Molecular Features Governing the Stability and Specificity of Functional Complex Formation by Mycobacterium tuberculosis CFP-10/ESAT-6 Family Proteins*

Received for publication, January 7, 2008, and in revised form, March 20, 2008 Published, JBC Papers in Press, April 21, 2008, DOI 10.1074/jbc.M800123200

Kirsty L. Lightbody, Dariush Ilghari, Lorna C. Waters, Gemma Carey, Mark A. Bailey, Richard A. Williamson, Philip S. Renshaw, and Mark D. Carr

From the Department of Biochemistry, Henry Wellcome Building, University of Leicester, Leicester, LE1 9HN, United Kingdom, the Iran Ministry of Health and Medical Education, MOHME Building, Sharake Qods, Tehran, Iran 1467664961, and the Department of Biosciences, University of Kent, Canterbury, Kent, CT2 7NJ, United Kingdom

The Mycobacterium tuberculosis complex CFP-10/ESAT-6 family proteins play essential but poorly defined roles in tuberculosis pathogenesis. In this article we report the results of detailed spectroscopic studies of several members of the CFP-10/ESAT-6 family. This work shows that the CFP-10/ESAT-6 related proteins, Rv0287 and Rv0288, form a tight 1:1 complex, which is predominantly helical in structure and is predicted to closely resemble the complex formed by CFP-10 and ESAT-6. In addition, the Rv0287-Rv0288 complex was found to be significantly more stable to both chemical and temperature induced denaturation than CFP-10-ESAT-6. This approach demonstrated that neither Rv0287-Rv0288 nor the CFP-10-ESAT-6 complexes are destabilized at low pH (4.5), indicating that even in low pH environments, such as the mature phagosome, both Rv0287-Rv0288 and CFP-10-ESAT-6 undoubtedly function as complexes rather than individual proteins. Analysis of the structure of the CFP-10-ESAT-6 complex and optimized amino acid sequence alignments of M. tuberculosis CFP-10/ESAT-6 family proteins revealed that residues involved in the intramolecular contacts between helices are conserved across the CFP-10/ESAT-6 family, but not those involved in primarily intermolecular contacts. This analysis identified the molecular basis for the specificity and stability of complex formation between CFP-10/ESAT-6 family proteins, and indicates that the formation of functional complexes with key roles in pathogenesis will be limited to genome partners, or very closely related family members, such as Rv0287/Rv0288 and Rv3019c/Rv3020c.

Tuberculosis remains one of the most significant infectious diseases of humans, with approximately one-third of the world’s population currently infected and resulting in over 1.5 million deaths annually (1, 2). Studies catalyzed by the completion of genome sequences for a number of closely related mycobacterial pathogens, including Mycobacterium tuberculosis (3), Mycobacterium leprae (4), and Mycobacterium bovis (5) have led to the identification of a number of essential protein virulence factors, such as the secreted proteins CFP-10, ESAT-6, and MPT70.

CFP-10 (Rv3874 or EsxB) and ESAT-6 (Rv3875 or EsxA) are members of a large family of mycobacterial proteins, which typically consist of about 100 amino acids and are characterized by their organization in pairs within the genome and by the conservation of a central WXG motif (6–8). In addition, several members of the family have been identified as potent T-cell antigens (9–12). The M. tuberculosis genome contains 22 pairs of genes coding for CFP-10/ESAT-6 family proteins, which are located at 11 genomic loci and are often preceded by genes for PE and PPE proteins (6). The genes for CFP-10 and ESAT-6 have been shown to be organized as a small operon, which is co-transcribed and translated (13), therefore it seems very likely that other genome pairs within this family will form similar operons. In addition, several pairs of CFP-10/ESAT-6 family proteins have already been shown to form tight complexes, including CFP-10/ESAT-6, Rv0287/Rv0288 (EsxG/EsxH), and Rv3019c/Rv3020c (EsxR/EsxS) (7, 14, 15), suggesting that complex formation will be a common feature across the family. For CFP-10 and ESAT-6, complex formation has been shown to result in both proteins adopting a stable tertiary structure, which is reflected in significantly increased resistance to chemical denaturation and protease digestion compared with the individual proteins, and undoubtedly represents the functional forms of CFP-10 and ESAT-6 (7, 16).

