

Asymmetry in catalysis by *Thermotoga maritima* membrane-bound pyrophosphatase demonstrated by a non-phosphorous allosteric inhibitor

Keni Vidilaseris¹, Alexandros Kiriazis², Ainoleena Turku², Ayman Khattab³,
Niklas G. Johansson², Teppo O. Leino², Paula S. Kiuru², Gustav Boije af Gennäs²,
Seppo Meri³, Jari Yli-Kauhaluoma², Henri Xhaard², Adrian Goldman^{1,4*}

¹Research Program in Molecular and Integrative Biosciences, University of Helsinki, Helsinki, Finland.

²Drug Research Program, Division of Pharmaceutical Chemistry and Technology, Faculty of Pharmacy, University of Helsinki, Helsinki, Finland.

³Malaria Research Laboratory, Immunobiology Research Program, Department of Bacteriology and Immunology, Haartman Institute, University of Helsinki, Helsinki, Finland.

⁴School of Biomedical Sciences and Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK.

*Correspondence and requests for materials should be addressed to A.G (email:
a.goldman@leeds.ac.uk)

Abstract

Membrane-bound pyrophosphatases are homodimeric integral membrane proteins that hydrolyse pyrophosphate into orthophosphates, coupled to the active transport of protons or sodium ions across membranes. They are important in the life cycle of bacteria, archaea, plants, and protist parasites, but no homologous proteins exist in vertebrates, making them a promising drug target. Here, we report the first non-phosphorous allosteric inhibitor (K_i of $1.8 \pm 0.3 \mu\text{M}$) of the thermophilic bacterium *Thermotoga maritima* membrane-bound pyrophosphatase and its bound structure at 3.7 \AA resolution together with the substrate analogue imidodiphosphate. The unit cell contains two protein homodimers, each binding a single inhibitor dimer near the exit channel, creating a hydrophobic clamp that inhibits the movement of β -strand 1–2 during pumping, and thus preventing the hydrophobic gate from opening. This asymmetry of inhibitor binding with respect to each homodimer provide the first clear demonstration of asymmetry in the catalytic cycle of membrane-bound pyrophosphatases.

Keywords: membrane-bound pyrophosphatase, X-ray crystallography, inhibitor, drug design

Introduction

Membrane-bound pyrophosphatases (mPPases) are a family of enzymes that hydrolyse pyrophosphate into two phosphates and couple this with proton and/or sodium transport across the membrane, thus creating an electrochemical gradient. These enzymes were initially discovered in photosynthetic bacteria and plants^{1,2,3}. They were later found in protist parasites, archaea, and certain species of bacteria, but do not occur in animals and humans^{4,5}. In various organisms, these proteins are important for the survival under diverse stress situations due to energy limitation, such as osmotic stress, anoxia, mineral deficiency, low temperature, and intense light⁶. In bacteria and archaea, mPPases reside in the cell membrane⁴, while in protists, algae and plants, the proteins can also be located in acidocalcisomes, the vacuole, and/or Golgi apparatus^{7,8}.

mPPases are homodimeric. Each monomer contains 15–17 transmembrane helices with a molecular weight of 70–81 kDa. So far, there are structures of only two mPPases: four from *Thermotoga maritima* (TmPPase) and two for mung bean (*Vigna radiata*: VrPPase). TmPPase structure has been solved in the resting state (TmPPase:Ca:Mg)⁹, with two phosphates bound (TmPPase:2P_i)⁹, with the substrate-analogue imidodiphosphate (IDP) bound (TmPPase:IDP)¹⁰, and with the phosphate analogue (WO₄)-bound (TmPPase: WO₄)¹⁰. VrPPase has been solved in the IDP-bound (VrPPase:IDP)¹¹ and single phosphate-bound states (VrPPase:P_i)¹⁰. In all the structures, the mPPase is a symmetric homodimer with each monomer consisting of 16 transmembrane helices (TMHs). These helices form two concentric layers with six helices (TMH5–6, 11–12, and 15–16) forming the inner layer and the other ten (TMH 1–4, 7–10, and 13–14) forming the outer layer. Each monomer consists of four regions: a hydrolytic centre, a coupling funnel, an ion gate, and an exit channel¹² (Fig. 1a). The hydrolytic centre and coupling funnel are comprised of TMH5–6, 11–12, and 15–16 (Fig. 1b). Upon binding of substrate, the active site is closed by a long loop between TMH5 and TMH6 and opened again after hydrolysis and ion pumping^{9,10,11}. In the structure of TmPPase:IDP, the Na⁺ binds within the membrane plane at the ionic gate between TMH6 and

TMH16. Na^+ is pentacoordinated by the carboxylate groups of D243^{6,50}, E246^{6,53}, and D703^{16,46}, the O⁶ of S247^{6,54} and with the main-chain carbonyl of D243^{6,50 10} (Fig. 1c). (Residues are numbered as in ¹², shown as $\text{X}\Sigma^{a,b}$ where X is the amino acid, Σ is amino acid position in TmPPase, a is the helix number and b is the amino acid position according to a central conserved residue in each helix).

Protist parasites such as *Plasmodium falciparum*, *Toxoplasma gondii*, *Trypanosoma brucei*, and *Leishmania donovani* all possess mPPases¹³, two types (H^+ -pump K^+ -dependent and H^+ -pump K^+ -independent mPPase) for *Plasmodium*¹⁴ and one type (H^+ -pump K^+ -dependent) for the others^{15,16,17}. The mPPases, along with a V-ATPase, maintain the ionic gradient across the acidocalcisome membrane, which is necessary for acidocalcisome function, such as osmotic homeostasis upon passage of the parasite from the insect vector into the mammalian bloodstream⁸. Indeed, knockdown/knockout of mPPases causes severe reduction in polyphosphate and the loss of acidocalcisome acidity, leading to the failure of the parasites to stabilize their intracellular pH upon exposure to external basic pH¹⁸. In addition, it has been shown that mPPase is essential for virulence of *T. gondii* in a mouse model¹⁹, and both mPPases are essential genes in *P. falciparum* (Totanes, *unpubl.*). mPPases are thus promising targets for structure-based drug design because they are essential, do not exist in multicellular animals, and structures in different conformations are known.

Non-hydrolysable PP_i analogues are not viable mPPase drug candidates, as they will inhibit the many human enzymes that produce or hydrolyse PP_i, from inorganic pyrophosphatases to polymerases and ectonucleotide pyrophosphatase/phosphodiesterases²⁰. Our goal is thus the discovery of drug-like small molecular weight compounds that would specifically inhibit mPPases from protist parasites¹³. Here, we report the first non-phosphorous non-substrate analogue low micromolar inhibitor of TmPPase, *N*-[(2-amino-6-benzothiazolyl)methyl]-1*H*-indole-2-carboxamide (ATC), its preliminary structure-activity relationships, and a 3.7 Å resolution structure

of ATC bound to TmPPase. The binding site of ATC is allosteric and furthermore asymmetric with respect to the structural homodimer, providing strong evidence that the substrate binding and catalysis events process asymmetrically.

Results

Structure overview

We solved the 3.7 Å resolution of TmPPase structure in complex with IDP and ATC (TmPPase:IDP:ATC) by molecular replacement using the TmPPase:IDP structure (PDB ID: 5LZQ)¹⁰ as the search model but with the IDP removed. Molecular replacement identified four TmPPase molecules (two dimers) in the asymmetric unit. After the first round of refinement, we observed positive $F_o - F_c$ density at 3.0 σ at the hydrolytic centre (Supplementary Fig. 1a) and at the ionic gate (Supplementary Fig. 1b) in all four monomers. We built the Mg₅IDP complex into the former and a sodium ion into the latter. We also observed positive 3.0 σ $F_o - F_c$ density near the exit channels of chains A and C that could be fit with ATC molecules. Two molecules of ATC (ATC-1 and ATC-2) could be fit at the interface between A and D (Fig. 2a and Supplementary Fig. 1c), but only one (ATC-3) between B and C, as the density there was weaker (data not shown). After refinement, manual rebuilding and manual placement of the IDP and ATC into the electron density, we were able to refine it to an $R_{\text{work}}/R_{\text{free}}$ of 22.5%/26.8% with acceptable stereochemistry (Table 1). Unlike previously solved TmPPase structures^{9,10}, there are four molecules (two dimers, AB and CD) in the asymmetric unit (Supplementary Fig. 2a). The root mean squared deviation (RMSD) per C α is 0.30 Å² for the AB versus CD dimer, and 0.245–0.304 Å² when comparing individual monomers (Supplementary Table 1).

There are thus two dimer interfaces in the crystal structure: the AB dimer described before^{9,10} and ATC-mediated AD and BC dimers (Supplementary Fig. 2). Are these latter dimers physiologically relevant? It is fairly clear that the answer is no. First, the interactions are weak for a protein interface (219 Å² buried) and the proteins interact tail to tail (Supplementary Fig. 2b), placing the extracellular loops together in a manner that would not allow a bilayer to form. Second, size-exclusion chromatography-multiangle laser light scattering (SEC-MALLS) analysis showed that the addition of ATC to the TmPPase:IDP complex (158.9 ± 0.1 kDa) did not change the

oligomerisation state of the protein (162.8 ± 3.7 kDa for TmPPase:IDP:ATC complex) (Supplementary Fig. 3). ATC clearly interacts as a dimer, binding to both the E₂S and E₂S₂ complexes in solution (Table 4). We have shown this both structurally and functionally, as we describe below.

The overall structure of TmPPase:IDP:ATC complex is very similar to the structure of the TmPPase:IDP complex with an RMSD/C α of 0.46 Å and 0.42 Å to dimers AB and CD, respectively. (In what follows, all structural alignments were made with respect to monomer A unless otherwise stated.) Even though the volume of the hydrolytic pocket in TmPPase:IDP:ATC is very similar to that of TmPPase:IDP (707 \AA^3 *versus* 720 \AA^3), based on the structural alignment on TMH13–14 (residues 542–629), there are some slight clockwise movements on the cytoplasmic side of the protein helices relative to the TmPPase:IDP structure (Fig. 1b). This causes a slight movement of the Mg₅IDP complex in the same direction as the helices. In addition, at the gate of the TmPPase:IDP:ATC structure, Na⁺ is tetracoordinated by D243^{6.50}, D246^{6.53}, D703^{16.46} and S247^{6.54} (Fig. 1c). This is in contrast to TmPPase:IDP, where Na⁺ appears to be pentacoordinated, with the fifth ligand being the main-chain carbonyl oxygen of D243^{6.50}. The C=O–Na⁺ distance is 2.8 Å, *versus* 3.6 Å in the TmPPase:IDP:ATC structure. This is mostly because the sodium ion appears to be translated by about ~1.2 Å (see Methods). We believe that these changes are structurally and functionally significant and related to the binding of ATC (see *Asymmetry* below).

