FEBS Letters

FEBS PRESS
science publishing by scientists

Inorganic pyrophosphatases of Family II: two decades after their discovery

Journal:	*FEBS Letters*
Manuscript ID	FEBSL-17-1024.R2
Wiley - Manuscript type:	Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Baykov, Alexander; Lomonosov Moscow State University, Belozersky Institute of Physico-Chemical Biology Anashkin, Victor; Lomonosov Moscow State University, Belozersky Institute of Physico-Chemical Biology Salminen, Anu; University of Turku, Department of Biochemistry Lahti, Reijo; University of Turku, Dept. of Biochemistry
Keywords:	CBS domain, family II inorganic pyrophosphatase, enzyme catalysis
Abstract:	Inorganic pyrophosphatases (PPases) convert pyrophosphate to phosphate and are present in any cell types. Soluble PPases belong to three non-homologous families, of which Family II is found in approximately a quarter of prokaryotic organisms, often pathogenic. Each subunit of dimeric canonical Family II PPases is formed by two domains connected by a flexible linker, with the active site located between the domains. The enzyme requires both magnesium and a transition metal ion (manganese or cobalt) for maximal activity and is the most active (kcat ≈ 10^4 s^{-1}) among all PPase types. Catalysis by Family II PPases requires four metal ions per substrate molecule, three of which form a unique trimetal center that coordinates the nucleophilic water and converts

Inorganic pyrophosphatases of Family II: two decades after their discovery

Alexander A. Baykov[1], Victor A. Anashkin[1], Anu Salminen[2], and Reijo Lahti[2]

[1] *Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Ak. Khokhlova 1, Bldg 40, 119992 Moscow, Russian Federation*

[2] *Department of Biochemistry, University of Turku, Vatselankatu 2, FIN-20500 Turku, Finland,*

Correspondence: Alexander A. Baykov (*baykov@genebee.msu.ru*) or Reijo Lahti (*reijo.lahti@utu.fi*)

Inorganic pyrophosphatases (PPases) convert pyrophosphate to phosphate and are present in all cell types. Soluble PPases belong to three non-homologous families, of which Family II is found in approximately a quarter of prokaryotic organisms, often pathogenic. Each subunit of dimeric canonical Family II PPases is formed by two domains connected by a flexible linker, with the active site located between the domains. The enzyme requires both magnesium and a transition metal ion (manganese or cobalt) for maximal activity and is the most active (kcat ≈ 10^4 s^{-1}) among all PPase types. Catalysis by Family II PPases requires four metal ions per substrate molecule, three of which form a unique trimetal center that coordinates the nucleophilic water and converts it to a reactive hydroxide ion. A quarter of Family II PPases contain an autoinhibitory regulatory insert formed by two cystathionine β-synthase (CBS) domains and one DRTGG domain. Adenine nucleotide binding either activates or inhibits the CBS domain-containing PPase (CBS-PPase), thereby tuning its activity and, hence, pyrophosphate level, in response to changes in cell energy status (ATP/ADP ratio).

Introduction

Pyrophosphate (PP_i), a byproduct and regulator of numerous biosynthetic reactions [1], is converted to metabolizable phosphate via the action of specific constitutive enzymes—inorganic pyrophosphatases (PPases; EC 3.6.1.1). Soluble PPases convert PP_i energy into heat, as opposed to membrane-bound PPases, which employ PP_i energy to transport H^+ or Na^+ across membranes in plants and some bacteria, archaea, and protists [2,3]. Both PPase types can also catalyze the reverse reaction of PP_i synthesis from P_i, but this activity does not seem physiologically important. Soluble PPases belong to three non-homologous families, I, II and III [4–6]. Family I PPases are found in all kingdoms of life, whereas Family II and Family III PPases are found in prokaryotes. Family I PPase was discovered nearly 100 years ago and is now among the best-studied enzymes. For a long time, it was considered the sole form of soluble PPase found in living cells. However, advances in whole-genome sequencing have led to the identification of several prokaryotic organisms that lack Family I PPase genes.

In 1998, two groups independently reported the discovery of a novel type of PPase [7,8], later termed Family II PPase. These PPases belong to the "DHH phosphoesterase superfamily" [9], named because of a conserved Asp-His-His motif in its member enzymes, such as exopolyphosphatase, cyclic AMPase and the single-stranded DNA exonuclease, RecJ [9]. PPases of Families I and III, with very rare exceptions [10], consist of single-domain subunits, whereas subunits of the first discovered Family II PPases are formed by two domains, DHH and DHHA2. We will refer to these Family II PPases as "canonical", because nearly 10 years later another subfamily with a more complex organization was discovered—the CBS-PPases [11]. These latter PPases contain a regulatory insert formed by two cystathionine β-synthase (CBS) domains, usually, but not always, intercalated by a DRTGG domain (named after its conserved Asp-Arg and Thr-Gly-Gly motifs). Family III PPases were discovered in 2009 [12] and belong to the

"haloacid dehalogenase (HAD) superfamily". They are by 3−4 orders of magnitude less active and much less characterized in comparison with Families I and II.

In this review, we discuss our current knowledge of the structure, catalytic mechanism, and regulation of Family II PPases. Where appropriate, we compare them with the best-studied Family I enzymes from *Saccharomyces cerevisiae* and *Escherichia coli*. A comparison is also made between canonical and CBS-PPases. A very brief overview of canonical Family II PPases was provided previously by Kajander et al. [5].

Distribution of Family II PPases

Family II PPases mainly occur in bacterial and archaeal lineages, including many human pathogens. Each cell usually contains only one soluble PPase, but a few microbes, such as *Vibrio cholera* and *Vibrio parahaemolyticus*, contain genes for both Family I and Family II PPases. Both *V. cholera* genes encode functional PPases when expressed in *E. coli*; however, only the Family I PPase protein is detected in cultivated *V. cholera* cells [13]. A combination of Family II and putative Family III PPase genes is found in *Bacillus subtilis* and *Clostridium perfringens*.

