

**Title.** New insights into the domain of unknown function DUF of EccC<sub>5</sub>, the pivotal ATPase providing the secretion driving force to the ESX5 secretion system.

**Authors:** Fernando Ceballos-Zúñiga<sup>1</sup>, Margarita Menéndez<sup>2,3</sup>, Inmaculada Pérez-Dorado<sup>1\*</sup>

## Affiliations

1. Groups of Crystallography and Structural Biology. Institute of Physical-Chemistry Blas Cabrera, Spanish Research Council, Serrano 119, 28006 – Madrid (Spain)

2. Group of Structure and Thermodynamics of Macromolecules. Institute of Physical-Chemistry Blas Cabrera, Spanish Research Council, Serrano 119, 28006 – Madrid (Spain)

3. Ciber of Respiratory Diseases, ISCIII, Sinesio Delgado 10, 28029 – Madrid (Spain)

\*Corresponding author

## Summary

Type VII secretion (T7S) systems, also referred to as ESAT6 secretion (ESX) systems, are molecular machines that have gained great attention due to their implication in cell homeostasis and host pathogen interactions in mycobacteria. The latter include important human pathogens such as *Mycobacterium tuberculosis* (Mtb), the etiological cause of human tuberculosis and a pandemic accounting for more than 1 million deaths every year. The ESX5 system is exclusively found in slow-growing pathogenic mycobacteria, where it mediates the secretion of a large family of virulence factors, the PE and PPE proteins. The secretion driving force is provided by EccC<sub>5</sub>, a

multidomain ATPase operating through four globular cytosolic domains, an N-terminal domain of unknown function (EccC<sup>DUF</sup>) and three FtsK/SpoIIIE ATPase domains. Recent structural and functional studies of ESX3 and ESX5 systems have revealed EccC<sup>DUF</sup> as an ATPase-like fold domain with potential ATPase activity, and whose functionality is essential for secretion. Here we report the crystal structure of *MtbEccC*<sub>5</sub><sup>DUF</sup> domain at 2.05 Å resolution, which unveils a nucleotide-free structure with degenerated *cis*-acting and *trans*-acting elements involved in ATP-binding and hydrolysis. Our crystallographic study, together with a biophysical assessment of *MtbEccC*<sub>5</sub><sup>DUF</sup> interaction with ATP/Mg<sup>2+</sup>, supports the absence of ATPase activity proposed for this domain. We show that this degeneration is also present in DUF domains of other ESX and ESX-like systems, which are likely to exhibit poor or null ATPase activity. Moreover, and based on an *in-silico* model of *MtbEccC*<sub>5</sub> N-terminal region, we propose that *MtbEccC*<sub>5</sub><sup>DUF</sup> is a degenerated ATPase domain that may have retained the ability to hexamerise. Observations that call the attention on DUF domains as structural elements with potential implications in the opening and closure of the membrane pore during the secretion process.

**Keywords.** ESX systems, T7SS, secretion, mycobacteria, ATPase activity, crystallography

## Introduction

Molecular machines are sophisticated protein complexes ubiquitously present in all cellular organisms (Miller & Enemark, 2016). They convert the chemical energy resulting from nucleoside triphosphate (NTP) hydrolysis into the mechanical force needed in numerous cellular events such as DNA replication, protein degradation, cell motility or protein secretion, among others (Miller & Enemark, 2016; Schmidt *et al.*, 2012; Famelis *et al.*, 2023; Crosskey *et al.*, 2020). Members of the superfamily of ATPases Associated with various cellular Activities (AAA +) are key constituents of many molecular machines, where the mechanical work is performed at the expense of adenosine triphosphate (ATP) hydrolysis (Miller & Enemark, 2016; Leipe *et al.*, 2002). Type VII secretion systems (T7SSs), also referred to as ESAT6 secretion (ESX) systems, are AAA + dependent molecular machines found in the *Actinomyete* phylum that have gained great attention due to their implication in cell homeostasis and host pathogen interactions in mycobacteria (Famelis *et al.*, 2023; Bunduc *et al.*, 2020; Houben *et al.*, 2014). The latter include devastating pathogenic species such as *Mycobacterium tuberculosis* (*Mtb*), the etiological cause of human tuberculosis (TB) and a pandemic responsible of more than 1 million deaths every year (World Health Organization, 2022). T7SSs are also found in bacteria belonging to the *Firmicutes* phylum (T7SSb), which comprise other important pathogens such as *Staphylococcus aureus*, *Listeria monocytogenes*, *Bacillus anthracis* or *Bacillus subtilis* (Zoltner *et al.*, 2016; Mietrach *et al.*, 2020).

Mycobacteria produces up to five ESX secretion systems (T7SSa), named from ESX1 to ESX5, consisting of large membrane complexes spanning the inner bacterial membrane. Among them, ESX5 is found almost exclusively in slow-growing

pathogenic mycobacteria, where it participates in nutrient uptake, intracellular colonization, and modulation of the immune response during infection through the secretion of a large family of protein effectors, the PE and PPE proteins (Bunduc *et al.*, 2020; Houben *et al.*, 2014). These roles have important implications in pathogen life cycle and virulence, thus pointing out at the ESX5 as a potential drug target (Bunduc *et al.*, 2021). Efforts aiming at the structural-functional characterization of mycobacterial ESX secretion systems, have resulted in the structures of ESX5 from *Mtb* and *Mycobacterium xenopi* (*Mxp*), and ESX3 from *Mycobacterium smegmatis* (*Msm*), which have enlightened the architecture of the pore complex (Famelis *et al.*, 2019; Poweleit *et al.*, 2019; Bunduc *et al.*, 2021; Beckham *et al.*, 2021)(Fig. 1A). The latter can be defined as a hexamer of protomers, where each protomer is formed by four different membrane proteins, EccB:EccC:EccD:EccE, interacting with a 1:1:2:1 stoichiometry. Among them, EccC is a pivotal component of the pore complex providing the secretion driving force via ATP hydrolysis (Famelis *et al.*, 2019; Poweleit *et al.*, 2019; Bunduc *et al.*, 2021; Beckham *et al.*, 2021). In ESX2-5 systems, EccC consists of two N-terminal transmembrane (TM) helices connected to two additional helices (stalk region) and a domain of unknown function (EccC<sup>DUF</sup>), which are followed by three FtsK/SpoIIIE AAA + ATPase domains, here referred to as D1, D2 and D3 (EccC<sup>D1-3</sup>) (Famelis *et al.*, 2023; Bunduc *et al.*, 2021) (Fig. 1B). This multi-domain organization is also shared by EccC orthologues (EssC) found in *Firmicutes* (Mietrach *et al.*, 2020), which differs in ESX1, where EccC is replaced by two functionally equivalent fragments: EccC<sub>1a</sub>, containing of the DUF and D1 domains; and EccC<sub>1b</sub> containing D2 and D3 domains (Famelis *et al.*, 2023; Houben *et al.*, 2014). EccC/EssC enzymes are expected to operate through a mechanism involving their hexamerisation. In line with this, the cytosolic region of *Mtb*EccC<sub>5</sub> has been observed

to transient from an extended/open state to a contracted/closed state, where the enzyme multimerises to form the cytosolic chamber that is expected to accommodate the effectors to be secreted (Bunduc *et al.*, 2021) (Fig. 1A).

D1-3 domains exhibit a three-layer  $\alpha$ - $\beta$ - $\alpha$  core structure typically found in prokaryotic AAA + enzymes, which is frequently followed by a C-terminal alpha-helical lid domain missing in EccC/EssC enzymes (Zoltner *et al.*, 2016). The hexameric architecture observed in ESX3 and ESX5 complexes is in line with the functional oligomeric state frequently found among prokaryotic AAA +. The latter include EccC close relatives, FtsK and SpoIIIE ATPases, reported to function as hexamers, and where ATPase sites locate at the interface between two adjacent protomers (Miller & Enemark, 2016; Leipe *et al.*, 2002; Bunduc *et al.*, 2021; Rosenberg *et al.*, 2015). In such cases, both subunits provide with *cis*- and *trans*-acting catalytic elements required to form the active site. The former includes the Walker A motif, the Walker B motif and Sensor 1, which typically locate at the loop connecting  $\beta$ 1 and  $\alpha$ 1 (P-loop), the C-terminus of the  $\beta$ 3 strand, and the loop connecting  $\beta$ 4 and  $\alpha$ 4, respectively. The Walker A motif, also referred to as P-loop, consists of a glycine-rich sequence ('G<sub>1</sub>xxG<sub>2</sub>xG<sub>3</sub>K[S/T]' in FtsK homologues) involved in the stabilization of the phosphate group of the nucleotide via highly conserved lysine and serine/threonine residues that enable the correct orientation of the  $\gamma$ -phosphate required for ATP hydrolysis (Miller & Enemark, 2016; Leipe *et al.*, 2002). The Walker B motif provides two acidic residues ('hhhhDE' in FtsK homologues, where 'h' is any hydrophobic amino acid) of which, the second performs as the catalytic base activating the water involved in the nucleophilic attack of the  $\gamma$ -phosphate (Miller & Enemark, 2016; Leipe *et al.*, 2002). Sensor 1 is a key catalytic element consisting of a polar residue, typically an asparagine but it can be serine, threonine, or aspartate as well. It locates between both Walker motives to assist in the

correct orientation of the catalytic glutamate towards the nucleophilic water. *Trans*-acting elements are positively charged amino acids, mostly (but not exclusively) arginine residues, being thus referred to as arginine fingers or sensors (Miller & Enemark, 2016; Leipe *et al.*, 2002). The latter include the Arg finger, which consists of an arginine (or lysine) residue located at the end of the  $\alpha 4$  helix that orients towards the neighbouring active site to form contacts with the  $\gamma$ -phosphate. Through this interaction, Arg finger favours the transition state of hydrolysis, thus playing an important role in ATP hydrolysis and, in some cases, in the enzyme multimerization (Miller & Enemark, 2016; Ogura *et al.*, 2004). The other two *trans*-acting motifs in AAA + enzymes are Sensors 2 and 3, which are generally arginines involved in sensing and/or stabilising the ATP/ADP-bound state, and promoting conformational changes coupled to these states or nucleotide hydrolysis (Miller & Enemark, 2016; Li *et al.*, 2015).