ATC binds in a cleft beside the exit channel

During refinement, we observed extra density near chains A and D (Supplementary Fig. 1c), into which we were able to model ATC. ATC is a rigid multiring system with only two free torsion angles, which we modeled in *trans* geometry, the most prevalent in solution (Supplementary Info. NMR spectra). We could thus, even at this low resolution (Supplementary Fig. 1c), place two molecules of ATC into the (F_o–F_c) electron density map. There is some (F_o–F_c) density in an

equivalent region near chains C and B into which a single ATC could be fitted, but the density there is too poor to fit an ATC dimer, presumably due to molecular disorder in the crystal. We thus focus on **ATC-1** and **ATC-2**, which bind to chains A and D as representing the $\text{TmPPase}_2(\text{Mg}_5\text{PP}_i)_2(\text{ATC}_2)_1$ complex, where the ATC dimer binds asymmetrically to TmPPase.

The cleft binds two ATC molecules stacked tail-to-tail through π -stacking interactions (Fig. 2). Strands β 1–2 (loop6–7), loop8–9, and loop12–13 near the exit channel of chain A form a hydrophobic cleft with loops 2–3 and 4–5 on chain D (Fig. 3a). (A similar cleft forms between chains C and B.) In terms of buried surface area; ATC-1 buries 121 \AA^2 on chain A and 92 \AA^2 on chain D, while ATC-2 buries 98 \AA^2 on chain A and 94 \AA^2 on chain D (Table 2).

ATC has a 2-aminothiazole group, which is a weak base ($\text{p}K_a 4.23 \pm 0.03^{21}$) reported in both neutral and protonated forms in the PDB and CSD. We modeled the 2-aminothiazole groups in the protonated state because both ATCs in the AD interface form ion pairs/hydrogen bonds with chain A (Table 2, Fig. 2b). Chain A D351 (O δ 2) forms a salt bridge with the 2-aminothiazole on ATC-2 while chain A G528 (C=O) forms a bifurcated H-bond to the 2-aminothiazole on ATC-1 (N19 and N20; see Supplementary Fig. 1d for atom numbering of ATC) (Table 2, Fig. 2b). In chain D, on the other hand, the chief interactions do not seem as strong: a single hydrogen bond is formed, the N157 O δ 1 hydrogen bonds N20 on ATC-2, and the remaining interaction are hydrophobic stacking of W164 on ATC-1, and the indole of ATC-2 sandwiched between W74 and W164 and stacking on Y161 (Table 2, Fig. 2c). The interactions are likely due to crystal packing as the binding of ATC to the protein did not change the protein oligomerization in solution (Supplementary Fig. 3). We therefore focus on the interactions of the ATC dimer with chain A.

Asymmetry

There is recent evidence of functional asymmetry in K^+ -dependent mPPases in the presence of excess potassium²². Here we show the first structural evidence of asymmetry. The IDP-bound state has the most symmetrical structure (the lowest rmsd values) between each monomer compared to others (Table 3, Supplementary Fig. 4a). However, the binding of ATC makes the dimeric protein asymmetric, especially in the loops that interact with the inhibitor (loops 8–9, β 1–2 and 12–13). The maximum rmsd is 1.6 Å (Table 3). Structural alignment between monomers of other states (the resting state, $2P_i$ -bound state and WO_4^- bound state) of those loops also shows more structural symmetry. In addition, there is a change in the binding mode of the Na^+ at the ionic gate (see above).

ATC binds as a dimer in chain A, holding the β 1–2 loop (Q268, K269, I270, and Q277), loop 8–9 (Q348, D351, and V352), and loop 12–13 (G528, P530, and P531) together (Fig. 3a; Supplementary Fig. 4b, c). Structural comparison of the loops that interact with ATC with the equivalent loops in TmPPase:IDP showed that the β 1–2 and 12–13 loops in chain A move the most, with an RSMD per $C\alpha$ of 1.15 and 0.80 Å respectively (Supplementary Table 2). This is due to the interactions with the ATC dimer (Fig. 3b, Table 2). In contrast, the RMSDs/ $C\alpha$ for the other loops that interact with ATC are smaller (Table 3). When all the TmPPase states are compared, it is clear that the β 1–2 loop moves closer/away relative to loop 12–13 and the dimer interface of the protein (Fig. 3c). In particular, the binding of ATC causes the β 1–2 strand of chain A to move away from the dimer interface relative to its position in TmPPase:IDP and TmPPase: P_i to roughly the same angle as in the resting and P_i structures (Fig. 3c, inset bottom view; periplasmic side) – when the exit channel must be closed. The movement of chain A β 1–2 due to the binding of the ATC dimer thus creates a hydrophobic clamp that locks the TmPPase exit channel in a closed state after substrate binding. In contrast, it is reasonable to assume that ATC does not bind to chain B because the β 1–2 strand is

moving away from the loop12-13, closer to its position in the resting state (Table 3 and Supplementary Fig. 5a). This movement appears to prevent ATC binding because it leads to the protrusion of Q268 and Q277 that occlude the ATC binding site (Supplementary Fig. 5b).

Pharmacology of ATC and analogues at TmPPase

Together with ATC, we identified four of its analogues (compound **2**, compound **3**, compound **4**, compound **5** (Fig. 4a). A set of four analogues (**6**, **7**, **8**, and **9**), three of them brominated, were then designed for structural studies. Their IC₅₀s were at least 20 times worse than the original hit ATC (Fig. 4b, c), and they did not help us solve the structure. They do, however, provide some insight into the structure-activity relationship (SAR) for this family of compounds on TmPPase.

The suggested binding modes also explain the structure-activity relationships (SARs) of the compounds presented in Figure 4. First, they highlight that the hydrogen bonding functionality of the indole ring is important; compounds **2** and **3** that lack such a functionality are inactive. In the TmPPase:IDP:ATC structure, there are two asparagine residues (N163 and Q268) near the indole nitrogen of ATC, which would explain not only the lack of activity of compounds **2** and **3**, but also the tolerance of both hydrogen bond donor (nitrogen; compound **5**) and acceptor (oxygen; compound **4**) functionalities in this position. Second, the aromatic nature of the indole ring seems important for activity: compounds **4** and **5** that include suitable hydrogen bonding functional groups but not a bulky ring structure are approximately 10-fold less active than ATC. The indole rings of ATC-**1** and ATC-**2** form π - π stacking interactions with each other, and removing the benzene ring from the structure weakens this interaction. Third, the compounds **6**, **7** and **8** with bromine substitutions are from 10-fold to 100-fold weaker binders than ATC, suggesting the importance of the unsubstituted indole ring. The bromine substitution might weaken the π - π stacking interactions observed in the structure by altering the shape and location of the π -electron cloud, but also directly clash with the loops 2–3 and 4–5; the latter is especially probable with the weakest brominated

compound **8**. Finally, benzimidazole substitution instead of an indole yields fully inactive compound **9**. This might again be due to the loss of the observed stacking of the indole rings of ATC-**1** and ATC-**2**: the benzimidazole group could be both protonated at physiological pH, leading to repulsive interactions.

ATC binds tightly to the E₂S form

As ATC is a potent inhibitor, we further characterized its effect on the rate of substrate (PP_i) hydrolysis using a range of compound concentrations (0.0 – 4.0 μ M). We performed the kinetic assay using PP_i concentration of 0 – 1714 μ M at 71°C with 2 minutes of reaction time (Fig. 4d). The simultaneous analysis of the data for all inhibitor and substrate concentrations is consistent with a model in scheme 1 with equation (1) and n=2 (Table 4), with the exception that K_{I1} is not fit (when we included it, the residuals did not improve, and its mean value is >0.5 mM; data not shown). The standard deviation of the residuals (Sy.x) for the dimer fit is 2.0 for scheme 1 where n=2, but 4.3 when n=1, and the curves clearly do not fit (data not shown). This means that ATC binds as a dimer to TmPPase. ATC clearly binds tightest to the E₂S complex ($K_{I2} = 1.8 \pm 0.3 \mu$ M) and the binding to E₂S₂ is 5-fold weaker ($10.0 \pm 1.1 \mu$ M) and no binding to E₂ (Table 4). ATC is thus an uncompetitive dimeric inhibitor with a clear preference for binding to the single-substrate bound form. After the binding of substrate to monomer A (E₂S), monomer B of TmPPase showed higher affinity to the substrate to form E₂S₂ complex as the α value is very low (0.12). However, after both monomers bind substrate (E₂S₂), the hydrolysis rate decreases by twenty-fold ($\beta = 0.05$) compared to the rate of hydrolysis of the E₂S complex ($k_p = 239.2 \pm 32.2 \mu$ mol P_i/mg/min) (Table 4).

Overall, it is clear that TmPPase shows positive cooperativity for substrate binding but negative cooperativity for hydrolysis, one where only the E₂S form binds the ATC inhibitor tightly. In the crystal structure we solved, ATC is bound to the E₂S₂ form, not to the E₂S form. This is probably

because the concentration of ATC added for crystallization was higher (1 mM) than the binding affinity of ATC to the E₂S₂ form (10.0 μM) or because the ATC binding to E₂S form can still induce the conformational change in the other monomer that allows further substrate binding (see Discussion). This is consistent with the fact that we were unable to crystallise the TmPPase:ATC complex except in the presence of IDP as a competitive inhibitor. ATC is thus the first non-phosphorus and the first allosteric inhibitor identified for mPPase.

Use of ATC in designing inhibitors of parasitic mPPases

ATC would be a potential lead as an antiparasitic agent if it were active against protozoan parasites like *Plasmodium* spp. Preliminary studies with isolated *P. falciparum* membranes showed that ATC had an IC₅₀ around 80 μM (Vidilaseris, *unpubl.*), *i.e.* at least 40-fold lower than in our test system *T. maritima* (Fig. 4a). In the *P. falciparum* survival assays, ATC did not show any anti-plasmodial activity (Supplementary Fig. 6). On the other hand, knockout of mPPases by RNAi in *T. brucei*²³ and in *P. falciparum* (Totanes, *unpubl.*) demonstrate that functional mPPase is required for survival and virulence.

