Comparative structural analysis of Rv2557 and Rv2558, two hypothetical proteins from *Mycobacterium tuberculosis* found in the human granuloma during persistence
3.1 Introduction

Rv2557 and Rv2558 are up regulated in *in vitro* tuberculosis persistence models designed to mimic the latent stage (Betts *et al.*, 2002) but are not necessary for long-term survival (Bhavna *et al.*, 2006). Loss of the genes also did not affect *M. tuberculosis* virulence in a SCID mouse model (Bhavna *et al.*, 2006), although it has been found to be expressed in human macrophages (Fenhalls *et al.*, 2002). *In situ* hybridization studies demonstrated that the genes were expressed in human necrotic and non-necrotic granulomas (Fenhalls *et al.*, 2002) (Fig. 3.1). Taken together these studies indicate that these genes might play an important role in the physiological adaptation of *M. tuberculosis* for persistence.

In this present work, cloning, over expression, purification and structural analysis of the recombinant Rv2557 and Rv2558 proteins have been reported. Subsequently structural differences in the two proteins have been probed through thermodynamics calculations. Characterization of the proteins has also been done by thermal denaturation and chemical denaturation.
Non-necrotic Granuloma

Macrophage, positive for all the mRNAs

Giant cell, positive for all the mRNAs

Lymphocyte negative for mycobacterial mRNAs

Macrophage in the transition zone of a necrotic granuloma
positive for iniB, kasA and Rv2557/8 mRNAs

Macrophage negative for the mRNAs

Giant cell negative for the mRNAs

Necrotic Granuloma

Figure 3.1: Schematic diagram of mycobacterial gene expression within a non-necrotic tuberculous granuloma. In a nonnecrotic granuloma consisting predominantly of macrophages, lymphocytes, and giant cells, mycobacteria expressing rpoB, narX, ica, Rv2557 and/or Rv2558, iniB, and kasA mRNAs were detected by using nonradioactive in situ hybridization. The transition zone of the necrotic granuloma still contains a small number of macrophages positive for Rv2557 and/or Rv2558, iniB, and kasA mRNA. (Fenhalls et al., 2002).
Chapter 3

3.2 Materials and Methods

3.2.1. Strains, vectors and chemicals

The *E. coli* strain DH-5α was used for construction of clones. The *E. coli* strain C-41 (DE3) which contains T7 RNA polymerase gene and *E. coli* strain XL-1 blue which contains T5 RNA polymerase gene were used for protein expression. The pET-23a plasmid and *E. coli* C-41 (DE3) were obtained from Novagen (Madison, WI, USA). *E. coli* strain XL-1 blue was obtained from Stratagene. The pQE30 and pQE32 were purchased from Qiagen. Restriction endonucleases, isopropyl thio-β-D-galactopyranoside (IPTG), and T4 DNA ligase were obtained from Invitrogen. All chemicals used in the study were purchased from Sigma-Aldrich Chemicals Co. and were of highest purity available. All chromatographic columns were purchased from GE Healthcare Biosciences.

3.2.2 Phylogenetic tree, sequence analysis and secondary structure analysis

Protein sequences of Rv2557 and Rv2558 were retrieved from Tuberculist (http://genolist.pasteur.fr/Tuberculist). ClustalX (Thompson et al., 1997) was used to perform the multiple alignments with proteins of related *Mycobacterium* sp. and to construct a phylogenetic tree. The output was obtained after bootstrapping by neighbor-joining algorithm. The methods used for general secondary-structure prediction were PHD (Rost et al., 1994) and GORIV (Sen et al., 2005).

3.2.3 Cloning of Rv2557 and Rv2558

The two genes have about 89% sequence identity at the DNA level and this consequently makes the amplification by PCR techniques difficult. The genes were therefore amplified by nested-PCR techniques and this allowed for prevention of cross-annealing of gene specific primers. In the first step the genomic regions containing the genes Rv2557 and Rv2558 were PCR amplified from *M. tuberculosis* genomic DNA using the nested primers: (Table 3.1). These primers were designed using the flanking DNA sequences at 5’ and 3’ end of both genes available in the database (http://genolist.pasteur.fr/Tuberculist). The nested PCR products (~735
bases pair for Rv2557 and ~770 base pair for Rv2558) were used for further PCR amplification of gene specific PCR of Rv2557 (675 base pair) and Rv2558 (711 base pair) using gene specific primers: (Table 3.1) containing BamHI and HindIII restriction sites. The amplified PCR product of Rv2557 was digested with BamHI and HindIII and cloned into T5-based expression vector, pQE32 at the same sites. This attaches a tag comprising of the residues MRGSHHHHHH to the N-terminal end. The PCR product of Rv2558 was digested with BamHI and HindIII and cloned into T5-based expression vector pQE30 at the same sites. The clones were screened by PCR amplification and restriction digestion. In pQE30, the expression level of Rv2558 is very low, so for increasing the level of expression of Rv2558, the digested PCR product of Rv2558 was cloned into a T7-based expression vector pET23a at the same sites. In pET23a a tag comprising of the residues ACGRTRAHHHHHH was attached at the C-terminal end.

3.2.4 Over-expression and purification

The clone of Rv2557 was transformed into E. coli host cells XL-1blue compatible for the T5-based expression plasmids. Single colonies were grown at 37 °C in 10 ml of LB media containing 100 μg/ml ampicillin to an A₆₀₀ of 0.4, 0.6 and greater than 1.0 induced with 0.3mM to 1mM isopropyl-1-thio-β-D-galactopyranoside, and grown for further 4-8 hrs. The level of induction was monitored using 12% SDS-PAGE. Over-expression was achieved in XL1-Blue by inducing with 0.3mM of IPTG at an A₆₀₀ of 0.4. One-liter medium of LB containing 100μg/ml ampicillin were inoculated with seed culture of XL-1blue containing Rv2557 and grown at 37°C until the A₆₀₀ reached 0.4. The cultures were induced by the addition of 0.3 mM IPTG and grown for 4hrs. The IPTG induced cells were harvested, resuspended in buffer A (50 mM Tris, pH 7.2, 150 mM NaCl, 10 mM Imidazole), and sonicated. The cell lysate was centrifuged at 14000 RPM for 20min at 4°C to remove the cell debris The supernatant was loaded on a Ni²⁺-NTA column pre-equilibrated with Buffer A (50 mM Tris, pH 7.2, 150 mM NaCl, 10 mM Imidazole). The column was washed with 5 column volumes of buffer A. The protein was eluted using a linear gradient of buffer B containing 500 mM Imidazole. The protein was eluted at the gradient from 50-65% Imidazole.
Table 3.1: Primers used in cloning of Rv2557 and Rv2558