A quarter of the nearly 4,600 completely sequenced prokaryotic genomes in the KEGG database [14] contain Family II PPase genes, as identified by BLAST search. Of the total of 865 unique Family II PPase sequences, 499 are found in the Firmicutes (367 of which are in the Bacilli and 112 in the Clostridia) and 228 belong to the Proteobacteria. Of note, the Firmicutes represent only 18.5% of the total number of completely sequenced genomes. Approximately one quarter (205) of unique Family II PPase sequences are CBS-PPases, which are also predominantly found among the Firmicutes (120 sequences), but with a different distribution— 106 sequences belong to the Clostridia and none to the Bacilli. Similarly, all 19 CBS-PPases found in the Proteobacteria belong to the Deltaproteobacteria, whereas 208 canonical Family II PPases are distributed among various classes. The majority of CBS-PPases contain both CBS domains and a DRTGG domain, whereas the enzymes from 10 organisms (*Moorella thermoacetica*, *Syntrophomonas wolfei*, *Syntrophothermus lipocalidus*, *Eggerthella lenta*, *Eggerthella sp.*, *Ethanoligenens harbinense*, *Megasphaera elsdenii*, *Gordonibacter pamelaeae*, *Coriobacteriaceae bacteria* and *Chthonomonas calidirosea*) lack the DRTGG domain. Genomes of 18 organisms contain genes for two Family II PPase types in different combinations.

All 41 unique sequences of Family II PPases found in 253 completely sequenced archaeal genomes belong to the Euryarchaeota. Among them, nine CBS-PPases are found in methane-producing organisms, whereas 32 canonical PPase sequences are also observed in other classes.

While it was initially believed that all Family II PPases belong to prokaryotic organisms, the current genome database lists ten canonical Family II PPase and two CBS-PPase sequences found in eukaryotic genomes. The corresponding organisms include seven protists (*Emiliania huxleyi*, *Fragilariopsis cylindrus*, *Giardia lamblia* [two CBS-PPase sequences], *Guillardia theta*, *Salpingoeca rosetta*, *Symbiodinium minutum* and *Thalassiosira pseudonana*) and four green algae (*Bathycoccus prasinos*, *Micromonas commoda*, *Ostreococcus lucimarinus* and *Ostreococcus tauri*).

Catalytic properties

Canonical Family II PPases, which have a k_{cat} of ~10^4 s^{-1} for PP_i hydrolysis and 10^2 for PP_i synthesis [8, 15], are the most active among all PPase types. CBS-PPases are less active by 1–3 orders of magnitude in PP_i hydrolysis because of the autoinhibitory action of the regulatory insert (see below).

All PPases are Mg^{2+}-dependent enzymes, but Family II PPases additionally require a transition metal ion, such as Mn^{2+} or Co^{2+}, for maximal activity. Different canonical Family II PPases favor Mn^{2+} [7,16] or Co^{2+} [17] as the best second cofactor, but all so far characterized CBS-PPases appear to prefer Co^{2+} [11, 18]. Interestingly, the relative effectiveness of Mn^{2+} and Co^{2+} in canonical Family II PPases may reverse in the PP_i synthesis reaction [15]. The transition

metal ions bind with affinities characteristic of metalloenzymes at one site per subunit, and their removal requires prolonged incubation with EDTA [16,19]. Noteworthy in this regard, organisms containing Family II PPases accumulate transition metal ions [20,21]. It is thus highly likely that they serve as physiological cofactors of Family II PPases, although direct evidence for this is absent. One or two additional sites that bind Mn^{2+} and Mg^{2+} with millimolar affinities have been detected in canonical Family II PPases of *Streptococcus gordonii* and *B. subtilis* in the absence of substrate [16,22]; an additional Mg^{2+} ion is brought to the enzyme as part of a $MgPP_i$ complex—the true substrate. In the cell, Mg^{2+} ions appear to occupy all sites except that containing a transition metal ion. Ca^{2+}, a strong antagonist of Mg^{2+} and inhibitor of all other PPases, can replace Mg^{2+} as activator of Mn^{2+}-bound canonical Family II PPases, conferring ~10% of their maximal activity [16].

Family II PPases appear to be the most specific among different PPases and do not hydrolyze ATP in the presence of any metal ion [16]; this is in contrast to Family I enzymes, whose activity towards ATP in the presence of transition metal ions amounts to 1% of their Mg^{2+}-supported PPase activity [23]. Family II PPases also convert tripolyphosphate [16] and imidodiphosphate (PNP) [24]—the only alternative substrates of this family—but at rates that are 10^4–10^5 times lower than those for PP_i. In contrast, Family I PPases hydrolyze tripolyphosphate only 2–10 times slower than PP_i [23,25]. The strict specificity of Family II PPases is explained by their weak interaction with the electrophilic phosphate group in substrates other than PP_i (for instance, γ-phosphate in ATP), which destabilizes the transition state [23].

Fluoride inhibits Family I PPases at micromolar concentrations [26] by replacing the nucleophilic water molecule [27,28]. The effect of fluoride on Family II enzymes strongly depends on the metal cofactor in the tight binding site. Mn/Co-enzymes are inhibited weakly ($K_i \approx 10$ mM) by fluoride [15,17], but if the transition metal is replaced by Mg^{2+}, fluoride binds 1,000-times tighter [15], achieving an affinity characteristic of Family I enzymes. C-substituted derivatives of methylene bisphosphonate, which are non-hydrolyzable PP_i analogs, bind to Family II PPases 2−3 orders of magnitude more weakly than to Family I enzymes, whereas PNP binds with similar affinity, regardless of the metal cofactor bound [29]. A distinctive feature of CBS-PPases is their sensitivity to adenine nucleotides, as considered below.

Structure

Each subunit of canonical Family II PPases is formed by two domains, whereas CBS-PPases contain four or five domains (Fig. 1). The degree of pairwise sequence identity in canonical PPases varies considerably, averaging 46% for DHH domains and 37% for DHHA2 domains. In CBS-PPases, DHH domain sequences are more highly conserved (49% average pairwise identity) than DHHA2 domain sequences (30% identity). The CBS and DRTGG domains are less conserved, with 27% identities for both. Importantly, all Family II PPases contain 10 invariable charged residues in the active site (see below). In contrast, no single residue is fully conserved in CBS1, CBS2, or DRTGG domains.