To explain these findings and study the potential specificity determinants at the binding sites of ATC-1 and ATC-2 across mPPases, we retrieved the sequences of 16 mPPases from protozoan parasites (Supplementary Fig. 7). Within these sequences, the areas corresponding to the loops 2–3, 4–5, 6–7 (*i.e.* β1–2 strand), 8–9 and 12–13 showed a high degree of variation (Fig. 5a and Supplementary Fig. 7). As the binding site of ATC-1 and ATC-2 is located on these loops (6–7, 8–9 and 12–13 of chain A), it is unlikely that ATC would bind to the mPPases of pathogenic parasites as it does in the current TmPPase structure. To validate this, we built a homology model of *P. falciparum* mPPase (PfPPase; Fig. 5b). The loops 4–5, 6–7 and 12–13 of the modelled structure are notably shorter than those of TmPPase: the ATC binding site is not present, as loop 6–7 contributes about 2/3 of the interface (Table 2). This not only explains why ATC does not inhibit PfPPase, but

also supports the notion it functions asymmetrically, preventing a full catalytic cycle from occurring by binding to one chain and preventing the exit channel from opening.

Discussion

The first novel inhibitors of integral membrane pyrophosphatasases

mPPases are potential targets for anti-protozoan drug design as they do not occur in humans and other mammals, are essential in *P. falciparum* (Totanes, *unpubl.*) and render *T. gondii* non-infectious¹⁹. Bisphosphonate derivatives (PP_i analogues) can inhibit mung bean mPPase²⁴ and prevent the proliferation of *T. gondii*²⁵ and other protozoan parasites^{26,27}. However, they also inhibit human enzymes such as farnesyl pyrophosphate synthase, so they are not viable as anti-parasitic drug candidates (for a review, see ¹³). This study reports the first non-phosphorus inhibitors of mPPase, which were identified through a screening process. The best compound (ATC), binds and inhibits both the E₂S and E₂S₂ complexes, with inhibition constants K_{i2} and K_{i3} of 1.8 and 10.0 μ M, respectively (Table 4). The overall IC₅₀ is 1.7 μ M. It appears to act in an allosteric manner. This mode of inhibition has been observed for an inhibitor of soluble PPase from *Mycobacterium tuberculosis* (MtsPPase)²⁸, but to our knowledge, this is the first one observed in mPPase.

Identification of alternate inhibitory mechanisms and evidence for allostery in mPPases

ATC clearly binds beside the exit channel (Fig. 2 and Fig. 3). Similar types of inhibition occur in other membrane proteins. For instance, the voltage-dependent K_v11.1 (hERG) potassium channel that mediates the repolarizing current in the cardiac action potential is easily inhibited by a wide variety of drugs, leading to drug-induced long QT symptom and lethal arrhythmia. The molecules, such as quinolone and macrolide antibiotics, bind to hydrophobic pockets on the side of the selectivity filter^{29,30}. Recently, a structural study found inhibitors that targeted a hitherto-unknown allosteric transmembrane site in MsbA, an ATP-binding cassette (ABC) transporter. They lock MsbA in an inward facing conformation that prevents transition to the outward-facing conformation during lipopolysaccharide transport³¹. The inhibitor is wedged in the membrane-exposed pocket

formed between TMH4, TMH5, and TMH6 of each subunit. This binding uncoupled the nucleotide-binding domains (NBDs), which adopt an asymmetrical conformation.

Similar interactions to the tail-to-tail π -stacked ATC dimer that binds to TmPPase have been observed in soluble protein-inhibitor complexes as well. For example, Stornaiuolo, Kloe³² found an isoquinoline derivative that formed a π -stack of three identical molecules that bound to one binding site of acetylcholine-binding protein. A quinazoline derivative formed a π -stack of five molecules to the lysine-specific demethylase 1 (LSD1) — RE1-silencing transcription factor co-repressor (CoRESt) complex³³, and a phenylpyrrole derivative forms a head-to-tail π -stacked dimer that binds in the two-fold cavity in soluble hexameric *M. tuberculosis* PPase. Similarly to the TmPPase complex, this inhibitor appears to act uncompetitively²⁸, but the interaction clearly has 1:1 stoichiometry, with the binding site is 20 Å from the nearest pyrophosphate²⁸. The asymmetric binding of the ATC dimer to one monomer of TmPPase is similar in general.

A recent study has shown that all mesophilic mPPases show substrate inhibition with increased K_m and decreased V when two molecules of substrate are bound in the presence of 50 mM K⁺²². The binding of S to E₂S is weaker than to E₂, and k_{cat} for E₂S₂ is lower than for E₂S. They interpreted this to mean that substrate binding to one monomer causes a conformational change in the other that inhibits its interaction with the substrate. Our ATC-bound structure (Fig. 2 and Fig. 3) and inhibition data (Fig. 4 and Table 4) are similar: an ATC *dimer* binds only to one monomer in the TmPPase₂:(Mg₅PPi)₂ dimer. The second site is unavailable in the substrate-bound form, and neither monomer binds ATC before substrate binds. However, our data differ from theirs in that binding of the first substrate molecule appears to be much weaker than binding of the second as α is 0.125, while the hydrolysis rate in the E₂S₂ complex is twenty-fold slower (Table 4).

How does ATC exert its effects? After substrate binds, its binding to chain A *via* loops 6–7, 8–9 and 12–13 creates a hydrophobic clamp that locks the exit channel in the closed state and TMH12

in the down state. We posit that the mechanistic implications of these events are as follows. $\beta 1$ extends from TMH6, which contains D243^{6.50}, D246^{6.53}, and D247^{6.54}, all of which coordinate Na^+ in the gate. In the ATC-bound structure, Na^+ is coordinated by four oxygens from three Asp side chains and a serine, not five as in the TmPPase:IDP structure. Furthermore, in our current structure, D246^{6.53}, D703^{16.46}, and Na^+ move up to 1.2 Å from their positions in the TmPPase:IDP structure (Fig. 1c). This corresponds to a new state that, we believe, keeps the exit channel closed. Our current model¹⁰ posits that, during ion pumping, TMH12 moves downward by at least 2 Å⁹. Comparing the relative positions of the ATC-bound and resting states suggests that this movement of TMH12 makes loop12-13 move towards the dimeric interface of the protein and TMH13 move “up” towards the cytoplasmic side (Fig. 3d). This would induce equal conformational changes in monomer B in the TmPPase:IDP. However, ATC binding locks monomer A TMH12 in the “down” state. Because of its “down” position, loop12–13 and TMH13 can still trigger the conformational changes of monomer B into a state that binds substrate more tightly. However, further catalysis is impossible because monomer B cannot undergo a full catalytic cycle: substrate can bind, but the full motions required for catalysis are impossible. Diagnostic of these motions is the conformation that leads to binding of ATC. In agreement with the structure, the kinetic data shows that ATC dimer only binds to one of the monomers, and binds most tightly to the TmPPase₂(Mg₅PP_i) state (Table 4).

The requirement for potassium for maximal activity and for substrate inhibition is also explained by these structures. The amino acid responsible for K^+ dependency (A495^{12.46} in TmPPase) is located in TMH12, which is directly linked to the dimer interface (loop12-13 and TMH13). We suggest that, in the presence of K^+ , the downward movement of TMH12 upon substrate binding and ion pumping induces conformational changes in the monomer-monomer interface by the “away” movement of loop12-13 and upward movement of TMH13 (Fig. 4d). This is then propagated to other TMHs. In the absence of K^+ , such coupling does not occur. These changes increase affinity

and activity. They also provide a possible structural explanation for the Na^+/H^+ pumping enzymes^{22,34}, where H^+ -pumping is not inhibited by 100 mM Na^+ and 50 mM K^+ . One way this might happen is through a conformation similar to that observed here: the binding of PP_i to monomer A and pumping of, say, a proton would drive changes in the loop12-13 and TMH 13 such that the conformation of monomer B became suitable for binding PP_i and pumping Na^+ .

In conclusion, we propose the following model for TmPPase catalysis and its inhibition by ATC (Fig. 6). Substrate binding to monomer A leads to pumping in monomer A and induces a conformational change in monomer B that increases its affinity for PP_i . After PP_i binds in monomer B, PP_i hydrolysis occurs in monomer A and ion pumping in monomer B. The product of catalysis (2P_i) is released from monomer A, allowing hydrolysis and subsequent product release in monomer B. The ATC dimer can only bind when PP_i is bound to monomer A. Nonetheless, the A: PP_i form can still induce the conformational change in monomer B that allows PP_i binding. However, since the ATC binding locks monomer A in the down state, no further catalysis can occur.

Our novel structure of TmPPase with IDP and ATC has opened up new routes to drug discovery by serendipitously demonstrating that there are alternative ways to inhibit mPPase. Although the ATC-binding interface is not preserved (Fig. 5), our data nevertheless indicate that binding at the exit channel can inhibit mPPases. The work shows how extracellular loop binding regions can have significant effects on intracellular substrate binding and *vice versa*. Finally, by stabilizing asymmetric complexes, ATC provides a route into determining other steps along the reaction pathway and thus deriving a full catalytic model – one that may unify the Na^+/H^+ and Na^+ pumbers as outlined above.

Methods

Discovery of ATC and its analogues

The screening process that led to the discovery of compound **1** (ATC) will be described in more detail elsewhere. Briefly, we tested compounds using a recently developed 96-well plate assay with TmPPase as a model system³⁵. The screening collections included compounds synthesized in our laboratories. The identification of ATC as an inhibitor was followed by the test of analogues **2–5**. ATC and its analogues were initially developed for their antimicrobial, antiviral and anticancer activities as reported elsewhere³⁶.

Compound synthesis

The purity of all the compounds tested was above 95%. The synthesis of ATC (compound **1**) and compounds **2–5** has been described elsewhere³⁶. The synthesis of compounds **6–9** is described in supplemental material. For synthesis, all reactions were carried out using commercially available starting materials (Sigma-Aldrich, Schnelldorf, Germany; BioFine International Inc, Vancouver, Canada, Fluorochem, Hedfield, UK) and solvents without further purification. Column chromatography was performed with an automated Biotage high performance flash chromatography Isolera One (Uppsala, Sweden) using a 0.1-mm path length flow cell UV-detector/recorder module (fixed wavelength 254 nm). Analytical thin layer chromatography (TLC) was carried out using 0.2-mm silica gel plates (silica gel 60, F₂₅₄, Merck KGaA, Darmstadt, Germany). Nuclear magnetic resonance spectra (¹H NMR and ¹³C NMR) were recorded on a Bruker Ascend 400 (Bruker Corporation, Billerica, Massachusetts, USA). ¹H NMR at 400 MHz and ¹³C NMR at 100 MHz. High resolution mass spectra (HRMS) were measured on a Waters Synapt G2 (Waters Corporation, Milford, Massachusetts, USA) and reported for the molecular ions [M+H]⁺.