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’→3’)</th>
<th>Application and coordinate of amplicons</th>
</tr>
</thead>
<tbody>
<tr>
<td>N57-F</td>
<td>5’GCCCGTGGAGGAAGGTAA TGATTGACC 3’</td>
<td>Nested forward primer of Rv2557. (28 bp upstream from 5’ end of Rv2557 (2877.072))</td>
</tr>
<tr>
<td>N57-R</td>
<td>5’CAGACTAGTTGGGCGGCCCGTCATGGCC 3’</td>
<td>Nested reverse primer of Rv2557 (30 bp downstream from 5’ end of Rv2557 (2677.746))</td>
</tr>
<tr>
<td>N58-F</td>
<td>5’ CGAGCGGCCTCGCCGTAGGCATTGCCA 3’</td>
<td>Nested forward primer of Rv2558 (28 bp upstream from 5’ end of Rv2557 (2877.831))</td>
</tr>
<tr>
<td>N58-R</td>
<td>5’GCCTCTCGGGGAAGCCAGCTCCGTCAGAC 3’</td>
<td>Nested reverse primer of Rv2558 (30 bp downstream from 5’ end of Rv2557 (2678.541))</td>
</tr>
<tr>
<td>57F</td>
<td>5’ GCAGGATCCATGACCGGTGGCGCCACCGGGGC 3’</td>
<td>Gene specific forward (sense) primer of Rv2557 (BamHI site)</td>
</tr>
<tr>
<td>57R</td>
<td>5’GCCGAAGCTTGACCAGCTCGGGTACCCGTTGCG3’</td>
<td>Gene specific reverse (antisense) primer of Rv2557 (HindIII site)</td>
</tr>
<tr>
<td>58F</td>
<td>5’GCAAGGATCTTTGCCAGGCTAGCGGTTGGAG 3’</td>
<td>Gene specific forward (sense) primer of Rv2558 (BamHI site)</td>
</tr>
<tr>
<td>58R</td>
<td>5’ACGAAGCTTCAGACCGCTCGGTACCCG 3’</td>
<td>Gene specific reverse (antisense) primer of Rv2558 (HindIII site)</td>
</tr>
</tbody>
</table>
The clone of Rv2558 was transformed into E. coli strain C-41(DE3) compatible for the T-7-based expression plasmids. Single colonies were grown at 37°C in 10 ml of LB media containing 100µg/ml ampicillin to an A₆₀₀ of 0.6 and induced 1 mM isopropyl-1-thio-β-D-galactopyranoside, and grown for further 8 hrs. The level of induction was monitored using 12% SDS-PAGE. One-liter medium of LB containing 100µg/ml ampicillin were inoculated with a 5 ml culture of C-41(DE3) containing Rv2558 and grown at 37°C until the A₆₀₀ reached 0.6. The cultures were induced by the addition of 1.0 mM IPTG and grown for 8 hrs. The IPTG induced cells were harvested, resuspended in buffer A (50 mM Tris, pH 8.0, 50 mM NaCl, 10 mM Imidazole), and sonicated. The cell lysate was centrifuged at 14000 RPM for 20min at 4°C to remove the cell debris. The supernatant was loaded on a Ni²⁺-NTA column pre-equilibrated with Buffer A (50 mM Tris pH 8.0, 50 mM NaCl, 10 mM Imidazole). The column was washed with 5 column volumes of buffer A. The protein was eluted using a linear gradient of buffer B containing 800 mM Imidazole. The protein was eluted at the gradient from 50-65% Imidazole (400-520 mM).

3.2.5 Circular Dichroism Measurements

The CD measurements were carried out using a Jasco J810 spectropolarimeter calibrated with ammonium (+)-10-ncamphorsulfonate and the results were expressed as relative ellipticity and plotted as percentage values. The CD spectra were obtained at protein concentrations of 2 µM for far-UV CD, respectively, with a 2-mm cell at 25°C. The values obtained were normalized by subtracting the baseline recorded for the buffer having the same concentration of denaturant under similar conditions. Each spectrum was an average of three scans to minimize errors. Per residue molar ellipticity was calculated from the observed ellipticities according to the equation:

$$[	heta] = \theta_0 M_r / I c$$

where $[\theta]$ is per residue molar ellipticity in deg cm² dmol⁻¹; $\theta_0$, observed ellipticity in mdeg; $M_r$, average molecular weight of an amino acid in the protein; $I$, path-length in centimeters; $c$, protein concentration in mol/l. Percentage of secondary structure was calculated using the web-based program K2D (http://www.embl-heidelberg.de/~andrade/k2d/).
3.2.6 Fluorescence spectroscopy

Fluorescence spectra were recorded with PerkinElmer LS5B spectroluminescence meter in a quartz cell with a 5-mm path length. All samples were incubated for 4 hours under specified conditions before recording the spectra. For tryptophan fluorescence measurements the excitation wavelength was 285 nm and the spectra were recorded from 300 nm to 400 nm. The protein concentration used was 0.5 μM and the measurement was carried out at 25°C.

3.2.7 Thermal denaturation

Thermal denaturation of the proteins was monitored by following the change in molar ellipticity at 222 nm as a function of the temperature on a Jasco J810 spectropolarimeter equipped with a peltier temperature controller system. The measurements were carried out in 25 mM Tris buffer, pH 7.5. Samples were heated at a constant rate of 1°C/min in a 2 mm cell. The folded fraction of protein at any temperature was determined as follows:

\[
\frac{([\theta]_{\text{obs}} - [\theta]_{\text{den}})}{([\theta]_{\text{n}} - [\theta]_{\text{den}})}
\]

where \([\theta]_{\text{obs}}\) denotes the ellipticities at any temperature, \([\theta]_{\text{den}}\) at highest temperature, \([\theta]_{\text{n}}\) at lowest temperature, respectively.