CFP-10 and ESAT-6, as well as several other CFP-10/ESAT-6 family members, are found in the short-term culture filtrate of M. tuberculosis cells, despite lacking a conventional signal sequence (10, 17–20). Recent studies indicate that secretion of the CFP-10/ESAT-6 complex is an active process involving a specialized transport system formed from the products of several flanking genes, including the putative membrane ATPases Rv3870 and Rv3871 and the predicted transmembrane protein Rv3877 (21–24). Secretion of CFP-10 and ESAT-6 also appears to depend upon proteins encoded by the distant Rv3613c-Rv3616c operon (25, 26). Analogous gene clusters coding for homologues of essential elements of the CFP-10/ESAT-6 secretory complex are also found surrounding other pairs of CFP-10/ESAT-6 family genes, including Rv0287/Rv0288 (EsxG/
Features and Properties of CFP-10-ESAT-6 Family Complexes

EcSH, Rv1792/Rv1793 (EcSM/EcSN) and Rv3890c/Rv3891c (EcSC/EcSD) (6). Interestingly, the genome of M. leprae, which probably contains the minimal gene set required for a pathogenic mycobacterium (4), has retained functional copies of both the CFP-10/ESAT-6 and Rv0287/Rv0288 gene clusters, which clearly emphasizes the importance of these proteins to mycobacterial pathogens.

Despite the clear importance of the CFP-10/ESAT-6 family proteins in mycobacterial virulence and pathogenesis, we have relatively little understanding of their precise functions and molecular mechanisms of action. The structural and surface features of the CFP-10/ESAT-6 complex, as well as the ability to specifically bind to the surface of monocyte and macrophage cells, strongly suggests a key role in pathogen to host cell signaling (27). However, it is not known whether complexes formed by other CFP-10/ESAT-6 family proteins will have similar roles, or whether they mediate a diverse range of functions. For example, recent studies have shown that expression of the Rv0287/Rv0288 gene cluster (from Rv0282 to Rv0292) is induced under conditions of iron and zinc starvation in vitro, suggesting a possible role in the scavenging or uptake of iron and zinc (28, 29). Regulation of Rv0287/Rv0288 by IdeR (iron-dependent regulator protein) and ZurB (zinc uptake regulator protein) may also suggest that members of the CFP-10/ESAT-6 protein family are expressed under different physiological conditions, or at various stages throughout infection. Further detailed investigation of the structural and functional features of the complexes formed by CFP-10/ESAT-6 family proteins is clearly a priority, including the need to understand the features governing the specificity of complex formation between CFP-10-related proteins and their ESAT-6 partners, as well as to investigate their stability under conditions found in vivo.

In this article we report the characterization of the structural properties and features of the Rv0287-Rv0288 complex, and compare these with the CFP-10-ESAT-6 complex. We demonstrate that both complexes remain stable under conditions comparable with those expected in vivo, further emphasizing that the functional forms of CFP-10/ESAT-6 family proteins are highly likely to be as complexes formed between CFP-10 and ESAT-6 related genome partners. In addition, combined analysis of the sequence conservation between CFP-10/ESAT-6 family proteins and the structure determined for the CFP-10-ESAT-6 complex has allowed the identification of key residues involved in complex stability and specificity for CFP-10/ESAT-6 family proteins.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—CFP-10 and ESAT-6 were cloned, expressed, and purified as described by Renshaw et al. (7). The genes encoding Rv0287 and Rv0288 were cloned into pET28a-based Escherichia coli expression vectors and were expressed as full-length proteins with an N-terminal His tag (14). Both Rv0287 and Rv0288 were purified from inclusion bodies, which were initially washed four times in 50 mM Tris, 10 mM EDTA, 0.5% Triton X-100 buffer at pH 8.0 and then resolubilized in 25 mM sodium phosphate, 6 M guanidine hydrochloride (GdnHCl)3 buffer at pH 7.4 (1 mM EDTA and 100 μM 4-(aminoethyl)benzenesulfonyl fluoride). The two proteins were refolded by dialysis against 25 mM sodium phosphate, 200 mM sodium chloride, 30 mM imidazole buffer at pH 7.4 (100 μM 4-(aminoethyl)benzenesulfonyl fluoride) and then purified by nickel affinity chromatography. Rv0287 and the Rv0287-Rv0288 also required a final gel filtration chromatography step on a 120-ml Superdex 75 16/60 prepacked column (GE Healthcare) to remove the remaining contaminants.