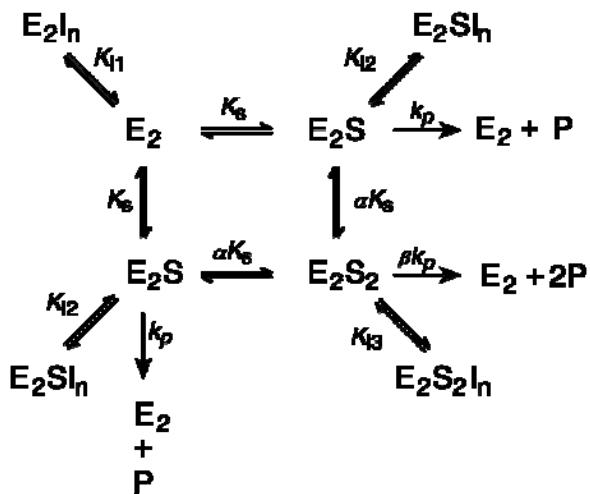
Protein expression and purification

Expression and purification of TmPPase are described elsewhere^{37,38}. In brief, His-tagged TmPPase in pRS1024 plasmid under the control of the PMA1-promoter was freshly transformed into *Saccharomyces cerevisiae* strain BJ1991. The cells were grown overnight in 250 ml of selective synthetic complete drop-out (SCD) media and then added to 740 ml of 1.5× YP media containing 2% glucose. The cells were further grown at 30 °C for 8h, collected by centrifugation (4,000 rpm, 10 min) and lysed at 4 °C using a bead beater with 0.2 mm glass beads. The membrane fraction was collected by ultracentrifugation (100,000 × g, 45 min) and the pellets were resuspended in buffer containing 50 mM MES-NaOH pH 6.5, 20% (v/v) glycerol, 50 mM KCl, 5.2 mM MgCl₂, 1.33 mM dithiothreitol (DTT), 2 µg ml⁻¹ (w/v) pepstatin-A (Sigma) and 0.334 mM PMSF (Sigma). The membranes were solubilized using the ‘hot-solve’ method³⁸ at 75 °C for 1.5 h in solubilisation buffer (50 mM MES-NaOH pH 6.5, 20% (v/v) glycerol, 5.33% (w/v) *n*-dodecyl-β-D-maltopyranoside (DDM) (Anatrace)). After centrifugation to remove denatured proteins, KCl (to a final concentration of 0.3 M) and 2 ml of Ni-NTA beads (Qiagen) were added and incubated at 40 °C for 1.5 h to each 40 ml of the solubilized protein, and then loaded into an Econo-Pac® column (Bio Rad). Then the column was washed with 2 × CV (column volume) of washing buffer (50 mM MES-NaOH pH 6.5, 20% (v/v) glycerol, 50 mM KCl, 20 mM imidazole pH 6.5, 5 mM MgCl₂, 1 mM DTT, 2 mg ml⁻¹ (w/v) pepstatin-A, 0.2 mM PMSF and 0.05% DDM (Anatrace) and eluted with 2 × CV of (50 mM MES-NaOH pH 6.5, 3.5% (v/v) glycerol, 50 mM KCl, 400 mM imidazole pH 6.5, 5 mM MgCl₂, 1 mM DTT, 2 mg ml⁻¹ (w/v) pepstatin-A, 0.2 mM PMSF and 0.5% octyl glucose neopentyl glycol (OGNPG, Anatrace), 1 mM DTT, 2 mg ml⁻¹ (w/v) pepstatin-A, 0.2 mM PMSF and 0.05% DDM (Anatrace)).

Activity measurement and kinetic analysis

TmPPase activity, compound screening and kinetic experiments were done using the molybdenum blue reaction method in a purified protein solubilised in DDM³⁵. Initial reaction rates of TmPPase were determined using PP_i as substrate and ATC as inhibitor at varying concentrations. The concentration of MgCl₂ and Na₄PP_i required to maintain 5 mM free Mg²⁺ at pH 8.0 were approximated as described previously³⁹. The reaction was done in reaction buffer (60 mM Tris-HCl pH 8.0, 5 mM free Mg²⁺, 100 mM KCl, 10 mM NaCl) at 71 °C for 2 min.

For analysis, we used a variant of Artukka, Luoto²², which assumed different kinetic behavior of each monomer active sites of the mPPase dimer, with the addition of inhibitor (Scheme 1). E₂ represents the dimeric enzyme, S represents substrate (Na₄PP_i), I represents inhibitor (ATC), n represents the amount of ATC binds to the enzyme, K_s is a microscopic Michaelis—Menten constant, and k_p is per-site maximum rates for the substrate complex. Alpha (α) is the factor relating the binding of the first substrate molecule (K_s) to the binding of the second, and β the factor relating the rate of hydrolysis of the E₂S complex (k_p) to that of the E₂S₂ complex, and K_{I1}, K_{I2}, K_{I3} the inhibition constants for binding to the E₂, E₂S and E₂S₂ complexes.



Scheme 1. Substrate binding, hydrolysis and inhibition at two active sites of a dimeric TmPPase.

For the mechanism in scheme 1 with n=2 (*i.e.* only inhibitor dimers bind), the rate equation (1)⁴⁰ is:

$$v = \frac{(2k_p + 2\beta k_p [S]/\alpha K_s)}{2(1+[I]^2/K_{I2}) + (K_s/[S])(1+[I]^2/K_{I1}) + ([S]/\alpha K_s)(1+[I]^2/K_{I3})} \quad \text{equation (1)}$$

where [S] and [I] correspond to substrate and inhibitor concentrations, respectively and the equilibrium and Michaelis-type constants are as in scheme 1. The kinetic data were fit together to obtain all parameter values and their standard errors by nonlinear regression as a function of inhibitor (0, 1.0, 2.0, and 4.0 μM) and substrate (0 – 1714 μM) concentrations using Prism 6.0 (GraphPad software).

Crystallization, structure determination and analysis

For crystallization, the purified protein was buffer exchanged to the crystallization buffer (50 mM MES-NaOH pH 6.5, 3.5% (v/v) glycerol, 50 mM KCl, 5 mM MgCl₂, 2 mM DTT and 0.5% OGNPG) on a Micro Bio-Spin 6 column (Bio-Rad) and then diluted to a concentration of 10 mg ml⁻¹. Prior to crystallization, 1 mM imidodiphosphate (IDP) and 1 mM ATC were added, the solution incubated on ice for 10 min, and centrifuged for 20 min (16,000 $\times g$, 4 °C). Crystallization trials were done at 22 °C by sitting drop vapour-diffusion method using commercial screens, MemGoldTM and PGA screenTM (Molecular Dimensions) in MRC 2-well crystallisation plates (Swisscli) with a Mosquito robot (TPP Labtech) and the drops were monitored using the minstrel DT UV imaging system (Formulatrix). Crystal hits appeared in the MemGoldTM screen under different conditions. Crystallization conditions were then optimised to get better crystals by varying pH, buffer, salt and PEG concentrations using the vapour-diffusion method in 1 μl + 1 μl (protein-mother liquor) drops in a 24-well plate at room temperature. Harvestable crystals appeared within a week and were frozen after soaking in cryo-protectant (mother liquor + 20% glycerol) or directly from mother liquor. The best diffracting crystal appeared from a solution containing 0.1 M MES pH 6.5, 0.1 M NaCl, 33% polyethylene glycol (PEG) 400, and 4% ethylene glycol.

X-ray diffraction data were collected at ESRF Grenoble (France) on the ID29 beamline at 100 K on a Pilatus 6M detector. 1400 images were collected at an oscillation angle of 0.1°. The data were merged and scaled using X-ray Detector Software (XDS)⁴¹ and the structure solved by molecular replacement with Phaser⁴² using the IDP-bound state of TmPPase structure (5LZQ)¹⁰ as the search model. Phaser found a unique solution in space group P2₁2₁2₁ with four monomers in the asymmetric unit forming two dimers. The structure was built and refined using phenix.refine⁴³ and Coot⁴⁴. The final model has R_{work}/R_{free} of 22.5%/26.8% with 94.1% of residues in the favoured and 5.7% in the allowed region of the Ramachandran plot (Table 1).

The structural interface and assembly of the protein were analysed using the PDBe PISA server (<http://www.ebi.ac.uk/pdbe/pisa/>)⁴⁵. Internal pocket analysis was done using the Splitpocket server (<http://pocket.med.wayne.edu/patch/>)^{46,47}. Structural alignments were done in PyMOL with the default setting using the monomer structure, except if stated differently. All structural figures were produced using PyMOL (<https://pymol.org/2/>).

Comparison with pathogenic mPPases

The sequences of TmPPase and 16 pathogenic mPPases were aligned using Clustal Omega software⁴⁸, and conservation analysis was conducted and visualized with the ConSurf server^{49,50} with default parameters. The initial pairwise sequence alignment of TmPPase and *P. falciparum* mPPase (VP1; PfPPase; sequence identity 42%) was produced with Discovery Studio 4.5 software⁵¹ using the Align123 algorithm. Homology modelling was carried out with MODELLER 9v14⁵² with default settings in the Discovery Studio platform. The pairwise alignment was revised during the course of model building; in total nine models were constructed in three modelling rounds. The Modeller DOPE scores did not differ significantly among these models, and a representative model was selected to minimize strain in the loops connecting transmembrane helices.

