjanjoki, P., Fabrichniy, I., Kasho, V., Cooperman, B., Goldman, A., Baykov, A. and Lahti, R. (2001) Probing essential water in yeast pyrophosphatase by directed mutagenesis and fluoride inhibition measurements. *J Biol Chem.* **276**, 434-441
- 40 Belogurov, G., Fabrichniy, I., Pohjanjoki, P., Kasho, V., Lehtihuhta, E., Turkina, M., Cooperman, B., Goldman, A., Baykov, A. and Lahti, R. (2000) Catalytically important ionizations along the reaction pathway of yeast pyrophosphatase. *Biochemistry.* **39**, 13931-13938
- 41 Baykov, A., Fabrichniy, I., Pohjanjoki, P., Zyryanov, A. and Lahti, R. (2000) Fluoride effects along the reaction pathway of pyrophosphatase: evidence for a second enzyme. pyrophosphate intermediate. *Biochemistry.* **39**, 11939-11947
- 42 Kuranova, I., Polyakov, K., Smirnova, E., Höhne, W., Lamzin, V. and Meijer R. (2003) Three-Dimensional Structure of *Saccharomyces cerevisiae* Inorganic Pyrophosphatase Complexed with Cobalt and Phosphate Ions. *Crystallography Reports.* **48**, 953-958
- 43 Cooperman, B., Panackal, A., Springs, B. and Hamm, D. (1981) Divalent metal ion, inorganic phosphate, and inorganic phosphate analogue binding to yeast inorganic pyrophosphatase. *Biochemistry.* **20**, 6051-6060
- 44 Springs, B., Welsh, K. M. and Cooperman, B. S. (1981) Thermodynamics, kinetics, and mechanism in yeast inorganic pyrophosphatase catalysis of inorganic pyrophosphate: inorganic

- phosphate equilibration. *Biochemistry*. **20**, 6384-6391
- 45 Zyryanov, A., Shestakov, A., Lahti, R. and Baykov, A. (2002) Mechanism by which metal cofactors control substrate specificity in pyrophosphatase. *Biochem J*. **367**, 901-906
- 46 Zyryanov, A. B., Pohjanjoki, P., Kasho, V. N., Shestakov, A. S., Goldman, A., Lahti, R. and Baykov, A. A. (2001) The electrophilic and leaving group phosphates in the catalytic mechanism of yeast pyrophosphatase. *J Biol Chem*. **276**, 17629-17634
- 47 Kuhn, N., Wadeson, A., Ward, S. and Young, T. (2000) *Methanococcus jannaschii* ORF mj0608 codes for a class C inorganic pyrophosphatase protected by Co^{2+} or Mn^{2+} ions against fluoride inhibition. *Arch Biochem Biophys*. **379**, 292-298
- 48 Ahn, S., Milner, A., Fütterer, K., Konopka, M., Ilias, M., Young, T. and White, S. (2001) The "open" and "closed" structures of the type-C inorganic pyrophosphatases from *Bacillus subtilis* and *Streptococcus gordonii*. *J Mol Biol*. **313**, 797-811
- 49 Aravind, L. and Koonin, E. (1998) A novel family of predicted phosphoesterases includes *Drosophila* prune protein and bacterial RecJ exonuclease. *Trends Biochem Sci*. **23**, 17-19
- 50 Merckel, M., Fabrichniy, I., Salminen, A., Kalkkinen, N., Baykov, A., Lahti, R. and Goldman, A. (2001) Crystal structure of *Streptococcus mutans* pyrophosphatase: a new fold for an old mechanism. *Structure*. **9**, 289-297
- 51 Fabrichniy, I., Lehtiö, L., Salminen, A., Zyryanov, A., Baykov, A., Lahti, R. and Goldman, A. (2004) Structural studies of metal ions in family II pyrophosphatases: the requirement for a Janus ion. *Biochemistry*. **43**, 14403-14411
- 52 Rantanen, M., Lehtiö, L., Rajagopal, L., Rubens, C. and Goldman, A. (2007) Structure of the *Streptococcus agalactiae* family II inorganic pyrophosphatase at 2.80 Å resolution. *Acta Crystallogr D Biol Crystallogr*. **63**, 738-743
- 53 Fabrichniy, I., Lehtiö, L., Tammenkoski, M., Zyryanov, A., Oksanen, E., Baykov, A., Lahti, R. and Goldman, A. (2007) A trimeric site and substrate distortion in a family II inorganic pyrophosphatase. *J Biol Chem*. **282**, 1422-1431