The Protein Data Bank contains 11 closely related structures of six canonical Family II PPases. The two domains, each belonging to the α/β class, are well separated in the structures. All structures except one (PDB code: 1WPM) contain divalent metal ions; three structures (2HAW, 2IW4 and 2ENX) additionally contain the pyrophosphate analog, PNP. All enzymes are dimeric in crystals, in accord with solution data [16]. Subunits contact each other through N-terminal DHH domains that form an extended β-sheet buttressed by the interaction between loops connecting two β-strands [30,31] (Fig. 2A,B). The DHHA2 domain, connected by a flexible linker, can occupy different positions. In the "closed" conformation (Fig. 2A), the active site is formed at the domain interface, allowing substrate analog binding. Domain closure and opening appear to be essential for catalytic function [32,33]. The Family II PPase from *Streptococcus aureus* adopts the closed conformation in the absence of substrate, which causes a

further induced-fit conformational change in the loop containing a conserved Arg-Lys-Lys motif [34].

Metal binding sites are found in the DHH domain, whereas the substrate recruits ligands from both domains [24] (Fig. 2C). Four Asp and two His residues form the metal-binding sites M1 and M2, and the third metal ion is bound through two Asp residues. The electrophilic phosphate group (P2) interacts through its terminal oxygen atoms with three metal ions. The DHHA2 domain provides an Arg residue that binds the leaving phosphate group (P1), and two Lys residues that interact with both P1 and P2. The metal ion that comes with PP_i also interacts with both phosphate groups and has no protein ligands. The bridging N/O atom of PNP/PP_i interacts with a His residue in the DHH domain, explaining the low affinity of Family II PPases for diphosphonates. In total, four metal ions per substrate molecule participate in catalysis. Because of the abundance of metal ions and phosphate groups, the active site of Family II PPases may be thought of as a mini-mineral, as noted previously in connection with the Family I enzymes [35].

The metal ions found in the DHH domain form a triangle that contains a water molecule nearly in the plane of the triangle and in a suitable position for nucleophilic attack on P2 [24]. The water oxygen is coordinated to all three metal ions, which is only possible if the water is converted to a highly nucleophilic hydroxide ion. This M_3-water center is a unique water-activating structure not found in other enzymes. A comparison of different structures showed that the nucleophilic water moves 1 Å away from the P2 site upon substrate binding [24]. This rearrangement requires one of the metal ion (M2) to change its coordination from five-coordinate bipyramidal to six-coordinate octahedral [24,36]. This finding explains the need for a transition metal ion, which, unlike Mg^{2+}, has a variable coordination number. In Family I PPases, the nucleophilic water molecule is coordinated by only two Mg^{2+} ions [27]; as a consequence, it is less activated (ionized), explaining the lower activity of these PPases. His residues are absent in the active site of these PPases.

In addition to its role in catalysis, the transition metal ion M2 dramatically stabilizes the catalytically competent dimeric form of canonical Family II PPases by a not yet understood mechanism [16]. In terms of ΔG, the stabilization exceeds 7 kcal/mol, making the dimer the predominant form at any feasible protein concentration. Mg^{2+} is much less efficient in dimer stabilization [16].

The complete three-dimensional (3D) structure of a CBS-PPase still remains to be elucidated. Tuominen et al. [37] have determined crystal structures of the isolated regulatory part (residues 66–306) of *C. perfringens* CBS-PPase, comprised of two CBS domains and one DRTGG domain (Fig. 3A). The regulatory part dimerizes by forming CBS1-CBS1′, CBS2-CBS2′ and DRTGG-DRTGG′ contacts. Two interacting pairs of CBS domains (Bateman modules) form a disk-like structure (CBS module), characteristic of CBS domain-containing proteins [38,39]. The structures of the CBS-PPase regulatory part contain AMP or diadenosine tetraphosphate (Ap_4A) bound to the CBS domains in different modes [37]. AMP is bound in each monomeric unit at the interface between its CBS domains (Fig. 3B), whereas one Ap_4A molecule occupies both AMP-binding sites [37]. The conformational states of the AMP- and Ap_4A-bound CBS modules are significantly different, explaining the different effects of the nucleotides on enzyme activity (see below).

Based on overall sequence similarity of the catalytic domains of CBS-PPases and canonical PPases and conservation of all active site residues, a structure of a full-size *C. perfringens* CBS-PPase was modeled [37]. The model suggests that the regulatory and catalytic parts are well separated in space, consistent with their abilities to retain functionality (catalytic activity [18] or adenine nucleotide binding [37]) when produced separately.

Mechanism of catalysis

Catalysis by PPases of Families I and II occurs through direct attack of a water molecule without formation of a phosphorylated intermediate. The reaction steps in the direction of hydrolysis include $MgPP_i$ binding and subsequent hydrolysis on the enzyme, and stepwise removal of two P_i molecules as HPO_4^{2-} and/or $MgHPO_4$ complexes (Scheme 1). The rate constants for all reaction steps in Scheme 1 in both directions (Table 1) were estimated from combined measurements of the PP_i concentration dependence of PP_i hydrolysis and P_i concentration dependencies of the rates of medium PP_i synthesis, oxygen exchange between $[^{18}O]P_i$ and $[^{16}O]H_2O$, and the amount of enzyme-bound PP_i in the equilibrated PP_i–P_i system [15]. This approach, which makes use of steady-state ^{18}O exchange instead of rapid kinetic methods, was devised [40] and further elaborated [41] in studies of Family I PPases. The ^{18}O exchange results from the dynamic reversal of PP_i synthesis in the active site, resulting in one ^{18}O atom being replaced in P_i by ^{16}O in each synthesis/hydrolysis cycle [42]. In the exchange experiments, highly O^{18}-enriched P_i was incubated with PPase in an $[^{16}O]H_2O$-containing medium in the presence of the metal cofactors. Following the incubation, the distribution of the P_i species containing from zero to four ^{16}O atoms was measured with a mass spectrometer and used to calculate the rate constants k_2, k_{-2} and k_3 that govern the exchange reaction at saturating P_i concentration [15].