Fluorescence Spectroscopy—Intrinsic tryptophan fluorescence spectra of Rv0287, Rv0288, and the Rv0287-Rv0288 complex were acquired on a PerkinElmer LS50B luminescence spectrometer at a scan rate of 150 nm/min. The proteins were excited at 280 nm and fluorescence emission was monitored between 300 and 450 nm. The final spectra were corrected for the buffer background and were the result of a smoothed average of 10 accumulations. The individual proteins and the Rv0287-Rv0288 complex were analyzed at 0.5 to 3.0 μM in 25 mM sodium phosphate, 200 mM sodium chloride buffer at pH 6.5. Complex formation between Rv0287 and Rv0288 was investigated by monitoring the shift in the maximum intrinsic tryptophan fluorescence of Rv0288 at increasing molar ratios of Rv0287 to Rv0288 (0.1 to 2.25:1).

Circular Dichroism Spectroscopy—Far UV circular dichroism (CD) spectroscopy was used to determine the secondary structure of the individual Rv0287 and Rv0288 proteins and the Rv0287-Rv0288 complex using a Jasco J-715 spectropolarimeter. All CD spectra were obtained from protein samples dissolved in 25 mM sodium phosphate, 100 mM sodium chloride buffer at pH 6.5 with protein concentrations of 5.0 to 15.0 μM. Spectra were recorded from 180 to 250 nm at a scan speed of 30 nm/min and a response time of 1 s, with each spectrum representing the sum of 10 accumulations. All spectra were acquired at 25 °C in a 1-mm path length cell. Spectra were corrected for the buffer and converted to molar CD per residue. The CD data were analyzed using the CD Pro package to provide estimates of the secondary structure content (30).

NMR Spectroscopy—15N/1H HSQC spectra were acquired from 0.35-ml samples of ~0.8 mM Rv0287-Rv0288 complex in 25 mM sodium phosphate, 100 mM sodium chloride, 100 μM 4-(aminoethyl)benzenesulfonyl fluoride, 0.02% sodium azide buffer at pH 6.5. The NMR spectra were recorded from complexes containing either uniformly 15N-labeled Rv0287 and unlabeled Rv0288, or vice versa, to reduce spectral overlap. The NMR experiments were collected at 35 °C on a 800 MHz Bruker Avance spectrometer equipped with a triple resonance (1H/15N/13C) cryoprobe and processed using the Topspin package (Bruker).

Chemical Denaturation—The conformational stability of Rv0288 and the Rv0287-Rv0288 complex were assessed by determining their resistance to chemical denaturation by GdnHCl. This was monitored by following the shift in the wavelength of maximum intrinsic fluorescence as the concentration of GdnHCl was increased from 0 to 3 M. The samples were allowed to equilibrate for at least 30 min prior to fluorescence

3 The abbreviation used is: GdnHCl, guanidine hydrochloride.
analysis, as described above. Rv0288 and the Rv0287-Rv0288 complex were analyzed at 0.5 M in 25 mM sodium phosphate, 200 mM sodium chloride buffer at pH 6.5.

Temperature Denaturation—The stability of the Rv0287-Rv0288 complex to temperature denaturation was determined by following changes in the CD spectra. The changes in the intensity of the maximum negative peak at 208 nm were recorded as a function of increasing temperature (from 5 to 85 °C), with the temperature rising incrementally at 1 °C/min. All CD spectra were obtained from protein samples at ~15.0 μM in 25 mM sodium phosphate, 100 mM sodium chloride buffer at pH 6.5. The CD spectra were collected in a 1-mm path length cell and corrected for the buffer background and converted to molar CD per residue. Data for the CFP-10-ESAT-6 complex were collected as described by Renshaw et al. (7).

pH Stability—The effect of pH on the stability of both Rv0287-Rv0288 and CFP-10-ESAT-6 complexes was investigated by determining GdnHCl denaturation curves at pH 6.5, 5.5, and 4.5, as described above. The complexes were analyzed at 0.5 μM in either 25 mM sodium phosphate, 100 mM sodium chloride at pH 6.5 or 25 mM sodium acetate, 100 mM sodium chloride at pH 5.5 and 4.5.