- 54 Parfenyev, A., Salminen, A., Halonen, P., Hachimori, A., Baykov, A. and Lahti, R. (2001) Quaternary structure and metal ion requirement of family II pyrophosphatases from *Bacillus subtilis*, *Streptococcus gordonii*, and *Streptococcus mutans*. *J Biol Chem*. **276**, 24511-24518
- 55 Zyryanov, A., Tammenkoski, M., Salminen, A., Kolomiytseva, G., Fabrichniy, I., Goldman, A., Lahti, R. and Baykov, A. (2004) Site-specific effects of zinc on the activity of family II pyrophosphatase. *Biochemistry*. **43**, 14395-14402
- 56 Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. **28**, 27-30
- 57 Harding, M. (2001) Geometry of metal-ligand interactions in proteins. *Acta Crystallogr D Biol Crystallogr*. **57**, 401-411
- 58 Rajagopal, L., Clancy, A. and Rubens, C. (2003) A eukaryotic type serine/threonine kinase and phosphatase in *Streptococcus agalactiae* reversibly phosphorylate an inorganic pyrophosphatase and affect growth, cell segregation, and virulence. *J Biol Chem*. **278**, 14429-14441
- 59 Nováková, L., Bezousková, S., Pompach, P., Spidlová, P., Sasková, L., Weiser, J. and Branny, P. (2010) Identification of multiple substrates of the StkP Ser/Thr protein kinase in *Streptococcus pneumoniae*. *J Bacteriol*. **192**, 3629-3638
- 60 Finn, R., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J., Gavin, O., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E., Eddy, S. and Bateman, A. (2010) The Pfam protein families database. *Nucleic Acids Res*. **38**, D211-222
- 61 Kemp, B. (2004) Bateman domains and adenosine derivatives form a binding contract. *J Clin Invest*. **113**, 182-184
- 62 Bateman, A. (1997) The structure of a domain common to archaeobacteria and the homocystinuria disease protein. *Trends Biochem Sci*. **22**, 12-13
- 63 Ignoul, S. and Eggermont, J. (2005) CBS domains: structure, function, and pathology in human proteins. *Am J Physiol Cell Physiol*. **289**, C1369-1378
- 64 van der Heide, T. and Poolman, B. (2002) ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep*. **3**, 938-943
- 65 Hedstrom, L. (2009) IMP dehydrogenase: structure, mechanism, and inhibition. *Chem Rev*. **109**, 2903-2928
- 66 Chen, T. (2005) Structure and function of clc channels. *Annu Rev Physiol*. **67**, 809-839
- 67 Lucas, M., Encinar, J., Arribas, E., Oyenarte, I., García, I., Kortazar, D., Fernández, J., Mato, J., Martínez-Chantar, M. and Martínez-Cruz, L. (2010) Binding of S-methyl-5'-thioadenosine and S-adenosyl-L-methionine to protein MJ0100 triggers an open-to-closed conformational change in its CBS motif pair. *J Mol Biol*. **396**, 800-820
- 68 Gómez-García, I., Oyenarte, I. and Martínez-Cruz, L. (2010) The crystal structure of protein MJ1225 from *Methanocaldococcus jannaschii*

- shows strong conservation of key structural features seen in the eukaryal gamma-AMPK. *J Mol Biol.* **399**, 53-70
- 69 Hattori, M., Tanaka, Y., Fukai, S., Ishitani, R. and Nureki, O. (2007) Crystal structure of the MgtE Mg²⁺ transporter. *Nature.* **448**, 1072-1075
- 70 Mahmood, N., Biemans-Oldehinkel, E. and Poolman, B. (2009) Engineering of ion sensing by the cystathionine beta-synthase module of the ABC transporter OpuA. *J Biol Chem.* **284**, 14368-14376
