Published in final edited form as:

J Mol Biol. 2008 November 14; 383(3): 603–614. doi:10.1016/j.jmb.2008.08.047.

Structure of *Staphylococcus aureus* EsxA suggests a contribution to virulence by action as a transport chaperone and/or adaptor protein

Ramasubramanian Sundaramoorthy, Paul K. Fyfe, and William N. Hunter*

Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK

Abstract

Staphylococcus aureus pathogenesis depends on a specialized protein secretion system, ESX-1, that delivers a range of virulence factors to assist infectivity. We report the characterization of two such factors, EsxA and EsxB; small acidic dimeric proteins carrying a distinctive WXG motif. EsxA crystallized in triclinic and monoclinic forms and high-resolution structures were determined. The asymmetric unit of each crystal form is a dimer. The EsxA subunit forms an elongated cylindrical structure created from side-by-side α -helices linked with a hairpin bend formed by the WXG motif. Approximately 25% of the solvent accessible surface area of each subunit is involved in interactions, predominantly hydrophobic, with the partner subunit. Secondary structure predictions suggest that EsxB displays a similar structure. The WXG motif helps to create a shallow cleft at each end of the dimer, forming a short β -sheet-like feature with an N-terminal segment of the partner subunit. Structural and sequence comparisons, exploiting biological data on related proteins found in *Mycobacteria tuberculosis* suggest that this family of proteins may contribute to pathogenesis by transporting protein cargo through the ESX-1 system exploiting a C-terminal secretion signal and / or are capable of acting as adaptor proteins to facilitate interactions with host receptor proteins.

Keywords

adaptor protein; chaperone; helical bundle; secretion system; virulence factor

Introduction

The commensal Gram-positive bacterium *Staphylococcus aureus* lives on human skin and in nostrils. Under certain conditions it represents a serious problem in hospitals especially for immunocompromised and surgery patients¹ and the number of deaths attributed to *S. aureus* infection is comparable to that of acquired immune deficiency syndrome.² Increasing levels of antibiotic resistance contribute significantly to this statistic and create an urgent need for the development of new treatments.³ Thankfully, current developments in this area look promising with new drug candidates progressing through clinical trials.⁴ Nevertheless, it would be highly desirable to have access to a vaccine to complement new drug treatments, and an understanding of the molecular determinants of virulence has the potential to underpin such a development.⁵

*Address correspondence to: Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, Tel. +44 (0)1382-385745; Fax. +44 (0)1382-345764; w.n.hunter@dundee.ac.uk.

The ability of *S. aureus* to adhere to the surface of human cells is key to infection, subsequently leading to cell invasion and / or destruction.⁶ Pathogenesis is dependant upon the secretion of a wide range of exoproteins and virulence factors,⁷ such as fibronectin-binding proteins.^{8,9} To extrude proteins Gram-positive bacteria use a general secretion (Sec) pathway, which exploits an N-terminal signal sequence to assist translocation,¹⁰ and a twin-arginine translocation (Tat) pathway.⁷ Of particular interest in our research is a third secretion system discovered in *Mycobacteria* and directly involved in pathogenesis of *Mycobacterium tuberculosis*¹¹ and *S. aureus*.¹²

Two virulence factors of *M. tuberculosis*, termed ESAT-6 (Early secreted antigenic target-6 kDa) and CFP-10 (Culture Filtrate Protein-10 kDa), elicit a strong T-cell immune response and stimulate the production of γ -interferon.^{13,14} These proteins were identified as the products of two open reading frames (*esat-6* and *cfp-10*) in the region of difference (RD1 or ESAT-6 locus), a segment of DNA present in virulent *M. bovis* but absent from the BCG (Bacille Calmette-Guérin) strain used as a live vaccine and from avirulent *M. microti* strains.¹⁵⁻¹⁷ The re-introduction of RD1 into *M. bovis* BCG restores ESAT-6 and CFP-10 expression, and reinstates virulence and immunogenicity.¹⁷⁻¹⁹ The RD1 segment contains additional genes flanking *esat-6* and *cfp-10* encoding snm1 (Rv3870), snm2 (Rv3871) and snm4 (Rv3877).²⁰⁻²² These three, membrane anchored proteins together help form the non-Sec dependent ESX-1 (ESAT-6 secretion) system and are required for virulence.²² Two of the proteins, snm1 and snm2, carry FtsK/SpoIIIE domains (FSD),²¹ a distinctive protein sub-structure often associated with ATPase activity. CFP-10 and ESAT-6 are representative of a protein family, observed in many Gram-positive bacteria, all comprising approximately 100 amino acids and carrying a Trp-X-Gly signature motif.²¹ Strikingly, there are similarities also in flanking genes and encoded proteins and it was suggested,²³ and later proven²¹ that this locus encodes a secretion system. Similar loci are present in *S. aureus*,¹² and termed the ESX-1 secretion system (Ess). Here, genes encoding an ESAT-6-like protein, EsxA (Ess extracellular A) and the CFP-10-like protein EsxB (Ess extracellular B) are found. The flanking genes encode four membrane-bound proteins, EsaA, EssA, EssB, and EssC of which the last three are conserved in *Mycobacteria* and required for secretion of *S. aureus* EsxA (*SaEsxA*) and EsxB (*SaEsxB*).¹² EssC is homologous to snm1 and snm2 proteins of *M. tuberculosis*.¹²

We set out to establish if accurate structural data could provide insight into the biological role of these proteins and now describe the production and characterization of recombinant *SaEsxA* and *SaEsxB* together with the high-resolution structure of *SaEsxA* determined in two crystal forms. Sequence and structural comparisons, in conjunction with previous studies on the related *M. tuberculosis* virulence factors, allow us to describe structural features that appear important for the integrity of the structure of this protein family and to propose functional contributions to pathogenesis.

Results and discussion

SaEsxA and *SaEsxB* form homodimers

Recombinant expression systems to produce *SaEsxA* (Mwt ~ 11.0 kDa, 97 residues) and *SaEsxB* (Mwt ~ 11.5 kDa, 104 residues) were prepared and both proteins purified to homogeneity with a good yield of approximately 8 mg L⁻¹ of bacterial culture. SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry were used to confirm the high level of purity and integrity of the products (data not shown). Gel filtration was applied as the final purification step and indicated that each protein formed a homodimer, of approximate molecular weight 23 kDa with no evidence for any monomer in solution. The

high stability of such a quaternary structure is evident by the observation that dimer bands were even present in the SDS-PAGE gels (data not shown).

