Published in final edited form as:
Proteins. 2014 January ; 82(1): 159–163. doi:10.1002/prot.24351.

Crystal structure of the N-terminal domain of $EccA_1$ ATPase from the ESX-1 secretion system of *Mycobacterium tuberculosis*

Jonathan M. Wagner, **Timothy J. Evans**, and **Konstantin V. Korotkov***
Department of Molecular & Cellular Biochemistry and Center for Structural Biology, University of Kentucky, Lexington, Kentucky 40536, USA

Abstract

$EccA_1$ is an important component of the type VII secretion system (T7SS) that is responsible for transport of virulence factors in pathogenic mycobacteria. $EccA_1$ has an N-terminal domain of unknown function and a C-terminal AAA+ (ATPases associated with various cellular activities) domain. Here we report the crystal structure of the N-terminal domain of $EccA_1$ from *Mycobacterium tuberculosis*, which shows an arrangement of six tetratricopeptide repeats that may mediate interactions of $EccA_1$ with secreted substrates. Furthermore, the size and shape of the N-terminal domain suggest its orientation in the context of a hexamer model of full-length $EccA_1$.

Keywords

Rv3868; tetratricopeptide repeat; TPR domain; AAA+ ATPase; type VII secretion system

INTRODUCTION

Mycobacterium tuberculosis employs an army of secreted proteins to subvert the immune system during infection. In order to transport these virulence factors across an unusual two-membrane cell envelope mycobacteria use the specialized type VII secretion system (T7SS).[1] The T7SS machinery is an ~1500 kDa complex composed of a membrane pore and associated proteins including membrane-associated and cytosolic ATPases that are thought to provide the energy for transporting protein cargoes across the membrane. In the genome of *M. tuberculosis* there are five homologous T7SS clusters named ESX-1 to ESX-5. The most intensely studied region, ESX-1, includes the Rv3868 gene encoding the AAA+ ATPase $EccA_1$. Like other AAA+ ATPases $EccA_1$ forms oligomers, possibly hexamers, and hydrolyzes ATP *in vitro*.[2] Furthermore, its ATPase activity promotes virulence *in vivo* through increased mycolic acid synthesis.[3] In addition $EccA_1$ has been shown to be essential *in vivo* for specific targeting and secretion of EspC and other co-secreted virulence factors such as ESAT-6/CFP-10 through the ESX-1 system.[4]

Sequence analysis of $EccA_1$ indicates that it contains a C-terminal ATPase domain and a tetratricopeptide repeat (TPR) containing N-terminal domain of unknown function [Fig. 1(A)]. However, there is not yet any structure of an AAA+ ATPase containing a TPR domain, nor is there any reported structure with significant primary sequence identity with the $EccA_1$ N-terminal domain. To gain insight into the function of $EccA_1$ and related T7SS ATPases, we solved the structure of the N-terminal TPR domain of $EccA_1$ from *M. tuberculosis*.

*Correspondence to: Konstantin V. Korotkov, Department of Molecular & Cellular Biochemistry, University of Kentucky, 741 South Limestone, Lexington, KY 40536, USA. Phone: 859-323-5493. Fax: 859-323-5504. kkorotkov@uky.edu.

METHODS

Expression, purification and crystallization

The gene fragment corresponding to the N-terminal domain of $EccA_1$, residues 1–280, was PCR amplified from genomic DNA of *M. tuberculosis* H37Rv and cloned into a modified pET-28b vector (EMD Millipore) to encode an N-terminal His_6-tag followed by a tobacco etch virus (TEV) protease cleavage site. The resulting vector was transformed into Rosetta (DE3) cells (EMD Millipore) for expression. The cells were grown in Luria broth at 37°C to OD_{600} 0.6 and induced for 4 h with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside at 25 °C. Cells were harvested by centrifugation and resuspended in buffer containing 20 mM Tris-HCl pH 8.4, 300 mM NaCl, and 20 mM imidazole. The resuspended cells were lysed using an EmulsiFlex-C5 microfluidizer (Avestin) and $EccA_1$ was purified from the soluble fraction of the lysed cells using Ni-nitrilotriacetic acid agarose (Qiagen) column followed by His_6-tag cleavage with TEV protease. $EccA_1$ was further purified by size-exclusion using a Superdex200 column (GE Healthcare) in buffer containing 10 mM Tris-HCl pH 8.4, 200 mM NaCl. Crystallization screens were performed using the vapor diffusion method in a hanging drop 96-well plate format with JCSG Core Suites I–IV (Qiagen). The optimized $EccA_1$ crystals were grown in sitting drop format using 0.1 M sodium citrate pH 5.6, 1.0 M lithium sulfate, 0.5 M ammonium sulfate.