- 71 Miller, M., Schwarzenbacher, R., von Delft, F., Abdubek, P., Ambing, E., Biorac, T., Brinen, L., Canaves, J., Cambell, J., Chiu, H., Dai, X., Deacon, A., DiDonato, M., Ellsiger, M., Eshagi, S., Floyd, R., Godzik, A., Grittini, C., Grzechnik, S., Hampton, E., Jaroszewski, L., Karlak, C., Klock, H., Koesema, E., Kovarik, J., Kreuzsch, A., Kuhn, P., Lesley, S., Levin, I., McMullan, D., McPhillips, T., Morse, A., Moy, K., Ouyang, J., Page, R., Quijano, K., Robb, A., Spraggon, G., Stevens, R., van den Bedem, H., Velasquez, J., Vincent, J., Wang, X., West, B., Wolf, G., Xu, Q., Hodgson, K., Wooley, J. and Wilson, I. (2004) Crystal structure of a tandem cystathionine-beta-synthase (CBS) domain protein (TM0935) from *Thermotoga maritima* at 1.87 Å resolution. *Proteins.* **57**, 213-217
- 72 Dong, A., Xu, X., Edwards, A., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., Filippova, E., Gu, J., Guthrie, J., Ignatchenko, A., Joachimiak, A., Klostermann, N., Kim, Y., Korniyenko, Y., Minor, W., Que, Q., Savchenko, A., Skarina, T., Tan, K., Yakunin, A., Yee, A., Yim, V., Zhang, R., Zheng, H., Akutsu, M., Arrowsmith, C., Avvakumov, G., Bochkarev, A., Dahlgren, L., Dhe-Paganon, S., Dimov, S., Dombrovski, L., Finerty, P. J., Flodin, S., Flores, A., Gräslund, S., Hammerström, M., Herman, M., Hong, B., Hui, R., Johansson, I., Liu, Y., Nilsson, M., Nedyalkova, L., Nordlund, P., Nyman, T., Min, J., Ouyang, H., Park, H., Qi, C., Rabeh, W., Shen, L., Shen, Y., Sukumard, D., Tempel, W., Tong, Y., Tresagues, L., Vedadi, M., Walker, J., Weigelt, J., Welin, M., Wu, H., Xiao, T., Zeng, H. and Zhu, H. (2007) In situ proteolysis for protein crystallization and structure determination. *Nat Methods.* **4**, 1019-1021
- 73 King, N., Lee, T., Sawaya, M., Cascio, D. and Yeates, T. (2008) Structures and functional implications of an AMP-binding cystathionine beta-synthase domain protein from a hyperthermophilic archaeon. *J Mol Biol.* **380**, 181-192
- 74 Ragunathan, P., Kumarevel, T., Agari, Y., Shinkai, A., Kuramitsu, S., Yokoyama, S. and Ponnuraj, K. (2008) Crystal structure of ST2348, a CBS domain protein, from hyperthermophilic archaeon *Sulfolobus tokodaii*. *Biochem Biophys Res Commun.* **375**, 124-128
- 75 Sharpe, M., Gao, C., Kendall, S., Baker, E. and Lott, J. (2008) The structure and unusual protein chemistry of hypoxic response protein 1, a latency antigen and highly expressed member of the DosR regulon in *Mycobacterium tuberculosis*. *J Mol Biol.* **383**, 822-836
- 76 Oliveriusová, J., Kery, V., Maclean, K. and Kraus, J. (2002) Deletion mutagenesis of human cystathionine beta-synthase. Impact on activity, oligomeric status, and S-adenosylmethionine regulation. *J Biol Chem.* **277**, 48386-48394
- 77 Steinberg, G. and Kemp, B. (2009) AMPK in Health and Disease. *Physiol Rev.* **89**, 1025-1078
- 78 Jentsch, T., Steinmeyer, K. and Schwarz, G. (1990) Primary structure of Torpedo marmorata chloride channel isolated by expression cloning in *Xenopus* oocytes. *Nature.* **348**, 510-514
- 79 Zhang, R., Evans, G., Rotella, F., Westbrook, E., Beno, D., Huberman, E., Joachimiak, A. and Collart, F. (1999) Characteristics and crystal structure of bacterial inosine-5'-monophosphate dehydrogenase. *Biochemistry.* **38**, 4691-4700
- 80 van der Heide, T. and Poolman, B. (2000) Osmoregulated ABC-transport system of *Lactococcus lactis* senses water stress via changes in the physical state of the membrane. *Proc Natl Acad Sci U S A.* **97**, 7102-7106