The application of fluorescence, circular dichroism, and nuclear magnetic resonance spectroscopy methods indicate that ESAT-6 in isolation is a monomeric molten globule protein, CFP-10 a monomeric and unstructured protein, but when mixed together they associate tightly to form a stable heterodimeric complex.^{24,25} The NMR derived structure of the CFP-10/ESAT-6 complex has been reported.²⁶ Based on these observations it was previously accepted that *SaEsxA* and *SaEsxB* would form a heterodimeric assembly.¹² However, in our case we were unable to find any evidence for a *SaEsxA:SaEsxB* heterodimer following incubation of the two proteins together under similar conditions employed for the *M. tuberculosis* proteins. We note the recent report that under the specific conditions of yeast two-hybrid experiments CFP-10 and ESAT-6 form homodimers,²⁷ an observation not described by other investigators.

Structural analysis of *SaEsxA* and model quality

Crystallization trials of homodimeric *SaEsxA* and *SaEsxB* were carried out individually, and also with a *SaEsxA:SaEsxB* mixture. *SaEsxB* did not crystallize but two highly ordered crystal forms of *SaEsxA* were obtained. Triclinic crystal form I and monoclinic form II crystals diffracted to 1.4 Å and 1.9 Å resolution respectively. The form II crystals were obtained from the *SaEsxA:SaEsxB* mixture. The structure of *SaEsxA* form I was solved by multi-wavelength anomalous dispersion (MAD) methods exploiting the signal from a selenomethionine (SeMet) derivative. This model was then used in molecular replacement (MR) calculations to determine the structure of form II. Refinement statistics are presented in Table 1 and indicate that the analyses have produced acceptable high-resolution models with all residues displaying ϕ/ψ combinations within the allowed regions of a Ramachandran plot.

Both crystal forms contain a dimer per asymmetric unit with subunits labeled A and B. Residues 4-87 in subunit A of both crystal forms are ordered, while in subunit B the electron density is well defined for residues 4-92 of form I and for all but the C-terminal residue of form II. Two additional N-terminal residues (Ser-His) are observed in subunit B of form II. These are remnants of the proteolysis carried out to remove the histidine-tag.

Several prominent features in electron and difference density maps were observed for both crystal forms. These corresponded to the major peaks also seen in anomalous difference Fouriers calculated with data from crystal form I (data not shown). On the basis of difference density and anomalous difference density peak heights, the presence of 80 mM zinc acetate and 100 mM sodium cacodylate in the crystallization medium, coordination geometry,²⁸ and successful refinement these peaks were assigned as Zn^{2+} ions in both crystal forms and a single cacodylate in form I. Seven Zn^{2+} ions were located in form II, five in form I and we note that these five binding sites, on the surface of the homodimer, are conserved between the two forms (data not shown). The binding of Zn^{2+} ions is mainly achieved using glutamate and aspartate side chains, with several water molecules also involved. In form II, Lys73 NZ from both subunits coordinate a Zn^{2+} . Lysine is not a common ligand for cation binding but there is precedent for such an interaction in high resolution structures.²⁸ In form I, a cacodylate ion interacts with Lys73 of subunit A and one of the Zn^{2+} ions (data not shown).

The superposition of subunits within each crystal form yields an r.m.s.d (root-mean-square-deviation) of 0.3 Å for 85 Ca atoms (data not shown). Superpositions of subunit A of form I with subunit A or B of form II results in an r.m.s.d of 0.7 Å (data not shown). Such a high

degree of structural similarity means that it is only necessary to detail one structure and we concentrate on form II since it contains the more ordered C-terminus.

Structure of SaEsxA

SaEsxA is a small acidic protein of 97 amino acids (theoretical pI 4.6), and crystallized as a dimer in both crystal forms. The *SaEsxA* monomer is composed of two ($\alpha 1$, $\alpha 2$) side-by-side anti-parallel helices folded to create an elongated elliptical cylinder (Fig. 1A) of approximate dimensions 15 Å x 24 Å x 80 Å. The *SaEsxA* primary structure displays a heptad repeat where the first and fourth positions comprise hydrophobic residues, and the other positions are mainly hydrophilic. This pattern is recognized as a signature for interlocking α -helices.^{29,30}

Approximately 85% of the residues contribute to the two sections of secondary structure. A flexible N-terminal segment, lying across $\alpha 2$, leads to Pro8 which then initiates $\langle 1$. Ten turns of helix extend to Trp43 and then a short hairpin bend links onto $\langle 2$, which has thirteen turns of helix (Gly45 to Asn93). This second helix is distorted at turn four by Pro60 (Fig. 1A). The C α atoms of $\langle 1$ and $\langle 2$ prior to Pro60 are separated by approximately 9 Å. After the distortion imposed by Pro60 the distance increases to 12 Å. The monomer is slightly curved with one side presenting a concave hydrophobic surface while the other displays a polar, acidic surface (Fig. 1B, C).

The dimerization of *SaEsxA* (Fig. 2) is dominated by hydrophobic interactions with approximately 75% of the interface constructed by non-polar residues. Each monomer has a dimerization interface area of ~1830 Å² (calculated using the Protein Interfaces, Surfaces and Assemblies service at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)),³¹ This interface uses approximately 25% of the solvent accessible surface area of a single subunit. The few residues that form polar interactions (five hydrogen bonds and two salt bridges) directly implicated in dimerization are concentrated at the C- and N-terminus and the hairpin bend. At the N-terminal section, the main chain of Met3 and Lys5 form a short β -sheet-like interface using two hydrogen bonds with the main chain of residues Asn42 and Glu44 of the neighboring subunit hairpin region (LysA5 N and MetA3 O with AsnB42 O and GluB44 N respectively are shown in Fig. 3). Tyr18 OH and Gln36 NE2 of both subunits form two hydrogen bonds with each other and salt bridges between Lys14 and Glu38 are also observed. GlnA75 OE1 accepts a hydrogen bond donated from ArgB50 NH2. In addition to these stabilizing interactions formed between the amino acids, the carboxylate side chains of GluA10 and GluB38 are involved in a Zn²⁺-mediated interaction (data not shown).

The hairpin bend between $\alpha 1$ and $\alpha 2$ contains the WXG motif (Trp43-Glu44-Gly45, Fig. 1A, 3). The side chain of Trp43 is positioned between Phe48 on the same polypeptide and Lys5 of the partner subunit. The indole moiety also forms van der Waals interactions with Ile39 of the same subunit, Met6, Ile11, Thr79 and Val83 from the partner to stabilize one end of the *SaEsxA* dimer (data not shown). In most of the 226 putative EsxA/B sequences identified in the Pfam database,³² (PF06013) the central residue of the tripeptide motif is large and hydrophilic. In 28% of the sequences the central residue is an acidic aspartate or glutamate and in 11 % asparagine or glutamine. A basic residue (arginine, histidine or lysine) occupies 12 % and serine 20 % of the central positions. The notable exceptions are all but one of the sequences derived from *Mycobacteria* species where the motif is strictly conserved as Trp-Gly-Gly. In subunit A the carboxylate side chain of Glu44 is 2.8 Å from the carbonyl of Ala40 suggesting that protonation has occurred to allow a hydrogen bond to form (data not shown). This is plausible given the acidic conditions under which crystals were obtained (pH ~ 5.5). In subunit B, Glu44 adopts a different conformation and interacts

with a Zn^{2+} ion, which in turn is linked to a symmetry related molecule (data not shown). Gly45 completes the tight hairpin turn into $\alpha 2$.