As Table 1 makes clear, PP_i hydrolysis on the enzyme (k_2) proceeds faster than product removal (k_3 and k_4) with both families; however, the rate constants for all steps following substrate binding (k_2, k_3, k_4) are greater for Family II PPases. While the difference in k_2 is explainable in terms of the greater polarization of metal-bound nucleophilic water, the greater values of k_5 and (probably) k_7 in Family II PPases may result from acceleration of the conformational change that accompanies product release. PP_i-binding amino acid residues belong to different domains; thus, the enzyme–PP_i complex resembles a loaded spring, whose release is triggered by conversion of PP_i to P_i [15].

The P_i molecule containing the electrophilic P2 phosphorus is released from the active site first in both PPase types, as indicated by the fact that the partition coefficient, $P_c = k_{-2}/(k_{-2} + k_3)$, derived from the kinetics of the P_i/water oxygen exchange, is independent of medium P_i concentration [15,40,42]. The value of P_c, which characterizes the probability that bound P_i will form PP_i rather than leave the enzyme, is greater for Family I PPases (0.3–0.7) [42] than for Family II PPases (0.02–0.12) [15].

There are a number of other significant differences between Family I and Family II PPases. First, although the reaction with Family I enzymes had been initially described by Scheme 1 [41], later studies revealed an additional EM_4PP isomerization step [26] that most likely reflects expulsion of one of the two water molecules that initially bridge metals M1 and M2 [27]. The other water molecule becomes coordinated to two metal ions, making it a powerful nucleophile. This step is absent in Family II enzymes, in which the water is always coordinated to three metal ions and is ready for nucleophilic attack. Second, the acid bases used to protonate the leaving P1 group are different—His imidazole and Tyr hydroxyl in Family I and II, respectively—and they protonate different oxygen atoms. Third, Family II PPases require four metal ions for catalysis, whereas Family I PPases are active with three metal ions; binding the fourth metal results in partial inhibition [43]. Fourth, PP_i synthesis in the active site is less favored in Family II PPases than in Family I PPases. Based on the corresponding equilibrium constant (k_{-2}/k_2), approximately 1% of P_i bound to Family II PPases is converted to PP_i in the equilibrated P_i–PP_i system, compared with 20% in Family I PPases.

Perhaps, the most interesting mechanistic difference is that the hydrolysis mechanism appears to have an appreciably dissociative character in Family II PPases. By interacting with P2 oxygens, the unique trimetal center induces distortions in the substrate (PNP) by lengthening the N–P1 bond to 1.8 Å, shortening the N–P2 bond to 1.5 Å, and increasing the P–N–P angle by 10° (to 143°) [24]. The positively charged Lys–Lys–His cluster located around the bridging N may also plays a role in the substrate distortion; this cluster is also unique to Family II PPases. Although a purely dissociative mechanism involving metaphosphate formation is hardly possible

because the nucleophilic water is already bound in the transition state (Fig. 2C), the observed ground state substrate distortions suggest that breakage of the bridging O–P1 bond precedes terminal O–P2 bond formation with the nucleophile. In Family I enzymes, bond breakage and formation appear to occur simultaneously (S_N2 mechanism) [27].

Regulation

PPase activity is the key factor controlling the cellular concentration of PP_i, which is by far greater than predicted for an equilibrated $PP_i \rightleftarrows P_i$ system [1], inconsistent with high PPase activity measured in *in vitro* tests. Therefore PPase should itself be under strict control in the cell. However, surprisingly little is known about how this control of PPase activity is achieved.

CBS-PPases are the only PPases that are regulated by a well-established universal mechanism based on adenine nucleotide binding to the auxiliary CBS domains. AMP and ADP inhibit whereas ATP and, especially, Ap_nAs, the stress-associated "alarmones" containing 3–6 phosphate units [44], activate CBS-PPases several fold [11,18,44]. The activation results from favorable changes in both k_{cat} and K_m values [44]. The monoadenine nucleotides bind with micromolar affinity whereas Ap_nAs bind with nanomolar affinity [18,45]. It thus seems that the nucleotide-binding sites are always occupied *in vivo*, but the distribution of the bound nucleotides will depend on their cellular concentrations, which in turn depend on cellular energy status. Hence, CBS-PPases will consume PP_i more efficiently at high ATP concentrations when biosynthetic reactions proceed faster and produce more PP_i.

There are other factors that contribute to the CBS-PPase regulatory mechanism. Substrates interact cooperatively with the active sites in the two subunits [18], and the regulatory adenine mononucleotides bind cooperatively to the CBS domain pairs in different subunits [18]. Since the cooperativity is positive in both cases, it makes CBS-PPase activity more sensitive to changes in adenine nucleotide and pyrophosphate concentrations. Ap_nAs bind non-cooperatively but very tightly [45], which allows them to activate CBS-PPases *in vivo* under low-energy stress conditions and thereby relieve PP_i-mediated inhibition of biosynthetic reactions.

The detailed structural mechanism of CBS domain-mediated regulation remains to be determined, but it is best interpreted in terms of internal inhibition by the CBS domains, as is the case in other CBS domain-containing enzymes, such as mammalian cystathionine β-synthase [46,47] and AMP-dependent protein kinase [48]. The concept of internal inhibition is supported by several lines of evidence. First, CBS-PPases are 1−3 orders of magnitude less active than canonical Family II PPases, despite having identical active sites. Second, adenine nucleotide binding to, or mutations in, the CBS domains can either inhibit or activate CBS-PPase [11,18,49], enhancing or diminishing the internal inhibition. Finally, some mutations even reverse the effects of nucleotides on CBS-PPases from inhibition to activation [11].

The above features imply "crosstalk" among all catalytic and regulatory sites in the dimeric enzyme. A recent study [50] identified an Asn residue located in the DHH domain between the active site and subunit interface as lying at the 'crossroads' of information paths connecting all sites within the enzyme dimer. Replacement of this Asn residue with Ser was shown to abolish the kinetic cooperativity in *Desulfitobacterium hafniense* CBS-PPase and modify the effect of Ap_4A on it. This is, however, only a starting point on the way to understanding in detail the regulatory mechanism in CBS-PPases.