The first structural and functional studies conducted by Rosenberg and co-workers in the EccC from *T. curvata* (TcrEccC) showed that the D1 domain is an active ATPase exhibiting fully conserved Walker A ('GxxGxGK[S/T]') and Walker B ('hhhhDE') motives with respect to FtsKs, as close canonical orthologues, including the Arg-finger *trans*-acting motif (Rosenberg *et al.*, 2015). This and later studies carried out on EccCs/EssCs from *Mtb* and *S. aureus* showed that Walker A and B motives are, however, degenerated in D2 and D3, which aligns with non-detectable or poor catalytic activities observed for those domains, which are proposed to play a regulatory role (Rosenberg *et al.*, 2015; Wang *et al.*, 2020; Zoltner *et al.*, 2016). More recently, the structural characterization of the *Msm*ESX3 complex revealed that the N-terminal domain of unknown function (DUF) exhibits an ATPase-like fold, also observed in the DUF domains of *Mtb*ESX5 and *Mxp*ESX5 complexes reported afterward (Famelis *et*

*al.*, 2019; Bunduc *et al.*, 2021; Beckham *et al.*, 2021). The authors describe the presence of a Walker B motif ('hhhhDD'<sup>320</sup>) in *MsmEccC*<sub>3</sub><sup>DUF</sup>, partially conserved with respect to FtsK orthologs ('hhhhDE'), and how mutation of each aspartate to alanine obliterates secretion. These observations highlight the essentiality of *MsmEccC*<sub>3</sub><sup>DUF</sup> in secretion, and they point to a potential ATPase activity in this domain (Famelis *et al.*, 2019). The structure of *MsmEccC*<sub>3</sub><sup>DUF</sup> is however devoid of ATP, analogously to *MtbEccC*<sub>5</sub><sup>DUF</sup> and *MxpEccC*<sub>5</sub><sup>DUF</sup> structures that exhibit valine and isoleucine residues replacing the catalytic amino acid, respectively (Famelis *et al.*, 2019; Bunduc *et al.*, 2021; Beckham *et al.*, 2021). These observations expose a variable degeneration of the Walker B motif across different EccC<sup>DUF</sup> domains that may impair the ATPase activity to a different extent. In addition, *MsmEccC*<sub>3</sub><sup>DUF</sup>, *MtbEccC*<sub>5</sub><sup>DUF</sup>, and *MxpEccC*<sub>5</sub><sup>DUF</sup> domain structures reveal a non-canonical arrangement of the Walker A motif, lacking the typical P-loop structure that is replaced by an extended alpha-helical conformation. Interestingly, this arrangement is observed in the structure of the EssC<sup>D3</sup> domain of *S. aureus* (*SrsEssC*<sup>D3</sup>, PDB entry 6TV1) also devoid of ATP, and exhibiting an ATP-binding affinity within the low-millimolar range (Mietrach *et al.*, 2020).

The structural and functional information available in T7SSs leave outstanding questions regarding the molecular basis underlying the ATPase activity proposed for the DUF domain. In this sense, here we report the crystallographic structure of *MtbEccC*<sub>5</sub><sup>DUF</sup> domain at 2.05 Å resolution, which provides an unambiguous model showing a nucleotide-free structure with degenerated *cis*-acting and *trans*-acting elements involved in ATP-binding and hydrolysis. Our high-resolution structure, together with a biophysical assessment of *MtbEccC*<sub>5</sub><sup>DUF</sup> interaction with ATP/Mg<sup>2+</sup> *in vitro*, supports the absence of ATPase in this domain. These results are in line with the *in-silico* analysis carried out in other EssC and EccC enzymes, which reveals the

presence of degenerated DUF domains in other mycobacterial and non-mycobacterial T7S systems that are likely to exhibit null or deficient ATPase activity. These findings suggest that DUF domains play a different role in the secretion process that, based on an *in-silico* model of *MtbEccC*<sub>5</sub> and mutagenesis studies reported in *MsmESX3* (Famelis *et al.*, 2019), we propose is related to the aperture of the membrane-pore complex during the secretion process.



## Material and methods

### 3.1. Protein expression and purification

Protein constructs containing the DUF domain of EccC<sub>5</sub> (EccC<sub>5</sub><sup>DUF</sup>, residues 1-417) from *Mycobacterium tuberculosis* were designed for their recombinant production. Residues 17-118 and 167-198 were replaced by 'GSSG' and 'GSG' sequences, respectively, in order to remove the transmembrane (TM) and stalk regions and enable its production as a soluble protein. The construct contains a N-terminal 6His-tag followed by a SUMO-tag, and a PreScission 3C cleavage site. The DNA coding for the construct cloned into a pET-28a plasmid (between NcoI and XhoI restriction sites) was purchased to GenScript with codon optimization for *E. coli* expression (pET-EccC<sub>5</sub><sup>DUF</sup>). Expression of EccC<sub>5</sub><sup>DUF</sup> was performed in *E. coli* BL21 (DE3) Star (Invitrogen) using 2xTY media supplemented with kanamycin (50 µg/ml). Cells were grown at 37 °C until an OD<sub>600 nm</sub> of 0.8 and then cooled down to 16 °C for an hour. Expression was then induced at 16 °C by the addition of 1 mM IPTG for 20 h. Cells were harvested by centrifugation at 5251 *g* for 20 min at 4 °C, and cell pellets were stored at -20 °C.

Cell pellets were thawed and resuspended in 20 mM Tris-HCl pH 7.5 and 500 mM NaCl, 10 mM imidazole (buffer A), supplemented with 5 mM MgCl<sub>2</sub>, 1 mM MnCl<sub>2</sub> and 1 µg/ml DNase (Sigma). This step was performed at RT while subsequent purification steps were carried out at 4 °C. Cells were disrupted by sonication and the clarified extract was loaded into a 1 ml HisTrap HP column (Cytiva) previously equilibrated in buffer A. Protein samples were eluted using an imidazole gradient (from 10 mM to 500 mM) and then buffer exchanged to 20 mM Tris-HCl pH 7.5, 150 mM NaCl (Buffer B) using a PD-10 desalting column (Cytiva). Samples were diluted to 1 mg/ml in buffer B prior to overnight cleavage with the 3C protease. The cleaved protein was

concentrated using Amicon 10 KDa MWCO concentrators (Millipore), and further purified by size exclusion chromatography using a Superdex 200 (16/60) column (Cytiva) preequilibrated on buffer B or 20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM MgCl<sub>2</sub> (buffer C), for crystallographic or ITC experiments, respectively. Pure protein fractions obtained from the size-exclusion chromatography (SEC) were pooled, concentrated, and flash-frozen in liquid nitrogen for their storage at -80 °C. Protein sample purity and quantification were assessed using PAGE-SDS and UV-Vis absorbance at 280 nm, respectively.

### 3.2. Protein crystallization, structure resolution and analysis

All crystallization experiments were carried out using the sitting-drop vapour diffusion method at 20 °C in 96-well MRC plates (Hampton Research) and employing an Oryx8 robot (Douglas Instruments). Initial crystals of EccC<sub>5</sub><sup>DUF</sup> were obtained in the F2 condition of the Index screen. After optimization, the best crystals grew in 300 nl droplets formed by mixing 150 nl of protein at 10 mg/ml and 150 nl of precipitant condition (25 % (w/v) of PEG MME 2K, 300 mM of trimethylamine N-oxide and Tris-HCl pH 8). Crystals were flash-cooled in liquid nitrogen without cryoprotection for X-ray data collection, which enabled to measure a complete X-ray data set at 100K using synchrotron radiation at ALBA (Barcelona, Spain) at 2.05 Å resolution. The data set was indexed and integrated with XDS (Kabsch, 2010), and scaled and reduced using AIMLESS (Evans & Murshudov, 2013). Crystals belonged to the *P*1 space group, with unit cell dimensions  $a=48.30$  Å,  $b=52.24$  Å,  $c=57.44$  Å,  $\alpha=88.39$  °,  $\beta=84.60$ °,  $\gamma=80.17$  ° (Table 1). Structure resolution was carried out using the molecular replacement method with PHASER (McCoy *et al.*, 2007), and the coordinates of EccC<sub>5</sub><sup>DUF</sup> structure (1-416 residues) solved by cryo-EM (PDB entry 7NPT) (Bunduc *et al.*, 2021) as the

search model. A single and unambiguous solution was obtained, which consisted of two EccC<sub>5</sub><sup>DUF</sup> molecules in the asymmetric unit. The initial model was subjected to alternate cycles of model building with COOT (Emsley *et al.*, 2010), and refinement with PHENIX (Adams *et al.*, 2010) and BUSTER (Bricogne G. *et al.*, 2017) using NCS restrictions. Electron density maps allowed us to model almost completely both molecules of the asymmetric unit and 287 water molecules. Residues 275-283 and 311-315 (in chain A), and 274-283 (Chain B), were not modelled due to poor electron density observed for these residues. The geometry of the final model was validated using MOLPROBITY (Chen *et al.*, 2009). Dihedral angles were analysed with COOT (Emsley *et al.*, 2010) and MOGUL (Version 2020.3.0). Figures were generated using PYMOL (Version 2.0; Schrödinger). The atomic coordinates and structure factors have been deposited in the Protein Data Bank under the accession code 8RIN.