3.2.8 Urea and guanidinium chloride denaturation

Recombinant Rv2557 and Rv2558 (2 μM) were dissolved in 25 mM Tris buffer, pH 7.6, containing, 50 mM sodium sulfate in the presence of increasing concentrations of urea or GdmCl and incubated for 4 hr at 25 °C. These samples were then used for taking various measurements in CD and fluorescence. Data was fitted according to equation

\[
\frac{([\theta]_{\text{obs}} - [\theta]_{\text{den}})}{([\theta]_{\text{n}} - [\theta]_{\text{den}})}
\]

where \([\theta]_{\text{obs}}\) denotes the ellipticities at specific concentration, \([\theta]_{\text{den}}\) at highest denaturant concentration, \([\theta]_{\text{n}}\) at lowest denaturant, respectively.
3.2.9 Analysis of the denaturation curves

According to the established theory of linear free energy changes (Bolen and Santoro, 1988), each unfolding step has the following relationship to the standard free energy changes:

\[ \Delta G_1^0 = \Delta G_i + m_i [D] \tag{4} \]

In which \( \Delta G_i \) is the free energy change at each step and \( m_i \) represents the dependence of the \( \Delta G_i^0 \) on the denaturant concentration \([D]\). The equilibrium constant \( K_i \) then can be given as

\[ K_i = \exp\{ -\Delta G_1^0 \cdot m_i [D] / R T \} \tag{5} \]

Where \( R \) is the gas constant and \( T \) is the absolute temperature in K. The overall equation that describes the three-state chemical denaturant induced unfolding is given as

\[
Y_{obs} = Y_N + Y_1 \exp\left( -\frac{(\Delta G_1^0 - m_1 [D])}{R T} \right) + Y_D \exp \left( -\frac{(-((\Delta G_1^0 - m_1 [D]) + (\Delta G_2^0 - m_2 [D]))}{R T} \right)
\]

Where \( Y_{obs} \) is the observed spectral parameter and \( Y_N, Y_1 \) and \( Y_D \) are the signals of the native, intermediate and denatured states, respectively, given by linear regression of spectral points at those states

For two state unfolding process Equation. (5) gets simplified to

\[
Y_{obs} = \frac{Y_N + Y_D \exp\left( -\frac{-(\Delta G_1^0 - m [D])}{R T} \right)}{1 + \exp\left( -\frac{-(\Delta G_1^0 - m [D])}{R T} \right)} \tag{6}
\]

3.2.10 Analytical gel filtration

Gel-filtration experiments were carried out on Superdex 75 HR (10/30 column) on AKTA FPLC (GE healthcare) calibrated by various low molecular mass standards from GE healthcare. The marker proteins were Albumin (Rs 35 Å, Mw 67 kDa), Ovalbumin (Rs 30 Å, Mw 43 kDa), Chymotrypsinogen (Rs 20 Å, Mw 25 kDa) and Ribonuclease A (Rs 16 Å, Mw 13.7 kDa). The Column was equilibrated and run in desired buffer at 25°C. Rv2557 and Rv2558 (50 µM) were incubated in different
denaturant buffers at 25°C for 2hrs. The sample (200 μL) was loaded on the column and run at 25°C, at a flow rate of 0.3 ml/min, and detection at 280 nm.

The relative elution volume was calculated as:

\[ K_{av} = \frac{V_e - V_0}{V_t - V_0} \]  

(8)

Where \( V_e \) is the elution volume of the center of a protein peak, \( V_0 \) is the void volume (determined by elution of blue dextran), and \( V_t \) is the total elution volume. For deconvolution of gel filtration peaks, Peak fit (Systat Software, Inc) software was used. Laurent and Killander solution (T.C. Laurent and J. Killander, 1964) for the calculation of the Stokes radius was used where \( \sqrt{(-\log K_{av})} \) is plotted against the Stokes radius.

3.2.11 pH-dependent structural change

Rv2557 and Rv2558 dissolved in (0.1 M) buffer of varying pH, 3 to 11 citrate (pH 3–5), sodium phosphate (pH 6–8), and glycine-KOH (pH 9–11) was incubated for 4 hrs at 25 °C before the measurements were made through CD and Fluorescence.

3.2.12 Crystallization

3.2.12.1 Preparation

Control of excessive nucleation by filtration and general cleanliness has been shown to be conducive to the formation of better crystals (Cudney et al., 1994) & (Blow et al., 1994). All buffers, crystallization reagents, solutions were therefore filtered. Protein solutions were centrifuged at 10,000g for 5 min to settle extraneous debris prior to setting up drops. Microscopic cover slips used in vapor diffusion method needed special cleaning treatments. These cover slips were thoroughly washed with cleaning solution, rinsed with Milli Q, dried and then coated with dilute siliconizing solution.

3.2.12.2 Crystallization trials

Preliminary crystallization trials of his-tagged protein were carried out using hanging drop vapor diffusion using a “Sparse matrix” formulation of 50 solutions (Jancarik and Kim, 1991a;), “Crystal screen 2” (Jancarik and Kim, 1991a) & (Cudney et al., 1994;) and the recently developed “Mixed precipitant
screens" (Majeed et al., 2003). The protocol utilized in sparse matrix evaluates 50 unique combinations involving the pH, salts, precipitants, and other additives. Initial crystallization trials of native enzyme were set up by mixing 2 µl of 10-30 mg/ml protein in 50 mM Tris-HCl, pH 7.2, 50 mM NaCl, 5 mM EDTA and 2 mM β-mercaptoethanol with 1 µl volume of reservoir solution using hanging drop vapor diffusion method at different temperatures ranging from 4°C to 37°C. The initial crystallization trials and subsequent variations in different physical parameters led to the identification of some lead conditions. These were further optimized by modulating different parameters to improve the size and suitability for diffraction experiments.

### 3.2.12.2.1 Amount of protein

Crystallizations were set-up at varying protein concentrations from 2mg ml-1 to 8 mg ml-1 by the hanging drop vapor diffusion method. in addition varying drop volumes were used to explore suitable crystallization conditions.