The residues that coordinate Zn^{2+} are Glu9, Glu10, Asp23, Glu38, Glu44, Glu52, Glu70, Asp81, Glu85 and Lys73 (data not shown). Four of these are conserved in EsxA/B sequences. Glu9 and Glu10 are conserved in 44 % and 33 % of the sequences respectively. Glu70 and Asp81 are conserved in about 25 % of the sequences. These residues do not create specific metal ion binding environments on the EsxA dimer but rather participate with symmetry related molecules in the crystal lattice to bind the ions.

Comparison of *SaEsxA* with *SaEsxB*, *M. tuberculosis* ESAT-6 and CFP-10

A sequence alignment (MUSCLE)³³ of *SaEsxA* with ESAT-6, CFP-10 and *SaEsxB* is presented in Fig. 4. *SaEsxA* shares 12 % sequence identity with ESAT-6, 14 % with CFP-10 and 13 % with *SaEsxB*. *SaEsxB* shares 8% sequence identity with ESAT-6 and 13% with CFP-10. Such a low level of sequence conservation immediately suggests that caution must be exercised in comparisons. In this, we are helped by access to structural data and despite the low sequence conservation the *SaEsxA* and CFP-10/ESAT-6 structures are similar. However, there are also important differences.

A representative conformer in the deposited ensemble of NMR structures for the CFP-10/ESAT-6 heterodimer²⁶ is model 9 and a superposition of 80 C α atoms of one subunit of *SaEsxA* with each subunit of the CFP-10/ESAT-6 heterodimer gives an r.m.s.d of 2.0 Å. One overlay, of *SaEsxA* with ESAT-6 is depicted in Fig. 5. *SaEsxA* shares a similar helix-hairpin-helix topology with the CFP-10/ESAT-6 heterodimer. The main differences between these structures lie at the C-terminus and an extra turn in $\alpha 1$ prior to the hairpin in ESAT-6. In contrast to both CFP-10 and ESAT-6, which display a flexible, poorly defined C-terminus, the C-terminal segment of *SaEsxA* retains a helical structure to Asn93.

Secondary structure prediction of *SaEsxB* (PSIPRED)³⁴ suggested, with a high level of confidence, two α -helices between residues Lys13 and Ile42, Gly49 (of the WXG motif) and Gln97 (data not shown). Despite the low level of sequence homology between *SaEsxA* and *SaEsxB* this would strongly suggest similar structures. It is not common to find proline embedded in an α -helical structure and a distinctive feature of *SaEsxA* is the presence of Pro60, which creates a distortion in $\alpha 2$. In *SaEsxB* Pro72 is predicted to occur in $\alpha 2$ and may well produce a distortion in that protein although at a different position, two turns of helix further along compared to *SaEsxA*. There are no proline residues in CFP-10/ESAT-6 and the helices appear more regular.

The low level of sequence conservation in the EsxA/EsxB family and the complication of now knowing which are homo or heterodimers renders it difficult to definitively assign structural features related to dimerization. However, inspection of the 226 EsxA/EsxB sequences in Pfam family PF06013 identified that *SaEsxA* Pro60 is conserved in 40 % of the entries (data not shown). This residue and the specific distortion it imparts to $\alpha 2$ may represent a signature for homodimer formation. Experimental data on quaternary structure and the ability to be transported through the ESX-1 system on range of orthologues will be required to confirm or refute this suggestion.

Mutational analysis of *M. tuberculosis* ESAT-6 has identified several residues required for targeting, secretion, complex formation with CFP-10 and immunogenicity.³⁵ These residues are, in general, structurally conserved in *SaEsxA*, marked on the sequence alignment presented in Fig. 4 and on the molecular structures in Fig. 5. A total of sixteen residues were mutated, three at the N-terminus, three within $\alpha 1$, two in the hairpin region and eight in $\alpha 2$.

A major conclusion from that study is that changes likely to perturb the dimer structure have a pronounced affect on virulence.

Virulence was abolished in ESAT-6 Leu28Ala, Leu29Ala, Trp43Arg, Gly45Thr Val90Arg and Phe94Gln mutants. These mutations can be divided into three groups according to the proposed structural contributions of the specific residues. Leu28 and Leu29 are in the middle of $\alpha 1$, buried at and important contributors to the dimer interface. The corresponding residues in *SaEsxA* are Ile28 and Leu29, and with hydrophobic side-chains conserved they are likely to play a similar role in oligomerization and maintenance of the helical structure of *SaEsxA*. Trp43 and Gly45 of the WXG motif are conserved in *SaEsxA* and positioned on the hairpin loop linking $\alpha 1$ and $\alpha 2$. Mutations such as these would be predicted to disrupt the dimer. ESAT-6 Val90 and Phe94 are conserved as Leu90 and Phe94 in *SaEsxA*. These residues are part of the C-terminal sequence important for secretion and will be discussed later.

In *SaEsxA*, Phe55, Gln56, Ala63, and Gln67 align with four of the mutated residues Gln55, Gln56, Asn66, and Asn67. In ESAT-6 these polar residues are all exposed on the surface of the dimeric assembly. In *SaEsxA*, Gln56 and Gln67 are likewise solvent exposed but the others, Phe55 and Ala63 are buried and contribute hydrophobic interactions to stabilize the dimer. The difference between the two proteins here may be a result of the distortion to *SaEsxA* $\alpha 2$ caused by Pro60, a distortion that is absent from ESAT-6. Three other mutations were made to ESAT-6 (Gln4Leu, Ala14Arg, Met83Ile). These changes introduced amino acids that are conserved in other ESAT-6 family members and it is perhaps not surprising that these mutations had no apparent effect on ESAT-6 activity.³⁵ In *SaEsxA* the corresponding residues are Ile4, Lys14, and Val83. Of these only Lys14 appears to be significant from a structural viewpoint; the aliphatic part of the side chain contributing to the hydrophobic interactions that stabilize the dimer and the amine group a salt bridge with Glu38 of the partner subunit as described earlier.