Sequence Alignments—CFP-10 and ESAT-6 family proteins were aligned using ClustalW with a Gonnet scoring matrix, a gap opening penalty of 10, and a gap extension penalty of 0.2 (31). Due to very high levels of sequence similarity between some members of the CFP-10/ESAT-6 protein family, Rv0287 and Rv0288 were chosen to represent both Rv0287/Rv0288 (esxG/esxH) and Rv3019c and Rv3020c (esxR/esxS). Similarly, Rv2346c (esxO) and Rv2347c (esxP) were selected to represent five pairs of highly sequence-related proteins (MTINY/QILSS families), including Rv1037c/Rv1038c (esxI/esxJ), Rv1197/Rv1198 (esxK/esxL), Rv1792/Rv1793 (esxM/esxN), and Rv3619c/Rv3620c (esxV/esxW).

RESULTS AND DISCUSSION

Characterization of Complex Formation between Rv0287 and Rv0288—We have previously reported yeast two-hybrid studies showing that Rv0287 and Rv0288, like CFP-10 and ESAT-6, interact with each other to form a tight complex (14). We have now used changes in the wavelength of maximum intrinsic tryptophan fluorescence (λ_max) for Rv0288 induced by Rv0287...
binding to determine the stoichiometry and affinity of this interaction (Fig. 1). The $\lambda_{max}$ of the individual Rv0288 protein is 349 nm, indicating that the four tryptophan residues (Trp$^{43}$, Trp$^{54}$, Trp$^{58}$, and Trp$^{94}$) are predominantly exposed to the solvent. Rv0287 does not contain any tryptophan residues hence, the fluorescence spectra only directly monitor changes in Rv0288. A shift in the wavelength of maximum fluorescence is observed as the ratio of Rv0287 to Rv0288 is increased from 0:1 to 1:1, but further addition of Rv0287 has no significant effect, which strongly suggests the formation of a 1:1 complex. The formation of a heterodimer was confirmed by the line widths observed in $^{15}$N/$^1$H HSQC spectra of the Rv0287-Rv0288 complex, and by the complexes behavior on gel filtration chromatography. The $\lambda_{max}$ of the Rv0287-Rv0288 complex (343 nm) is significantly blue shifted compared with Rv0288 alone (349 nm), suggesting that at least one of the four tryptophan residues from Rv0288 becomes less solvent exposed on formation of the complex. Based on the structure of the CFP-10-ESAT-6 complex (27), it is likely that the highly conserved Trp$^{43}$ residue, which is located in the tight turn between the two helices of ESAT-6, will be largely buried within the hydrophobic core of the Rv0287-Rv0288 complex. Similarly, both Trp$^{54}$ and Trp$^{58}$ are also expected to form part of the hydrophobic core of the complex, located at the interface between Rv0287 and Rv0288. In contrast, the C-terminal Trp$^{94}$ is likely to be fully exposed to the solvent. Thus, the blue shift in tryptophan fluorescence on complex formation probably reflects Trp$^{43}$, Trp$^{54}$, and Trp$^{58}$ occupying a less polar environment within the protein complex compared with the individual Rv0288 protein.

The fluorescence measurements were carried out at a Rv0288 concentration of 0.5 $\mu$M and the lack of any significant shift in $\lambda_{max}$ after the addition of an equimolar amount of Rv0287 indicates not only that the two proteins form a 1:1 complex, but also that the vast majority (>90%) of the two proteins are complexed at this concentration. This suggests a tight interaction between

**FIGURE 4. Structure of the CFP-10-ESAT-6 complex and positions of key residues within the mini hydrophobic core of ESAT-6.** A, ribbon representation of the backbone structure of the CFP-10-ESAT-6 complex, with CFP-10 in red and ESAT-6 in blue (27). B, orientation of key amino acid side chains (Leu$^{36}$, Trp$^{43}$, Tyr$^{51}$, Val$^{54}$, and Gln$^{55}$) within the mini hydrophobic core formed at the tight turn between the two helices in ESAT-6. The backbone of the protein in this region is shown in gray.