Data collection and structure determination

Crystals were soaked in crystallization solution supplemented with 20% glycerol and flash cooled in liquid nitrogen. To obtain heavy atom derivatives, the crystals were soaked for brief time periods, 30 s to 2 min, in cryo-protectant solution supplemented with compounds from Heavy Atom Screens (Hampton Research).[5] Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Data were processed and scaled using XDS and XSCALE.[6]

The structure of $EccA_1$ was determined by single wavelength anomalous diffraction (SAD) method using data from a single crystal soaked in the presence of $KAu(CN)_2$. Initial Au sites were found using SHELXD[7] at resolution 3.3 Å. Additional Au sites were found and refined using SAD protocol in PHASER.[8] Following density modification in Parrot,[9] an initial model was built using Buccaneer.[10] After manual rebuilding in Coot,[11] the improved model was refined using REFMAC5[12] against a ‘native’ dataset at resolution 2.0 Å. The ‘native’ dataset was collected from a crystal soaked in the presence of samarium acetate. A preliminary data analysis showed a lack of strong anomalous signal in this dataset, however, three partially occupied samarium ions were identified by peaks in anomalous difference maps and were included in the final model. The final rounds of refinement were performed applying eight translation, libration and screw-rotation displacement (TLS) groups determined by the TLSMD server.[13] The final model contains two $EccA_1$ monomers (residues 1-273) in the asymmetric unit, three samarium ions, six sulfate ions and 453 water molecules. The structure was validated using Coot and the Molprobity server (http://molprobity.biochem.duke.edu). Coordinates and structure factors have been deposited at the Protein Data Bank with accession code 4F3V.

Sequence alignments were done using ClustalW2 (http://www.clustal.org) and rendered using the ESPript server (http://espript.ibcp.fr). Structural illustrations were prepared using PyMol (http://www.pymol.org).

RESULTS AND DISCUSSION

The three dimensional structure of the N-terminal domain (NTD) of *M. tuberculosis* $EccA_1$, residues 1-280, was determined to 2.0 Å resolution. The protein crystallized in space group $P2_12_12_1$ with two molecules in the asymmetric unit (Table I). Crystallographic phases were experimentally determined by single wavelength anomalous diffraction method using apotassium dicyanoaurate derivative. The final model includes residues 1-273 of $EccA_1$ in both molecules. The two monomers adopt highly similar structures with an r.m.s.d. of 0.52 Å between the two chains. In the asymmetric unit the two molecules of $EccA_1$ dimerize via a non-crystallographic two-fold axis through surfaces on the edge of TPR motifs 2 and 3. This dimerization interface probably represents a non-physiological interaction. Analysis using the PISA server[14] shows that the interface between the two chains buries 2200 $Å^2$ of surface area. However, a calculated PISA score of 0.110 for the interface indicates that it probably does not represent a physiological interaction. Furthermore, Arg62 and Arg63 from each chain form bridging interactions mediated by a sulfate ion from crystallization solution. Because the two chains are nearly identical and probably do not dimerize in the context of full-length $EccA_1$ we limit our discussion to a monomer of $EccA_1$.

The 12 anti-parallel α helices of $EccA_1$ arrange into 6 tandem TPR motifs of approximately 34 residues each [Fig. 1]. The TPRs associate through hydrophobic interactions between consecutive helices and together form a right-handed superhelix with a pitch of ~60 Å, and a width of ~40 Å that is characteristic of TPR domains[15]. A DALI search ranked the *Pseudomonas aeruginosa* type IV pili system protein PilF as the closest characterized bacterial homolog of $EccA_1$ with an r.m.s.d of 3.5 Å and 9% sequence identity over 173 aligned residues (PDB 2HO1).[16,17] Despite low homology between two proteins, the TPR helices of $EccA_1$ superpose with PilF remarkably well, however, unlike PilF, $EccA_1$ has a β finger insertion after TPR2 that occupies the concave groove of the TPR superhelix [Fig. 1]. This is the same groove commonly used for protein-protein interactions by the TPR motif proteins.[18] PilF contains a conserved Asn ladder within its binding groove that is thought to promote peptide binding through formation of bidentate hydrogen bonds to substrate backbone.[16,19] However, the Asn ladder is missing in $EccA_1$, and in its place are hydrophobic residues that form a hydrophobic core with β finger insert residues. This hydrophobic core suggests that the insert is a permanent feature within the putative binding groove. It may be that the insert participates in protein-protein interactions by forming β strand complementation interactions with the substrates in an extended conformation within the concave groove. Alternatively, proteins may interact with the side of the TPR bundle instead of the canonical binding groove. This latter binding mechanism by a TPR protein is illustrated by the $p67^{phox}$ - Rac complex structure (PDB 1E96), which contains a similar β finger motif in its concave groove.[20]