CFP-10 interacts with the C-terminal domain of the FtsK/SpoIIIE ATP-dependent transporter protein smn2.²² The conjugation of the seven C-terminal residues of *M. tuberculosis* CFP-10 to *Saccharomyces cerevisiae* ubiquitin and heterologous expression led to secretion of the yeast protein into the culture indicating that these residues comprise a secretion signal. The level of protein secreted was however reduced significantly indicating that other factors contribute to the recognition and translocation mechanism.³⁶ Site-directed mutagenesis targeting these seven residues identified that mutation of CFP-10 Leu94, Met98, Gly99 or Phe100 to alanine abolishes interaction with smn2 and prevents secretion whilst similar alterations to ESAT-6 have no effect.³⁶ The conclusion therefore is that a key determinant for secretion resides on the C-terminus of only one component and this explains the requirement for heterodimerization of CFP-10/ESAT-6 in *M. tuberculosis*.

The sequence alignment shown in Fig. 4 indicates a significant degree of conservation at the C-terminus of CFP-10, *SaEsxA* and *SaEsxB*. CFP-10 Leu94 and Gly99 are strictly conserved in the *S. aureus* proteins and Phe100 conservatively replaced by leucine. Met98 in CFP-10 is replaced by phenylalanine and glutamine in *SaEsxA* and *SaEsxB* respectively. It appears that both *SaEsxA* and *SaEsxB* carry the C-terminus translocation signal, which means that each protein is transported and acts independently as a virulence factor with no requirement for heterodimerization. It may be significant that the C-terminus of *SaEsxA* adopts a helical conformation (Fig. 3), a structural feature that may be important in the molecular recognition of EsxA by the transport machinery.

Under acidic conditions such as those found in phagosomes, ESAT-6 can dissociate from CFP-10 and bind, with high specificity, to liposomes containing cholesterol and

phosphatidylcholine.³⁷ ESAT-6 can also disrupt artificial lipid bilayers but not when complexed to CFP-10,³⁸ and this property appears linked to a concomitant increase in α -helical content.²⁵ It has been suggested that the complex with CFP-10 is required to facilitate secretion with regulation of ESAT-6 cytolytic activity provided by the requirement for it to be released from the partner subunit. Renshaw *et al.*, did not observe any cytolytic activity in fluorescence based binding assays,²⁶ however the pH used in the experiments may not have been sufficiently acidic to promote subunit dissociation, and hence the lytic activity of ESAT-6 may have been masked. It remains to be shown if either *SaEsxA* or *SaEsxB* display similar properties.

Biological implications

Dimeric CFP-10/ESAT-6 proteins are secreted to the culture supernatant and are not found in the cytosol, nor are they associated with membrane fractions.²⁶ This suggests to us that this protein family is not a component of the transporter complex but may fulfill other roles. For example, the homo (in the case of *S. aureus* proteins) or CFP-10/ESAT-6 heterodimers carry a C-terminal segment that interacts with the transport system. Since these proteins are secreted the weight of evidence suggests they are more likely to be cargo, a transport module or chaperone to assist export of protein by the outer membrane ESX-1 secretion system.

Our study has revealed that *SaEsxA* forms a homodimeric four-helix bundle and that *SaEsxB* forms a homodimer likely of similar structure. The polar surface of the *SaEsxA* dimer suggests it is unlikely to be a membrane pore forming protein. The ProKnow server (<http://proknow.mbi.ucla.edu>) was used seeking clues to infer function, based on structure and sequence. The highest gene ontology score for molecular function was for a structural molecule and for a biological process of protein folding. Further analysis, by searching nearest structural neighbors against the Protein Data Bank using the DALI server,³⁹ and structure analysis using the Profunc server⁴⁰ identified the closest structural homologues of the *SaEsxA* dimer as vinculin,⁴¹ the talin rods that bind to vinculin,⁴² and paxillin-binding focal adhesion proteins⁴³ (data not shown). These proteins are involved in the maintenance of cell structure integrity and adhesion events. The *SaEsxA* monomer also gives a structural match to the BtuB porin-binding domain of the colicin E3 receptor.⁴⁴ In BtuB, hydrophobic residues displayed on a hairpin bend bind the hydrophobic core of the membrane bound porin.

Indeed, the structural features of *SaEsxA*, the tightly bound helical section with a hydrophilic surface, and the extended C-terminus containing exposed hydrophobic residues suggest that EsxA, or indeed the ESAT-6 family could be adaptor proteins, capable of bridging bacterial proteins to host cell surface receptors, that facilitate localization of virulence factors and exotoxins or indeed any secreted factor that contributes to establishing an infection.

We have shown that *SaEsxA* binds divalent cations and these were assigned as Zn^{2+} as discussed above. This observation may simply reflect the presence of these cations in the crystallization medium. We note that some of the liganding residues are, to a degree conserved in EsxA/EsxB family members and an association with cations may facilitate interactions with other metal binding proteins to support an adaptor role. This is an area that warrants further investigation.

The *M. tuberculosis* genome is predicted to encode more than 20 paired CFP-10/ESAT-6 paralogs, of which several have been found in culture supernatant.³⁶ This suggests that the numerous gene duplication events that have occurred have given *M. tuberculosis* a high degree of redundancy with respect to these particular proteins. The formation of

heterodimers with a CFP-10 type protein carrying the secretion signal provides a mechanism whereby a diverse range of proteins might be assisted through the ESX-1 system. The C-terminal sequences of some CFP-10 paralogs lack the Met-X-Phe tripeptide important for secretion in *M. tuberculosis* and for these we suggest that heterodimer formation with paralogs carrying the correct sequence is required to ensure interaction with the ESX-1 secretion system or that separate ESX-1-like systems drive their secretion. A plethora of secretion chaperones such as these would provide a highly specialized mechanism of translocating different virulence factors and proteins required to establish an infection or if acting as an adaptor system they would provide diversity in terms of recognition for host cell targets.

The WXG motif and nearby interactions between the subunits result in a shallow cleft positioned close to the important helical C-terminal signal sequence in *SaEsxA*. The strict conservation of this motif suggests an important role. This tripeptide is involved in dimer formation and the distinctive hairpin bend structure that is observed. In certain pathogenic bacteria, assembly of the bacterial flagellum is dependent on small, acidic, dimeric polypeptides such as FliS, FlgN and FliT that function as export chaperones to transport proteins across the outer membrane. These proteins display similarities to the chaperones that support the delivery of cytotoxic effector proteins into eukaryotic cells by the bacterial type III secretion system. The structure of one such chaperone, FliS, reveals a compact four-helix bundle that binds its target protein in an extended conformation.⁴⁵ We speculate that *EsxA* might function in similar fashion and that the shallow cleft at the WXG motif may represent a peptide recognition feature by which cargo proteins might be acquired for transport. One possible model is that N-terminus of the partner subunit could adopt or be forced into a different conformation allowing partner proteins to bind at the WXG motif.

The next and important stage of research in this area is to identify exactly which substrates are exported, investigate interactions with *EsxA/B* and the mechanism of transport. With the new structural information presented here we would then be well placed to further dissect the ESX-1 system.