**FIGURE 5.** Multiple sequence alignment showing the conservation of amino acids in members of the M. tuberculosis CFP-10/ESAT-6 protein family, with CFP-10-related proteins grouped at the top and ESAT-6 related proteins at the bottom. Rv0287 (EsxG) and Rv0288 (EsxH) also represent the closely related Rv3019c/Rv3020c (EsxR/EsxS) (>80% identity). Similarly, Rv2346c and Rv2347c represent the highly related MTINY and QILSS proteins (>90% identity). The sequences were aligned using ClustalW with a Gonnet scoring matrix, a gap opening penalty of 10 and a gap extension penalty of 0.2 (31). The residues are highlighted as follows: aliphatic residues with hydrophobic side chains (Ala, Ile, Leu, and Val) in blue, aromatic residues (His, Phe, Trp, and Tyr) in light blue, positively charged residues (Arg and Lys) in red, negatively charged residues (Asp and Glu) in dark purple, and polar residues (Asn and Gln) in light purple. Small polar residues (Cys, Ser, and Thr) are colored in green and the remaining small residues (Gly and Pro) are yellow. Colors were applied using Jalview 2.3 (40), with a conservation visibility score of 30%. The positions of the helices are indicated by the bars above the sequences (x-helices in dark gray and 3_10-helices in light gray). Conserved residues at positions a and d within the helical wheel diagrams are shown by colored triangles (blue and red, respectively) under the sequence alignments. Similarly, key residues involved in stabilizing the mini hydrophobic core within ESAT-6 are highlighted by the solid circles.
the two proteins with an estimated dissociation constant ($K_d$) for the complex of 5 nM or lower.

The secondary structure of the Rv0287/H18528 Rv0288 complex and the individual proteins were determined by far UV circular dichroism (CD) spectroscopy. The spectra obtained for the isolated Rv0287 protein (Fig. 2) are typical of an unstructured, random coil polypeptide and analysis with CDPro (30) indicates only 1% $\alpha$-helical and 11% $\beta$-sheet structure. In contrast, the spectra for Rv0288 and the Rv0287/H18528 Rv0288 complex (Fig. 2) are typical of proteins with significant helical structure, which are characterized by intense negative CD peaks at 208 and 221 nm (32). Rv0288 and the complex were estimated by CDPro (30) to contain 24% $\alpha$-helix (18% $\beta$-sheet and 58% unstructured) and 50% $\alpha$-helix (23% $\beta$-sheet and 27% unstructured), respectively.

However, the CD spectra were acquired from Rv0287 and Rv0288, which both contain an extra 20 residues at the N terminus from the His6 tag and linker. These residues are expected to be unstructured, which clearly suggests that the helical content of the regions corresponding to Rv0288 and the Rv0287/H18528 Rv0288 complex are significantly higher. If the secondary structures are corrected for the presence of an unstructured N-terminal His tag, the helical content is about 30% for Rv0288 and 60% for the complex. The helical content of the Rv0287/H18528 Rv0288 complex is similar to that reported for the CFP-10/ESAT-6 complex (70%) (7), which together with significant sequence similarity suggests that both complexes form similar structures (27).

Characterization of the Stability of the Rv0287/Rv0288 Complex and Individual Proteins—The CD studies described above clearly show that the isolated Rv0287 exists as an unfolded random coil polypeptide, however, the stability of Rv0288 and the Rv0287/Rv0288 complex were assessed by determining their
resistance to chemical and heat-induced denaturation, which were monitored by changes in intrinsic tryptophan fluorescence and CD spectra. The chemical (Fig. 3) and temperature (data not shown) denaturation curves obtained for Rv0288 both show a gradual, non-cooperative unfolding of the protein, which indicates that Rv0288 contains no stable tertiary structure and the spectral changes seen probably reflect the unfolding of isolated helical regions. The presence of significant helical structure but no apparent resistance to denaturation implies that Rv0288 exists as a molten globule-type structure. Interestingly, ESAT-6 in the absence of its binding partner CFP-10 also exists in a molten globule state (7, 33). Examination of the structure of the CFP-10-ESAT-6 complex reveals a mini hydrophobic core within the region of ESAT-6 forming the tight turn between the pair of anti-parallel helices, as illustrated in Fig. 4. The structure of this region of ESAT-6 is clearly stabilized by favorable van der Waals interactions between the side chains of Leu36, Trp43, Tyr51, Val54, and Gln55. The equivalent region of CFP-10 in the complex is not stabilized by a comparable network of interactions, which strongly suggests that the formation of this mini hydrophobic core is necessary to induce partial folding of ESAT-6, and probably explains why isolated ESAT-6 alone forms a molten globule, whereas CFP-10 is a random coil polypeptide. The importance of this region of ESAT-6 in complex formation has been highlighted by several recent studies (34, 35) in which the highly conserved tryptophan residue within the WXG motif was mutated to arginine in both ESAT-6 (W43R) and CFP-10 (W43R). The substitution in ESAT-6 had a profound effect, completely abolishing complex formation with CFP-10, but the substitution in CFP-10 had no significant effect on complex formation. This strengthens the proposal that partial folding of ESAT-6, mediated by formation of the mini hydrophobic core, is essential for assembly of the CFP-10-ESAT-6 protein complex.