References:

1. Baltscheffsky H., von Stedingk L.-V., Heldt H.-W., Klingenberg M. Inorganic pyrophosphate: formation in bacterial photophosphorylation. *Science* **153**, 1120-1122 (1966).
2. Karlsson J. Membrane-bound potassium and magnesium ion-stimulated inorganic pyrophosphatase from roots and cotyledons of sugar beet (*Beta vulgaris L.*). *Biochim. Biophys. Acta* **399**, 356-363 (1975).
3. Rea P. A., Kim Y., Sarafian V., Poole R. J., Davies J. M., Sanders D. Vacuolar H^+ - translocating pyrophosphatases: a new category of ion translocase. *Trends Biochem. Sci.* **17**, 348-353 (1992).
4. Baykov A. A., Malinen A. M., Luoto H. H., Lahti R. Pyrophosphate-Fueled Na^+ and H^+ Transport in Prokaryotes. *Microbiol. Mol. Biol. Rev.* **77**, 267-276 (2013).
5. Serrano A., Perez-Castineira J. R., Baltscheffsky M., Baltscheffsky H. H^+ -PPases: yesterday, today and tomorrow. *IUBMB Life* **59**, 76-83 (2007).
6. Kajander T., Kellosalo J., Goldman A. Inorganic pyrophosphatases: one substrate, three mechanisms. *FEBS Lett.* **587**, 1863-1869 (2013).
7. Baltscheffsky M., Schultz A., Baltscheffsky H. H^+ -proton-pumping inorganic pyrophosphatase: a tightly membrane-bound family. *FEBS Lett.* **452**, 121-127 (1999).
8. Docampo R., Souza W. d., Miranda K., Rohloff P., Moreno S. N. J. Acidocalcisomes? conserved from bacteria to man. *Nat. Rev. Microbiol.* **3**, 251-261 (2005).
9. Kellosalo J., Kajander T., Kogan K., Pokharel K., Goldman A. The structure and catalytic cycle of a sodium-pumping pyrophosphatase. *Science* **337**, 473-476 (2012).
10. Li K.-M., *et al.* Membrane pyrophosphatases from *Thermotoga maritima* and *Vigna radiata* suggest a conserved coupling mechanism. *Nat. Comm.* **7**, 13596 (2016).
11. Lin S.-M., *et al.* Crystal structure of a membrane-embedded H^+ -translocating pyrophosphatase. *Nature* **484**, 399-403 (2012).
12. Tsai J.-Y., Kellosalo J., Sun Y.-J., Goldman A. Proton/sodium pumping pyrophosphatases: the last of the primary ion pumps. *Curr. Opin. Struct. Biol.* **27**, 38-47 (2014).
13. Shah N. R., Vidilaseris K., Xhaard H., Goldman A. Integral membrane pyrophosphatases: a novel drug target for human pathogens? *AIMS Biophysics* **3**, 171-194 (2016).
14. McIntosh M. T., Drozdowicz Y. M., Laroiya K., Rea P. A., Vaidya A. B. Two classes of plant-like vacuolar-type H^+ -pyrophosphatases in malaria parasites. *Mol. Biochem. Parasitol.* **114**, 183-195 (2001).
15. Drozdowicz Y. M., *et al.* Isolation and characterization of TgVP1, a type I vacuolar H^+ - translocating pyrophosphatase from *Toxoplasma gondii*. The dynamics of its subcellular

localization and the cellular effects of a diphosphonate inhibitor. *J. Biol. Chem.* **278**, 1075-1085 (2003).

16. Rodrigues C. O., Scott D. A., Docampo R. Presence of a vacuolar H^+ -pyrophosphatase in promastigotes of *Leishmania donovani* and its localization to a different compartment from the vacuolar H^+ -ATPase. *Biochem. J.* **340** (Pt 3), 759-766 (1999).
17. Rodrigues C. O., Scott D. A., Docampo R. Characterization of a vacuolar pyrophosphatase in *Trypanosoma brucei* and its localization to acidocalcisomes. *Mol. Cell. Biol.* **19**, 7712-7723 (1999).
18. Lemercier G., et al. A pyrophosphatase regulating polyphosphate metabolism in acidocalcisomes is essential for *Trypanosoma brucei* virulence in mice. *J. Biol. Chem.* **279**, 3420-3425 (2004).
19. Liu J., et al. A vacuolar- H^+ -pyrophosphatase (TgVP1) is required for microneme secretion, host cell invasion, and extracellular survival of *Toxoplasma gondii*. *Mol. Microbiol.* **93**, 698-712 (2014).
20. Zhang Y., et al. Structural Isosteres of Phosphate Groups in the Protein Data Bank. *J. Chem. Inf. Model.* **57**, 499-516 (2017).
21. Forlani L., Maria P. D. Tautomerism of aminothiazoles. pKBH⁺ Values of 2-aminothiazoles and of some model imines. *J. Chem. Soc., Perkin Trans. 2* **0**, 535-537 (1982).
22. Artukka E., Luoto H. H., Baykov A. A., Lahti R., Malinen A. M. Role of the potassium/lysine cationic center in catalysis and functional asymmetry in membrane-bound pyrophosphatases. *Biochem. J.* **475**, 1141-1158 (2018).
23. Lemercier G., et al. A vacuolar-type H^+ -pyrophosphatase governs maintenance of functional acidocalcisomes and growth of the insect and mammalian forms of *Trypanosoma brucei*. *J. Biol. Chem.* **277**, 37369-37376 (2002).
24. Gordon-Weeks R., Parmar S., Davies T. G., Leigh R. A. Structural aspects of the effectiveness of bisphosphonates as competitive inhibitors of the plant vacuolar proton-pumping pyrophosphatase. *Biochem. J.* **337** (Pt 3), 373-377 (1999).
25. Rodrigues C. O., et al. Vacuolar proton pyrophosphatase activity and pyrophosphate (PPi) in *Toxoplasma gondii* as possible chemotherapeutic targets. *Biochem. J.* **349** Pt 3, 737-745 (2000).
26. Martin M. B., et al. Bisphosphonates inhibit the growth of *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania donovani*, *Toxoplasma gondii*, and *Plasmodium falciparum*: a potential route to chemotherapy. *J. Med. Chem.* **44**, 909-916 (2001).
27. Martin M. B., et al. Activity of bisphosphonates against *Trypanosoma brucei rhodesiense*. *J. Med. Chem.* **45**, 2904-2914 (2002).

28. Pang A. H., Garzan A., Larsen M. J., McQuade T. J., Garneau-Tsodikova S., Tsodikov O. V. Discovery of allosteric and selective inhibitors of inorganic pyrophosphatase from *Mycobacterium tuberculosis*. *ACS Chem. Biol.* **11**, 3084-3092 (2016).
29. Bischoff U., Schmidt C., Netzer R., Pongs O. Effects of fluoroquinolones on HERG currents. *Eur. J. Pharmacol.* **406**, 341-343 (2000).
30. Wang W., MacKinnon R. Cryo-EM structure of the open human ether-à-go-go related K⁺ channel hERG. *Cell* **169**, 422-430.e410 (2017).
31. Ho H., *et al.* Structural basis for dual-mode inhibition of the ABC transporter MsbA. *Nature* **557**, 196-201 (2018).
32. Stornaiuolo M., *et al.* Assembly of a π - π stack of ligands in the binding site of an acetylcholine-binding protein. *Nat. Comm.* **4**, 1875 (2013).
33. Speranzini V., *et al.* Polymyxins and quinazolines are LSD1/KDM1A inhibitors with unusual structural features. *Sci. Adv.* **2**, (2016).
34. Luoto H. H., Baykov A. A., Lahti R., Malinen A. M. Membrane-integral pyrophosphatase subfamily capable of translocating both Na⁺ and H⁺. *Proc. Natl. Acad. Sci. USA* **110**, 1255-1260 (2013).
35. Vidilaseris K., Kellosalo J., Goldman A. A high-throughput method for orthophosphate determination of thermostable membrane-bound pyrophosphatase activity. *Anal. Methods* **10**, 646-651 (2018).
36. Montalvão S., Leino T. O., Kiuru P. S., Lillsunde K.-E., Yli-Kauhaluoma J., Tammela P. Synthesis and Biological Evaluation of 2-Aminobenzothiazole and Benzimidazole Analogs Based on the Clathrodin Structure. *Arch. Pharm.* **349**, 137-149 (2016).
37. Kellosalo J., Kajander T., Palmgren M. G., Lopez-Marques R. L., Goldman A. Heterologous expression and purification of membrane-bound pyrophosphatases. *Protein Expr. Purif.* **79**, 25-34 (2011).
38. López-Marqués R. L., *et al.* Large-scale purification of the proton pumping pyrophosphatase from *Thermotoga maritima*: A “Hot-Solve” method for isolation of recombinant thermophilic membrane proteins. *Biochim. Biophys. Acta* **1716**, 69-76 (2005).
39. Baykov A. A., Bakuleva N. P., Rea P. A. Steady-state kinetics of substrate hydrolysis by vacuolar H⁺-pyrophosphatase. *Eur. J. Biochem.* **217**, 755-762 (1993).
40. Segel I. H. *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady-state enzyme systems*. John Wiley & Sons (1975).
41. Kabsch W. XDS. *Acta Crystallogr. D: Biol. Crystallogr.* **66**, 125-132 (2010).
42. McCoy A. J., Grosse-Kunstleve R. W., Adams P. D., Winn M. D., Storoni L. C., Read R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).

43. Adams P. D., *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. Sect. D: Biol. Crystallogr.* **66**, 213-221 (2010).
44. Emsley P., Lohkamp B., Scott W. G., Cowtan K. Features and development of Coot. *Acta Cryst. Sect. D: Biol. Crystallogr.* **66**, 486-501 (2010).
45. Krissinel E., Henrick K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774-797 (2007).
46. Tseng Y. Y., Dupree C., Chen Z. J., Li W.-H. SplitPocket: identification of protein functional surfaces and characterization of their spatial patterns. *Nucleic Acids Res.* **37**, W384-W389 (2009).
47. Yuan T. Y., Li W. H. Identification of protein functional surfaces by the concept of a split pocket. *Proteins* **76**, 959-976 (2009).
48. Sievers F., *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
49. Ashkenazy H., *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344-W350 (2016).
50. Landau M., *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299-302 (2005).
51. Discovery Studio Modeling Environment, Release 4.5. (ed^(eds)). Dassault Systèmes, Biovia (2015).
52. Sali A., Blundell T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815 (1993).

Acknowledgements

This work was supported by the grants from the Jane and Aatos Erkko Foundation and the BBSRC (BB/M021610) to Adrian Goldman, the Academy of Finland (No. 308105) to Keni Vidilaseris and (No. 310297) to Henri Xhaard and (279755) to Seppo Meri, and University of Helsinki funds to Gustav Boije af Gennäs. The authors thank Violeta Manole for her technical help, Nina Sipari for help with the MS-analyses, and Juho Kellosalo, Anssi Malinen and Peter Henderson for fruitful discussions during the project.