### 3.3. Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) assays were carried out to estimate the inflection temperature ( $T_i$ ) associated with unfolding transitions of the EccC<sub>5</sub><sup>DUF</sup> construct in the absence and in the presence of nucleotide/Mg<sup>2+</sup>. The experiments consisted of monitoring the variation in the emission of fluorescence of tryptophan residues, buried and exposed, at a wavelength of 350 nm and 330 nm, respectively, using the Tycho NT.6 (NanoTemper). DSF experiments were set up to a final volume of 10  $\mu$ l of buffer C containing the EccC<sub>5</sub><sup>DUF</sup> construct at 2  $\mu$ M in the absence or in the presence of ADP-AlF<sub>3</sub> at 4 mM. Mixtures of protein with the ATP analogue were incubated for 15 min before DSF measurements. Graph representation and analysis were performed using the Tycho interface of analysis. Three independent runs were

used in each case for calculating mean  $T_i$  values and the corresponding standard deviations.

### 3.4. Isothermal Titration Calorimetry

ATP binding and ATPase activity by the EccC<sub>5</sub><sup>DUF</sup> domain were assessed by isothermal titration calorimetry (ITC) using an ITC-VP instrument from GE Instruments. EccC<sub>5</sub><sup>DUF</sup> at 61  $\mu$ M was loaded in the cell at 25 °C, and it was titrated with ATP at 1.3 mM (in the syringe) using 15 injections, first and second of 1 and 10  $\mu$ l and the remaining of 20  $\mu$ l at intervals of 5 minutes. Both protein and ATP were prepared in buffer C. Analysis of results and final figures were carried out using the Origin 7 ITC Software.

### 3.5. Primary, secondary and tertiary structure analysis *in silico*

A comparative analysis of the putative ATP/Mg<sup>2+</sup>-binding site across DUF domains, and comparison with canonical ATPases, was performed using sequence alignment with Clustal Omega Multi Sequence Alignment (Sievers *et al.*, 2011); and by structure superimposition in PYMOL (Version 2.0; Schrödinger) using experimental coordinates deposited in the Protein Data Bank and models predicted by AlphaFold (Jumper *et al.*, 2021) available at Uniprot. ATPases used for the sequence alignment correspond to: FtsK from *Pseudomonas aeruginosa* (PrgFtsK) (Uniprot entry Q9I0M3; PDB entry 2IUT (Massey *et al.*, 2006)); HerA DNA translocase from *Sulfolobus solfataricus* (Ss/HerA) (UniProt entry Q97WG8; PDB entry 4D2I (Rzechorzek *et al.*, 2014)), MCM helicase from *Pyrococcus furiosus* (PfrMCM) (UniProt entry Q8U3I4; PDB entry 4R7Y (Miller *et al.*, 2014)), Vps4 from *Metallosphaera sedula* (MsdVps4) (UniProt entry A4YHC5; PDB entry 4D81 (Caillat *et al.*, 2015)); EccC from *Thermomonospora*

*curvata* (*TcrEccC*) (UniProt entry D1A4G7); *MtbEccCa*<sub>1</sub> (UniProt entry P9WNB3); *MtbEccC*<sub>2</sub> (UniProt entry O05450); *MtbEccC*<sub>3</sub> (UniProt entry P9WNA9); *MtbEccC*<sub>4</sub> (UniProt entry P9WNA7); *MtbEccC*<sub>5</sub> (UniProt entry P9WNA5; PDB entries 7NPT and 7NPR (Bunduc *et al.*, 2021)); *MxpEccC*<sub>5</sub><sup>DUF</sup> (UniProt entry I0RZI0; PDB entry 7B9S (Beckham *et al.*, 2021)); *MsmEccCa*<sub>1</sub> (UniProt entry A0A653FCM4), *MsmEccC*<sub>3</sub> (UniProt entry A0QQ40; PDB entry 6SGX (Famelis *et al.*, 2019)); *MsmEccC*<sub>4</sub> (UniProt entry A0QSN0); EccC from *Geobacillus thermodenitrificans* (*GthEccC*) (UniProt entry A4IKE7); EssC from *Staphylococcus aureus* (*SrsEssC*) (UniProt entry Q932J9; PDB entry 6TV1); EssC from *Clostridium sp.* (*CspEssC*) (UniProt entry A0A3R6STG9), EssC *Streptococcus oralis* (*SrlEssC*) (UniProt entry A0A656Z0M6) and YukB from *Bacillus subtilis* (*BsbYukB*) (UniProt entry C0SPA7). Structural analysis of dihedral angles and their occurrence in the Cambridge Crystallographic Data Centre (CCDC) database was performed using the software MOGUL (Version 2020.3.0).

## 4. Results and discussion

### 4.1. High-resolution structure of EccC<sub>5</sub><sup>DUF</sup> domain

The overall crystallographic structure of the *MtbEccC*<sub>5</sub><sup>DUF</sup> domain (Fig. 2A) consists of an  $\alpha$ - $\beta$ - $\alpha$  sandwich highly homolog to the  $\alpha/\beta$  fold found in the core of AAA + domains. The core structure consists of a six-stranded parallel  $\beta$ -sheet ( $\beta$ 5- $\beta$ 1- $\beta$ 4- $\beta$ 3- $\beta$ 2- $\beta$ 2'), further extended by an additional two-stranded anti-parallel sheet ( $\beta$ 6- $\beta$ 7). This  $\beta$ -sheet core is flanked by two helices ( $\alpha$ 1 and  $\alpha$ 2) on one side, and four helices ( $\alpha$ 2'A,  $\alpha$ 2'B,  $\alpha$ 3, and  $\alpha$ 4) on the other side. The C-terminal  $\alpha$ -helical lid present in other prokaryotic AAA + domains is missing in the *MtbEccC*<sub>5</sub><sup>DUF</sup> structure, which is flanked by two parallel helices ( $\alpha$ A and  $\alpha$ B) and a two-stranded anti-parallel sheet ( $\beta$ A- $\beta$ B) that connects with the stalk domain. This structure is conserved in both *MtbEccC*<sub>5</sub><sup>DUF</sup> molecules present in the asymmetric unit, with an rmsd 0.173 Å for 218 Ca atoms. Superimposition of crystallographic and cryo-EM (*MtbEccC*<sub>5</sub><sup>CRYO</sup>, PDB entry 7NPR (Bunduc *et al.*, 2021)) models shows overall structure conservation (rmsd=0.671 Å for 198 Ca atoms) except for the N-terminal region (residues 1-16), present but not visible in the crystallographic model, and residues 124-129 and 166-209, connecting to TM and stalk regions, which are truncated in the crystallographic construct (Fig. 2SM). However, despite the overall fold conservation, other differences between our structure and *MtbEccC*<sub>5</sub><sup>CRYO</sup> are found as will be described below.