- 81 Kluijtmans, L., Boers, G., Stevens, E., Renier, W., Kraus, J., Trijbels, F., van den Heuvel, L. and Blom, H. (1996) Defective cystathionine beta-synthase regulation by S-adenosylmethionine in a partially pyridoxine responsive homocystinuria patient. *J Clin Invest.* **98**, 285-289
- 82 Maclean, K., Gaustadnes, M., Oliveriusová, J., Janosik, M., Kraus, E., Kozich, V., Kery, V., Skovby, F., Rüdiger, N., Ingerslev, J., Stabler, S., Allen, R. and Kraus, J. (2002) High homocysteine and thrombosis without connective tissue disorders are associated with a novel class of cystathionine beta-synthase (CBS) mutations. *Hum Mutat.* **19**, 641-655
- 83 Urreiziti, R., Balcells, S., Rodés, M., Vilarinho, L., Baldellou, A., Couce, M., Muñoz, C., Campistol, J., Pintó, X., Vilaseca, M. and Grinberg, D. (2003) Spectrum of CBS mutations in 16 homocystinuric patients from the Iberian Peninsula: high prevalence of T191M and absence of I278T or G307S. *Hum Mutat.* **22**, 103
- 84 Kennan, A., Aherne, A., Palfi, A., Humphries, M., McKee, A., Stitt, A., Simpson, D., Demtroder, K., Orntoft, T., Ayuso, C., Kenna, P., Farrar, G. and Humphries, P. (2002) Identification of an IMPDH1 mutation in autosomal dominant retinitis pigmentosa (RP10) revealed following comparative microarray

- analysis of transcripts derived from retinas of wild-type and Rho(-/-) mice. *Hum Mol Genet.* **11**, 547-557
- 85 Bowne, S., Sullivan, L., Blanton, S., Cepko, C., Blackshaw, S., Birch, D., Hughbanks-Wheaton, D., Heckenlively, J. and Daiger, S. (2002) Mutations in the inosine monophosphate dehydrogenase 1 gene (IMPDH1) cause the RP10 form of autosomal dominant retinitis pigmentosa. *Hum Mol Genet.* **11**, 559-568
- 86 Konrad, M., Vollmer, M., Lemmink, H., van den Heuvel, L., Jeck, N., Vargas-Poussou, R., Lakings, A., Ruf, R., Deschênes, G., Antignac, C., Guay-Woodford, L., Knoers, N., Seyberth, H., Feldmann, D. and Hildebrandt, F. (2000) Mutations in the chloride channel gene CLCNKB as a cause of classic Bartter syndrome. *J Am Soc Nephrol.* **11**, 1449-1459
- 87 Kornak, U., Kasper, D., Bösl, M., Kaiser, E., Schweizer, M., Schulz, A., Friedrich, W., Delling, G. and Jentsch, T. (2001) Loss of the CIC-7 chloride channel leads to osteopetrosis in mice and man. *Cell.* **104**, 205-215
- 88 Cleiren, E., Bénichou, O., Van Hul, E., Gram, J., Bollerslev, J., Singer, F., Beaverson, K., Aledo, A., Whyte, M., Yoneyama, T., deVernejoul, M. and Van Hul, W. (2001) Albers-Schönberg disease (autosomal dominant osteopetrosis, type II) results from mutations in the C1CN7 chloride channel gene. *Hum Mol Genet.* **10**, 2861-2867
- 89 Frattini, A., Pangrazio, A., Susani, L., Sobacchi, C., Mirolo, M., Abinun, M., Andolina, M., Flanagan, A., Horwitz, E. M., Mihci, E., Notarangelo, L. D., Ramenghi, U., Teti, A., Van Hove, J., Vujic, D., Young, T., Albertini, A., Orchard, P. J., Vezzoni, P. and Villa, A. (2003) Chloride channel C1CN7 mutations are responsible for severe recessive, dominant, and intermediate osteopetrosis. *J Bone Miner Res.* **18**, 1740-1747
- 90 Lloyd, S., Gunther, W., Pearce, S., Thomson, A., Bianchi, M., Bosio, M., Craig, I., Fisher, S., Scheinman, S., Wrong, O., Jentsch, T. and Thakker, R. (1997) Characterisation of renal chloride channel, CLCN5, mutations in hypercalcaemic nephrolithiasis (kidney stones) disorders. *Hum Mol Genet.* **6**, 1233-1239
- 91 Carballo-Trujillo, I., Garcia-Nieto, V., Moya-Angeler, F. J., Antón-Gamero, M., Loris, C., Méndez-Alvarez, S. and Claverie-Martin, F. (2003) Novel truncating mutations in the CIC-5 chloride channel gene in patients with Dent's disease. *Nephrol Dial Transplant.* **18**, 717-723