Author Contributions

Conceptualization, K.V., A.G., H.X., J.Y.-K., G.B.G; Methodology, K.V., A.G., A.T., A.K., N.G.J., T.O.L.; Investigation, K.V., A. K., A.T, Ay.K.; Resources, A.G, H.X, S.M., J.Y-K.; Writing – Original draft, K.V., A.T., H.X., A.K; Writing-Review and Editing, K.V., A.K, A.T., Ay.K., N.G.J., T.O.L., P.S.K., G.B.G, S.M., J.Y-K., H.X, A.G.; Visualization, K.V., A.T., Ay.K.; Supervision and Project Administration, A.G., H.X, G.B.G, J.Y-K, S.M.; Funding Acquisition, A.G., H.X, K.V, J.Y-K., G.B.G, S.M.

Accession codes: The atomic coordinates and structure factors of the TmPPase:IDP:ATC complex have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID ****).

Declaration of Interest

All authors declare no conflict of interest in this paper.

Figure Legends

Figure 1. Overview of TmPPase structure. **(a)** Monomer showing the location of the hydrolytic centre, coupling channel, ion gate, and the exit channel **(b)** Top view of the superposition of TmPPase:IDP:ATC (wheat) and TmPPase:IDP complex (cyan) structure showing relative TMHs movement (arrow) upon the binding of ATC. **(c)** Superposition of the gate region between two structures (TmPPase:IDP:ATC (wheat) and TmPPase:IDP complex (cyan)). D246^{6.53}, D703^{16.46}, and Na⁺ slightly moving away (arrow) at the same direction relative to their positions in TmPPase:IDP structure. Violet-purple coloured sphere is the Na⁺ of TmPPase:IDP structure.

Figure 2. ATC interactions with TmPPase domains. **(a)** ATC located on the interface of chain A (wheat) and D (grey) in a hydrophobic cleft. Pale-green: carbon, blue: nitrogen, red: oxygen and yellow: sulphur. **(b)** Residues in chain A that are important for the interactions with the ATC dimer. **(c)** Residues on chain D that are important for the interactions with the ATC dimer. Dashed lines represent hydrogen bonds.

Figure 3. Loop movement caused by the interactions with ATC. **(a)** ATC dimer interaction with β 1–2 strand (deep blue), loop8–9 (red), and loop12–13 (cyan). ATC is depicted in space-filling model (pale-green: carbon, blue: nitrogen, red: oxygen and yellow: sulphur). **(b)** Superposition of TmPPase:IDP (cyan) with TmPPase:IDP:ATC structure (wheat) and TmPPase:Ca:Mg (green) showing the large side movement of β 1–2 strand (arrow). ATC is depicted as in **a**. **(c)** Superposition of TmPPase structures in all different states showing the closer/away movement of β 1–2 strand relative to the loop12-13 and the dimer interface of the protein. Colouring scheme from the left to right: phosphate analogue (WO₄)-bound state (TmPPase: WO₄) = pink; resting state (TmPPase:Ca:Mg) = green; two product bound-state (TmPPase:2P_i) = orange; IDP-bound state (TmPPase:IDP)= cyan; IDP:ATC-bound state (TmPPase:IDP:ATC) = wheat. **(d)** Superposition of

TmPPase structures showing the downward movement of TMH12 and away movement of loop12-13 from the resting state (green) to the IDP:ATC-bound form (wheat) of TmPPase. The green-to-wheat coloured arrow shows the movement.

Figure 4. Compound library and inhibition activity against TmPPase. **(a)** Screening hits. **(b)** Analogues developed for structural studies. **(c)** Inhibition curve of ATC, compounds **6–9**. **(d)** Kinetic inhibition plot for the inactivation of TmPPase at four different concentrations of ATC. The solid lines represent nonlinear regression fits with the residuals standard deviation (Sy.x) of 2.0. All data shown as mean \pm SD with n=3 replicates.

Figure 5. The conservation of ATC binding site among mPPases. **(a)** The structure of TmPPase coloured according to the sequence conservation among the set of 18 pathogenic mPPases. The pale-green surface indicates the location of the ATC binding site. **(b)** Comparison of the ATC binding sites of the crystal structure of TmPPase (wheat) and a homology model of PfPPase (teal). Carbon atoms of ATC are presented as pale-green sticks. Blue: nitrogen, red: oxygen and yellow: sulphur. View as in Figure 2A.

Figure 6. Catalytic scheme of TmPPase and its inhibition by ATC. The binding of substrate (PPi; two joined red circles) to monomer A (wheat colour) **(a)** leads to Na^+ pumping (blue circle and blue arrow for the pumping direction) and induces a conformational change in monomer B (pink colour) **(b)** which increase its affinity for substrate binding. The Na^+ pumping causes the substrate hydrolysis (yellow lightning symbol for this event) in monomer A, while the substrate binding in monomer B leads to its Na^+ pumping **(c)**. After hydrolysis, the monomer A released its hydrolysis products (2P_i; red circles) while monomer B proceed to the hydrolysis event **(d)** and then moves back to the state in **a**. The ATC dimer binds to its binding site after substrate binding and Na^+ pumping in monomer A **(e)** which lock this monomer in the “down” state. Even though no

hydrolysis occurred in monomer A due to the ATC binding, its conformational change still able to induces a conformational change in monomer B that increase its substrate affinity (**f**). The numbering with white background corresponds to transmembrane helix number.

Table 1. X-ray data collection and refinement statistics.

Data Collection	
Space group	P 2 ₁ 2 ₁ 2 ₁
Cell dimensions	
<i>a, b, c</i> (Å)	98.133, 141.63, 252.06
Source	ESRF - ID29
Wavelength (Å)	0.9677
Resolution (Å)	47.41 – 3.7 (3.832 – 3.7)
R-merge	0.1201 (1.451)
<i>I</i> / σ (<i>I</i>)	9.90 (1.24)
Completeness (%)	99.35 (99.95)
Multiplicity	5.3 (5.6)

Refinement	
Resolution (Å)	47.41 – 3.7
No. of total reflections	203154 (21201)
No. of unique reflections	38085 (3794)
R _{work} (%)/R _{free} (%)	22.5/26.8
No. of atoms	21075
Protein	20929
ligands	145
No. of chains (ASU)	4
B-factors (Å ²)	
Protein	134.56
Ligands/Ion	144.93
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angle (°)	0.62
Ramachandran statistics	
Favored (%)	94.12
Allowed (%)	5.71
Outliers (%)	0.17

Statistics for the highest-resolution shell are shown in parentheses.

Table 2. Major interactions of ATC with chains A and D

Interaction type	Compound	Chain A		Chain D	
		Residue	BSA* (Å ²)	Residue	BSA* (Å ²)
Hydrophobic/ Aromatic	ATC-1	Q268	46.3	M69	15.0
		K269	34.3	T73	15.9
		I270	21.4	W74	15.3
		P530	19.1	N163	15.7
	ATC-2			W164	30.1
		Q268	25.6	W74	45.5
		277	22.2	Y161	32.0
Possible hydrogen bond	Compound	D351	28.6	F169	16.2
		V352	21.1		
	Total BSA* (Å ²)		218.7		185.7
	ATC-1 (N19)		G528 (C=O)	3.6	
Salt bridge	ATC-1 (N20)		G528 (C=O)	3.8	
	ATC-2			N157 (Oδ1)	3.6
Salt bridge	ATC-2 (N19)	D351 (Oδ2)	3.0		

* BSA: Buried surface area: the solvent-accessible surface area (cut-off = 15 Å²) of the corresponding residue that is buried upon interface formation.

Table 3. Root mean square deviation (rmsd) of chain A and B of loops in different TmPPase structures.

Loop	Rmsd (Å) ¹				
	TmPPase:IDP:ATC (AB)	5lzQ (AB)	5lzR (AB)	4AV3 (AB)	4av6 (AB)
2-3 (71-75)	0.382	0.184	0.444	0.331	0.261
4-5 (153-172)	0.905	0.242	0.579	0.369	0.322
β1-2 (6-7) (266-276)	1.630	0.486	1.237	0.613	0.503
8-9 (342-359)	0.743	0.405	0.695	0.337	not applicable ²
12-13 (520-540)	0.442	0.367	0.406	0.455	0.364

¹ Comparisons were done by superimposing helix 7 (281-311), which moves least between the structures

² Missing loop from residue 349-357

Table 4. Kinetic parameters of TmPPase activity and ATC inhibition constants

Kinetic parameter	Value
k_p	$239.2 \pm 32.2 \mu\text{mol P}_i/\text{mg/min}$
β	0.05 ± 0.005
βk_p	$12.2 \pm 1.03 \mu\text{mol P}_i/\text{mg/min}$
K_s	$666.2 \pm 110.2 \mu\text{M}$
α	0.12 ± 0.05
αK_s	$80.7 \pm 19.3 \mu\text{M}$
K_{I2}	$1.8 \pm 0.3 \mu\text{M}$
K_{I3}	$10.0 \pm 1.1 \mu\text{M}$