The proposed ATP-binding site should localise at the N-terminus of the  $\alpha$ 1 helix and the C-terminus of the  $\beta$ 3 strand, expected to allocate the P-loop and the Walker B motif, respectively. Sequence analysis of the Walker A shows a highly degenerated motif ('GEREQVL<sup>231</sup>') exhibiting a 1-aa shorter sequence missing the conserved G<sub>2</sub> and G<sub>3</sub> glycines, and lacking the Lys and Thr/Ser residues key for ATP/Mg<sup>2+</sup>

stabilisation (V230 and L231 in *MtbEccC5<sup>DUF</sup>*) (Fig. 2B). Inspection of *MtbEccC5<sup>DUF</sup>* crystal model shows an ATP-devoid structure where the degenerated Walker A motif folds as an extended  $\alpha 1$  helix instead of the standard P-loop conformation. In this arrangement,  $\alpha 1$  would clash with the nucleotide molecule, as reveals the structural superimposition of *MtbEccC5<sup>DUF</sup>* with FtsK from *P. aeruginosa* (*PrgFtsK*), a close-homologue canonical ATPase. To understand the structural basis causing this extended  $\alpha 1$  structure, we investigated the impact of G<sub>2</sub> and G<sub>3</sub> variations in *MtbEccC5<sup>DUF</sup>*, as invariant flexible features involved both turns of the P-loop motive. To do so, we modelled *in silico* the replacement of G<sub>3</sub> by alanine in *PrgFtsK* (G471A), which shows how the newly added C $\beta$  gives place to an unfavoured dihedral angle C<sub>i</sub>-N<sub>i+1</sub>-C $\alpha$ <sub>i+1</sub>-C $\beta$ <sub>i+1</sub> ( $\theta = 49^\circ$ ) between S470 and A471 residues (Fig. 2D). This is supported by the poor occurrence of this torsion angle when searched in the CCDC database (Fig. 2SM). This disfavoured conformation is prevented in *MtbEccC5<sup>DUF</sup>* by a flip of the E228-Q229 peptide bond leading to the prolonged alpha-helical structure observed in  $\alpha 1$ . Analogously, the absence of G<sub>2</sub> should foster an alpha-helical conformation to avoid non-favourable dihedral angles between the two adjacent non-glycine amino acids. Altogether, this points to G<sub>2</sub> and G<sub>3</sub> variations as structural factors responsible for the extended  $\alpha 1$  helix observed in *MtbEccC5<sup>DUF</sup>*. Inspection of the Walker B motif also shows conformational deviations in *MtbEccC5<sup>DUF</sup>*, which includes a shorter  $\beta 3$ - $\alpha 3$  loop and a hydrophobic residue (V329) replacing the catalytic base (Fig. 2B). V329 rests partially buried between  $\beta 3$ - $\alpha 3$  and  $\beta 4$ - $\alpha 4$  loops in a conformation well stabilised by an H-bond network with nearby residues. This drags the Walker B away from the P-loop by  $> 2 \text{ \AA}$  compared to *PrgFtsK*, thus providing the room necessary to allocate the extended  $\alpha 1$ . This non-canonical arrangement of both Walker motives is further stabilized by numerous direct and water-mediated interactions. These include an H-

bond between the side-chain amide group of Q229 and the main-chain nitrogen of E226, and a salt-bridge formed between R227 and D328 side chains, stabilising the extended  $\alpha 1$  structure, which are well defined in the electron density maps (Fig. 2SM). Moreover, the short bond distances measured for the salt-bridge interaction indicate this is a strong contact that, altogether, should further challenge structural rearrangements of this non-productive conformation observed in  $\alpha 1$  for ATP binding. Superimposition of crystal and cryo-EM structures of *MtbEccC*<sub>5</sub> shows overall conservation of the non-canonical arrangement observed in the Walker motives in both models (Fig. 2SM). However, significant differences were found in structural features shaping the architecture of the degenerated ATP-binding site, supported by the electron density and cryo-EM maps (Fig. 2SM). These comprise many direct or water-mediated H-bonds stabilising Walker and Sensor 1 motives, including the interaction between Q229 and E226 that is too long to form an H-bond in *MtbEccC*<sub>5</sub><sup>CRYO</sup>; or the water-mediated contact established between R227 and D258 side chains, which seems to assist R227 side chain to adopt an optimal orientation to interact with D328 and was not observed in the cryo-EM model. Moreover, contrasting with *MtbEccC*<sub>5</sub><sup>CRYO</sup>, the crystal structure of *MtbEccC*<sub>5</sub><sup>DUF</sup> shows a bidentate salt-bridge between R227 and D328 side chains well defined on the electron density maps that contributes to optimise the interactions stabilising the non-canonical configuration of the Walker A motif (Fig. 2SM). Thus, our structure provides an unambiguous model defining the non-canonical ATP-binding site of *MtbEccC*<sub>5</sub> at high resolution, which has allowed us to acquire a detailed description and understanding of key structural features contributing to its degenerated arrangement.

We next analysed the regions expected to allocate the *trans*-acting elements in *MtbEccC*<sub>5</sub><sup>DUF</sup> (Fig. 2SM). No basic residues were observed at the end of  $\alpha 4$  as



candidates to act as Arg-finger. Only R362 was found in the vicinity and further down the  $\alpha 4$  sequence. However, its quite buried position in the structure makes it a poor candidate for functioning as an Arg finger. The helical C-terminal lid, habitually found in prokaryotic AAA + clades, that should allocate Sensor 2 is missing in *MtbEccC5*<sup>DUF</sup>. Besides, inspection of the  $\alpha 3$  helix shows no basic residues in this region as potential Sensor 3. Taking all these observations together, our structural analysis of *MtbEccC5*<sup>DUF</sup> shows the degeneration and absence of *cis*- and *trans*-acting elements in *MtbEccC5*<sup>DUF</sup>, respectively, which would support the lack of ATPase activity in this domain. These observations are in line with our failed attempts to co-crystallise *MtbEccC5*<sup>DUF</sup> with ATP/Mg<sup>2+</sup>, and structures devoid of nucleotide and magnesium obtained instead.

## 4.2. ATP-binding and ATP-hydrolysis analysis

The structural arrangement of the putative ATP-binding site of *MtbEccC5*<sup>DUF</sup> resembles the nucleotide-binding pocket observed in the structure of the *SrsEssC*<sup>D3</sup> domain (PDB entry 6TV1), reported to bind ATP with an affinity in the low millimolar range (Mietrach *et al.*, 2020) (Fig. 3A). This poor but detectable binding points to potential structural rearrangements of the ATP-binding site of *SrsEssC*<sup>D3</sup> enabling the interaction with the nucleotide. Therefore, we analysed the interaction of *MtbEccC5*<sup>DUF</sup> with ATP/Mg<sup>2+</sup> in order to explore a possible reorganization of the ATP-binding site that might enable nucleotide binding and hydrolysis in solution using two different biophysical approaches.

Firstly, the interaction of *MtbEccC5*<sup>DUF</sup> with ADP-AlF<sub>3</sub> (used as ATP analogue) was assessed by DSF in the presence of magnesium. Despite the high ADP-AlF<sub>3</sub>:*EccC5*<sup>DUF</sup> concentration used (~ 4mM), the experiments showed that the difference between *T<sub>i</sub>*

values in the presence ( $59.13 \pm 0.04$  °C) and in the absence ( $60.46 \pm 0.04$  °C) of nucleotide was slightly negative (around  $-1.3$  °C) (Fig. 3B) and very close to the device's error limit, thus pointing to the absence of binding. Nucleotide interaction and ATP hydrolysis were next assessed by ITC in the presence of magnesium. The heat released upon titration of ATP into *MtbEccC<sub>5</sub><sup>DUF</sup>* was comparable to the thermal effect of ATP dilution in the range of concentrations tested (Figure 3C). Additionally, no thermal power deflection (proportional to the potential substrate concentration in the cell) was observed after each ATP injection, as would be expected if nucleotide was being hydrolysed (Menéndez, 2020). Altogether ITC and DSF results supports the lack of ATPase activity by *MtbEccC<sub>5</sub><sup>DUF</sup>*.

### 4.3. Nucleotide-binding site in other EccC<sup>DUF</sup> domains

Given the structural degeneration observed in the *MtbEccC<sub>5</sub><sup>DUF</sup>* structure, and its observed inability to interact with ATP using X-ray crystallography, DSF and ITC, we decided to examine the putative ATP-binding site in other DUF domains. With this purpose, homologue EccC<sup>DUF</sup> and orthologue EssC<sup>DUF</sup> domains from mycobacterial and non-mycobacterial species were used for sequence alignment (Fig. 4A), which showed degeneration of the Walker A motif in all the domains inspected. Analogously to *MtbEccC<sub>5</sub><sup>DUF</sup>*, these motifs present a 1-aa shorter sequence, where conserved residues G<sub>2</sub> and G<sub>3</sub> are missing and substituted by a non-glycine amino acid, respectively. Also, they lack both the lysine and threonine/serine residues required for ATP/Mg<sup>2+</sup> interaction. DUF domains inspected also exhibit degeneration of their Walker B motives, where the catalytic glutamate is replaced by other hydrophobic or polar amino acids (Rosenberg *et al.*, 2015; Zoltner *et al.*, 2016; Wang *et al.*, 2020). Interestingly, analysis of *MxpEccC<sub>5</sub><sup>DUF</sup>* and *MsmEccC<sub>3</sub><sup>DUF</sup>* structures showed that their

Walker motifs adopt a similar arrangement to the one observed in *MtbEccC5*<sup>DUF</sup>, including an extended  $\alpha 1$  helix (Fig. 4A). Analogously to *MtbEccC5*<sup>DUF</sup>, the prolonged  $\alpha 1$  helix can be explained by the amino acid variations observed at G<sub>2</sub> and G<sub>3</sub> positions with respect to canonical ATPases, which should elicit a helical conformation versus the P-loop structure. This structural reorganisation was also found in the structural models predicted for *MtbEccC1a*<sup>DUF</sup>, *MtbEccC2*<sup>DUF</sup> and *MtbEccC4*<sup>DUF</sup> and other six DUF domains, all of which present the same extended conformation of  $\alpha 1$  (Fig. 4 and 4SM). Inspection of Walker motifs in *MxpEccC5*<sup>DUF</sup> and *MsmEccC3*<sup>DUF</sup> structures also reveal an arginine-aspartate salt-bridge connecting  $\alpha 1$  with the C-terminus of  $\beta 3$  (Fig. 4B). The salt-bridge observed in *MxpEccC5*<sup>DUF</sup> involves an arginine at the N-terminus of  $\alpha 1$ , analogously to *MtbEccC5*<sup>DUF</sup>, while the salt bridge in *MsmEccC3*<sup>DUF</sup> is formed by an arginine substituting the canonical threonine/serine of the Walker A motif (Fig. 4A). Out of the fifteen sequences analysed, five mycobacterial DUF domains exhibit an arginine/lysine and aspartate that could potentially form a salt bridge, also predicted in the structural models of *MtbEccC2*<sup>DUF</sup> and *MtbEccC4*<sup>DUF</sup>, where it is likely to contribute to stabilise this unproductive arrangement for ATP-Mg<sup>2+</sup> recognition (Fig. 4B). Thus, our analysis shows that the non-canonical structure observed in *MtbEccC5*<sup>DUF</sup> Walker motives is present in the cryo-EM structures of *MxpEccC5*<sup>DUF</sup> and *MsmEccC3*<sup>DUF</sup>, as well as in the predicted models available for other DUF of T7Sa and T7Sb systems. Altogether these observations support that DUFs are degenerated ATPase domains exhibiting a non-functional structure for ATP-binding and thus a poor or null ATPase activity.