### 3.2.12.2 Temperature and pH

Crystallizations were set-up with 6 mg/ml -10mg/ml protein in duplicates at 4°C and at room temperature (25°C) to see effect of temperature. In addition finer scans of pH of buffers (ΔpH =0.5 ) was used to further optimization.

### 3.2.12.3 Seeding Technique

Further attempts at improvement of crystals were carried out by use of Microseeding to hanging drop vapor diffusion conditions. Microseeding experiments, few micro crystals were crushed with the help of capillary jets. The crushed crystals were then transferred to an eppendorf and mixed in 10 µl well solution. Centrifugation at 6000 x g for 5 min to settle the large nuclei and ant precipitate or debris. The microcrystals were then picked with the help of a fine hair and transferred to nuclei to low supersaturating solutions to perform streak seeding or Microseeding respectively.
3.3 Results and discussion

3.3.1 In silico analysis of Rv2557 and Rv2558

The two proteins have homologs only within mycobacteria (Figure 3.2) and have been classified as conserved hypothetical proteins in the databases. A striking feature of Rv2557 and Rv2558 is the amino acid composition. The amino acid composition of the both proteins is unusual with respect to that of the whole proteome of *M. tuberculosis* H37Rv (Fig. 3.2). The proportion of charged amino acids (D, E, K, R) and hydrophobic amino acids is higher than the average of whole proteome.

3.3.2 Nested PCR and cloning of genes

The two genes have 89% similarity between them at the DNA level (Fig. 3.3 A) and this consequently makes the amplification by PCR techniques difficult. The genes were therefore amplified by nested-PCR techniques as detailed in the Methods section and this also allowed for prevention of cross-annealing (3.3 B). The PCR products were eluted using PCR gel extraction kit and cloned into T7 and T5 RNA polymerase-based bacterial expression vectors, *pET23a* and *pQE30* between BamH1 and HindIII respectively (Fig. 3.4). Clones were confirmed through restriction digestion and PCR. The integrity of the clones was verified by sequencing.

3.3.3 Expression and purification of recombinant His$_6$-tagged-proteins

The recombinant proteins were purified based on its N-terminal His$_6$-tag (Rv2557) and C-terminal His$_6$-tag (Rv2558) by affinity chromatography using a Ni-NTA Hi-trap chelating Sepharose column, as described in materials and methods. The purification conditions were standardized by optimizing pH, salt and imidazole concentration. Attempts to purify the protein (Rv2557) at low salt concentration (50 mM NaCl) gave non-specific contaminant proteins. Increasing salt concentration up to 200 mM NaCl greatly removed all the non-specific contaminant proteins. The fractions eluted were pooled and imidazole was later removed by dialyzing overnight in 2 L of buffer 20 mM Tris (7.5), 50 mM sodium sulphate for further experiments.
Chapter 3

Protein quantification was performed as described by Bradford (Bradford, 1976) using BSA as standard and purity of proteins were analyzed by 12% SDS-PAGE (Fig. 3.5A & B). The yield was in the range of 10-12 mg/liter and 5-6 mg/ml for Rv2557 and Rv2558 respectively.
Figure 3.2:

(A) Multiple sequence alignment of amino acid sequences of Rv2557 and Rv2558 of *M. tuberculosis*, and respective homologs viz. Mb2580, Mb2588 of *M. bovis*; Mul1752 of *M. ulcerans*, Mav3433 of *M. avium*, Mvan4767 of *M. vanbaalenii* and Mflv4652 of *M. gilvum*. The alignment was carried out using ClustalX.

(B) Phylogenetic analysis of Rv2557 and Rv2558. The unrooted tree was constructed using the n-j algorithm. The analysis was based on distance relation parameters. The bootstrap values (%) are shown at the nodes.
A) Gene sequence alignment of Rv2557 and Rv2558

B) Nested-PCR procedure used for amplification of the genes
Figure 3.4: Cloning of Rv2557 and Rv2558

(A) PCR of Rv2557. Lane 1, 1kb DNA ladder. Lane 2, nested PCR fragment of Rv2557. Lane 3, specific PCR fragment.

(B) Cloning of Rv2557 in pQE32. Lane 1, 1kb DNA ladder. Lane 2, PCR fragment amplified by clone. Lane 3, restriction digestion of full length in pQE32.

(C) PCR of Rv2558: Lane 1, 1kb DNA ladder. Lane 2, nested PCR fragment of Rv2558. Lane 3, specific PCR fragment.

(D) Cloning of Rv2558 in pET23a: Lane 1, 1kb DNA ladder. Lane 2, PCR fragment amplified by clone. Lane 3, restriction digestion of full length in pET23a.
3.3.4 Molecular weight and subunit structure of Rv2557 and Rv2558

The molecular masses of the purified recombinant Rv2557 and Rv2558 were determined under non-dissociating conditions (Andrews, 1965) using the data of gel filtration experiments. A Superdex S-75 column was used for the size exclusion chromatography experiments. The column was calibrated using molecular weight standards prior to the experiments. An examination of the elution profile revealed a single peak for both proteins, albeit with a slight difference in the retention volumes viz. 12.10 ml and 12.25 ml, for the two proteins (Fig. 3.6). When the relative elution volume was plotted as a function of log of molecular masses, the molecular masses of 30 kDa and 28 kDa were obtained for Rv2557 and Rv2558, respectively. From the primary amino acid sequence, the molecular masses of 24.2 kDa and 25.7 kDa were obtained for Rv2557 and Rv2558 respectively. The results of the studies on subunit masses along with the size exclusion chromatography as reported above demonstrate that both Rv2557 and Rv2558 exist as monomer under physiological conditions. Furthermore, as both the proteins have almost similar molecular masses but significant differences in their retention volumes, it suggests that the proteins differ significantly in their molecular dimensions. Rv2558 has a slightly more compact conformation (Stokes radius 22.1 nm) compared to Rv2557 (Stokes radius 22.6 nm) under the experimental conditions.

