Chapter -2

Functional and Structural characterization of Rv3868, An essential hypothetical protein of ESX-1 system
2.1 Introduction

Earlier yeast two-hybrid and genetic experiments proposed the interaction of CFP-10 with the PPE-68 protein Rv3873 (Teutschbeina et al., 2006). It has also been hypothesized that it might mediate the formation of the recently observed homo-dimers (Teutschbeina et al., 2006) and hetero-dimers in the ESAT-6 and CFP-10 proteins (Renshaw et al., 2002; Stanley et al., 2003), a step that might require chaperone activity. The PPE-68 protein Rv3873 is suggested to be a gating component of the ESX-1 system and regulates the secretion of the ESAT-6–CFP-10 complex (Gao et al., 2004). Rv3868, on the other hand, is hypothesized to be the chaperone or a source of energy (ATPase activity) required for the export of the factors. Structural and functional characterization of Rv3868 is certainly important to understand its role in ESX-1 mediated secretion and to exploit its potential as a novel drug target.

The detailed characterization of the protein reported here is the first for a protein from the CbxX/CfqX sub-family of AAA-ATPases (Neuwald et al., 1999). Sequence analysis shows that it is conserved among a small group of largely hypothetical proteins in mycobacterium (Maier et al., 2000). The structural, functional and regulatory properties were studied in detail and reveal that the protein is a novel AAA-ATPase. Individual domains of the protein were identified by limited proteolysis and their roles in the function of the protein have also been delineated and characterized. Biochemical studies, in silico modeling and mutational studies have led to a structural model for the protein action. NMR studies to probe possible interactions with CFP-10 have been carried out. Experiments to probe possible chaperone functions using substrates like lysozyme and citrate synthase are also reported. The NMR studies and chaperone-like assays suggest that the protein is not likely to be a chaperone and does not interact with CFP-10 as has been hypothesized earlier (Teutschbeina et al., 2006). The overall picture that emerges from present work is that Rv3868 functions as a novel ATPase with a co-factor induced ‘open-close’ mechanism.
2.2 Materials and Methods

2.2.1 Multiple sequence alignment and construction of Phylogenetic tree

The sequence of Rv3868 was downloaded from the Tuberculist web site at (http://genolist.pasteur.fr/TubercuList/). The multiple sequence alignment and neighbour-joining phylogenetic tree (*Dendrogram*) for the different families of proteins was calculated using the ClustalX package (Thompson *et al.*, 1997). Sequences of proteins from different ATPase families were downloaded from the Swissprot database (http://www.expasy.ch/sprot/).

2.2.2 Cloning, expression and purification of Rv3868, CT-Rv3868 and CFP-10

The full-length *Rv3868* gene from *M. tuberculosis* H37Rv was amplified using the pfX DNA polymerase (*Invitrogen*) with forward and reverse primers (Table 2.1). The C-terminal domain of *Rv3868* (330-481) (to be called CT-Rv3868) was amplified from the *Rv3868* PCR product using the primers (Table 2.1). *Rv3868* was cloned into *pET23a* (*Novagen*) using NdeI and HindIII. CT-Rv3868 was cloned into *pET23a* using BamHI and HindIII. Full-length *Rv3868* and CT-Rv3868 were expressed in BL-21 (DE3) cells (0.5 mM IPTG; O.D. 0.6; 30°C). The cells were harvested by centrifugation, resuspended in 40 ml lysis buffer A (50 mM Tris-HCl, 200 mM NaCl, pH 7.5) and lysed by sonication. Centrifugation at 14,000 rpm was followed by a filtration step using a 0.22μm filter before loading onto a 5 ml Ni-Hi Trap column equilibrated in buffer A. The column was initially washed with lysis buffer and subsequently with the same buffer containing 40 and 80 mM imidazole, respectively. The proteins were eluted with 15 ml buffer B containing 200 mM imidazole for Rv3868 and 400 mM imidazole for CT-Rv3868. The samples containing protein were pooled and dialyzed against buffer (50 mM Tris-HCl, 200 mM NaCl, pH 7.5). The plasmid *pET28b-efp10* (Meher *et al.*, 2006) was grown in M9 medium using (N15) ammonium sulphate as sole nitrogen source supplemented with kanamycin and induced at 1.0 O.D. with 0.5 mM IPTG. The culture was harvested after 3 h of growth at 37°C. The C-terminal His-tagged protein was purified under denaturing condition.
Ni-NTA column. The protein was refolded by extensive dialysis against buffer containing 25 mM Na$_2$HPO$_4$, 100 mM NaCl, 1 mM EDTA pH 6.5. Finally the labeled protein was dialyzed against NMR buffer (20 mM NaH$_2$PO$_4$ 50 mM NaCl, 0.1% NaN$_3$, pH 6.5)

2.2.3 Cloning, expression and purification of CT-Rv3868$^{P336A}$, CT-Rv3868$^{T338A}$, CT-Rv3868$^{K340A}$, CT-Rv3868$^{R348A}$