#### 4.4. Hexameric model of *MtbEccC5* N-terminal region

The lack of ATPase function in EccC<sup>DUF</sup> domains here proposed raises important questions about their role in the secretion process. The relevance of these questions is particularly stressed by site-directed mutagenesis studies conducted on *MsmESX3*, where mutations of both aspartate residues of the Walker B motif of *MsmEccC3*<sup>DUF</sup> to alanine (D319A and D320A) were shown to abrogate secretion (Famelis *et al.*, 2019). This observation leads us to speculate that these mutations might affect protein:protein interactions that are key for secretion, and would be established between the DUF domain and other cytosolic components, or between adjacent DUF domains. Given the conserved ATPase-like fold of DUF, it is plausible to expect that this domain has retained the ability to multimerise, analogously to FtsK homologues and as proposed for EccC<sup>D1-3</sup> domains as well (Rosenberg *et al.*, 2015; Miller & Enemark, 2016). This hypothesis aligns with the size-exclusion chromatography profile observed during the production of *MtbEccC5*<sup>DUF</sup>, which indicates the presence of monomeric, dimeric and hexameric species (Fig. 5SM). To evaluate this hypothesis, and based on the multimerised structure of *PrgFtsK* (PDB entry 2IUU), we modelled the structure of the N-terminal region of *MtbEccC5* in an hexameric state, including TM, stalk and DUF domains, using the crystal coordinates of *MtbEccC5*<sup>DUF</sup> and *MtbEccC5*<sup>CRYO</sup> (PDB entry 7NPR) (Fig. 5A). According to this model, *MtbEccC5*<sup>DUF</sup> domains would form a closed ring, with stalk and TM regions located perpendicular to the DUF domains. Analysis of the interface between adjacent DUFs shows that both Walker A, B and Sensor1 motives are located near the neighbouring monomer, but they do not participate in inter-domain contacts. However, direct interactions via residues at  $\beta$ 2- $\alpha$ 2 loop,  $\alpha$ 2 and  $\beta$ 5-  $\beta$ 6 loop regions were identified. Interestingly, the  $\beta$ 2- $\alpha$ 2 region locates adjacent to Walker A and B motives, with which it forms direct interactions through several residues that include R227 and D328 (Fig. 5B). Based on

these interactions, mutations in R227 and D328 and/or other amino acids of both Walker motives might impact the structure of the  $\beta$ 2- $\alpha$ 2 region and, by extension, *MtbEccC*<sub>5</sub><sup>DUF</sup> multimerisation. Of note, this would explain as well the lack of secretion observed in both Walker B mutants, D319A and D320A, of *MsmEccC*<sub>3</sub><sup>DUF</sup> (Famelis et al., 2019), that correspond to Walker B residues D328 and V329 in *MtbEccC*<sub>5</sub>, respectively. Regarding TM and stalk regions, they are located perpendicularly, at the edge of the ring formed by *MtbEccC*<sub>5</sub><sup>DUF</sup> domains, with the TM regions of opposite monomers separated by a distance of ~86 Å. This disposition notably contrasts with EccC<sub>3</sub> and EccC<sub>5</sub> organization observed in *MsmESX*3, *MtbESX*5 and *MxpESX*5 structures, where the TM regions multimerise to form a helical bundle that connects, via stalk region, with the DUF domains laying away from each other (Famelis et al., 2019; Bunduc et al., 2021; Beckham et al., 2021) (Fig. 5C). Of note, the TM helical bundle locates at the centre of the membrane region where it contributes to a close conformation to the membrane pore. Thus, the aperture of TM and stalk regions depicted by our model suggests a potential conformational change of the N-terminal region of *MtbEccC*<sub>5</sub><sup>DUF</sup> that would explain how the membrane pore opens through the hexamerisation of the DUF domain (Fig. 5C). Altogether, our *MtbEccC*<sub>5</sub> model suggests that DUF hexamerisation may play a critical role in the aperture of the membrane pore. This multimerisation would be triggered by the interplay of DUFs with EccC<sup>D1-3</sup> ATPase domains, thus linking ATP hydrolysis and effector recognition with the opening of the membrane pore necessary for secretion. Certainly, the function here proposed for DUF domains will request further studies to be verified, and raises additional outstanding questions that require our attention, such as how ATPase hydrolysis, effector recognition and multimerization of D1-3 ATPase domains are

coupled with the proposed DUF hexamerisation and the aperture of the membrane pore.

## 5. Conclusions

Here we report the crystallographic structure of the *MtbEccC*<sub>5</sub><sup>DUF</sup> domain at high resolution, which reveals an ATP/Mg<sup>2+</sup>-free structure with both high degeneration and absence of *cis* and *trans*-acting motives required for ATP-binding and hydrolysis, respectively. Among the most remarkable features, we find the absence of the catalytic glutamate required for the nucleophilic attack over the ATP, and a non-canonical fold of the Walker A motif. Instead of the typical P-loop conformation, the Walker A motif adopts an extended alpha-helical arrangement that would clash with the nucleotide. Although overall conservation of this arrangement is also observed in *MtbEccC*<sub>5</sub><sup>CRYO</sup>, our crystal structure provides unseen and unambiguously defined interactions shaping the non-canonical ATP-binding site observed. Moreover, our structural analysis of *MtbEccC*<sub>5</sub><sup>DUF</sup> has led us to identify the structural features underlying this arrangement. They include the variations of G<sub>2</sub> and G<sub>3</sub> observed in the Walker A motif of *MtbEccC*<sub>5</sub><sup>DUF</sup>, which introduce geometrical restraints that favour the extended α1 helical structure instead of the canonical P-loop conformation. We also validate the lack of interaction of *MtbEccC*<sub>5</sub><sup>DUF</sup> with the nucleotide in solution using DSF and ITC, which aligns with our futile attempts to co-crystallise this domain with ATP-Mg<sup>2+</sup> that further supports the inability of the DUF domain to bind ATP. Our study also show that the degenerated ATP-binding site observed in *MtbEccC*<sub>5</sub><sup>DUF</sup> is present in the experimental and predicted structures of other homologue domains from T7Sa and T7Sb systems, which supports the notion that DUFs are degenerated ATPase domains, non-functional for nucleotide interaction/hydrolysis. DUF domains emerge,

however, as pivotal elements in the secretion process, as illustrated by the mutagenesis studies of the Walker B motif of *MsmEccC*<sub>3</sub><sup>DUF</sup> leading to abrogation of the *MsmESX3* secretion. Putting all these observations together and based on a structural model of *MtbEccC*<sub>5</sub> in a hexameric form, we proposed DUF domains as key structural elements involved in the opening of the membrane pore during the secretion process. This hypothesis, which requires of further studies to be validated, also opens new important questions such as how ATPase hydrolysis, effector recognition and multimerization of D1-3 ATPase domains couple with the proposed hexamerisation of DUF domains and the aperture of the membrane pore.

## 6. Acknowledgments

This work was funded by the *Atracción de Talento (Modalidad 1)* program from the Regional Government of Madrid (Spain) (reference 2019-T1/BMD-14774). We also thank the Synchrotron Radiation Facility ALBA (Barcelona, Spain) for the time allocated for X-ray data collection under proposal 2022096995, and BL13-XALOC beamline staff for their assistance in data collection. M. Menendez was supported by the CIBER of Respiratory Diseases (CIBERES), an initiative from the Spanish Institute of Health Carlos III (ISCIII). We also want to thank Dr. Lourdes Infantes San Mateo from the IQF-CSIC, for her input in the structural analysis presented in this work.