3.3.5 Rv2557 and Rv2558 adopt a predominantly helical structure

It is known that α-helical and β-sheet proteins exhibit characteristic far UV CD spectra respectively. α-helical proteins exhibit two minima at 222 nm and 208 nm while β-sheet containing proteins have a single minimum at 216 nm (Chen et al., 1972). Far-UV CD studies were carried out on both Rv2557 and Rv2558 to analyze the differences in the secondary structure that exist between the two proteins. For both Rv2557 and Rv2558, the far-UV CD spectra characteristic of a protein having helical structure were observed (Fig. 3.7A, Table 3.2) However, for similar molar concentrations of protein, a significantly higher ellipticity was observed for Rv2557 over the whole far-UV region (Fig. 3.7A). This observation suggests that Rv2557 has a significantly higher secondary structure as compared with Rv2558. It is interesting that Rv2557 was found to have a more compact conformation despite
having almost similar subunit molecular mass as Rv2558. One possible reason for this may be the presence of a significantly higher secondary structure or stabilization of secondary structure in Rv2557, which would result in different possible conformation. Deconvolution of the CD spectrum predicts both proteins have more α-helical comparative to β-sheet; this is in good agreement with different secondary structure prediction algorithms (Table 3.2).

3.3.6 Tryptophan Fluorescence of Rv2557 and Rv2558

According to the primary amino acid sequence, both proteins have 4 tryptophan residues, Rv2557 has tryptophan molecules at positions 17, 74, 106 and 127 and, whereas Rv2558 has tryptophan at position 7, 85, 118, 139. The fluorescence spectra of Rv2557 and Rv2558 are shown in Figure 4B. For Rv2557 the emission wavelength maxima for the tryptophan fluorescence were observed at about 340 nm, and for Rv2558 at 335 nm (Fig. 3.7B). The buried tryptophan residues in folded protein show fluorescence emission maxima at 330–335 nm, whereas on unfolding of protein the tryptophan fluorescence emission maxima shifts to about 355 nm. Hence, in Rv2557 the tryptophan molecule (s) is not completely buried but partially exposed to the solvent but in Rv2558 tryptophan molecule (s) is almost buried. The binding of metal ions like Ca$^{+2}$, Zn$^{+2}$ and Mg$^{+2}$ to the proteins was probed through fluorescence spectroscopy but the experiments rule out the presence of metal ion binding sites in the protein. Overall, the experiments suggest that both proteins are in the proper folded conformation.
Figure 3.5: SDS-PAGE analysis of *E. coli* lysate and the respective purified proteins. Panel (A) corresponds to Rv2557 while Panel (B) corresponds to Rv2558 purification steps respectively. *Lanes 1-6* in both panels represent molecular weight markers, supernatant of induced culture lysate, flow-through, 40 mM imidazole wash and 80 mM imidazole wash protein eluted in 300 mM imidazole from the nickel-nitrilotriacetic column respectively.

Figure 3.6: Size-exclusion chromatographic profiles for Rv2557 (*solid line*) and Rv2558 (*dashed line*) on the Superdex-75HR column at pH 7.5 and 25°C. The two insets represent the calibration curves of the column with respect to different molecular weight markers and stokes radius respectively.
Figure 3.7:

(A) Far-UV CD spectra of 1 μM of Rv2557 (▲) and 1 μM of Rv2558 (■). The samples were in 25 mM Tris, pH 7.5, 50 mM Na₂SO₄. The data were corrected with respect to the contributions of buffer as per routine practice.

(B) Tryptophan fluorescence emission spectra of 1 μM of Rv2557 (▲) and 1 μM of Rv2558 (■). The samples were in 25 mM Tris, pH 7.5 and 50 mM NaCl.
Table 3.2: **Secondary structure content of Rv2257 and Rv2558**

<table>
<thead>
<tr>
<th></th>
<th>Rv2557</th>
<th></th>
<th>Rv2558</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α-helix</td>
<td>β-sheet</td>
<td>Random coil</td>
<td>α-helix</td>
</tr>
<tr>
<td>Calculated from CD data</td>
<td>59</td>
<td>8</td>
<td>33</td>
<td>43</td>
</tr>
<tr>
<td>PHD</td>
<td>52</td>
<td>17</td>
<td>29</td>
<td>49</td>
</tr>
<tr>
<td>GOR IV</td>
<td>31</td>
<td>17</td>
<td>50</td>
<td>35</td>
</tr>
</tbody>
</table>
3.3.7 Thermal Denaturation of Rv2557 and Rv2558

The thermal unfolding of Rv2557 and Rv2558 were characterized by monitoring the loss of secondary structure of protein with temperature. Figure 3.8 summarizes the changes in CD ellipticity at 222 nm for Rv2557 and Rv2558 at increasing temperatures. For both proteins, a single sigmoidal transition was observed suggesting a cooperative thermal unfolding of native protein form native to unfolded state. Furthermore, for both proteins, transitions were observed in the temperature region 50–90 °C; however, it was centered at about 75 °C and 85 °C for Rv2558 and Rv2557 respectively. As a difference of about 10 °C in the \(T_m\) was observed between the two proteins, it demonstrates that Rv2557 has higher thermal stability compared to Rv2558. The biochemistry of many proteins from thermophilic sources is available and agrees well with our conclusions that Rv2557 and Rv2558 exhibit structural stability even at high temperatures. For example \(T.\ thermophilus\) RnaseH has a melting temperature of about 86 °C while archaeal histones exhibit melting temperatures around 75 °C. This agrees well with the melting temperatures of 85 and 75 °C respectively for Rv2557 and Rv2558 proteins (Table 3.3). This co-relates well with the higher structural compactness of the former protein. Subsequently, these results were further substantiated through comparative denaturation by guanidine hydrochloride.