Fig 1

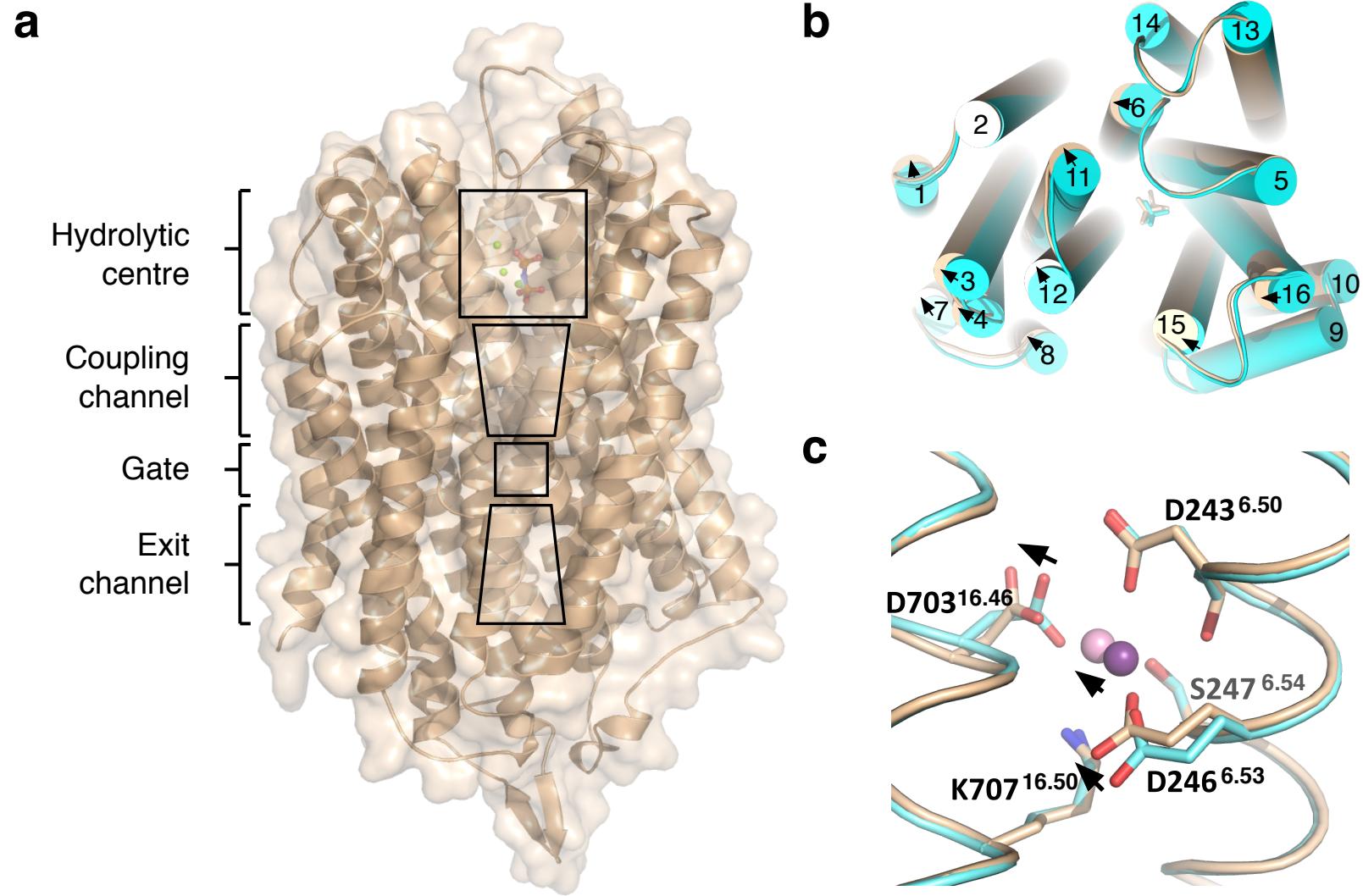


Fig. 2

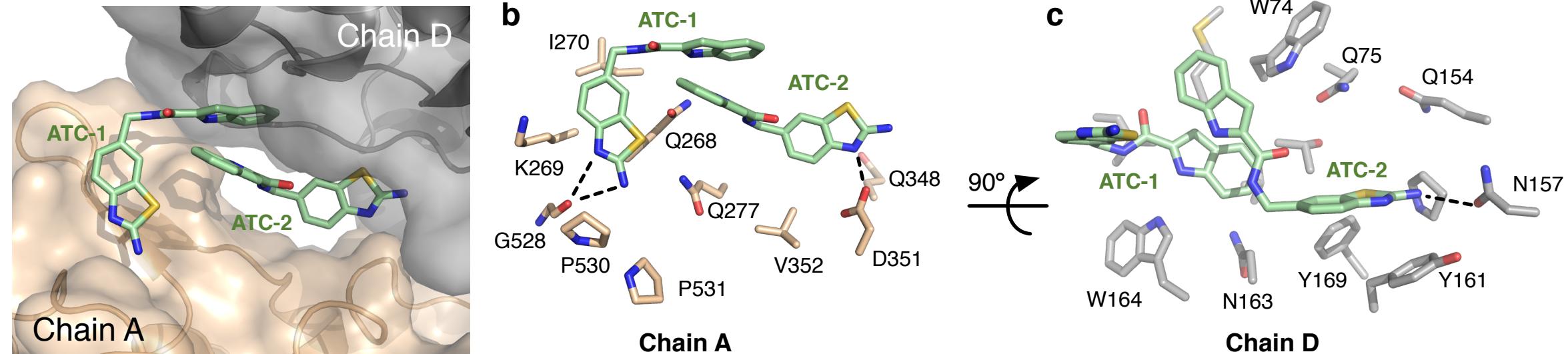
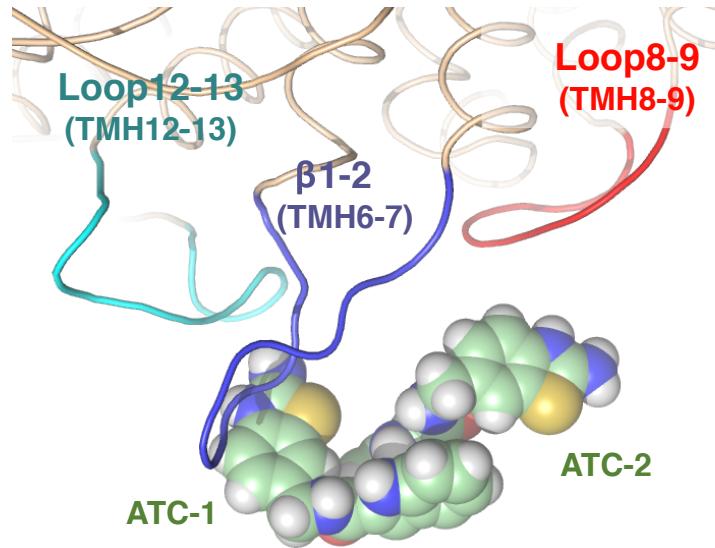
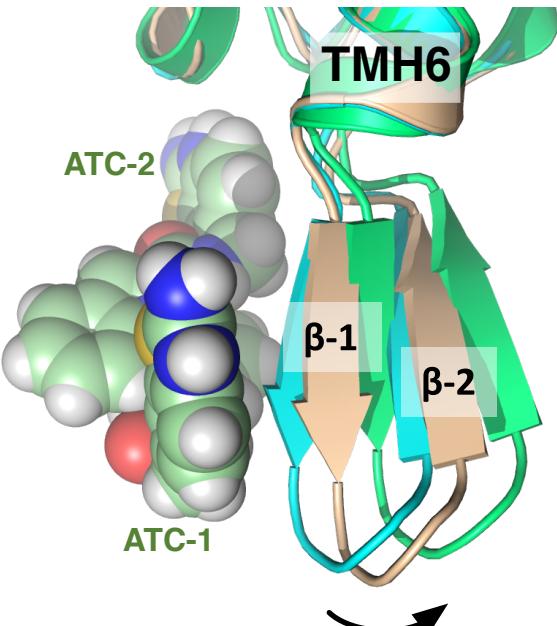


Fig 3

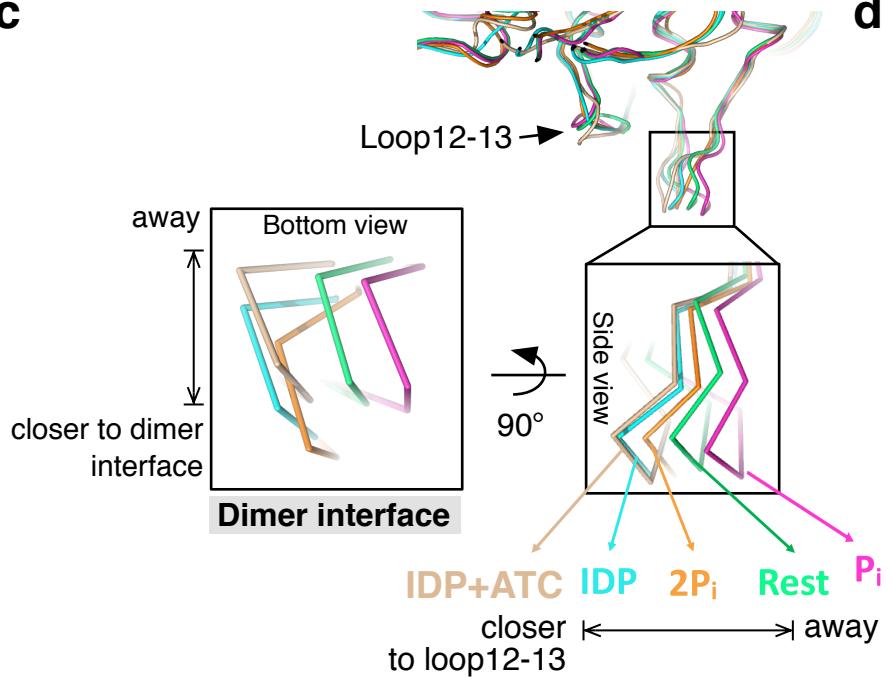
a



b



c



d

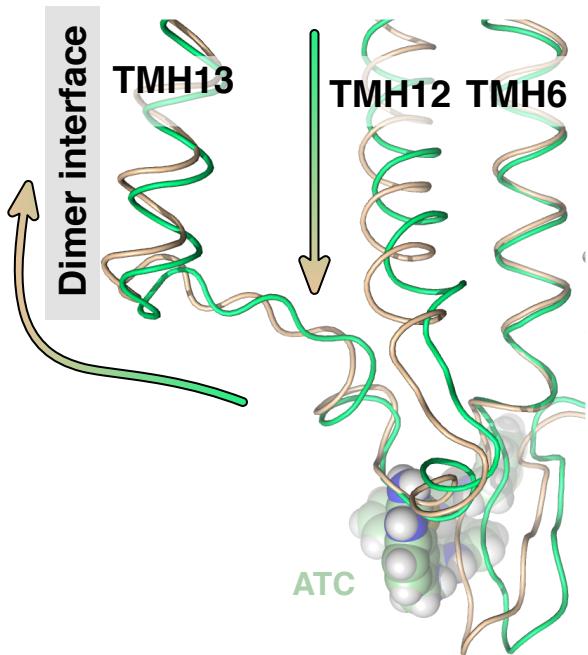
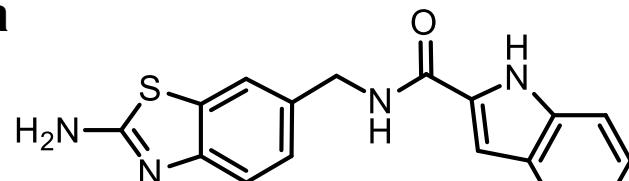
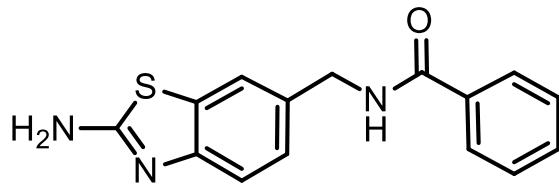