## Legends

**Figure 1.** (A) *Mtb*ESX5 model based on cryo-EM reported structures, showing the proposed conformational alteration of EccC<sub>5</sub> between an open/extended (left) and a close/contracted conformation (right). (B) Architecture of EccC<sub>5</sub> cytosolic region consisting of the DUF domain followed by three ATPase domains D1, D2 and D3 interconnected by connectors referred to as Linker1-3. (C) Superimposition of *SrsEssC*<sup>D3</sup> crystal structure (PDB entry 6VT1, in blue) with *MtbEccC*<sub>5</sub><sup>DUF</sup> (PDB entry 7NPR, dark grey), *MxpEccC*<sub>5</sub><sup>DUF</sup> (PDB entry 7B9S, light grey) and *MsmEccC*<sub>3</sub><sup>DUF</sup> (PDB entry 6SGX, white) cryo-EM models showing conservation of the overall fold (left) and similar arrangement of both Walker motives (right).

**Figure 2.** (A) Sequence alignment of Walker A and B motives of *MtbEccC*<sub>5</sub><sup>DUF</sup> with the canonical ATPase domain of *PrgFtsK*, *SsHerA*, *PfrMCM*, *MsdVps4*, and *MtbEccC*<sub>1-5</sub><sup>D1</sup>, showing the consensus sequence at the top. Key amino acids involved in nucleotide interaction/hydrolysis are highlighted in pink and blue, respectively. (B) A general view of the *MtbEccC*<sub>5</sub><sup>DUF</sup> crystal structure represented as green cartoons is shown (left), as well as details of the putative ATP-binding site (right). Salt-bridge and H-bond interactions are indicated as dashed lines in blue and black, respectively, as well as the bond distances measured between R227 and D328, and E226 and Q229. Walker A, B and Sensor 1 motives are coloured in pink, blue and orange respectively, showing key residues (sticks) and water molecules (red spheres). (C) Superimposition of *MtbEccC*<sub>5</sub><sup>DUF</sup> and *PrgFtsK* (PDB entry 4R7Y, in beige) structures represented as cartoons. Walker A, B and Sensor 1 motives are coloured in light (*MtbEccC*<sub>5</sub><sup>DUF</sup>) and dark (*PrgFtsK*) blue, pink and orange, respectively. In *PrgFtsK*, magnesium cation, ATP, and residues key for the interaction with the nucleotide are represented as a

purple sphere, lines, and sticks, respectively. In The R227-D328 salt bridge observed in *MtbEccC5*<sup>DUF</sup>, and H-bonds involved in ATP-Mg<sup>2+</sup> stabilization in *PrgFtsK* are represented as dashed lines in black and blue, respectively. (D) Detail of the Walker A motif in *MtbEccC5*<sup>DUF</sup> and *PrgFtsK*, showing the favourable dihedral angle C<sub>i</sub>-N<sub>i+1</sub>-C<sub>αi+1</sub>-C<sub>βi+1</sub> observed between E228 and Q229 in the former, versus the non-favourable dihedral angle generated between S470 and G471 of *PrgFtsK* when the glycine is substituted by an alanine residue.

**Figure 3.** (A) Superimposition of *MtbEccC5*<sup>DUF</sup> (in green) and *SrsEssC*<sup>D3</sup> (PDB entry 6VT1, in dark blue) crystal structures showing a detail of their degenerated ATP-binding sites. Residues occupying key positions for nucleotide/magnesium interaction and hydrolysis are labelled and represented as sticks. Walker A, B and Sensor 1 motives of *MtbEccC5*<sup>DUF</sup> are coloured following the same colour code as in Figure 2. (B) Superimposition of *MtbEccC5*<sup>DUF</sup> DSF profiles in the absence (control) and presence of ADP-AlF<sub>3</sub> (4 mM) and Mg<sup>2+</sup> (5 mM). The failure of ADP-AlF<sub>3</sub>/Mg<sup>2+</sup> to stabilize the domain structure against thermal denaturation was consistent with no nucleotide binding. (C) ITC titration of ATP (1.3 mM) into *EccC5*<sup>DUF</sup> (60 μM) and buffer C (black and blue data, respectively). Upper and lower panels show raw thermograms, and normalized heat effect per mole of ATP injected vs. ATP/*EccC5*<sup>DUF</sup> molar ratio, respectively.

**Figure 4.** Analysis of putative ATP/Mg<sup>2+</sup>-binding site in *EccC*<sup>DUF</sup> domains of ESX1-5 systems from *Mtb*, *Msm* and *Mxp*. (A) Sequence alignment of Walker A and B motives coloured in blue and pink, respectively. (B) Structural details of Walker A and B motives observed in *MtbEccC5*<sup>DUF</sup> (crystal structure), *MxpEccC5*<sup>DUF</sup> (PDB entry 7B9S), *MsmEccC3*<sup>DUF</sup> (PDB entry 6GSX), *MtbEccC1*<sup>DUF</sup> (AlphaFold model, Uniprot entry

P9WNB3), *MtbEccC<sub>2</sub>*<sup>DUF</sup> (AlphaFold model, Uniprot entry O05450), and *MtbEccC<sub>4</sub>*<sup>DUF</sup> (AlphaFold model, Uniprot entry P9WNA7) represented in cartoon and following the same colour code as in (A). Amino acids substituting key residues for ATP-Mg<sup>2+</sup> interaction, or involved in H-bond/salt-bridge interactions are depicted as sticks.

**Figure 5.** (A) Structural comparison of EccC<sub>5</sub> arrangement observed in *MtbESX5* structure (PDB entry 7NPR) with our model of *MtbEccC<sub>5</sub>* N-terminal region multimerising as a hexamer. A surface representation of *MtbEccC<sub>5</sub>* molecules is shown, with transmembrane, stalk and DUF domains coloured in yellow, grey and green, respectively. Walker A, B and Sensor1 motives are coloured in cyan, pink, and orange respectively. A zoom view showing the molecular interface between adjacent DUF monomers in our model is shown at the right box, where identified contact regions have been highlighted in yellow. (B) Details of the interface between DUF1 (grey) and DUF2 (colour code as in A) adjacent molecules is shown, with residues establishing contacts between both monomers and with Walker A and B motives represented as sticks, and their corresponding interactions are indicated as dashed lines. Water molecules are represented as red spheres. (C) Model showing the aperture of the TM region of *MtbEccC<sub>5</sub>* associated to the proposed multimerization of *MtbEccC<sub>5</sub>*<sup>DUF</sup>, where this domain would transit from a non-multimerised form to an hexameric state to enable the opening of the membrane pore.

**Table 1. X-ray crystallography statistics**

	<i>MtbEccC<sub>5</sub></i> <sup>DUF</sup>
Data collection statistics	
Space group	<i>P</i> 1
Number of complexes in ASU	2
<i>a</i> , <i>b</i> , <i>c</i> (Å)	48.30, 52.24, 57.44
α, β, γ (°)	88.39, 84.60, 80.17
Wavelength (Å)	0.9792
Beamline	BL13-XALOC
Resolution range (Å)	57.18 - 2.05 (2.11 - 2.05)
<i>R</i> <sub>merge</sub>	0.055 (0.354)
<i>R</i> <sub>pim</sub>	0.035 (0.242)
Number of total reflections	114509 (6654)
Number of unique reflections	32869 (2122)
Mean ( <i>I</i> )/SD( <i>I</i> )	9.6 (2.6)
CC <sub>1/2</sub>	0.998 (0.869)
Completeness (%)	95.1 (77.9)
Multiplicity	3.5 (3.1)
Wilson plot B (Å <sup>2</sup> )	29.5
Refinement	
Number of atoms	4219
Protein	3932
Water	287
<i>R</i> <sub>work</sub> (%)	0.19
<i>R</i> <sub>free</sub> (%)	0.22
<B> (Å <sup>2</sup> )	34.59
rms bond length deviation (Å)	0.008
rmsd angle deviation (°)	0.88

Values in parentheses correspond to the highest resolution shell.

$$R_{\text{merge}} = \frac{\sum (|I_h| - \langle I_h \rangle)}{\sum \langle I_h \rangle}$$

$$R_{\text{meas}} = \frac{\sum \sqrt{(n_h/n_h - 1)} (|I_h| - \langle I_h \rangle)}{\sum \langle I_h \rangle}$$

$$R_{\text{pim}} = \frac{\sum \sqrt{(1/n_h - 1)} (|I_h| - \langle I_h \rangle)}{\sum \langle I_h \rangle}$$

## Bibliography

- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst D* **66**, 213–221.
- Beckham, K. S. H., Ritter, C., Chojnowski, G., Ziemianowicz, D. S., Mullapudi, E., Rettel, M., Savitski, M. M., Mortensen, S. A., Kosinski, J. & Wilmanns, M. (2021). *Sci Adv* **7**, eabg9923.
- Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P, Sharff A., Smart O.S., Vonrhein C. & Womack T.O. (2017). BUSTER v.2.11.2, <https://www.globalphasing.com/>.
- Bunduc, C. M., Bitter, W. & Houben, E. N. G. (2020). *Annu Rev Microbiol* **74**, 315–335.
- Bunduc, C. M., Fahrenkamp, D., Wald, J., Ummels, R., Bitter, W., Houben, E. N. G. & Marlovits, T. C. (2021). *Nature* **593**, 445–448.
- Caillat, C., Macheboeuf, P., Wu, Y., McCarthy, A. A., Boeri-Erba, E., Effantin, G., Göttlinger, H. G., Weissenhorn, W. & Renesto, P. (2015). *Nat Commun* **6**, 8781.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2009). *Acta Cryst D* **66**, 12–21.
- Crosskey, T. D., Beckham, K. S. H. & Wilmanns, M. (2020). *Prog Biophys Mol Biol* **152**, 25–34.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst D* **66**, 486–501.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst D* **69**, 1204–1214.