Experimental procedures

Reagents

Chemicals of the highest quality available were sourced from Sigma-Aldrich and VWR International except where stated otherwise. The genes encoding *S. aureus* *EsxA* and *EsxB* were obtained by PCR amplification from genomic DNA (Strain SA 113 - ATCC 35556). Oligonucleotide primers for *EsxA* (forward, 5' ACG **CAT ATG** GCA ATG ATT AAG ATG AGT CCA 3' and reverse, 5' ACG **CTC GAG** TTA TTG CAA ACC GAA ATT ATT 3') and *EsxB* (forward, 5' ACG **CAT ATG** GGT GGA TAT AAA GGG ATT AAA 3' and reverse 3' ACG **CTC GAG** TCA TGG GTT CAC CCT ATC AAG 5') were designed to introduce unique restriction sites for *NdeI* and *XhoI*, respectively (shown in bold). The PCR products were blunt-end ligated into pUC18 (SureClone; GE Healthcare) then subcloned into a modified pET-15b vector (Novagen) which produces a histidine-tagged protein with a Tobacco Etch Virus (TEV) cleavage site. Native protein was obtained from expression in *E. coli* BL21 (DE3) GOLD cells (Novagen). Cells were grown in LB supplemented with 100 µg.ml⁻¹ carbenicillin. Expression was induced at 25 °C with the addition of 1 mM isopropyl-D-thiogalactopyranoside, and growth continued for a further 8 hours. Selenomethionine derivative *SaEsxA* was prepared using the methionine auxotroph strain *E. coli* B834. Cells were grown in minimal media supplemented with normal amino acids, except that L-methionine was replaced with L-SeMet using an established protocol.⁴⁶

Purification and crystallisation

Cells were harvested by centrifugation, resuspended in buffer A 100 mM Tris-HCl (pH 7.5) 5 mM benzamidine, 5 mM lysozyme, 100 units DNase, 5 mM phenylmethylsulphonyl fluoride, and left on ice for 15 minutes before lysis in a One-Shot cell disruptor (Constant Cell Disruption Systems) at 33 Kpsi. The insoluble debris was separated by centrifugation (53,000 g, 10 min, 4 °C), the supernatant passed through a 0.2 µm filter then applied to a 5 ml HisTrap HP column (Amersham Biosciences) that had been pre-charged with Ni²⁺ and eluted with a linear imidazole gradient. Fractions were analyzed by SDS-PAGE, and those that contained a high level of purified EsxA or EsxB were pooled and dialyzed against buffer A. The proteins were then incubated with a His-tagged TEV protease at 30 °C for 4 hours to remove the His-tag, and the sample passed through a HisTrap HP column to remove the cleaved components, the protease and uncleaved sample. The cleaved protein retains a Gly-Ser-His sequence prior to the initiating Met1. The flow-through was collected, and passed through a size exclusion chromatography column equilibrated with buffer A (HiLoad 16/60, Superdex 75 Prep grade, Amersham Biosciences). This column had previously been calibrated with molecular weight standards, blue dextran (> 2,000 kDa), bovine serum albumin (67 kDa), carbonic anhydrase (29.5 kDa) and cytochrome c (12.5 kDa, GE Healthcare; data not shown).

The EsxA and EsxB containing fractions were pooled, and buffer exchanged into 100 mM sodium cacodylate (pH 5.5) then concentrated to 10 mg ml⁻¹ (5kD centrifugal concentrator, Vivascience). Protein concentration was determined spectrophotometrically using a theoretical extinction coefficient of 6990 M⁻¹cm⁻¹.⁴⁷ The full incorporation of SeMet and purity of the sample was confirmed by matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) and SDS-PAGE.

Two crystal forms were obtained for *Sa*EsxA. Triclinic blocks (form I) with dimensions 0.3 × 0.2 × 0.1 mm, were grown at 20 °C by vapor diffusion in hanging drops assembled from 2 µl of protein solution and 2 µl of reservoir (0.1M sodium MES (2-morpholinoethanesulfonic acid) pH 5.5, 80 mM zinc acetate, 8 % w/v polyethylene glycol 800). SeMet labeled protein crystals were grown under similar conditions, except that the drops were maintained at 4 °C. The second crystal form of *Sa*EsxA was obtained during attempts to form a EsxA:EsxB complex following methods applied to prepare the heterodimeric *M. tuberculosis* CFP-10/ESAT-6.^{25,26} *Sa*EsxA and *Sa*EsxB were mixed in equimolar ratio and incubated for 30 minutes prior to setting up crystallization trials. The crystallization conditions were similar to those described above, leading to formation of monoclinic crystals. Crystals of both forms appeared after several days.

These crystals were cryo-protected using 20% PEG 300 then cooled in a stream of gaseous nitrogen at -173°C. The form I crystals were characterized in-house on a Rigaku Micromax 007 X-ray generator equipped with a Raxis IV++ image plate and stored under nitrogen. Multi-wavelength anomalous dispersion diffraction data were measured from a single SeMet labeled crystal on ID29 with an ADSC Q315 Charge Couple Device detector at the European Synchrotron Radiation Facility, Grenoble, France. The data were processed using MOSFLM⁴⁸ and SCALA⁴⁹, which are part of the Collaborative Computational Project number 4 suite of programs,⁵⁰ and relevant statistics are shown in Table 1. Data for the form II crystals were measured in-house then integrated and scaled using DENZO and SCALEPACK.⁵¹

Structure determination and refinement

The structure of SeMet substituted *Sa*EsxA was solved using MAD data, collected from a single form I crystal, in the program SHELX.⁵² A strong anomalous dispersion signal was

observed to 2.0 Å resolution and applying this resolution limit allowed the location of two selenium positions out of the six expected for a dimer in the asymmetric unit. Two metal ion sites were also identified, as determined after structure solution, but these were not included in calculation of the first phase set. The phase set was used with the automated electron density map interpretation software Arp-Warp,⁵³ employing default settings, and phases extended to 1.6 Å resolution. The figure-of-merit increased from 0.5 to 0.8. However, only about one-third of the structure was modeled, mainly a single helix as poly-alanine. Areas of the map for which no model had been constructed were near the metal ion binding sites and the electron density for the divalent cations and cacodylate may have complicated automated map interpretation and subsequent assignment of sequence to structure. Nevertheless, the resulting high resolution electron density map was of excellent quality and the remainder of model was built using COOT.⁵⁴ Refinement by maximum-likelihood restrained least-squares methods was then carried out against the remote (λ^3) dataset in Refmac5⁵⁵ and the resolution extended to 1.4 Å. These calculations were interspersed with inspection of electron and difference Fourier maps calculated with the CCP4 suite and incorporation of water molecules and ions. Metal ions were assigned as Zn^{2+} and a cacodylate. We cannot rule out the possibility that another divalent cation such as Ni^{2+} is present but the concentration of Zn^{2+} in the crystallization mixture (80 mM) gives us confidence in the assignment.