Table 3.3: Melting temperature (Tm) of proteins from different sources

<table>
<thead>
<tr>
<th>Protein</th>
<th>Species</th>
<th>Tm(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcRnaseH</td>
<td>Mesophilic</td>
<td>66 ± 1</td>
</tr>
<tr>
<td>TtRnaseH</td>
<td>Thermophilic</td>
<td>86 ± 1</td>
</tr>
<tr>
<td>Rnase Sa</td>
<td>Mesophilic</td>
<td>48.4 ± 0.3</td>
</tr>
<tr>
<td>Barnase</td>
<td>Mesophilic</td>
<td>53.2</td>
</tr>
<tr>
<td>rHFoB (Archaeal Histones)</td>
<td>Thermophilic</td>
<td>74.8 ± 0.2</td>
</tr>
<tr>
<td>Rv2557</td>
<td>Mesophilic</td>
<td>85 ± 1</td>
</tr>
<tr>
<td>Rv2558</td>
<td>Mesophilic</td>
<td>75 ± 1</td>
</tr>
<tr>
<td>Rv3290</td>
<td>Mesophilic</td>
<td>50 ± 1</td>
</tr>
</tbody>
</table>
Figure 3.8:
Thermal denaturation of proteins measured by loss of CD ellipticity at 222 nm. The (▲) and (■) lines represent data for Rv2557 and Rv2558, respectively.
3.3.8 Guanidine hydrochloride induced denaturation.

The unfolding characteristics of Rv2557 and Rv2558 were studied by monitoring the GdmCl induced changes in the secondary structure and the tryptophan fluorescence of the proteins. Time dependent changes in the structural parameters at increasing GdmCl concentrations (0.5 M, 1.0 M and 3.0 M) were monitored to standardize the incubation times required to achieve equilibrium under these conditions. Under all the conditions studied, the changes occurred within a maximum of 6 hr at 16 °C with no further alterations in the value up to 24 hours. These observations suggest that a minimum time of about 6 hr is sufficient for achieving equilibrium under any of denaturing conditions studied. Furthermore, under the experimental conditions the unfolding/refolding curves were fully reversible and the protein remained monomeric throughout the experiment.

To study the GdmCl-induced changes in the secondary structure of the Rv2557 and Rv2558, far-UV CD studies were carried out. Figure 3.9 A summarizes the effects of increasing GdmCl concentration on the CD ellipticity at 222 nm. The unfolding in presence of GdmCl is a two-state process with no intermediate state existing between 2 M and 4 M concentration for Rv2557 and between 1 M and 4 M concentration for Rv2558. Both proteins get completely denatured at 4 M GdmCl. The transition can be described adequately by a two-state reversible model $N \leftrightarrow U$ where $N$ and $U$ represent the native and unfolded state, respectively. By following this two state model, the data were fit to the two-state equation -6. The free energy change, $\Delta G^0$ and the m-value for the single transition from native to denatured state was obtained for Rv2557 and Rv2558 (Table 3.4). Comparison of $\Delta G$ (H$_2$O) and $C_m$ values clearly indicates different stabilization modes of both proteins. Compared to Rv2558, Rv2557 shows a 1.10 kcal/mol increase in the stabilization energy and a 0.45 M increase in $C_m$ values. The m-values of Rv2557 (1.5) and Rv2558 (1.3) indicate that only a few number of denaturant binding sites are exposed on unfolding for both proteins. Figure 3.9 B summarizes the change in the tryptophan emission wavelength ($\lambda_{max}$) of Rv2557 and Rv2558 after incubation with increasing concentrations of GdmCl. For Rv2557, a sigmoidal shift in emission wavelength maxima from 340 nm to 353 nm was obtained at increasing concentration from 2 to 3.0 M. For Rv2558, a sigmoidal change from 337 nm to 354 nm was obtained at increasing concentration.
from 1 M to 3 M. In both proteins no further changes were observed after 3 M GdmCl. This agrees well with the earlier CD experiments where also no unfolding intermediate was observed.

For studying the effect of GdmCl on the molecular dimensions, size exclusion studies were carried out. Figure 3.9 D summarizes the elution profile of Rv2557 and Rv2558 in the absence of and in the presence of 2.0 M GdmCl. For Rv2558, a single peak with a retention volume of 12.2 ml corresponding to the compact species of protein was observed. On increase in GdmCl concentration from 0 to 2 M, significant changes in the retention volume (10.66 ml) were observed. This suggests that a substantial alteration (60-70%) in the folded conformation of Rv2558 occurs on enhancement of GdmCl concentration. In the case of Rv2557, a partially unfolded (15-20%) form was observed at 2 M GdmCl. These results support that the secondary and tertiary structures of Rv2557 is more stabilized compared to Rv2558. Overall the results are in agreement with the thermal denaturation studies.

In order to explore whether the stabilization occurs primarily due to ionic or hydrophobic interactions, urea denaturation experiments were carried out on both proteins.
Chapter 3

Figure 3.9:

(A) Changes in CD ellipticity at 222 nm for Rv2557 (▲) and Rv2558 (■) on incubation with increasing concentrations of GdmCl. Data are fitted according to equation (2). Values observed for proteins in the absence of GdmCl taken as 100%.

(B) Changes in tryptophan fluorescence emission wavelength maximum of Rv2557 (▲) and Rv2558 (■) on incubation with increasing concentrations of GdmCl.

(C) Plots of the free energy of unfolding in the presence of denaturant versus the denaturant concentrations of Rv2557 (▲) and Rv2558 (■).

(D) GdmCl-induced alterations in the molecular dimension of Rv2557 (lower panel) and Rv2558 (upper panel). Size-exclusion chromatographic profiles for proteins (black) and on incubation with 2.0M GdmCl (grey) on a Superdex 75 H column at pH 7.0 and 25 °C. The columns were run with the same concentration of GdmCl in which the protein sample was incubated. The samples were incubated for 6 hrs in GdmCl before column chromatography.
3.3.9 Urea-induced denaturation

The urea-induced denaturation of Rv2557 and Rv2558 were studied by monitoring the changes in the secondary and tertiary structures at increasing urea concentrations. Figure 3.10A summarizes the changes in the far-UV CD signal of Rv2557 and Rv2558 on treatment with increasing concentrations of urea. Urea interfered with the absorption below 213 nm. The signal at 222 nm was therefore used to study changes in the secondary structures of Rv2557 and Rv2558. The unfolding in presence of urea is also two-state process and no intermediate was observed as before.