- Famelis, N., Geibel, S. & Van Tol, D. (2023). *Biol Chem* **404**, 691–702.
- Famelis, N., Rivera-Calzada, A., Degliesposti, G., Wingender, M., Mietrach, N., Skehel, J. M., Fernandez-Leiro, R., Böttcher, B., Schlosser, A., Llorca, O. & Geibel, S. (2019). *Nature* **576**, 321–325.
- Houben, E. N. G., Korotkov, K. V. & Bitter, W. (2014). *Biochimica et Biophysica Acta* **1843**, 1707–1716.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). *Nature* **596**, 583–589.
- Kabsch, W. (2010). *Acta Cryst D* **66**, 125–132.
- Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. (2002). *J Mol Biol* **317**, 41–72.
- Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B. K. & Gao, N. (2015). *Nature* **524**, 186–191.
- Massey, T. H., Mercogliano, C. P., Yates, J., Sherratt, D. J. & Löwe, J. (2006). *Mol Cell* **23**, 457–469.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J Appl Crystallogr* **40**, 658–674.
- Menéndez, M. (2020). *Encyclopedia of Life Sciences* **1**, 113–127.
- Mietrach, N., Damián-Aparicio, D., Mielich-Süss, B., Lopez, D. & Geibel, S. (2020). *J Bacteriol* **202**, 646–665.
- Miller, J. M., Arachea, B. T., Epling, L. B. & Enemark, E. J. (2014). *Elife* **3**, e03433.

- Miller, J. M. & Enemark, E. J. (2016). *Archaea* **2016**, ID 9294307.
- Ogura, T., Whiteheart, S. W. & Wilkinson, A. J. (2004). *Journal of Structural Biology* **146**, 106–112.
- Poweleit, N., Czudnochowski, N., Nakagawa, R., Trinidad, D., Murphy, K. C., Sassetti, C. & Rosenberg, O. S. (2019). *Elife* **8**, e52983.
- Rosenberg, O. S., Dovala, D., Li, X., Connolly, L., Bendebury, A., Finer-Moore, J., Holton, J., Cheng, Y., Stroud, R. M. & Cox, J. S. (2015). *Cell* **161**, 501–512.
- Rzechorzek, N. J., Blackwood, J. K., Bray, S. M., Maman, J. D., Pellegrini, L. & Robinson, N. P. (2014). *Nat Commun* **5**, 5506.
- Schmidt, H., Gleave, E. S. & Carter, A. P. (2012). *Nature Structural & Molecular Biology* **19**, 492–497.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). *Mol Syst Biol* **7**, 539.
- Wang, S., Zhou, K., Yang, X., Zhang, B., Zhao, Y., Xiao, Y., Yang, X., Yang, H., Guddat, L. W., Li, J. & Rao, Z. (2020). *Protein Cell* **11**, 124–137.
- World Health Organization (2022). Global Tuberculosis Report 2020, <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022>.
- Zoltner, M., Ng, W. M. A. V., Money, J. J., Fyfe, P. K., Kneuper, H., Palmer, T. & Hunter, W. N. (2016). *Biochem J* **473**, 1941–1952.

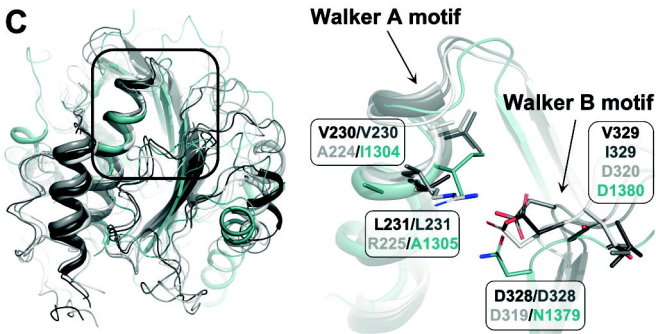
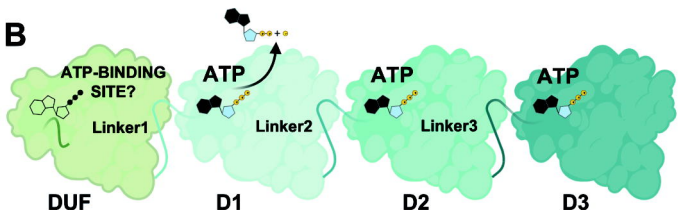
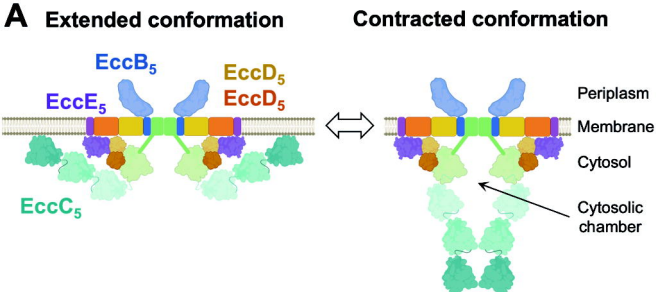
## Legends. Supplemental material figures

**Figure 2SM.** (A) Structural superimposition of both crystallographic and cryo-EM *MtbEccC<sub>5</sub><sup>DUF</sup>* models coloured in Figure 2 and grey, respectively. In the case of *MtbEccC<sub>5</sub><sup>CRYO</sup>*, N-terminal, stalk and TM regions are coloured in red, dark grey and yellow, respectively. A zoomed view showing the structural conservation of both the Walker A and B motives in both models is shown (right), including the salt-bridge R227-D328 and the H-bond interaction E226-Q229 indicated as dashed lines coloured in blue (crystal model) and grey (cryo-EM model). (B) Mogul search results obtained for torsion angles C<sub>S470</sub>-N<sub>A471</sub>-C<sub>α A471</sub>-C<sub>β A471</sub> of *PrgFtsK*, when G471 is substituted by an alanine, and C<sub>E228</sub>-N<sub>Q229</sub>-C<sub>α Q229</sub>-C<sub>β Q229</sub> in *MtbEccC<sub>5</sub><sup>DUF</sup>* crystal structure. (C) and (D) Detail of both Walker A and B motives in the *MtbEccC<sub>5</sub><sup>DUF</sup>* crystal structure showing the 2Fo-Fc electron density map (contoured at a sigma value of 1.0). Colour code as in Figure 1. (D) Detail of both Walker A and B motives in *MtbEccC<sub>5</sub><sup>DUF</sup>* model reported by cryo-EM showing the EM map at 3.27 Å resolution (PDB entry 7NPT) (contoured at a level value of 10.0). (E) Cartoon and surface representation *MtbEccC<sub>5</sub><sup>DUF</sup>*, showing details of the regions expected to allocate Arg-finger and Sensor 3 motives (highlighted in orange).

**Figure 4SM.** (A) Structural superimposition showing a detail of the Walker A and B motives predicted for *MtbEccC<sub>5</sub><sup>DUF</sup>*, *TcrEccC<sup>DUF</sup>*, *GthEccC<sup>DUF</sup>*, *SrsEssC<sup>DUF</sup>*, *BsbEssC<sup>DUF</sup>*, *CspEssC<sup>DUF</sup>*, and *SrlEssC<sup>DUF</sup>*. (B) Size-exclusion chromatography profile obtained during the purification of the *MtbEccC<sub>5</sub><sup>DUF</sup>* construct used in this study.

**Figure 5SM.** (A) Size-exclusion chromatography profile obtained during the purification of the *MtbEccC<sub>5</sub><sup>DUF</sup>* construct used in this study.





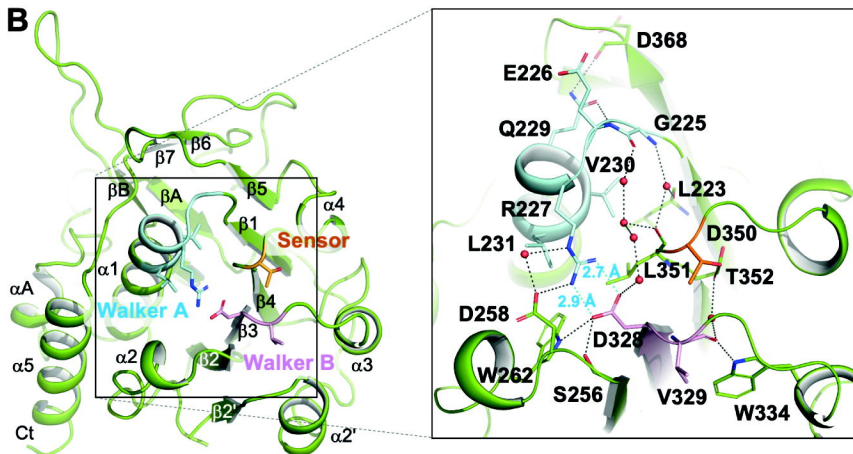
## Walker A

<i>MtbEccC</i> <sub>DUF</sub>	G	E	R	-	E	Q	V	L	(231)
<i>PrgFtsK</i>	G	T	T	G	S	G	G	K	(473)
<i>SslHerA</i>	S	T	A	G	S	G	K	S	(154)
<i>PfrMCM</i>	G	D	P	G	V	A	K	S	(1335)
<i>MsdVps4</i>	G	P	P	G	S	G	K	T	(145)
<i>TcvEccA</i> <sub>ATPaseI</sub>	G	A	T	G	S	G	K	S	(486)
<i>MtbEccC</i> <sub>ATPaseI</sub>	G	T	T	G	S	G	K	S	(486)
<i>MtbEccC</i> <sub>ATPaseI</sub>	G	T	S	G	S	G	K	S	(530)
<i>MtbEccC</i> <sub>3ATPaseI</sub>	G	M	T	G	S	G	K	S	(479)
<i>MtbEccC</i> <sub>ATPaseI</sub>	G	A	T	G	S	G	K	S	(437)
<i>MtbEccC</i> <sub>ATPaseI</sub>	G	T	T	G	S	G	K	S	(506)