All four mutants namely CT-Rv3868$^{P336A}$, CT-Rv3868$^{T338A}$, CT-Rv3868$^{K340A}$, CT-Rv3868$^{R348A}$ were constructed by using the primer pairs (Table 2.1) on pET23a-CT-Rv3868 template to construct the mutant by PCR based sense-antisense primer method (Sambrook, 1989). DNA sequencing confirmed the presence of the respective mutations and ruled out other unwanted changes to the sequence. The schematic diagram of the procedure used for site directed mutagenesis to create active site mutants is in Scheme 2.1. The purification of mutant enzymes was carried out in a manner similar to that for the native enzyme
### Table 2.1 Primers used for cloning of Rv3868, CT-Rv3868 and generation of CT-Rv3868 mutants

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv3868 Forward</td>
<td>5'-CGGGACCAACATATGACTGATCGCTTAGC-3'</td>
</tr>
<tr>
<td>Rv3868 Reverse</td>
<td>5'-TTGAAGCTTTTCTCTCATGTTGAGGTG-3'</td>
</tr>
<tr>
<td>CT-Rv3868 Forward</td>
<td>5'-CCAGGATCCATGATCTTCCACCGGCCC-3'</td>
</tr>
<tr>
<td>CT-Rv3868 Reverse</td>
<td>5'-TTGAAGCTTAGCTGCGGCAAATGACGTTGGC-3'</td>
</tr>
<tr>
<td>P336A sense</td>
<td>5'-TCACCGGACC GGCCG TACCCGGC-3'</td>
</tr>
<tr>
<td>P336A antisense</td>
<td>5'-GCCGGTACC GGCCG TCGTCG-3'</td>
</tr>
<tr>
<td>T338A sense</td>
<td>5'-CCCGGACC CGCCG TCGCCGCAAG-3'</td>
</tr>
<tr>
<td>T338A antisense</td>
<td>5'-CTTCCGGCACC CGCCG TCGCG-3'</td>
</tr>
<tr>
<td>K340A sense</td>
<td>5'-GCCC CGTACCGGGCGACCAGATCGCC-3'</td>
</tr>
<tr>
<td>K340A antisense</td>
<td>5'-CGCGATCGTGGTCGCGGCGGCAACCCCGGG-3'</td>
</tr>
<tr>
<td>R429A sense</td>
<td>5'-GATGGAGAACC CGGGAGACCCGGCTGCGTG-3'</td>
</tr>
<tr>
<td>R429A antisense</td>
<td>5'-CACCCAGCCGTCGCGTGTTCTCCATC-3'</td>
</tr>
</tbody>
</table>
Scheme 2.1: Procedure used for site directed mutagenesis. Initially, the first half fragment was generated using a pair of sense primer (oligo1) and an antisense primer (oligo3) containing the sequence for mutated residue. Oligo4 and an antisense primer oligo2 were used to generate the second half fragment subsequently. In the final step both PCR products in equimolar concentrations were used as a template with the oligo1 and oligo2 primer pairs to generate the desired mutated sequence. The final PCR product was digested with BamHI and HindIII and cloned into pET23a digested with the same enzymes.
2.2.4 ATPase activity assays

ATPase reactions were carried out in 30μl of ATPase buffer (25 mM Tris, pH 7.6; 5 mM MgCl₂) at 30°C for different time periods. Each reaction mixture contained 0.5 μCi of [γ-³²P] ATP. The reaction was stopped by the addition of 0.5 μl of 10% SDS. 1.0μl of each reaction was spotted on a TLC plate. The plate was developed in 0.5 M formic acid and 0.5 M LiCl and dried at 37°C. The percentage of ATP hydrolysis was calculated using the formula:

\[
\text{Percentage of ATP hydrolysis} = \frac{\text{quantity of } [\gamma-\text{³²P}] P}{\text{quantities of } [\gamma-\text{³²P}] P + [\gamma-\text{³²P}] \text{ATP}}} \times 100.
\]

The ATP hydrolysis value was corrected for background by subtracting the value obtained for a reaction mixture containing no protein. Colorimetric assays (Lanzetta et al., 1979) were performed to determine the ATPase activity of the CT-Rv3868. Except for specified variations, standard ATPase assays were carried out in the assay buffer containing 50 mM Tris-HCl (pH 8.0), 20 mM MgCl₂, 1 mM dithiothreitol, 0.5 mM ATP, and amount of protein for 15 min at 37°C. Briefly, CT-Rv3868 was added to 100 μl of assay buffer, the reaction was carried out at 30°C for 15 min, then 200 μl of dye buffer containing 6 mM ammonium heptamolybdate, 120 μM malachite green, 0.06% polyvinyl alcohol, and 4.25% sodium citrate was added. After 20 min of incubation at room temperature, 200 μl from each reaction was transferred to a 96-well plate, and the absorbance at 630 nm was measured. Reactions performed without protein were routinely included, and the activity, usually negligible, was subtracted from the respective experimental data. The inorganic phosphate released was calculated based on the absorbance standard curve established by KH₂PO₄ standards. CT-Rv3868 and NT-Rv3868 were affinity-purified to near homogeneity, used in the assays. All assays were repeated three times, and the average activity is reported. Kinetic parameters, \(K_m\), \(V_{\text{max}}\) and Hill coefficient, were derived using Prism 4.0 (GraphPad Software, Inc.).
2.2.5 Limited proteolysis and ESI-MS