It has been reported that one canonical Family II PPase, that of *Streptococcus agalactiae*, is reversibly phosphorylated by endogenous Stk1/Stp1 protein kinase/phosphatase, producing effects on cell behavior [51]. However, the effect of this phosphorylation on PPase activity has not been measured. Notably, phosphorylation of the Family I PPase from the flowering plant, *Papaver rhoeas*, suppresses PPase activity and is a key event in preventing self-fertilization [52]. These are, however, the only known cases of regulation via phosphorylation in the PPase world.

Concluding remarks

Of the four protein machines devised by Nature to hydrolyze PP_i, Family II PPases appear to be the most evolved in terms of catalytic efficiency and regulation. Their superiority arises in part from their domain organization and in part from their more efficient use of metal cofactors. Coordination of the nucleophilic water to a unique trimetal center converts the water molecule to a highly reactive hydroxide. Two correlated factors appear to contribute to the ability to hold three metal ions in a small space and avoid their mutual repulsion: recruitment of a transition metal ion and use of His as its strongly binding ligand. Other parts of the catalytic machinery are common for Family II and Family I PPases, presenting a remarkable example of convergent evolution.

The two-domain structure of the catalytic part allows rapid release of the reaction product and regulation of activity by auxiliary regulatory domains in CBS-PPases. The regulatory domains appear to act as internal inhibitors; which reaction steps they decelerate is not yet known, but will be an interesting question for future studies. The regulated CBS-PPases should increase the viability of the host organism. This prediction, together with the unique structure of the catalytic core of Family II PPases and their absence in higher eukaryotes, make Family II PPases attractive targets for combatting pathogenic organisms harboring these remarkable enzymes.

Abbreviations

Ap_nA, P^1,P^n-(diadenosine-5′) polyphosphate with *n* phosphate units; CBS, cystathionine β-synthase; CBS-PPase, Family II PPase containing CBS domains; PNP, imidodiphosphate; PPase, inorganic pyrophosphatase.

Funding

This work was supported by a grant from the Russian Science Foundation (14-14-00128).

Competing interests

The Authors declare that there are no competing interests associated with the manuscript.

References

1 Heinonen, J.K. (2001) Biological role of inorganic pyrophosphate. Kluwer Academic Publishers, London, United Kingdom

2 Baykov, A.A., Malinen, A.M., Luoto, H.H. and Lahti, R. (2013) Pyrophosphate-fueled Na^+ and H^+ transport in prokaryotes. *Microbiol. Mol. Biol. Rev*. **77**, 267–276

3 Tsai, J.-Y., Kellosalo, J., Sun, Y.-J. and Goldman, A. (2014) Proton/sodium pumping pyrophosphatases: the last of the primary ion pumps. *Curr. Opin. Struct. Biol*. **27**, 38–47

4 Baykov, A.A., Cooperman, B.S., Goldman, A. and Lahti, R. (1999) Cytoplasmic inorganic pyrophosphatases. *Progr. Mol. Subcell. Biol*. **23,** 127–150

5 Kajander, T.K., Kellosalo, J. and Goldman, A. (2013) Inorganic pyrophosphatases: one substrate, three mechanisms. *FEBS Lett*. **587**, 1863–1869

6 Huang, H., Patskovsky, Y., Toro, R., Farelli, J.D., Pandya, C., Almo, S.C., Allen, K.N. and Dunaway-Mariano, D. (2011) Divergence of structure and function in the haloacid dehalogenase enzyme superfamily: *Bacteroides thetaiotaomicron* BT2127 is an inorganic pyrophosphatase. *Biochemistry* **50**, 8937−8949

7 Young, T.W., Kuhn, N.J., Wadeson, A., Ward, S., Burges, D. and Cooke, G.D. (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**, 2563−2571

8 Shintani, T., Uchiumi, T., Yonezawa, T., Salminen, A., Baykov, A.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

9 Aravind, L. and Koonin, E.V. (1998) A novel family of predicted phosphoesterases includes *Drosophila* prune protein and bacterial RecJ exonuclease. Trends Biochem. Sci. 23, 17–19.
10 Yang, Y., Ko, T.P., Chen, C.C., Huang, G., Zheng, Y., Liu, W., Wang, I., Ho, M.R., Hsu, S.T., O'Dowd, B., Huff, H.C., Huang, C.H., Docampo, R., Oldfield, E. and Guo, R.T. (2016) Structures of trypanosome vacuolar soluble pyrophosphatases: antiparasitic drug targets. *ACS Chem. Biol.* **11**, 1362−1371
11 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
12 Lee, H.S., Cho, Y., Kim, Y.-J., Lho, T.O., Cha, S.-S., Lee, J.-H. and Kang, S.G. (2009) A novel inorganic pyrophosphatase in *Thermococcus onnurineus* NA1. *FEMS Microbiol. Lett.* **300**, 68–74
13 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 (Moscow)* **71**, 1202‑1207
14 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. **45**, D353‑D361
15 Zyryanov A. B., Vener A. V., Salminen A., Goldman A., Lahti R., Baykov A. A. (2004) Rates of elementary catalytic steps for different metal forms of the family II pyrophosphatase from *Streptococcus gordonii*. *Biochemistry* **43**, 1065-1074
16 Parfenyev A. N., Salminen A., Halonen P., Hachimori A., Baykov A. A., 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
17 Kuhn, N.J., Wadeson, A., Ward, S. and Young, T.W. (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
18 Salminen A., Anashkin V.A., Lahti M., Tuominen H.K., Lahti R., Baykov A.A. (2014) Cystathionine ß-synthase (CBS) domains confer multiple forms of Mg^{2+}-dependent cooperativity to Family II pyrophosphatases. *J. Biol. Chem*. **289**, 22865–22876
19 Solís, C., Celis, H., Romero, I., Isaac-Olivé, K., Andrade, E., Eslava-Rosas, F., Monroy-Rosales, J.H. and Cuevas-Moreno, R. (2011) Metal/protein ratio determination in the *Rhodobacter capsulatus* cytoplasmic pyrophosphatase enzyme by particle induced X-ray emission. *J. Microbiol. Meth*. **84**, 272–277
20 Charney, J., Fisher, W.P. and Hegarty, C.P. (1951) Managanese as an essential element for sporulation in the genus Bacillus. *J. Bacteriol*. **62**, 145–148
21 Martin, M.E., Byers, B.R., Olson, M.O.J., Salin, M.L., Arceneaux, J.E.L. and Tolbert, C. (1986) A *Streptococcus mutans* superoxide dismutase that is active with either manganese or iron as a cofactor. *J. Biol. Chem.* **261**, 9361–9367
22 Halonen P., Tammenkoski M., Niiranen L., Huopalahti S., Parfenyev A. N., Goldman A., Baykov A. A., Lahti R. (2005) Effects of active site mutations on the metal binding affinity, catalytic competence, and stability of the family II pyrophosphatase from *Bacillus subtilis*. *Biochemistry* **44**, 2088‑2096
23 Zyryanov A. B., Shestakov A. S., Lahti R., Baykov A. A. (2002) Mechanism by which metal cofactors control substrate specificity in pyrophosphatase. *Biochem. J*. **367**, 901‑906
24 Fabrichniy I. P., Lehtiö L., Tammenkoski M., Zyryanov A. B., Oksanen E., Baykov A. A., Lahti R., Goldman A. (2007) A trimetal site and substrate distortion in a family II inorganic pyrophosphatase. *J. Biol. Chem*. **282**, 1422–1431
25 Galizzi, M., Bustamante, J.M., Fang, J., Miranda, K., Soares-Medeiros, L.C., Tarleton, R.L. and Docampo, R. (2013) Evidence for the role of vacuolar soluble pyrophosphatase and inorganic polyphosphate in *Trypanosoma cruzi* persistence. *Mol. Microbiol*. **90**, 699–715