The residues forming the mini hydrophobic core in the turn region of ESAT-6 are fairly well conserved throughout the ESAT-6 protein family (Fig. 5), suggesting that all ESAT-6-related proteins are likely to form a molten globule-type structure in the absence of a CFP-10-related partner protein. The sequence alignment also shows that Rv0288 contains a third aromatic residue within the mini hydrophobic core region, with Val54 substituted by a tryptophan residue, which may result in greater stabilization of the molten globule.

In contrast to isolated Rv0288, the Rv0287-Rv0288 complex shows significant resistance to chemical denaturation and a
classical cooperative unfolding curve. The complex is stable to \( \sim 1.0 \text{ M GdnHCl} \) before cooperatively unfolding with a midpoint of around 1.7 \( \text{M GdnHCl} \) (Figs. 3 and 6A). Similarly, Fig. 6B shows that the Rv0287-Rv0288 complex resists heat-induced denaturation to at least 60 \( ^\circ \text{C} \) before cooperatively unfolding with a midpoint of about 70 \( ^\circ \text{C} \). The chemical and thermal denaturation curves observed for the complex are typical of a stable, folded protein complex (36). In common with CFP-10 and ESAT-6 this clearly suggests that the functional form of Rv0287 and Rv0288 is likely to be a 1:1 protein complex.

**Comparison of the CFP-10-ESAT-6 and Rv0287-Rv0288 Complexes**—The effect of increasing GdnHCl concentration on the wavelength of maximum intrinsic tryptophan fluorescence for Rv0287-Rv0288 and CFP-10-ESAT-6 in 25 \( \text{mM sodium phosphate, 100 mM sodium chloride at pH 6.5} \) is shown in Fig. 6A. Both complexes show initial stability to denaturation followed by cooperative unfolding, but Rv0287-Rv0288 is clearly more resistant to chemical denaturation by GdnHCl than CFP-10-ESAT-6, with midpoints of denaturation of about 1.7 \( \text{M} \) and 0.7 \( \text{M} \) GdnHCl, respectively. Likewise, temperature denaturation curves for both complexes indicate that the Rv0287-Rv0288 complex is also significantly more resistant to thermal denaturation (Fig. 6B). Rv0287-Rv0288 is stable to at least 60 \( ^\circ \text{C} \), with a denaturation midpoint of about 70 \( ^\circ \text{C} \), whereas CFP-10-ESAT-6 starts to unfold at about 45 \( ^\circ \text{C} \) and has a denaturation midpoint of around 55 \( ^\circ \text{C} \) (Fig. 6B). Similarities in sequence and secondary structure content strongly suggest that the two complexes will form very similar structures, however, the denaturation studies suggest that the Rv0287-Rv0288 complex contains some additional features that enhance its stability compared with CFP-10-ESAT-6, and may reflect functional differences.

It has recently been proposed that under the acidic conditions found within the phagosome the CFP-10-ESAT-6 complex may be destabilized, resulting in dissociation of the two proteins and the potential for the individual proteins to play important functional roles (37). The fluorescence and CD studies described above were carried out at pH 6.5, which reflects the conditions in phagosomes containing live mycobacteria (38). The chemical denaturation curves shown in Fig. 7 compare the stability of Rv0287-Rv0288 and CFP-10-ESAT-6 at pH 6.5, 5.5, and 4.5. The data clearly demonstrates that lowering the pH from 6.5 to 4.5 has no effect on the stability of the Rv0287-Rv0288 complex. Similarly, the CFP-10-ESAT-6 complex shows no decrease in stability and actually appears to be slightly more stable at pH 4.5 (Fig. 7B), with a denaturation midpoint of 0.9 \( \text{M GdnHCl} \) compared with 0.7 \( \text{M} \) at pH 6.5 and 5.5. The stability of both complexes from pH 6.5 to
4.5 strongly suggests that any possible functions within the phagosome are likely to be associated with the complexes rather than the individual proteins.

Close analysis of the structure of the CFP-10/ESAT-6 complex reveals an extensive hydrophobic contact surface between the two proteins (~1800 Å²), which is primarily stabilized by favorable van der Waals interactions (27). The interface also contains a small number of salt bridges (Glu¹⁴–Lys³⁸ and Glu⁷¹–Lys⁷⁷) and hydrogen bonds between CFP-10 and ESAT-6, respectively. The enhanced stability of the CFP-10/ESAT-6 complex could simply arise from a larger contact surface between the proteins resulting in a more extensive set of favorable van der Waals contacts, or alternatively, the interface may be very similar in area but feature an increased number of hydrogen bonds and salt bridges.