- 92 Gollob, M., Green, M., Tang, A., Gollob, T., Karibe, A., Ali Hassan, A., Ahmad, F., Lozado, R., Shah, G., Fananapazir, L., Bachinski, L., Roberts, R. and Hassan, A. (2001) Identification of a gene responsible for familial Wolff-Parkinson-White syndrome. *N Engl J Med.* **344**, 1823-1831
- 93 Blair, E., Redwood, C., Ashrafian, H., Oliveira, M., Broxholme, J., Kerr, B., Salmon, A., Ostman-Smith, I. and Watkins, H. (2001) Mutations in the gamma(2) subunit of AMP-activated protein kinase cause familial hypertrophic cardiomyopathy: evidence for the central role of energy compromise in disease pathogenesis. *Hum Mol Genet.* **10**, 1215-1220
- 94 Milan, D., Jeon, J., Looft, C., Amarger, V., Robic, A., Thelander, M., Rogel-Gaillard, C., Paul, S., Iannuccelli, N., Rask, L., Ronne, H., Lundström, K., Reinsch, N., Gellin, J., Kalm, E., Roy, P., Chardon, P. and Andersson, L. (2000) A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. *Science.* **288**, 1248-1251
- 95 Barnes, B., Marklund, S., Steiler, T., Walter, M., Hjälm, G., Amarger, V., Mahlapuu, M., Leng, Y., Johansson, C., Galuska, D., Lindgren, K., Abrink, M., Stapleton, D., Zierath, J. and Andersson, L. (2004) The 5'-AMP-activated protein kinase gamma3 isoform has a key role in carbohydrate and lipid metabolism in glycolytic skeletal muscle. *J Biol Chem.* **279**, 38441-38447
- 96 Costford, S., Kavaslar, N., Ahituv, N., Chaudhry, S., Schackwitz, W., Dent, R., Pennacchio, L., McPherson, R. and Harper, M. (2007) Gain-of-function R225W mutation in human AMPKgamma(3) causing increased glycogen and decreased triglyceride in skeletal muscle. *PLoS One.* **2**, e903
- 97 Proudfoot, M., Sanders, S., Singer, A., Zhang, R., Brown, G., Binkowski, A., Xu, L., Lukin, J., Murzin, A., Joachimiak, A., Arrowsmith, C., Edwards, A., Savchenko, A. and Yakunin, A. (2008) Biochemical and structural characterization of a novel family of cystathionine beta-synthase domain proteins fused to a Zn ribbon-like domain. *J Mol Biol.* **375**, 301-315
- 98 Schmidt-Rose, T. and Jentsch, T. (1997) Reconstitution of functional voltage-gated chloride channels from complementary fragments of CLC-1. *J Biol Chem.* **272**, 20515-20521
- 99 Estévez, R., Pusch, M., Ferrer-Costa, C., Orozco, M. and Jentsch, T. (2004) Functional and structural conservation of CBS domains from CLC chloride channels. *J Physiol.* **557**, 363-378
- 100 Bond, C. (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics.* **19**, 311-312
- 101 Meier, M., Janosik, M., Kery, V., Kraus, J. and Burkhard, P. (2001) Structure of human cystathionine beta-synthase: a unique pyridoxal 5'-phosphate-dependent heme protein. *EMBO J.* **20**, 3910-3916

- 102 Kery, V., Poneleit, L. and Kraus, J. (1998) Trypsin cleavage of human cystathionine beta-synthase into an evolutionarily conserved active core: structural and functional consequences. *Arch Biochem Biophys.* **355**, 222-232
- 103 Dutzler, R., Campbell, E., Cadene, M., Chait, B. and MacKinnon, R. (2002) X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature.* **415**, 287-294
- 104 Meyer, S. and Dutzler, R. (2006) Crystal structure of the cytoplasmic domain of the chloride channel ClC-0. *Structure.* **14**, 299-307
- 105 Meyer, S., Savaresi, S., Forster, I. and Dutzler, R. (2007) Nucleotide recognition by the cytoplasmic domain of the human chloride transporter ClC-5. *Nat Struct Mol Biol.* **14**, 60-67