26 Baykov, A.A., Fabrichniy, I.P., Pohjanjoki, P., Zyryanov, A.B., Lahti, R. (2000) Fluoride effects along the reaction pathway of pyrophosphatase. Evidence for a second enzyme·pyrophosphate intermediate. *Biochemistry* **39**, 11939‑11947
27 Heikinheimo P., Tuominen V., Ahonen A.-K., Teplyakov A., Cooperman B. S., Baykov A.A., Lahti R., Goldman A. (2001) Towards a quantum-mechanical description of metal assisted phosphoryl transfer in pyrophosphatase. *Proc. Natl. Acad. Sci. USA* **98**, 3121‑3126
28 Samygina, V.R, Moiseev, V.M., Rodina, E.V., Vorobyeva, N.N., Popov, A.N., Kurilova, S.A., Nazarova, T.I., Avaeva, S.M. and Bartunik, H.D. (2007) Reversible inhibition of *Escherichia coli* inorganic pyrophosphatase by fluoride: trapped catalytic intermediates in cryo-crystallographic studies. *J. Mol. Biol.* **366**, 1305‑1317
29 Zyryanov, A.B., Lahti, R. and Baykov, A.A. (2005) Inhibition of Family II pyrophosphatases by analogs of pyrophosphate and phosphate. *Biochemistry (Moscow)* **70**, 1099‑1103
30 Ahn, S., Milner, A. J., Futterer, K., Konopka, M., Ilias, M., Young, T. W., and White, S. A. (2001) The ”open” and “closed” structures of the type-C inorganic pyrophosphatases from *Bacillus subtilis* and *Streptococcus gordonii. J. Mol. Biol.* **313**, 797-811
31 Merckel, M. C., Fabrichniy, I. P., Salminen, A., Kalkkinen, N., Baykov, A. A., Lahti, R. and Goldman, A. (2001) Crystal structure of *Streptococcus mutans* pyrophosphatase: a new fold for an old mechanism. *Structure (Camb)* **9**, 289-297
32 Ilias, M. and Young, T.W. (2006) *Streptococcus gordonii* soluble inorganic pyrophosphatase: An important role for the interdomain region in enzyme activity. *Biochim. Biophys. Acta* **1764**, 1299–1306
33 Ilias, M., White, S.A. and Young, T.W. (2011) Isoleucine 259 and isoleucine 260 residues in *Streptococcus gordonii* soluble inorganic pyrophosphatase play an important role in enzyme activity. *J. Biosci. Bioeng*. **112**, 8–13
34 Gajadeera, C.S., Zhang, X., Wei, Y. and Tsodikov, O.V. (2015) Structure of inorganic pyrophosphatase from *Staphylococcus aureus* reveals conformational flexibility of the active site. *J. Struct. Biol*. **189**, 81–86
35 Cooperman, B.S., Baykov, A.A. and Lahti, R. (1992) Evolutionary conservation of the active site of soluble inorganic pyrophosphatase. *TIBS* **17**, 262–266
36 Fabrichniy, I. P., Lehtiö, L., Salminen, A., Zyryanov, A.B., Baykov, A.A., Lahti R. and Goldman, A. (2004) Structural studies on metal ions in family II pyrophosphatases: the requirements for a Janus ion. *Biochemistry* **43**, 14403–14411
37 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**, 410–413
38 Baykov A.A., Tuominen H.K. and Lahti, R. (2011) The CBS domain: a protein module with an emerging prominent role in regulation. *ACS Chem. Biol*. **6**, 1156–1163
39 Ereño-Orbea, J., Oyenarte, I. and Martìnez-Cruz, L.A. (2013) CBS domains: ligand binding sites and conformational variability. *Arch. Biochem. Biophys.* **540**, 70–81
40 Springs, B., Welsh, K.B. and Cooperman, B.S. (1981) Thermodynamics, kinetics, and mechanism in yeast inorganic pyrophosphatase catalysis of inorganic pyrophosphate: inorganic phosphate equilibration. *Biochemistry* **20**, 6384–6391
41 Baykov, A.A., Shestakov, A.S., Kasho, V.N., Vener, A.V. and Ivanov, A.H. (1990) Kinetics and thermodynamics of catalysis by the inorganic pyrophosphatase of *Escherichia coli* in both directions. *Eur. J. Biochem*. **194**, 879–887
42 Hackney, D. (1980) Theoretical analysis of distribution of [^{18}O]P_i species during exchange with water. Application to exchanges catalyzed by yeast inorganic pyrophosphatase. *J. Biol. Chem*. **255**, 5320–5328