A convenient way to visualize the interfaces between helices is to make use of helical wheel representations, as illustrated for the helical regions present in the CFP-10/ESAT-6 complex, which are assigned as seen in the high resolution structure (27). As expected for a four-helix bundle-type structure, the residues at positions a and d within the heptad helical repeat (a-b-c-d-e-f-g) form the hydrophobic interface between helices of CFP-10 and ESAT-6, with residues at positions b, c, e, and g shielding the hydrophobic core from the aqueous solution (39). The intramolecular interfaces stabilizing the helical hairpins formed by CFP-10 and ESAT-6 are primarily comprised of residues located at positions a, d, and e (Fig. 8). Interestingly, the helix-turn-helix structure of CFP-10 is stabilized by a number of salt bridges at positions e and b (Glu¹⁹–Arg⁷⁷, Asp³⁰–Lys⁶⁶, Glu³³–Lys⁶⁶, and Lys²⁶–Asp⁷⁰), whereas the contacts between the helices in ESAT-6 appear to be exclusively van der Waals in nature. The principal interactions found at the interface between CFP-10 and ESAT-6 are highlighted in Fig. 9, which clearly shows that residues at positions d and g within CFP-10 are able to interact with residues at positions g and c within ESAT-6. Residues located at positions g and c have the potential to form intermolecular salt bridges between CFP-10 and ESAT-6, but only two salt bridges are actually seen (Glu¹⁴–Lys³⁸ and Glu⁷¹–Lys⁷⁷). There is also the possibility of a salt bridge between Glu⁶⁸ and Lys⁵⁷, but the Glu⁷¹–Lys⁷⁷ interaction appears more favorable. Meher et al. (35) attempted to introduce additional salt bridges to the CFP-10/ESAT-6 complex by substituting Leu²⁵ and Phe²⁸ (both at position d) in CFP-10 with an arginine residue, and Leu²⁹ and Leu⁶⁵ (both at position a) in ESAT-6 with an aspartic acid residue. However, from examination of the structure of the complex and its representation in the helical wheel diagrams, it is clear that these
substitutions were highly unlikely to result in the formation of intermolecular salt bridges, as the residues at position a are predominantly involved in intramolecular interactions (Fig. 8).

A schematic representation of the predicted helical interfaces in the Rv0287-Rv0288 complex can be obtained using the reported structure of the CFP-10/ESAT-6 complex and the optimized sequence alignment for the CFP-10/ESAT-6 family shown in Fig. 5. The Rv0287-Rv0288 helical wheel diagram is shown in Fig. 10 and can be used to predict possible intra- and intermolecular interactions. The hydrophobic residues at positions a and d that lie at the center of both the intra- and intermolecular interfaces in the CFP-10/ESAT-6 complex are conserved in Rv0287 and Rv0288. However, the residues involved in stabilizing the intermolecular interface (positions g and c) show no significant conservation (Figs. 5 and 8–10). Consequently, the residues involved in the formation of salt bridges between CFP-10 and ESAT-6 are not conserved within Rv0287 and Rv0288. The helical wheel predictions for Rv0287 and Rv0288 suggest that intermolecular salt bridges could be formed between Lys21 and Glu31 and between Lys54 and Asp64 (Fig. 10), but do not predict the presence of substantially more stabilizing bridges in the Rv0287-Rv0288 complex. This suggests that the enhanced stability of Rv0287-Rv0288 arises from a slightly larger or more complimentary van der Waals interface between the proteins than is present in CFP-10/ESAT-6. The key interface residues at positions a and d are not only conserved in CFP-10/ESAT-6 and Rv0287/Rv0288, but across the whole protein family (Fig. 5), which suggests that all pairs of CFP-10/ESAT-6 proteins will form complexes with a structure close to that determined for CFP-10/ESAT-6.