$EccA_1$ belongs to the CbxX/CfxQ family of ATPases, however, the homology is limited to the AAA+ domain and the N-terminal TRP domain is unique for EccA family. The nearest homolog of $EccA_1$ with a known structure is Rubisco activase (PDB 3SYL and 3ZUH),[21] which is 46% identical to the C-terminal ATPase domain of $EccA_1$. Similar to Rubisco activase, $EccA_1$ may adopt a hexameric ring stabilized by contacts between the C-terminal domains [Fig. 2(B)].[2] Modeling the architecture of full-length $EccA_1$ using the structure of Rubisco activase (PDB 3ZUH) to align the $EccA_1$ monomers into a hexamer yielded insight into the possible conformations of the N- and C- terminal domains relative to each other. There is no obvious region of the C-terminal ATPase domain that might fit into the large groove of the N-terminal domain structure. This groove probably remains open and available for interactions with other proteins and/or substrates of the T7SS. Also, the length of the N-terminal TPR domain (~70 Å) is slightly shorter than the radius of the Rubisco activase hexamer. Thus, it is possible that the N-terminal domain lies across the top of the

ATPase domain without occluding the central pore. Importantly, this hexameric conformation could be modeled without occluding any of the ATP binding sites indicating that such a configuration is compatible with ATPase activity [Fig. 2(A)].

The sequence alignment of $EccA_1$ homologs from the ESX-1 clusters of mycobacteria shows that $EccA_1$ proteins are conserved with at least 74% pairwise sequence identity between family members [Fig. S1]. In contrast, the sequence identity ranges between 29–39% for EccA proteins from different ESX clusters: ESX-1, ESX-2, ESX-3 and ESX-5. This may indicate that EccA ATPases interact with diverse components of secretion system or substrates. However, the overall protein architecture with an N-terminal TPR domain and a C-terminal AAA+ domain is clearly present in all EccA proteins. The key elements of the AAA+ proteins including Walker A motif, Walker B motif, pore loop are conserved in $EccA_1$ family members [Fig. S1]. In addition, Tyr439 (*M. tuberculosis* $EccA_1$ numbering) may serve as sensor 1, whereas Arg519 and Arg522 may serve as sensor 2. Arg429 has been suggested as an Arg finger residue and Arg429Ala substitution affected ATP hydrolysis by the C-terminal domain of $EccA_1$.[2] Our hexamer model identified Arg456 as another candidate Arg finger residue. Indeed, Arg456 is located closer than Arg429 to the ATP-binding site of adjacent subunit.

In summary, the $EccA_1$ N-terminal domain adopts a TPR fold indicating it may mediate protein-protein interactions between the C-terminal ATPase domain and substrate proteins. Like other TPR motif proteins, $EccA_1$ may use its central concave groove to interact with protein cargos, but the presence of a β finger fixed within the groove raises questions about the exact nature of the interaction. Oligomerization interactions were not found in the structure indicating that it is the C-terminal domain that primarily mediates $EccA_1$ hexamerization. Finally, the dimensions of the N-terminal TPR domain suggest that along with the C-terminal domain it form a compact hexamer while maintaining a central pore that can open and close during ATP hydrolysis. What the mechanism may be for specific protein recognition and energy transfer to substrates passing through the T7SS remains to be investigated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Carol Beach for expert assistance with mass spectrometry. We acknowledge the University of Kentucky Proteomics Core and Protein Analytical Core that are partially supported by grants from the National Center for Research Resources (P20RR020171) and the National Institute of General Medical Sciences (P20GM103486) from the National Institutes of Health. We thank staff members of Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory, for assistance during data collection. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. This study was supported by a Center of Biomedical Research Excellence (COBRE) grantP20GM103486 (to KVK) from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Stoop EJM, Bitter W, van der Sar AM. Tubercle bacilli rely on a type VII army for pathogenicity. Trends Microbiol. 2012; 20:477–484. [PubMed: 22858229]
2. Luthra A, Mahmood A, Arora A, Ramachandran R. Characterization of Rv3868, an essential hypothetical protein of the ESX-1 secretion system in *Mycobacterium tuberculosis*. J Biol Chem. 2008; 283:36532–36541. [PubMed: 18974091]