- 106 Markovic, S. and Dutzler, R. (2007) The structure of the cytoplasmic domain of the chloride channel ClC-Ka reveals a conserved interaction interface. *Structure.* **15**, 715-725
- 107 Nayak, V., Zhao, K., Wyce, A., Schwartz, M., Lo, W., Berger, S. and Marmorstein, R. (2006) Structure and dimerization of the kinase domain from yeast Snf1, a member of the Snf1/AMPK protein family. *Structure.* **14**, 477-485
- 108 Polekhina, G., Gupta, A., van Denderen, B., Feil, S., Kemp, B., Stapleton, D. and Parker, M. (2005) Structural basis for glycogen recognition by AMP-activated protein kinase. *Structure.* **13**, 1453-1462
- 109 Day, P., Sharff, A., Parra, L., Cleasby, A., Williams, M., Hörer, S., Nar, H., Redemann, N., Tickle, I. and Yon, J. (2007) Structure of a CBS-domain pair from the regulatory gamma1 subunit of human AMPK in complex with AMP and ZMP. *Acta Crystallogr D Biol Crystallogr.* **63**, 587-596
- 110 Rudolph, M., Amodeo, G., Iram, S., Hong, S., Pirino, G., Carlson, M. and Tong, L. (2007) Structure of the Bateman2 domain of yeast Snf4: dimeric association and relevance for AMP binding. *Structure.* **15**, 65-74
- 111 Townley, R. and Shapiro, L. (2007) Crystal structures of the adenylate sensor from fission yeast AMP-activated protein kinase. *Science.* **315**, 1726-1729
- 112 Xiao, B., Heath, R., Saiu, P., Leiper, F., Leone, P., Jing, C., Walker, P., Haire, L., Eccleston, J., Davis, C., Martin, S., Carling, D. and Gamblin, S. (2007) Structural basis for AMP binding to mammalian AMP-activated protein kinase. *Nature.* **449**, 496-500
- 113 Jin, X., Townley, R. and Shapiro, L. (2007) Structural insight into AMPK regulation: ADP comes into play. *Structure.* **15**, 1285-1295
- 114 Hardie, D. and Hawley, S. (2001) AMP-activated protein kinase: the energy charge hypothesis revisited. *Bioessays.* **23**, 1112-1119
- 115 Scott, J., Hawley, S., Green, K., Anis, M., Stewart, G., Scullion, G., Norman, D. and Hardie, D. (2004) CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J Clin Invest.* **113**, 274-284
- 116 Bykova, E., Zhang, X., Chen, T. and Zheng, J. (2006) Large movement in the C terminus of CLC-0 chloride channel during slow gating. *Nat Struct Mol Biol.* **13**, 1115-1119
- 117 Pimkin, M., Pimkina, J. and Markham, G. (2009) A regulatory role of the Bateman domain of IMP dehydrogenase in adenylate nucleotide biosynthesis. *J Biol Chem.* **284**, 7960-7969
- 118 Hebeisen, S., Biela, A., Giese, B., Müller-Newen, G., Hidalgo, P. and Fahlke, C. (2004) The role of the carboxyl terminus in ClC chloride channel function. *J Biol Chem.* **279**, 13140-13147
- 119 Carr, G., Simmons, N. and Sayer, J. (2003) A role for CBS domain 2 in trafficking of chloride channel ClC-5. *Biochem Biophys Res Commun.* **310**, 600-605
- 120 Adams, J., Chen, Z. P., Van Denderen, B. J., Morton, C. J., Parker, M. W., Witters, L. A., Stapleton, D. and Kemp, B. E. (2004) Intrasteric control of AMPK via the gamma1 subunit AMP allosteric regulatory site. *Protein Sci.* **13**, 155-165
- 121 Bennetts, B., Rychkov, G., Ng, H., Morton, C., Stapleton, D., Parker, M. and Cromer, B. (2005) Cytoplasmic ATP-sensing domains regulate gating of skeletal muscle ClC-1 chloride channels. *J Biol Chem.* **280**, 32452-32458
- 122 Hong, S. P. and Carlson, M. (2007) Regulation of snf1 protein kinase in response to environmental stress. *J Biol Chem.* **282**, 16838-16845