A subset of data (5%) was used for R-free calculations to assist decisions during refinement. Strict non-crystallographic symmetry restraints and isotropic *B*-factor refinement were used during the initial rounds of refinement. Once the model was complete these non-crystallographic symmetry restraints were removed and anisotropic *B*-factor refinement carried out. The inclusion of anisotropic *B*-factors produced a reduction in *R*-work/*R*-free from 22/25 % to 16/21 %, improved the quality of the electron density map and served to reduce difference density features near to the metal ions. The stereochemistry of the model was inspected with PROCHECK.⁵⁶ The structure of form II was solved by MR using the refined form I structure as the search model in PHASER.⁵⁷ The structure was refined to 1.9 Å resolution combining least-squares calculations in Refmac5 with electron and difference density map inspection to guide model manipulation, water and cation identification in COOT. The resolution to which data were available did not justify the use of anisotropic *B*-factor refinement. Molecular images were prepared using PyMol[‡] and structural superpositions performed using SSM.⁵⁸

Data deposition

Atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org, with codes 2vs0 and 2vrz for crystal forms I and II respectively.

Acknowledgments

Funded by the Biotechnology and Biological Sciences Research Council (UK), [Structural Proteomics of Rational Targets] and The Wellcome Trust [grant numbers 082596 and 083481]. We are grateful to David Norman and Tracy Palmer for advice and to the European Synchrotron Radiation Facility for beam time and excellent staff support.

Abbreviations

APBS

Adaptive Poisson-Boltzmann Solver

[‡]DeLano, W.L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA. <http://pymol.sourceforge.net/http://pymol.sourceforge.net>.

BCG	Bacille Calmette-Guérin
CCP4	Collaborative Computational Project Number 4
CFP-10	Culture Filtrate Protein-10 kDa
DPI	Diffraction-Component Precision Index
ESAT-6	Early Secreted Antigenic Target-6 kDa
ESX-1	ESAT-6 secretion system
FSD	FtsK/SpoIIIE domains
MAD	multi-wavelength anomalous dispersion
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MES	2-morpholinoethanesulfonic acid
MR	molecular replacement
NCS	Non-Crystallographic Symmetry
NMR	nuclear magnetic resonance
RD1	Region of difference
r.m.s.d.	root-mean-square-deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sec	secretion
SeMet	L-selenomethionine
Tat	twin-arginine translocation

References

1. Archer GL. *Staphylococcus aureus*: a well-armed pathogen. Clin. Infect. Dis. 1998; 26:1179–1181. [PubMed: 9597249]
2. Bancroft EA. Antimicrobial resistance, it's not just for hospitals. J. Amer. Med. Assoc. 2007; 298:1803–1804.
3. Lowy FD. Antimicrobial resistance: the example of *Staphylococcus aureus*. J. Clin. Invest. 2003; 111:1265–1273. [PubMed: 12727914]
4. Projan SJ, Bradford PA. Late stage antibacterial drugs in the clinical pipeline. Curr. Opin. Microbiol. 2007; 10:441–446. [PubMed: 17950658]
5. Finlay BB, Falkow S. Common themes in microbial pathogenicity revisited. Microbiol Mol. Biol. Rev. 1997; 61:136–169. [PubMed: 9184008]
6. Jett BD, Gilmore MS. Internalization of *Staphylococcus aureus* by human corneal epithelial cells: role of bacterial fibronectin-binding protein and host cell factors. Infect. Immun. 2002; 70:4697–4700. [PubMed: 12117986]
7. Sibbald MJ, Ziebandt AK, Engelmann S, Hecker M, de Jong A, Harmsen HJ, Raangs GC, Stokroos I, Arends JP, Dubois JY, van Dijk JM. Mapping the pathways to staphylococcal pathogenesis by comparative secretomics. Microbiol. Mol. Biol. Rev. 2006; 70:755–88. [PubMed: 16959968]
8. Peacock SJ, Foster TJ, Cameron BJ, Berendt AR. Bacterial fibronectin-binding proteins and endothelial cell surface fibronectin mediate adherence of *Staphylococcus aureus* to resting human endothelial cells. Microbiology. 1999; 145:3477–3486. [PubMed: 10627045]
9. Roche FM, Downer R, Keane F, Speziale P, Park PW, Foster TJ. The N-terminal A domain of fibronectin-binding proteins A and B promotes adhesion of *Staphylococcus aureus* to elastin. J. Biol. Chem. 2004; 279:38433–38440. [PubMed: 15234962]

10. Economou A. Following the leader: bacterial protein export through the Sec pathway. *Trends Microbiol.* 1999; 7:315–320. [PubMed: 10431204]
11. Champion PA, Cox JS. Protein secretion systems in *Mycobacteria*. *Cell Microbiol.* 2007; 9:1376–1384. [PubMed: 17466013]
12. Burts ML, Williams WA, DeBord K, Missiakas DM. EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. *Proc. Natl. Acad. Sci. U S A.* 2005; 102:1169–1174. [PubMed: 15657139]
13. Sorensen AL, Nagai S, Houen G, Andersen P, Andersen AB. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect. Immun.* 1995; 63:1710–1717. [PubMed: 7729876]
14. Ulrichs T, Munk ME, Mollenkopf H, Behr-Perst S, Colangeli R, Gennaro ML, Kaufmann SH. Differential T cell responses to *Mycobacterium tuberculosis* ESAT6 in tuberculosis patients and healthy donors. *Eur. J. Immunol.* 1998; 28:3949–3958. [PubMed: 9862331]
15. Berthet FX, Rasmussen PB, Rosenkrands I, Andersen P, Gicquel B. A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). *Microbiology.* 1998; 144:3195–3203. [PubMed: 9846755]
16. Harboe M, Oettinger T, Wiker HG, Rosenkrands I, Andersen P. Evidence for occurrence of the ESAT-6 protein in *Mycobacterium tuberculosis* and virulent *Mycobacterium bovis* and for its absence in *Mycobacterium bovis* BCG. *Infect. Immun.* 1996; 64:16–22. [PubMed: 8557334]
17. Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR. Deletion of RD1 from *Mycobacterium tuberculosis* mimics *bacille Calmette-Guérin* attenuation. *J. Infect. Dis.* 2003; 187:117–123. [PubMed: 12508154]
18. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol. Microbiol.* 2002; 46:709–717. [PubMed: 12410828]
19. Pym AS, Brodin P, Majlessi L, Brosch R, Demangel C, Williams A, Griffiths KE, Marchal G, Leclerc C, Cole ST. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat. Med.* 2003; 9:533–539. [PubMed: 12692540]
20. Gey van Pittius NC, Gamielidien J, Hide W, Brown GD, Siezen RJ, Beyers AD. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome Biol.* 2001; 2:0044.
21. Pallen MJ. The ESAT-6/WXG100 superfamily - and a new Gram-positive secretion system? *Trends Microbiol.* 2002; 10:209–212. [PubMed: 11973144]
22. Stanley SA, Raghavan S, Hwang WW, Cox JS. Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proc. Natl. Acad. Sci. U S A.* 2003; 100:13001–13006. [PubMed: 14557536]
23. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature.* 1998; 393:537–544. [PubMed: 9634230]
24. Renshaw PS, Panagiotidou P, Whelan A, Gordon SV, Hewinson RG, Williamson RA, Carr MD. Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6*CFP-10 complex. Implications for pathogenesis and virulence. *J. Biol. Chem.* 2002; 277:21598–603. [PubMed: 11940590]
25. Meher AK, Bal NC, Chary KV, Arora A. *Mycobacterium tuberculosis* H37Rv ESAT-6-CFP-10 complex formation confers thermodynamic and biochemical stability. *FEBS J.* 2006; 273:1445–62. [PubMed: 16689931]
26. Renshaw PS, Lightbody KL, Veverka V, Muskett FW, Kelly G, Frenkiel TA, Gordon SV, Hewinson RG, Burke B, Norman J, Williamson RA, Carr MD. Structure and function of the

- complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO J.* 2005; 24:2491–2498. [PubMed: 15973432]
27. Teutschbein J, Schumann G, Möllmann U, Grabley S, Cole ST, Munder T. A protein linkage map of the ESAT-6 secretion system 1 (ESX-1) of *Mycobacterium tuberculosis*. *Microbiol. Res.* 2007 Epub ahead of print.
 28. Harding MM. Geometry of metal-ligand interactions in proteins. *Acta Crystallographica Section D-Biological Crystallography.* 2001; 57:401–411.
 29. Paliakasis CD, Kokkinidis M. Relationships between sequence and structure for the four-alpha-helix bundle tertiary motif in proteins. *Protein Eng.* 1992; 5:739–748. [PubMed: 1287654]
 30. Kamtekar S, Hecht MH. Protein Motifs. 7. The four-helix bundle: what determines a fold? *Faseb J.* 1995; 9:1013–1022. [PubMed: 7649401]
 31. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 2007; 372:774–797. [PubMed: 17681537]
 32. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A. Pfam: clans, web tools and services. *Nucleic Acids Res.* 2006; 34:D247–251. [PubMed: 16381856]
 33. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32:1792–97. [PubMed: 15034147]
 34. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000; 16:404–405. [PubMed: 10869041]
 35. Brodin P, de Jonge MI, Majlessi L, Leclerc C, Nilges M, Cole ST, Brosch R. Functional analysis of early secreted antigenic target-6, the dominant T-cell antigen of *Mycobacterium tuberculosis*, reveals key residues involved in secretion, complex formation, virulence, and immunogenicity. *J. Biol. Chem.* 2005; 280:33953–33959. [PubMed: 16048998]
 36. Champion PA, Stanley SA, Champion MM, Brown EJ, Cox JS. C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. *Science.* 2006; 313:1632–1636. [PubMed: 16973880]
 37. de Jonge MI, Pehau-Arnsudet G, Fretz MM, Romain F, Bottai D, Brodin P, Honoré N, Marchal G, Jiskoot W, England P, Cole ST, Brosch R. ESAT-6 from *Mycobacterium tuberculosis* dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. *J. Bacter.* 2007; 189:6028–6034.
 38. Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M, Eisenberg D, Russell RG, Derrick SC, Collins FM, Morris SL, King CH, Jacobs WR Jr. The primary mechanism of attenuation of bacillus Calmette-Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc. Natl. Acad. Sci. U S A.* 2003; 100:12420–12425. [PubMed: 14557547]
 39. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 1993; 233:123–138. [PubMed: 8377180]
 40. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res.* 2005; 33:89–93.
 41. Bakolitsa C, Cohen DM, Bankston LA, Bobkov AA, Cadwell GW, Jennings L, Critchley DR, Craig SW, Liddington RC. Structural basis for vinculin activation at sites of cell adhesion. *Nature.* 2004; 430:583–586. [PubMed: 15195105]
 42. Gingras AR, Vogel KP, Steinhoff HJ, Ziegler WH, Patel B, Emsley J, Critchley DR, Roberts GC, Barsukov IL. Structural and dynamic characterization of a vinculin binding site in the talin rod. *Biochemistry.* 2006; 45:1805–1817. [PubMed: 16460027]
 43. Hoellerer MK, Noble ME, Labesse G, Campbell ID, Werner JM, Arold ST. Molecular recognition of paxillin LD motifs by the focal adhesion targeting domain. *Structure.* 2003; 11:1207–1217. [PubMed: 14527389]
 44. Kurisu G, Zakharov SD, Zhalnina MV, Bano S, Eroukova VY, Rokitskaya TI, Antonenko YN, Wiener MC, Cramer WA. The structure of BtuB with bound colicin E3 R-domain implies a translocon. *Nat. Struct. Biol.* 2003; 10:948–954. [PubMed: 14528295]

45. Evdokimov AG, Phan J, Tropea JE, Routzahn KM, Peters HK, Pokross M, Waugh DS. Similar modes of polypeptide recognition by export chaperones in flagellar biosynthesis and type III secretion. *Nat. Struct. Biol.* 2003; 10:789–93. [PubMed: 12958592]
46. Hall DR, Gourley DG, Leonard GA, Duke EBH, Anderson LA, Boxer DH, Hunter WN. The high-resolution crystal structure of the molybdate-dependent transcriptional regulator (ModE) from *Escherichia coli*; a novel combination of domain folds. *EMBO J.* 1999; 18:1435–1446. [PubMed: 10075916]
47. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, MR.; Appel, RD.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker, JM., editor. *The Proteomics Protocols Handbook*. Humana Press; 2005. p. 571-607.
48. Leslie AG. Integration of macromolecular diffraction data. *Acta Crystallographica Section D-Biological Crystallography*. 1999; 55:1696–1702.
49. Evans P. Scaling and assessment of data quality. *Acta Crystallographica Section D-Biological Crystallography*. 2006; 62:72–82.
50. Collaborative Computational project number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D-Biological Crystallography*. 1994; 50:760–763.
51. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology “Macromolecular Crystallography”*. 1997; 276(Pt A):307–326.
52. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Crystallographica Section D-Biological Crystallography*. 2002; 58:1772–1779.
53. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 1999; 6:458–463. [PubMed: 10331874]
54. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography*. 2004; 60:2126–2132.
55. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography*. 1997; 53:240–255.
56. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 1993; 26:283–291.
57. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. *Acta Crystallographica Section D-Biological Crystallography*. 2005; 61:458–464.
58. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D-Biological Crystallography*. 2004; 60:2256–2268.
59. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U S A.* 2001; 98:10037–10041. [PubMed: 11517324]
60. Cruickshank DW. Remarks about protein structure precision. *Acta Crystallographica Section D-Biological Crystallography*. 1999; 55:583–601.