Fig 4

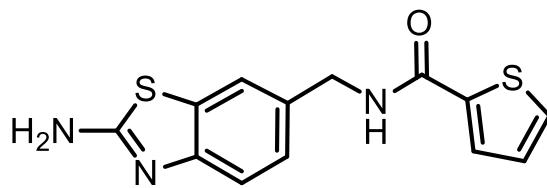
a



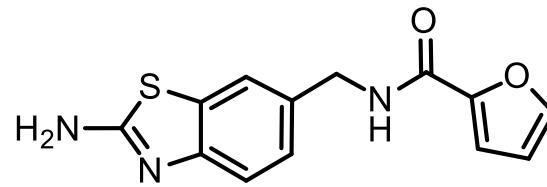
1 ATC ($IC_{50} = 1.7 \mu M$)



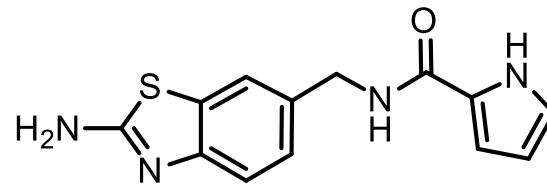
2 ($IC_{50} = > 500 \mu M$)



3 ($IC_{50} = > 500 \mu M$)

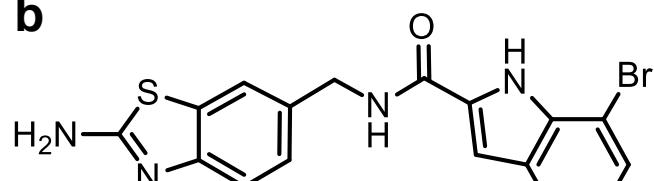


4 ($IC_{50} = 25.5 \mu M$)

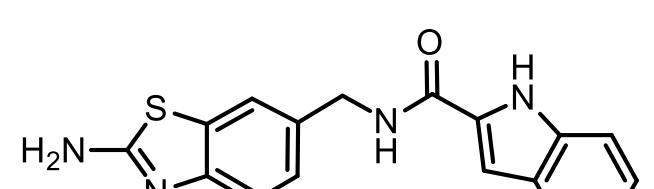


5 ($IC_{50} = 23.4 \mu M$)

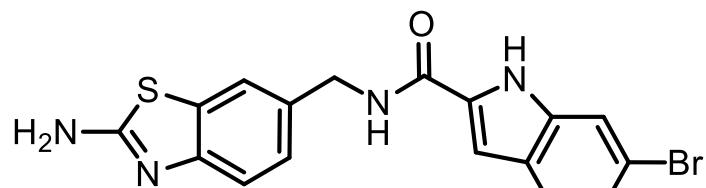
b



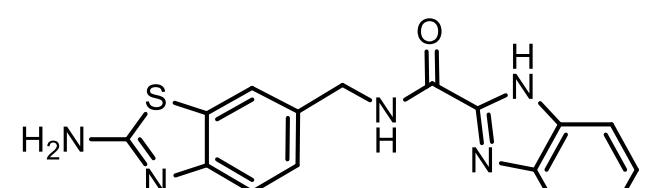
6 ($IC_{50} = 21.4 \mu M$)



7 ($IC_{50} = 58.8 \mu M$)

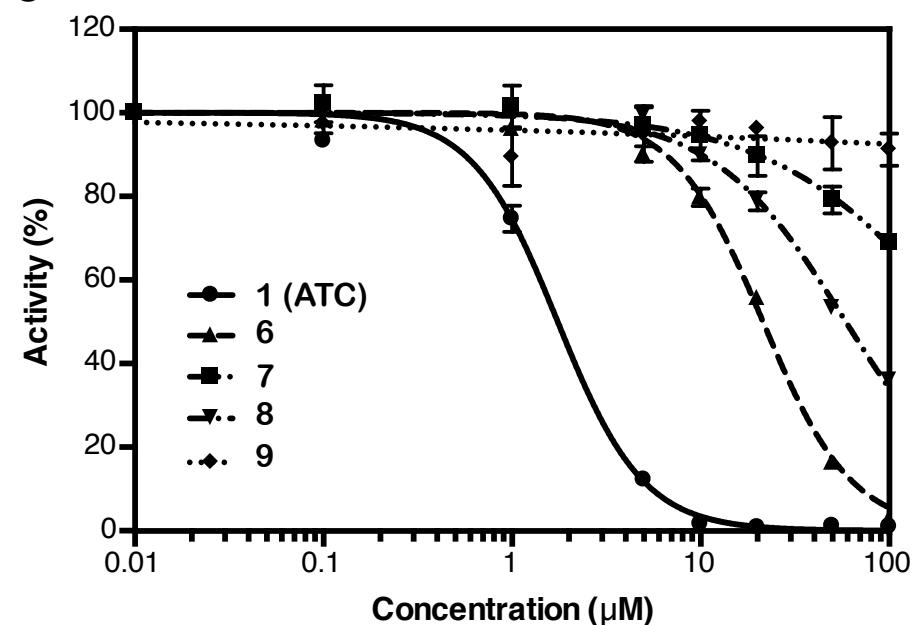


8 ($IC_{50} = 239 \mu M$)



9 ($IC_{50} = > 500 \mu M$)

c



d

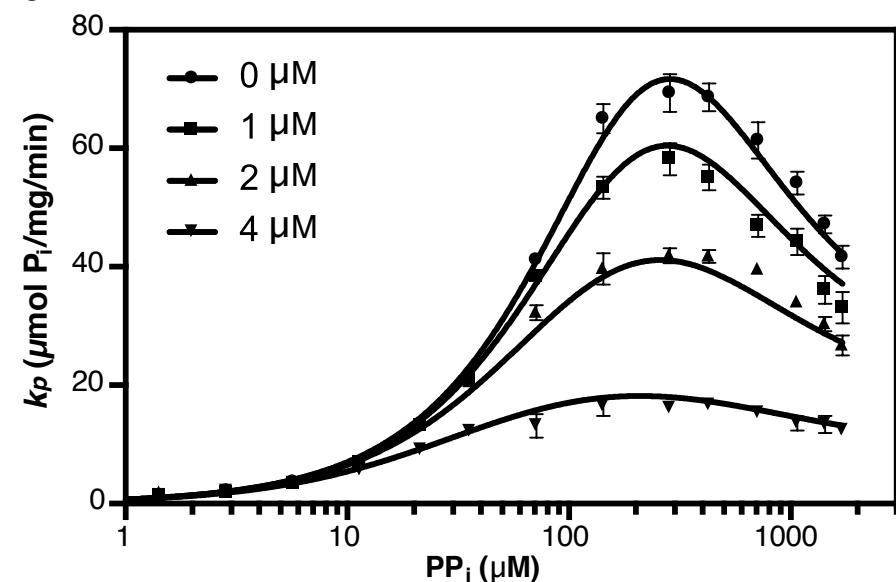


Fig 5

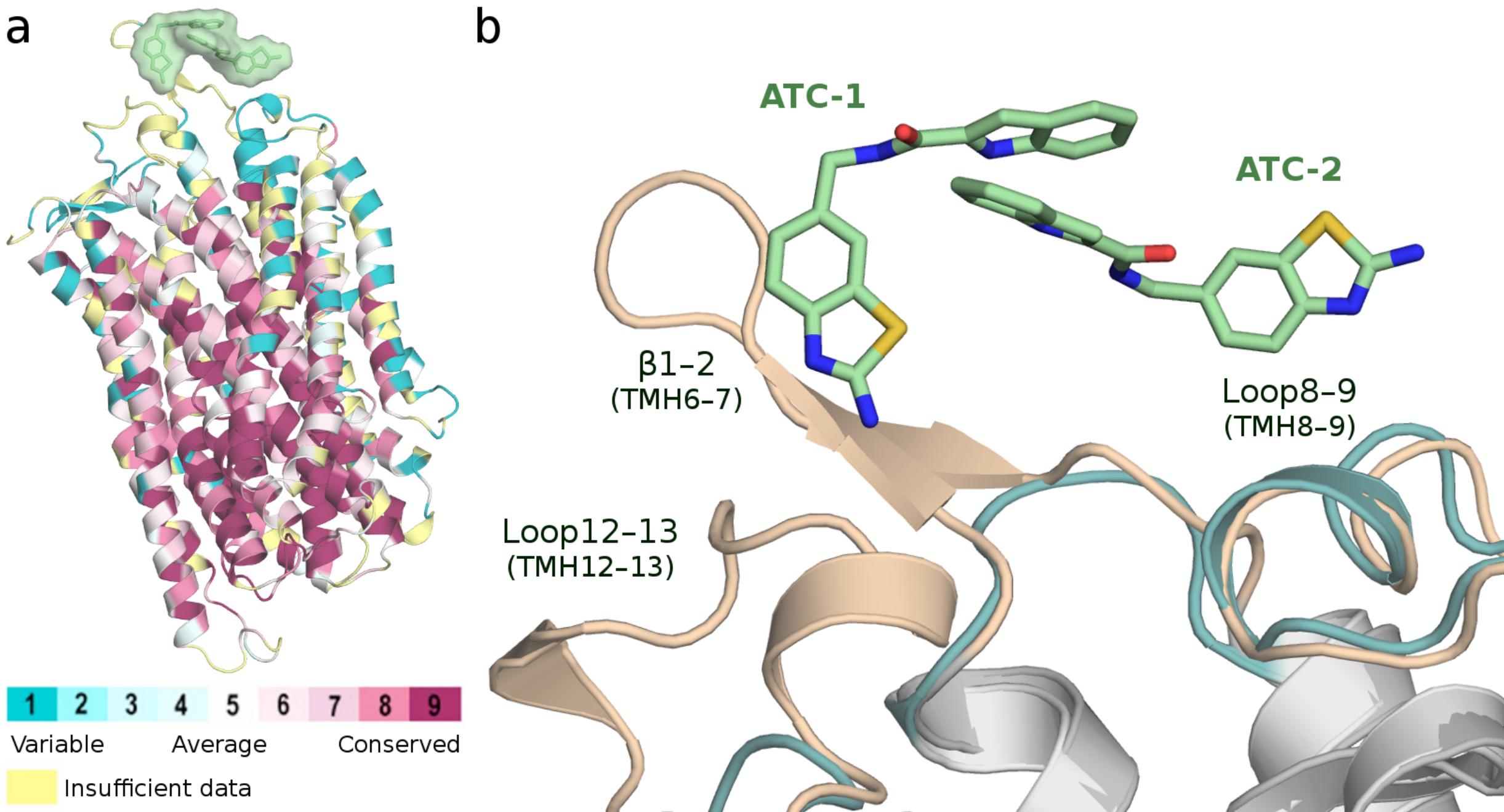


Fig 6