As discussed above, the residues at position d within the helices play a key role in complex formation, as they lie at the heart of van der Waals contacts stabilizing both the intra- and intermolecular interfaces. This essential structural role was confirmed by recent mutational studies, which revealed that L25R and F58R mutations in CFP-10, and L65D mutations in ESAT-6 prevented complex formation (32–34). The residues at positions d and g appear to determine the specificity of complex formation, which will limit the formation of non-genome partner complexes to closely related groups of proteins, such as Rv0287/Rv0288 and Rv3019c/Rv3020c and the MTINY/QILSS families. This is supported by previous yeast two-hybrid studies, which showed that Rv0287/Rv0288 and Rv3019c/Rv3020c were capable of forming tight complexes with closely related non-genome partners, but not with distantly related family members such as CFP-10 and ESAT-6 (14).

CONCLUSIONS

The work reported here clearly shows that Rv0287 and Rv0288 form a tight, 1:1 complex, which is significantly more stable than the related CFP-10/ESAT-6 complex. For both Rv0287/Rv0288 and CFP-10/ESAT-6 complex formation is coupled to folding of the constituent proteins, which as isolated proteins exist as random coil polypeptides (CFP-10 like) and molten globules (ESAT-6 related). This very strongly suggests that the functional form for all genome pairs of CFP-10/ESAT-6 family proteins will be a 1:1 complex. The Rv0287-Rv0288 and CFP-10/ESAT-6 complexes show no evidence of reduced stability at low pH (4.5), which is consistent with a principal role for van der Waals interactions in stabilizing the interfaces between helices and argues against a functional role for the individual proteins in the acidified phagosome. The residues located at positions c, d, and g within the helical regions form the intermolecular interface between the proteins, and govern the specificity of complex formation. The very limited conservation of the residues at positions c and g indicates that complex formation will be limited to only closely related groups of proteins, which significantly reduces the number of possible functional complexes. The structural similarities between Rv0287-Rv0288 and CFP-10/ESAT-6, together with the conservation of key residues across the CFP-10/ESAT-6 protein family, strongly suggests that all genome pairs of family members will form very similar structures, containing a pair of helix-turn-helix hairpins that lie anti-parallel to each other with an extensive hydrophobic contact surface.

REFERENCES

1. Stewart, G. R., Robertson, B. D., and Young, D. B. (2003) Nat. Rev. Microbiol. 1, 97–105
2. World Health Organization (2007) WHO Report 2007, WHO/HTM/TB/2007:376
3. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eglmeier, K., Gas, S., Barry, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Conner, R., Davies, R., Devlin, K., Feltwell, T., Gentleman, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Selton, J., Squares, R., Squares, S., Stützon, J. E., Taylor, K., Whitehead, S., and Barrett, B. G. (1998) Nature 393, 537–544
4. Cole, S. T., Eglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutherford, K. M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J. R., and Barrett, B. G. (2001) Nature 409, 1007–1011
5. Garnier, T., Eglmeier, K., Camus, J. C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P. R., Parkhill, J., Barrett, B. G., Cole, S. T., Gordon, S. V., and Hewinson, R. G. (2003) Proc. Natl. Acad. Sci. USA 100, 7877–7882
6. Gey van Pittius, N., Gamieldien, J., Hide, W., Brown, G., Siezen, R., and Beyers, A. (2001) Genome Biol. 2, 0044.0041–0044.0018
7. Renshaw, P. S., Panagiotidou, P., Whelan, A., Gordon, S. V., Hewinson, R. G., Williamson, R. A., and Carr, M. D. (2002) J. Biol. Chem. 277, 21598–21603
8. Pullen, M. J. (2002) Trends Microbiol. 10, 209–212
9. Sorensen, A. L., Nagai, S., Houen, G., Andersen, P., and Andersen, A. B. (1995) Infect. Immun. 63, 1710–1717
10. Alderson, M. R., Bement, T., Day, C. H., Zhu, L. Q., Molesh, D., Skeikey, Y. A., Coler, R., Lewisohn, D. M., Reed, S. G., and Dillon, D. C. (2000) J. Exp. Med. 191, 551–559
11. Dillon, D. C., Alderson, M. R., Day, C. H., Bement, T., Campos-Neto, A., Skeikey, Y. A., Vedvick, T., Badaro, R., Reed, S. G., and Houghton, R. (2000) J. Clin. Microbiol. 38, 3285–3290
12. Skjot, R. L., Brock, I., Arend, S. M., Munk, M. E., Theisen, M., Ottenhoff, T. H., and Andersen, P. (2002) Infect. Immun. 70, 5446–5453
13. Berthet, F. X., Rasmussen, B. P., Rosenkrands, I., Andersen, P., and Gicquel, B. (1998) Microbiology 144, 3195–3203