Figure 1.

The *SaEsxA* subunit. A. Ribbon diagram of subunit A. Pro60, which produces the kink on $\alpha 2$, is shown as black van der Waals spheres. The position of the conserved WxG motif is marked and Trp43 depicted as black van der Waals spheres. N and C-terminal positions are marked. B. A calculated electrostatic surface potential of subunit A (blue-positive, red-negative, white-neutral), showing the hydrophobic side of the subunit. The view is the same as 1A. C. A similar representation rotated 180° to show the hydrophilic side of the subunit. B and C were produced using PyMOL with APBS (Adaptive Poisson-Boltzmann Solver)⁵⁸ with electrostatic potential isocontours set at +5 kT/e (blue) and -5 kT/e (red).

Figure 2.

The *SaEsxA* dimer. A ribbon diagram with subunit colored red (subunit A) and blue (subunit B) with helices labeled. The orientation of subunit A is similar to that shown in Fig. 1. Trp43 and Pro60 residues are shown as black van der Waals spheres.

Figure 3.

The WXG motif and part of the dimer interface. A stereoview of one end of a *SaEsxA* dimer, with subunit A in red, and subunit B cyan. The subunits are shown as semi-transparent van der waals surfaces around the ribbon secondary structure. Residues are depicted as sticks colored according to atom type (N blue, O red, S yellow, C of subunit A pink and of subunit B cyan). The WXG motif from subunit B is shown interacting with the N-terminal section of subunit A to form a short span of β -sheet (hydrogen bonds are marked as magenta dotted lines). Also highlighted are Leu90, Ser91 and Gly95 from the C-terminus secretion signal.

Figure 4.

Primary and secondary structure of *SaEsxA* and comparison with *SaEsxB* and CFP-10 and ESAT-6. Cylinders on top of the sequence depict helices $\alpha 1$ - $\alpha 2$ assigned from the crystal structure of *SaEsxA*. The tryptophan and glycine residues in the WXG motif, present in all four sequences, are shown as white letters in black boxes. Residues strictly conserved in more than two sequences are shown as black letters in red boxes, and at least in two sequences are shown as black letters in gray boxes. Blue stars denote the position of mutations introduced into ESAT-6 which are discussed in text. Numbering is according to *SaEsxA*.

Figure 5. Comparison with ESAT-6. Stereo diagram showing the structural overlay of *SaEsxA* on ESAT-6. *SaEsxA* is depicted in red and ESAT-6 in gray. Residues of ESAT-6 mutated to investigate function and the corresponding structurally aligned residues of *SaEsxA* are shown in sticks and labeled. The partner subunit of the *SaEsxA* dimer is shown in van der Waals surface representation and colored cyan.

Table 1

Crystallographic statistics.

Structure	Form I λ^1 (peak)	Form I λ^2 (inflection point)	Form I λ^3 (remote)	Form II
Wavelength (Å)	0.97904	0.97930	0.97549	1.5418
Space group			<i>P</i> 1	<i>P</i> 21
<i>a</i> , <i>b</i> , <i>c</i> (Å) <i>α</i> , <i>β</i> , <i>γ</i> (°)			25.72, 38.99, 42.99 67.4, 84.6, 88.7	46.23, 27.64, 64.39 90.0, 105.4, 90.0
Resolution range (Å)	20-1.5	20-1.5	40-1.4	20-1.9
Unique reflections / Redundancy Anomalous redundancy	23341 / 3.7 1.9	23490 / 3.8 1.9	29022 / 3.8 / 1.9	13006 / 3.6 -
Completeness (%)	95.2 (93.2) / 90.1 (88.6)	96.0 (93.4) / 90.3 (90.0)	95.4(92.3) / 91.2 (86.4)	95.5 (69.0)
$\langle I/\sigma(I) \rangle$	18.6 (5.9)	17.4 (5.0)	16.1 (2.0)	11.2 (7.9)
<i>R</i> -sym (%) / <i>R</i> -anom (%)	5.2 (19.1) / 6.5 (13.9)	4.9 (20.7) / 5.4 (14.5)	6.1(65.0) / 5.9(24.3)	5.4 (14.0)
Wilson <i>B</i> (Å ²)	14.7	14.7	16.2	21.3
Protein residues / Solvent / Ions			167 / 175 / 7	183 / 117 / 7
<i>R</i> -work / <i>R</i> -free			16.6 / 21.7	18.3 / 22.4
Average <i>B</i> (Å ²) Overall / Protein / Solvent / Ions			21.1 / 19.1 / 35.5 / 28.2	21.2 / 20.8 / 31.2 / 27.8
r.m.s.d Bond lengths (Å) / angles (°)			0.012 / 1.281	0.015 / 1.380
DPI (Å)			0.07	0.20
Ramachandran plot Most favored region Additional allowed regions			98.7 1.3	98.2 1.8

Values in parentheses pertain to the highest resolution shell (width = 0.1 Å).

$R\text{-sym} = \frac{\sum_h |I_i(h) - \langle I(h) \rangle|}{\sum_h I_i(h)}$ where $I_i(h)$ is the intensity of the i^{th} measurement of reflection h and $\langle I(h) \rangle$ is the mean value of $I(h)$ for all measurements.

$R\text{-anom} = \frac{\sum_{hkl} |I(+) - I(-)|}{\sum_{hkl} (\langle I(+) \rangle + \langle I(-) \rangle)}$

$R\text{-work} = \frac{\sum_{hkl} ||F_o| - |F_c||}{\sum_{hkl} |F_o|}$, where F_o is the observed structure-factor amplitude and F_c the structure-factor amplitude calculated from the model. *R*-free is the same as *R*-work except only calculated using a subset, 5%, of the data that are not included in any refinement calculations.

Wilson *B* is derived from a plot of $\ln(\langle I_\theta \rangle / \theta)$ against $\sin^2 \theta / \lambda^2$. $\langle I_\theta \rangle$ is the average of the intensities in the range $\sin^2 \theta$ and θ is the value of θ for the particular mean value of θ . The slope of the resulting line is $-2B$. ($\langle I_\theta \rangle / \theta$).

B, the thermal parameter = $8\pi^2(\text{mean square displacement of the atomic vibration})$.

DPI = Diffraction-component precision index.⁶⁰