By applying the two-state equation to the molar ellipticity data at 222 nm, the standard free energy change for both proteins were calculated (Table 3.4). The denaturing action of urea was presumed to be based mainly on its ability to bind to the protein (Pace, 1986). The urea moiety is not expected to have major effects on the intermolecular and intramolecular electrostatic interactions in the protein because of its uncharged nature. Thus, the free energy of unfolding from the urea denaturation studies indicates the net stability of the protein. The sum of the stabilizing effect of hydrophobic interactions and the stabilizing effect of electrostatic attractions is therefore a measure of the protein’s stability. It has been suggested earlier that the slope \( m \) of the plots is a measure of the ability of the denaturant to unfold a protein (Greene and Pace, 1974), and furthermore the \( m \) value is proportional to the number of denaturant binding sites exposed on unfolding. The ratio of the \( m \) values between GdmCl and urea \( \left\{ \frac{m \text{(GdmCl)}}{m \text{(urea)}} \right\} \) denotes the mode of packing. When the numbers of electrostatic attractions were gradually decreased and, as a result, gradually increasing the contribution of hydrophobic packing relative to the contributions of electrostatic attractions, the ratio of the slopes proportionately decreased. A comparison of the \( \left\{ \frac{m \text{(GdmCl)}}{m \text{(urea)}} \right\} \) value denotes that in Rv2557, the N-U (1.0) stage is stabilized by hydrophobic interactions, compared to Rv2558 N-U (1.66), in which both electrostatic and hydrophobic interactions play a role in stabilization. These results can be rationalized in terms of the resultant effects of denaturation and electrostatic interactions. When electrostatic attractions predominate over repulsions, such as in Rv2558, the denaturing effect of urea may be partially opposed by stabilizing effects of electrostatic attractions, resulting in a shallower slope. Conversely, when the electrostatic repulsive forces dominate over the
attractive forces, as in Rv2557, the denaturing effect of urea acts in concert with the destabilizing effect of electrostatic repulsions, resulting in a steeper slope (Table 3.4). The urea-induced alterations in tertiary structure of recombinant proteins Rv2557 and Rv2558 were studied by monitoring the changes in the tryptophan fluorescence emission wavelength maxima at increasing concentrations of urea and these are summarized in Figure 3.10 B. No significant changes (339 nm to 341 nm) in tryptophan fluorescence emission wavelength maxima were observed with increasing urea concentrations up to 6 M urea for Rv2557 while an exponential change (341 nm to 348 nm) occurred between 6 and 8 M urea. No further changes were observed beyond 8 M urea concentrations. These observations suggest that treatment of the protein with increasing concentrations of urea leads to less movement of partially buried tryptophan compared to GdmCl denaturation in which the maximum emission was observed at 354 nm.

To probe for the effect of urea on the hydrodynamic dimensions of Rv2557 and Rv2558 and their unfolding intermediates, the gel-filtration profiles of the proteins under different experimental conditions were studied. Size exclusion chromatography separates proteins by differences in their hydrodynamic dimensions rather than by their molecular masses (Ackers, 1970). This approach has been successfully applied to determine the Stokes radius ($R_S$) values for proteins in different conformational states. Figure 3.10C and 3.11D presents gel filtration profiles of Rv2557 and Rv2558 respectively in solutions with different concentrations of urea. Rv2557 elutes as a single peak, whose elution volume corresponds to a protein with a $R_S$ of 22.6 Å (Table 3.5). When the urea concentration increases, the peak is substantially shifted to smaller elution volumes. This corresponds to a transition to a slightly more expanded denatured state. Rv2558 also elutes as a single peak, whose elution volume corresponds to a protein with $R_S$ of 22.1 Å. In the presence of the denaturant, the elution profiles of the Rv2558 (Fig. 3.10 C & 3.10 D) display the same trends as that of the Rv2557, except that changes in the elution profile occur at higher urea concentrations (Table 3.5). The appearance of a separate peak indicates the transformation of the protein into a more open conformation. These studies are in agreement with the earlier fluorescence denaturation studies where also similar results were obtained.
Figure 3.10:
(A) Changes in CD ellipticity at 222 nm for Rv2557 (▲) and Rv2558 (■) on incubation with increasing concentrations of urea. Data were fitted according to equation (2). Values observed for proteins in the absence of urea are taken as 100%. The inset shows the stability curves of Rv2557 (▲) and Rv2558 (■).
(B) Changes in tryptophan fluorescence emission wavelength maximum of Rv2557 (▲) and Rv2558 (■) on incubation with increasing concentrations of urea.
(C) And (D) Size-exclusion chromatographic profiles for Rv2557(C) and Rv2558 (D) and on incubation with increasing concentrations of urea on a Superdex-75 H column at pH 7.0 and 25 °C. Curves 1–4 represent profiles for proteins at pH 7.0 on incubation with 0, 2.0, 4.5 and 6.5 M urea, respectively. The columns were run with the same concentration of urea in which the protein sample was incubated. The samples were incubated for 6 hrs in urea before column chromatography.
3.3.10 pH-induced conformational changes

To probe for pH-induced structural changes in Rv2557 and Rv2558, the proteins were exposed to different pH values ranging from 3 to 11. The effects of changing pH on the secondary structure content of the proteins were studied. Figure 3.11 summarizes the effects of the pH on the CD signal at 222nm for Rv2557 and Rv2558. An approximately 20% change in the CD signal at 222 nm was observed between 7 and 11 pH units for both proteins. However, decreasing the pH below 7 resulted in a loss of secondary structure as indicated by the sigmoidal curve in Figure 3.11. At pH 4 a significant loss of secondary structure (60% for Rv2557 and 80% for Rv2558) was observed. These results demonstrate that secondary structure content of Rv2557 and Rv2558 are more resistant to change under alkaline pH conditions but are highly sensitive to acidic pH.