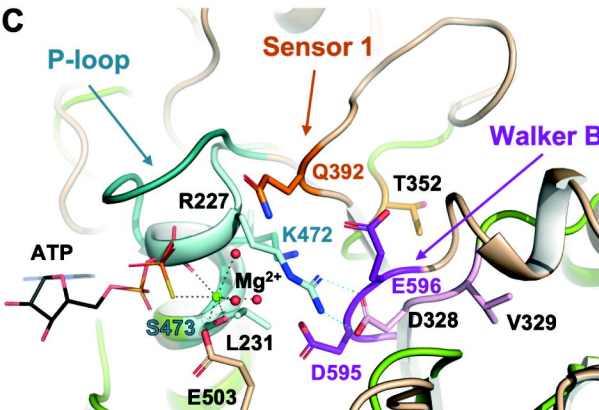
## Walker B

<i>MtbEccC<sub>DUF</sub></i>	V	I	I	A	D	V	(329)
<i>PrgFtsK</i>	V	V	V	V	D	E	(596)
<i>SslHerA</i>	I	A	V	I	E	E	(356)
<i>PfrMCM</i>	Y	A	L	I	D	E	(1760)
<i>MsdVps4</i>	I	I	F	I	D	E	(203)
<i>TcvEccC<sub>ATPaseI</sub></i>	F	I	V	L	D	E	(593)
<i>MtbEccC<sub>ATPaseI</sub></i>	F	V	V	V	D	E	(600)
<i>MtbEccC<sub>ATPaseI</sub></i>	L	V	I	V	D	E	(637)
<i>MtbEccC<sub>3</sub><sub>ATPaseI</sub></i>	F	V	V	A	D	E	(599)
<i>MtbEccC<sub>ATPaseI</sub></i>	F	I	V	V	D	E	(544)
<i>MtbEccC<sub>ATPaseI</sub></i>	F	V	V	V	D	E	(600)

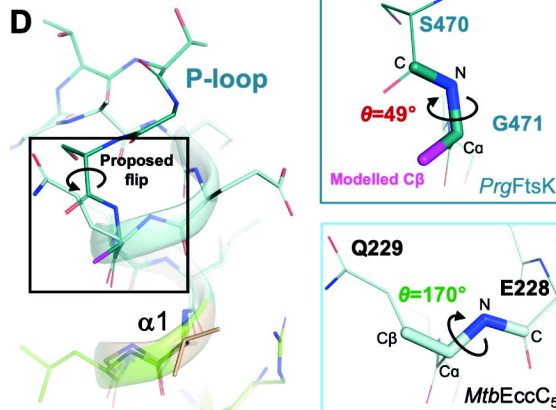
# B

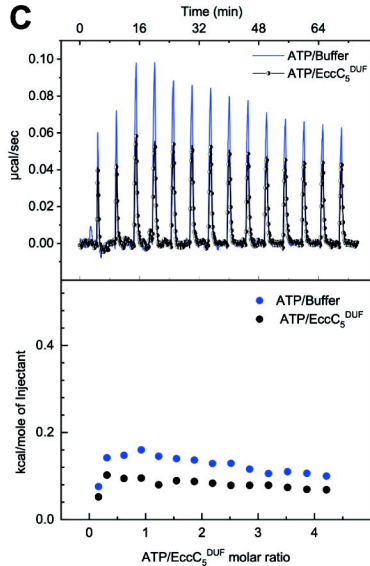
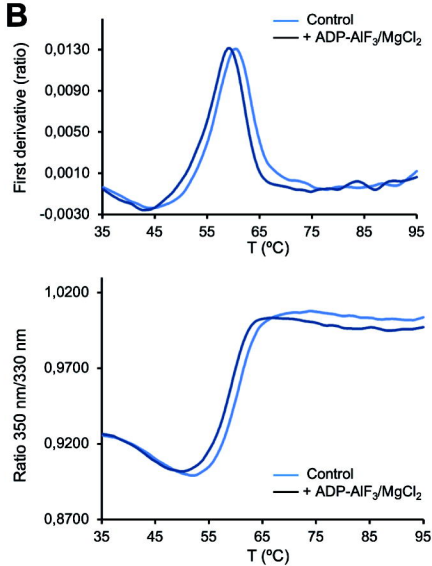
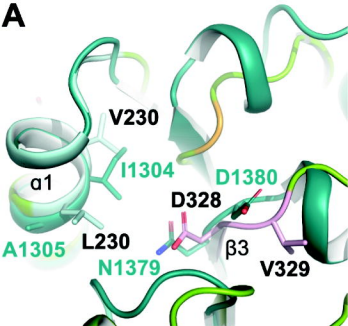


# C



## D





**A Walker A**

	G <sub>1</sub>	x	x	G <sub>2</sub>	x	G <sub>3</sub>	K	T	(aa)
								S	
<i>MtbEccC<sub>5</sub></i> <sup>DUF</sup>	G	E	R	-	E	Q	V	L	(231)
<i>MxpEccC<sub>5</sub></i> <sup>DUF</sup>	G	E	R	-	E	Q	V	L	(231)
<i>MtbEccCa<sub>1</sub></i> <sup>DUF</sup>	G	D	L	-	A	G	A	A	(235)
<i>MsmEccCa<sub>1</sub></i> <sup>DUF</sup>	G	D	P	-	E	A	A	E	(324)
<i>MtbEccC<sub>2</sub></i> <sup>DUF</sup>	S	E	D	L	D	R	V	R	(230)
<i>MtbEccC<sub>3</sub></i> <sup>DUF</sup>	G	E	R	-	A	Q	V	R	(224)
<i>MsmEccC<sub>3</sub></i> <sup>DUF</sup>	G	E	A	-	D	E	A	R	(225)
<i>MtbEccC<sub>4</sub></i> <sup>DUF</sup>	G	D	P	-	T	K	V	R	(216)
<i>MsmEccC<sub>4</sub></i> <sup>DUF</sup>	G	P	D	-	D	Q	V	A	(217)
<i>TcrEccC</i> <sup>DUF</sup>	G	D	P	-	K	A	V	Y	(224)
<i>GthEccC</i> <sup>DUF</sup>	G	K	R	-	S	I	V	N	(421)
<i>SrsEssC</i> <sup>DUF</sup>	G	A	R	-	H	L	I	L	(416)
<i>BsbEssC</i> <sup>DUF</sup>	G	K	S	-	Q	I	V	K	(431)
<i>CspEssC</i> <sup>DUF</sup>	G	S	E	-	K	G	R	F	(335)
<i>SrlEssC</i> <sup>DUF</sup>	G	P	R	-	N	L	V	L	(176)

**Walker B**

<i>MtbEccC<sub>5</sub></i> <sup>DUF</sup>	V	I	I	A	D	V	(329)
<i>MxpEccC<sub>5</sub></i> <sup>DUF</sup>	V	I	I	S	D	I	(329)
<i>MtbEccCa<sub>1</sub></i> <sup>DUF</sup>	V	V	V	V	D	L	(325)
<i>MsmEccCa<sub>1</sub></i> <sup>DUF</sup>	V	V	V	V	D	L	(323)
<i>MtbEccC<sub>2</sub></i> <sup>DUF</sup>	V	I	V	D	D	N	(342)
<i>MtbEccC<sub>3</sub></i> <sup>DUF</sup>	L	I	V	V	D	D	(320)
<i>MsmEccC<sub>3</sub></i> <sup>DUF</sup>	L	V	V	L	D	D	(320)
<i>MtbEccC<sub>4</sub></i> <sup>DUF</sup>	V	A	I	V	D	T	(298)
<i>MsmEccC<sub>4</sub></i> <sup>DUF</sup>	V	I	V	A	D	S	(274)
<i>TcrEccC</i> <sup>DUF</sup>	L	V	I	V	D	G	(319)
<i>GthEccC</i> <sup>DUF</sup>	F	V	F	V	A	D	(515)
<i>SrsEssC</i> <sup>DUF</sup>	V	F	V	I	T	D	(313)
<i>BsbEssC</i> <sup>DUF</sup>	V	F	V	I	Y	N	(525)
<i>CspEssC</i> <sup>DUF</sup>	V	V	F	L	Y	D	(424)
<i>SrlEssC</i> <sup>DUF</sup>	V	V	L	V	T	D	(274)