3. Joshi SA, Ball DA, Sun MG, Carlsson F, Watkins BY, Aggarwal N, McCracken JM, Huynh KK, Brown EJ. EccA1, a Component of the *Mycobacterium marinum* ESX-1 protein virulence factor secretion pathway, regulates mycolic acid lipid synthesis. Chem Biol. 2012; 19:372–380. [PubMed: 22444592]
4. DiGiuseppe Champion PA, Champion MM, Manzanillo P, Cox JS. ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. Mol Microbiol. 2009; 73:950–962. [PubMed: 19682254]
5. Sun PD, Radaev S, Kattah M. Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases. Acta Crystallogr D Biol Crystallogr. 2002; 58:1092–1098. [PubMed: 12077427]
6. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–132. [PubMed: 20124692]
7. Sheldrick GM. A short history of SHELX. Acta Crystallogr A. 2008; 64:112–122. [PubMed: 18156677]
8. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
9. Cowtan K. Recent developments in classical density modification. Acta Crystallogr D Biol Crystallogr. 2010; 66:470–478. [PubMed: 20383000]
10. Cowtan K. Completion of autobuilt protein models using a database of protein fragments. Acta Crystallogr D Biol Crystallogr. 2012; 68:328–335. [PubMed: 22505253]
11. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
12. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011; 67:355–367. [PubMed: 21460454]
13. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr. 2006; 62:439–450. [PubMed: 16552146]
14. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007; 372(3):774–797. [PubMed: 17681537]
15. Das AK, Cohen PTW, Barford D. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. EMBO J. 1998; 17:1192–1199. [PubMed: 9482716]
16. Kim K, Oh J, Han D, Kim EE, Lee B, Kim Y. Crystal structure of PilF: Functional implication in the type 4 pilus biogenesis in *Pseudomonas aeruginosa*. Biochem Biophys Res Comm. 2006; 340:1028–1038. [PubMed: 16403447]
17. Koo J, Tammam S, Ku SY, Sampaleanu LM, Burrows LL, Howell PL. PilF is an outer membrane lipoprotein required for multimerization and localization of the *Pseudomonas aeruginosa* Type IV pilus secretin. J Bacteriol. 2008; 190:6961–6969. [PubMed: 18776008]
18. D'Andrea LD, Regan L. TPR proteins: the versatile helix. Trends Biochem Sci. 2003; 28:655–662. [PubMed: 14659697]
19. Jinek M, Rehwinkel J, Lazarus BD, Izaurralde E, Hanover JA, Conti E. The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. Nat Struct Mol Biol. 2004; 11:1001–1007. [PubMed: 15361863]
20. Lapouge K, Smith SJM, Walker PA, Gamblin SJ, Smerdon SJ, Rittinger K. Structure of the TPR domain of p67(phox) in complex with Rac center dot GTP. Mol Cell. 2000; 6:899–907. [PubMed: 11090627]
21. Mueller-Cajar O, Stotz M, Wendler P, Hartl FU, Bracher A, Hayer-Hartl M. Structure and function of the AAA(+) protein CbbX, a red-type Rubisco activase. Nature. 2011; 479:194–199. [PubMed: 22048315]