- 123 Vanoye, C. and George, A. J. (2002) Functional characterization of recombinant human ClC-4 chloride channels in cultured mammalian cells. *J Physiol.* **539**, 373-383
- 124 De Angeli, A., Moran, O., Wege, S., Filleur, S., Ephritikhine, G., Thomine, S., Barbier-Brygoo, H. and Gambale, F. (2009) ATP binding to the C terminus of the *Arabidopsis thaliana* nitrate/proton antiporter, AtCLCa, regulates nitrate transport into plant vacuoles. *J Biol Chem.* **284**, 26526-26532
- 125 Biemans-Oldehinkel, E., Mahmood, N. and Poolman, B. (2006) A sensor for intracellular ionic strength. *Proc Natl Acad Sci U S A.* **103**, 10624-10629

- 126 McLean, J. E., Hamaguchi, N., Belenky, P., Mortimer, S. E., Stanton, M. and Hedstrom, L. (2004) Inosine 5'-monophosphate dehydrogenase binds nucleic acids *in vitro* and *in vivo*. *Biochem J.* **379**, 243-251
- 127 Brinsmade, S. and Escalante-Semerena, J. (2007) *In vivo* and *in vitro* analyses of single-amino acid variants of the *Salmonella enterica* phosphotransacetylase enzyme provide insights into the function of its N-terminal domain. *J Biol Chem.* **282**, 12629-12640
- 128 Campos-Bermudez, V. A., Bologna, F. P., Andreo, C. S. and Drincovich, M. F. (2010) Functional dissection of *Escherichia coli* phosphotransacetylase structural domains and analysis of key compounds involved in activity regulation. *FEBS J.* **277**, 1957-1966
- 129 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology.* **276**, 307-326
- 130 Langer, G., Cohen, S., Lamzin, V. and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc.* **3**, 1171-1179
- 131 Murshudov, G., Vagin, A. and Dodson, E. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* **53**, 240-255
- 132 Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47**, 110-119
- 133 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* **60**, 2126-2132
- 134 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.* **26**, 283-291
- 135 Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795-800
- 136 French, S. and Wilson, K. (1978) On the Treatment of Negative Intensity Observations. *Acta Crystallogr.* **A34**, 517-525
- 137 Padilla, J. and Yeates, T. (2003) A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr D Biol Crystallogr.* **59**, 1124-1130
- 138 Zwart, P., Grosse-Kunstleve, R. and Adams, P. (2005) Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 Newsletter.* **43**
- 139 Panjikar, S., Parthasarathy, V., Lamzin, V., Weiss, M. and Tucker, P. (2005) Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr D Biol Crystallogr.* **61**, 449-457
- 140 Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. and Tucker, P. A. (2009) On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr D Biol Crystallogr.* **65**, 1089-1097
- 141 Schneider, T. and Sheldrick, G. (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr.* **58**, 1772-1779
- 142 Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr D Biol Crystallogr.* **50**, 760-763
- 143 Cowtan, K. (1994) Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography. **31**, 34-38
- 144 Vagin, A. and Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Cryst.* **30**, 1022-1025
- 145 Potterton, E., Briggs, P., Turkenburg, M. and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr.* **59**, 1131-1137
- 146 Davis, I., Leaver-Fay, A., Chen, V., Block, J., Kapral, G., Wang, X., Murray, L., Arendall, W. r., Snoeyink, J., Richardson, J. and Richardson, D. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375-383