43 Baykov, A.A. and Shestakov, A.S. (1992) Two pathways of pyrophosphate hydrolysis and synthesis by yeast inorganic pyrophosphatase. *Eur. J. Biochem*. **206**, 463–470
44 Anashkin, V. A., Salminen, A., Tuominen, H. K., Lahti, R. and Baykov, A. A. (2015) Cystathionine β-synthase (CBS) domain-containing pyrophosphatase as a target for diadenosine polyphosphates in bacteria. *J. Biol. Chem*. **290**, 27594−27603
45 Varshavsky, A. (1983) Diadenosine 5′,5′′′-P^1,P^4-tetraphosphate: a pleiotypically acting alarmone. *Cell* **34**, 711–712
46 Kery, V., Poneleit, L. and Kraus, J. (1998) Trypsin cleavage of human cystathionine β-synthase into an evolutionarily conserved active core: structural and functional consequences. *Arch. Biochem. Biophys*. **355**, 222–232
47 Ereño-Orbea, J., Majtan, T., Oyenarte, I., Kraus, J.P. and Martínez-Cruz, L.A. (2013) Structural basis of regulation and oligomerization of human cystathionine β-synthase, the central enzyme of transsulfuration. *Proc. Natl. Acad. Sci. USA* **110**, 3790−3799
48 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 γ1subunit AMP allosteric regulatory site. *Protein Sci*. **13**, 155–165
49 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*. **435**, 497–504
50 Anashkin, V.A., Salminen, A., Vorobjeva, N.N., Lahti, R. and Baykov, A.A. (2016) Asparagine residue mediates intramolecular communication in nucleotide-regulated pyrophosphatase. *Biochem. J*. **473**, 2097−2107
51 Rajagopal, L., Clancy, A. and Rubens, C. E. (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
52 de Graaf, B.H.J., Rudd, J.J., Wheeler, M.J., Perry, R.M., Bell, E.M., Osman, K., Franklin, F.C.H. and Franklin-Tong, V.E. (2006). Self-incompatibility in *Papaver* targets soluble inorganic pyrophosphatases in pollen. *Nature* **444**, 490–493

Figure legends

Figure 1. Domain topologies in Family II PPases. **(A)** Canonical PPases. (**B**) and (**C**) CBS-PPases. The regulatory CBS-(DRTGG)-CBS insert usually replaces an 8-residue sequence after approximately 65th residue from the N-terminus in the DHH domain [19].

Figure 2. The structures of dimeric *B. subtilis* PPase. **(A)** Open conformation with two Mn^{2+} ions bound (PDB code: 1K23). **(B)** Closed conformation with four Mg^{2+} ions and PNP bound (PDB code: 2HAW). In one subunit, the domains are colored as in Figure 1A, and the metal ions (purple spheres) and PNP (black sticks) are shown; the other subunit is depicted in grey. Ten polar residues forming the active site are shown as green spheres. **(C)** Presumed coordination of pyrophosphate, four metal ions (M1–M4), and nucleophilic water/OH^- in the *B. subtilis* PPase active site, as deduced from the structure in (B) [24]. The dashed line separates the structures that belong to the DHH domain (below the line) and DHHA2 domain (above the line). Transition metal ion is bound in the M2 site; the other sites contain Mg^{2+}. Thin lines show noncovalent bonds. Panels (A) and (B) were created using PyMOL (PyMOL Molecular Graphics System, version 1.5.0.4; Schrödinger, LLC).

Figure 3. The structure of the regulatory part of *C. perfringens* CBS-PPase complexed with AMP (PDB code: 3L31). **(A)** The overall fold of dimeric regulatory part. Domains are labelled in one subunit; domain coloring is as in Figure 1C: CBS1, dark blue; CBS2, blue; DRTGG, tan; the other subunit is shown in grey. Two AMP molecules are shown in red. **(B)** A close-up view of the bound AMP molecule with its protein ligands. AMP carbon and phosphorus atoms are shown in green and orange, respectively. Dashed lines represent hydrogen bonds and ionic interactions. AMP binding is stabilized by hydrophobic interactions with CBS1 Met114 and CBS2 Tyr278 and hydrogen bonds to the main-chain atoms of CBS2 Val258 and Asn280 in the linker region and side-chain atoms of CBS1 Ser116, Ser118 and Asn119 and CBS2 Thr253 (residue numbering is for the full-size enzyme). Phosphate is additionally coordinated to Lys100 of the other subunit. The figure was created using PyMOL.

$$EM_3 + MPP \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} EM_4PP \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} EM_4P_2 \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} EM_3P + MP \underset{k_{-4}}{\overset{k_4}{\rightleftarrows}} EM_3 + MP + P$$

Scheme 1. Kinetics of catalysis by Family II PPases in both directions.
"M" denotes Mn^{2+}, Co^{2+} or Mg^{2+}; "PP" is pyrophosphate; and "P" is phosphate.

Table 1. Rate constants for individual reaction steps in the catalysis by the Family II PPase from *S. gordonii* (*sg*PPase) and Family I PPase from yeast (Y-PPase)

Rate constant	Mn-*sg*PPase [15]	Co-*sg*PPase [15]	Y-PPase [26]
k_1, $\mu M^{-1}s^{-1}$	350–7000	400	
k_{-1}, s^{-1}	≥60,000	70,000	
k_2, s^{-1}	17,000	17,000	1400
k_{-2}, s^{-1}	130	175	350
k_3, s^{-1}	6500	1300	800
k_{-3}, $mM^{-1}s^{-1}$	210	36	300
k_4, s^{-1}	≥7000	2600	960–3300
k_{-4}, $mM^{-1}s^{-1}$	200–1100	100	