![Graph showing pH-induced changes in secondary structure of Rv2557 and Rv2558](image)

**Figure 3.11:**
The pH-induced changes in secondary structure of Rv2557 (▲) and Rv2558 (■) monitored by following the changes in ellipticity at 222 nm obtained from far-UV CD curves by change in the pH of buffer. The insets, A (Rv2557) and B (Rv2558) show the far-UV scan at different pH values viz., 3 (■), 7 (*) and 11(▼) respectively.
Table 3.4: Thermodynamic parameters of unfolding of Rv2557 and Rv2558

<table>
<thead>
<tr>
<th>Protein</th>
<th>Thermal Denaturation</th>
<th>GdmCl Denaturation</th>
<th>Urea Denaturation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tm</td>
<td>$\Delta G^0$</td>
<td>-m</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(kcal/mol)</td>
<td>(kcal/mol/M)</td>
</tr>
<tr>
<td>Rv2557</td>
<td>75°C</td>
<td>4.87 ± 0.21</td>
<td>1.5 ± 0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>($N \leftrightarrow U$)</td>
<td></td>
</tr>
<tr>
<td>Rv2558</td>
<td>85°C</td>
<td>3.77 ± 0.99</td>
<td>1.3 ± 0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>($N \leftrightarrow U$)</td>
<td></td>
</tr>
</tbody>
</table>

N, native state; U, unfolded state
Table 3.5:
Molecular dimension of Rv2557 and Rv2558 in different conformational state

<table>
<thead>
<tr>
<th>Urea (M)</th>
<th>V_e (ml)</th>
<th>R_s (Å)</th>
<th>R_s/(R_o)_N</th>
<th>Urea (M)</th>
<th>V_e (ml)</th>
<th>R_s (Å)</th>
<th>R_s/(R_o)_N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>12.05</td>
<td>22.6</td>
<td>1.00</td>
<td>0.0</td>
<td>12.15</td>
<td>22.1</td>
<td>1.00</td>
</tr>
<tr>
<td>2.0</td>
<td>11.2</td>
<td>27.0</td>
<td>1.19</td>
<td>2.0</td>
<td>11.39</td>
<td>26.3</td>
<td>1.19</td>
</tr>
<tr>
<td>4.0</td>
<td>10.81</td>
<td>29.3</td>
<td>1.28</td>
<td>4.0</td>
<td>10.88</td>
<td>28.9</td>
<td>1.30</td>
</tr>
<tr>
<td>6.5</td>
<td>10.54</td>
<td>31.0</td>
<td>1.37</td>
<td>6.5</td>
<td>9.8, 10.43</td>
<td>34.8, 31.4</td>
<td>1.57, 1.42</td>
</tr>
</tbody>
</table>
3.4 Crystallization of Rv2557 and Rv2558

Crystallization trials of both proteins were performed as described in Section 3.2.12. Much success was not met in the crystallization of Rv2557 and Rv2558 with very tiny crystals (Fig. 3.12 & Fig. 3.13) conditions.

Rv2557
- 0.2M MgCl₂·6H₂O, 0.1M Tris pH 8.5, 3.4 M 1,6 hexandiol
- 0.01NiCl₂, 0.1Tris pH 8.5, 20%PEG Monomethyl ether 2000

Rv2558
- 0.5 M (NH₄)₂SO₄, 0.1 M Tri sodium citrate dihydrate pH5.6, 1.0 M LiSO₄·H₂O.
- 1.5 M (NH₄)₂SO₄, 0.1 M Tris pH 8.5, 12% Glycerol.
- 0.5 M NaCl, 0.01 M MgCl₂·6H₂O, 0.01 M Hexadecyltrimethylammonium bromide
- 0.1M HEPES pH7.5, 0.8 M NaH₂PO₄, 0.8 M KH₂PO₄

The further optimization by screening the constituent concentrations over a wide range, varying the overall pH and temperatures also did not bring any change in the quality of the crystals. Utilization of the nucleotide co-factors and combination of nucleotide and Mg²⁺ as additives in the crystallization solutions also was ineffective in bringing about a change in the size of the crystals. Additional inorganic and organic additives were also used in an attempt to improve crystalline conditions but the results were unaltered as above.

The primary step in the crystallization is the nucleation i.e., the formation of centers for the growth of crystals. A paradoxical situation that is often encountered during crystallization is that the optimal conditions for nucleation of the crystals are at times not the ideal ones to support their subsequent growth. This is because spontaneous nucleation is more likely to occur when the levels of super saturation are high, whereas slow, ordered growth of large crystals is favored by lower levels. A situation where the crystals do not grow beyond a limit needs can be tried for uncoupling the nucleation from growth to satisfy the distinctly different requirements.
of the two events. Seeding is one such technique that can be used for the separation of nucleation and growth (Thaller et al., 1985; Thaller et al., 1981). We utilized the Microseeding technique as described in section 3.12.3. Again no improvements in the quality of the crystals could be achieved.

Figure 3.12: Initial crystals of *M. tuberculosis* Rv2557
Crystallization of Rv2557 in 20%PEG Monomethyl ether 2000. Crystals grew at 22-24°C within a couple of days after continuous round of seeding.
Figure 3.13: Initial crystals of *M. tuberculosis* Rv2558

Crystallization of Rv2558 in 1.0M LiSO₄·H₂O. Crystals grew at 22-24°C within a couple of days after continuous round of seeding.
Figure 3.14: Initial crystals of *M. tuberculosis* Rv2558

Crystallization of Rv2558 in 1.5 M (NH₄)₂SO₄. Crystals grew at 22-24°C within a couple of days after further optimization.
3.5 Conclusion

Hypothetical proteins of the *M. tuberculosis* genome represent a wealth of information which needs to be understood to unravel the underlying biology. Biophysical characterization of the proteins will expectedly help the structure-function analysis, especially as several of these proteins are expected to play key roles during the pathogen life cycle. As part of a long term project on characterization of hypothetical mycobacterial proteins, two hypothetical proteins were purified which are apparently important for adaptation to persistence. These genes are highly restricted to mycobacteria with no known human homologs. The sequence analysis demonstrates an unusual amino acid composition for them in the *M. tuberculosis* genome with increased number of hydrophobic and charged residues. It is known generally from whole genome sequence and structural analysis of homologous proteins from mesophiles and thermophiles that the latter proteins exhibit a significant increase in the proportion of charged residues (Kumar et al., 2001) like the sequence properties of Rv2557 and Rv2558 reported here. Overall, Rv2557 appears to be structurally more stable compared to Rv2558 despite high sequence identity.

The analysis of denaturation profiles reveals that both proteins adopt different modes of structural stabilization. The comparative structural stability studies reveal that these proteins exhibit an almost thermophile like behaviour which might have arisen due to their presumed roles in adapting to conditions of stress and is also borne out by the unusual sequence features of the proteins. The studies also highlight interesting structural differences between the proteins despite sharing high sequence homology.
References