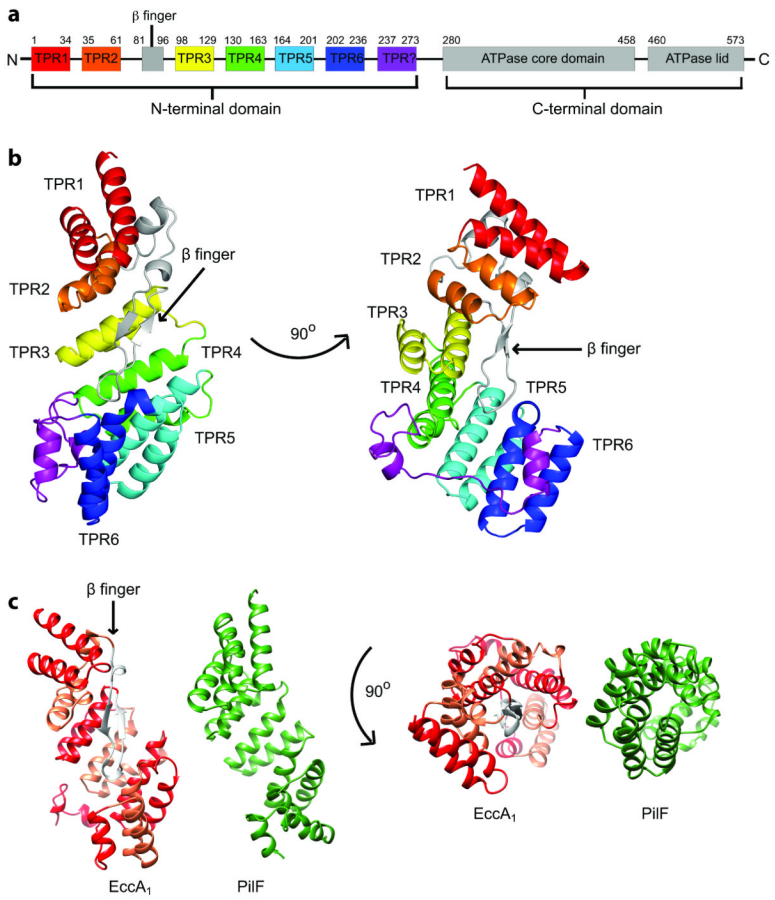

Figure 1.

$EccA_1$ N-terminal domain structure. (A) A schematic diagram of $EccA_1$'s TPR repeats, β finger insert, and C-terminal AAA+ ATPase domain. TPR motifs 1–6 are colored in rainbow colors from red to blue, whereas residues 237-273 are colored purple to highlight a capping C-terminal α helix or a possible TPR motif that was truncated in this construct. (B) Cartoon representation of $EccA_1$ N-terminal structure. TPR repeat motifs are colored as in (A). The β finger insert is highlighted in gray. (C) Structural conservation of the TPR superhelix motif between $EccA_1$ (alternating red / orange) and PilF (green). The β finger insert is highlighted in gray.

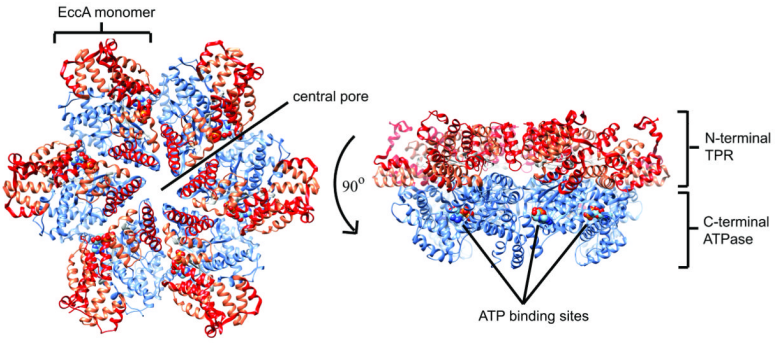

Figure 2.

Model of $EccA_1$ hexamer. (A) View looking down the central pore of full length $EccA_1$ modelled using the hexameric model of Rubisco activase (3ZUH). C-terminal AAA+ ATPase domains are colored blue. N-terminal domain TPR motifs are colored alternating red / orange. Bound ADP is shown as spheres. (B) Same as (A) after rotating the view 90°.

Table I

Data collection and refinement statistics.

	Native (PDB 4F3V)	$KAu(CN)_2$ derivative
Data collection		
Wavelength (Å)	1.0000	1.0000
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	73.23, 92.51, 105.71	74.32, 92.58, 105.76
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	29.6–2.00 (2.11–2.00)[a]	29.6–2.49 (2.63–2.49)
R_{sym}	0.101 (0.782)	0.122 (0.508)
$I / \sigma I$	14.2 (2.5)	13.9 (4.4)
Completeness (%)	99.8 (99.3)	99.4 (96.8)
Multiplicity	5.7 (5.7)	7.3 (7.2)
Anomalous completeness (%)		99.3 (96.1)
Anomalous multiplicity		3.9 (3.8)
Refinement		
Resolution (Å)	29.6–2.00	
No. reflections (total / free)	49226 / 2525	
R_{work} / R_{free}	0.174 / 0.211	
No. atoms		
Protein	4135	
Ligand/ion	33	
Water	453	
B-factors		
Protein	26.2	
Ligand/ion	56.4	
Water	36.3	
Wilson *B*	30.8	
R.m.s. deviations		
Bond lengths (Å)	0.011	
Bond angles (°)	1.286	
Ramachandran distribution (%)[b]		
Favored	98.5	
Outliers	0.0	

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

[b] Calculated using the MolProbity server (http://molprobity.biochem.duke.edu).