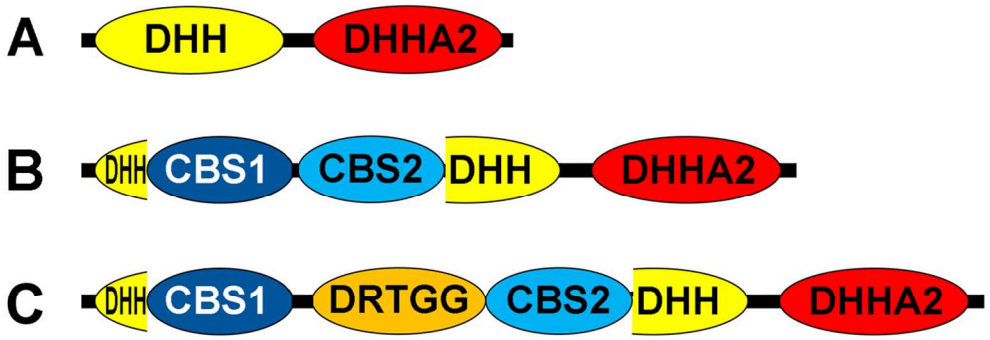

Fig. 1

Figure 1. Domain topologies in Family II PPases. **(A)** Canonical PPases. **(B)** and **(C)** CBS-PPases. The regulatory CBS-(DRTGG)-CBS insert usually replaces an 8-residue sequence after approximately 65th residue from the N-terminus in the DHH domain [19].

112x44mm (300 x 300 DPI)

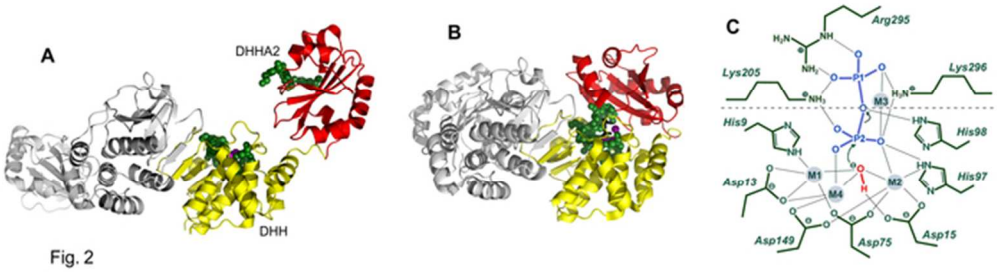

Fig. 2

Figure 2. The structures of dimeric *B. subtilis* PPase. **(A)** Open conformation with two Mn^{2+} ions bound (PDB code: 1K23). **(B)** Closed conformation with four Mg^{2+} ions and PNP bound (PDB code: 2HAW). In one subunit, the domains are colored as in Figure 1A, and the metal ions (purple spheres) and PNP (black sticks) are shown; the other subunit is depicted in grey. Ten polar residues forming the active site are shown as green spheres. **(C)** Presumed coordination of pyrophosphate, four metal ions (M1–M4), and nucleophilic water/OH^- in the *B. subtilis* PPase active site, as deduced from the structure in (B) [24]. The dashed line separates the structures that belong to the DHH domain (below the line) and DHHA2 domain (above the line). Transition metal ion is bound in the M2 site; the other sites contain Mg^{2+}. Thin lines show noncovalent bonds. Panels (A) and (B) were created using PyMOL (PyMOL Molecular Graphics System, version 1.5.0.4; Schrödinger, LLC).

59x16mm (300 x 300 DPI)

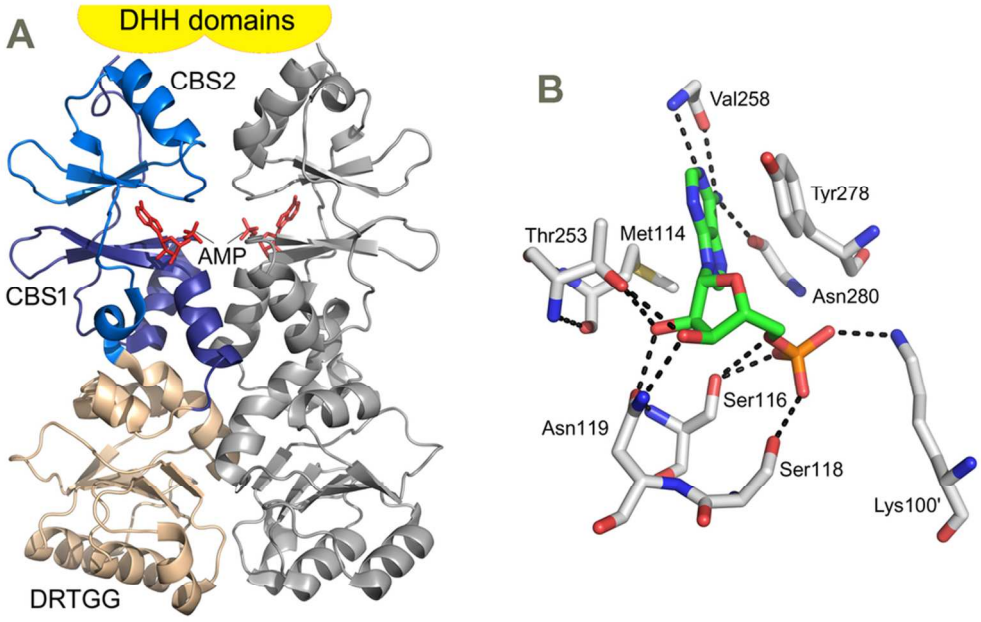

Fig. 3

Figure 3. The structure of the regulatory part of *C. perfringens* CBS-PPase complexed with AMP (PDB code: 3L31). **(A)** The overall fold of dimeric regulatory part. Domains are labelled in one subunit; domain coloring is as in Figure 1C: CBS1, dark blue; CBS2, blue; DRTGG, tan; the other subunit is shown in grey. Two AMP molecules are shown in red. **(B)** A close-up view of the bound AMP molecule with its protein ligands. AMP carbon and phosphorus atoms are shown in green and orange, respectively. Dashed lines represent hydrogen bonds and ionic interactions. AMP binding is stabilized by hydrophobic interactions with CBS1 Met114 and CBS2 Tyr278 and hydrogen bonds to the main-chain atoms of CBS2 Val258 and Asn280 in the linker region and side-chain atoms of CBS1 Ser116, Ser118 and Asn119 and CBS2 Thr253 (residue numbering is for the full-size enzyme). Phosphate is additionally coordinated to Lys100 of the other subunit. The figure was created using PyMOL.

86x60mm (300 x 300 DPI)