- 147 Winn, M., Isupov, M. and Murshudov, G. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr.* **57**, 122-133
- 148 Lehtonen, J., Still, D., Rantanen, V., Ekholm, J., Björklund, D., Ifthikhar, Z., Huhtala, M., Repo, S., Jussila, A., Jaakkola, J., Pentikäinen, O., Nyrönen, T., Salminen, T., Gyllenberg, M. and Johnson, M. (2004) BODIL: a molecular modeling environment for structure-function analysis and drug design. *J Comput Aided Mol Des.* **18**, 401-419
- 149 Jones, D. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* **292**, 195-202
- 150 Bryson, K., McGuffin, L., Marsden, R., Ward, J., Sodhi, J. and Jones, D. (2005) Protein structure prediction servers at University College London. *Nucleic Acids Res.* **33**, W36-38
- 151 Fiser, A. and Sali, A. (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* **374**, 461-491

- 152 Drake, H. and Daniel, S. (2004) Physiology of the thermophilic acetogen *Moorella thermoacetica*. *Res Microbiol.* **155**, 869-883
- 153 Zhou, P., Tian, F., Lv, F. and Shang, Z. (2009) Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins.* **76**, 151-163
- 154 Salminen, A., Parfenyev, A. N., Salli, K., Efimova, I. S., Magretova, N. N., Goldman, A., Baykov, A. A. and Lahti, R. (2002) Modulation of dimer stability in yeast pyrophosphatase by mutations at the subunit interface and ligand binding to the active site. *J Biol Chem.* **277**, 15465-15471
- 155 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* **372**, 774-797
- 156 Longenecker, K., Stamper, G., Hajduk, P., Fry, E., Jakob, C., Harlan, J., Edalji, R., Bartley, D., Walter, K., Solomon, L., Holzman, T., Gu, Y., Lerner, C., Beutel, B. and Stoll, V. (2005) Structure of MurF from *Streptococcus pneumoniae* co-crystallized with a small molecule inhibitor exhibits interdomain closure. *Protein Sci.* **14**, 3039-3047
- 157 Basavannacharya, C., Robertson, G., Munshi, T., Keep, N. and Bhakta, S. (2010) ATP-dependent MurE ligase in *Mycobacterium tuberculosis*: biochemical and structural characterisation. *Tuberculosis (Edinb).* **90**, 16-24
- 158 Márquez, J., Hasenbein, S., Koch, B., Fieulaine, S., Nessler, S., Russell, R., Hengstenberg, W. and Scheffzek, K. (2002) Structure of the full-length HPr kinase/phosphatase from *Staphylococcus xylosus* at 1.95 Å resolution: Mimicking the product/substrate of the phospho transfer reactions. *Proc Natl Acad Sci U S A.* **99**, 3458-3463
- 159 Consortium, U. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* **38**, D142-148
- 160 Salminen, A., Ilias, M., Belogurov, G. A., Baykov, A. A., Lahti, R. and Young, T. (2006) Two soluble pyrophosphatases in *Vibrio cholerae*: transient redundancy or enduring cooperation? *Biochemistry (Mosc).* **71**, 978-982
- 161 Park, S., Li, J., Pittman, J. K., Berkowitz, G. A., Yang, H., Undurraga, S., Morris, J., Hirschi, K. D. and Gaxiola, R. A. (2005) Up-regulation of a H⁺-pyrophosphatase (H⁺-PPase) as a strategy to engineer drought-resistant crop plants. *Proc Natl Acad Sci U S A.* **102**, 18830-18835
- 162 D'yakova, E. V., Rakitin, A. L., Kamionskaya, A. M., Baikov, A. A., Lahti, R., Ravin, N. V and Skryabin, K. G. (2006) A study of the effect of expression of the gene encoding the membrane H⁺-pyrophosphatase of *Rhodospirillum rubrum* on salt resistance of transgenic tobacco plants. *Doklady Akademii Nauk.* **409**, 844-846