2.0 mg/ml protein was subjected to limited proteolysis with trypsin, at a protease to protein ratio of 1:50 and 1:100 (w/w), for different time periods at 30°C. The protease reaction was stopped by adding PMSF to a final concentration of 1 mM in the reaction mixture, and the samples were analyzed on 12% SDS-PAGE. Digested product was purified by gel filtration chromatography and transferred to a PVDF membrane for N-terminal sequencing. The ESI-MS analysis was carried out using a MICRO-MASS QUATTRO II mass spectrometer (Micromass, Altricem, UK)

2.2.6 Tryptophan and Tyrosine fluorescence

Protein concentrations of 0.5 and 1 μM for full length and purified domains respectively were used. Fluorescence spectra were recorded using a Perkin Elmer Life Sciences LS 50B instrument with samples placed in a 5 mm path length quartz cell at 25°C. Excitation wavelength of 285 nm was used and the spectra were recorded between 300 and 400 nm to monitor tryptophan fluorescence. Tyrosine fluorescence was monitored by using an excitation wavelength of 274 nM.

2.2.7 Circular Dichroism measurements

The CD measurements were made using a Jasco J810 spectropolarimeter calibrated with non-hygroscopic ammonium (+)-10-camphor sulfonate and the results were expressed as relative ellipticity and plotted as percentage values. The CD spectra were obtained at protein concentrations of 0.5 μM for far-UV CD, 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 buffer and denaturant under similar conditions. Each spectrum was an average of three scans. Per residue molar ellipticity was calculated from the observed ellipticities according to the equation:

\[ [\theta] = \frac{\theta_0 M_r}{l c} \]  

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; \(l\), path-length in centimeters; \(c\), protein concentration in mol/l. Percentage of secondary structure was calculated using the K2D program (http://www.heidelberg.de/~andrade/k2d/).
2.2.8 Analytical gel filtration

Gel filtration experiments were carried out using a Superdex 200 HR 10/300 column on an AKTA-FPLC system (M/s GE Healthcare). The column was calibrated using molecular weight standard markers (M/s GE Healthcare). All experiments were carried out using 50 mM Tris, pH 7.5. Other parameters like salt and nucleotide concentrations were varied for the experiments. Typically, 500 µl of the sample was loaded on the column and run at 25°C at a flow rate of 0.3 ml/min, with detection at 280 nm.

The relative elution volume was calculated as:

\[ K_{AV} = \frac{V_e - V_0}{V_g - V_0} \]  \hspace{1cm} (2)

Where \( V_e \) is the elution volume, \( V_0 \) is the void volume determined by elution of Blue Dextran 2000 kDa and \( V_g \) is the geometric column volume. For deconvolution of gel filtration peaks, Peakfit (Systat Software, Inc.) software was used for determination of different oligomers in Rv3868.

2.2.9 Dynamic light scattering

The DLS experiments were carried out on a Zetasizer Nano ZS instrument (Malvern instruments). Data were acquired at 20°C over 10s, repeated 10 times and averaged. Ten such acquisitions were performed to give 1000s of data. The inbuilt software was used to fit the autocorrelation function using the cumulants method and to extract the approximate molecular weight.

2.2.10 Analysis of steady state nucleotide binding

Binding of nucleotides to the proteins were determined by monitoring the change in protein fluorescence upon addition of ligand. Measurements were carried out using a Perkin Elmer Life Sciences LS 50B spectrofluorimeter, [excitation 280 nm; emission 330 nm; slit widths 5 nm] for Rv3868 and NT-Rv3868 where tryptophan fluorescence was followed. In the case of the CT-Rv3868, an excitation wavelength of 274 nm was used along with a 5 nm slit width. Tyrosine emission was followed at 304 or 340 nm for lower order and high order oligomeric forms of the domain. Titrations were performed at 25 °C by the addition of ATP to 0.6 ml of 50
mM Tris (pH 7.5), 50 mM NaCl, and 5 mM MgCl₂ buffer containing different amounts of proteins. To avoid dilution effects, volume change during the titration was limited to 3% of total volume. Control titrations with buffer alone did not produce any significant change in emission signal. The $K_d$ value was calculated fitting the data to the equations 3 and 4. $\Delta F$ is the change in emission signal in the presence of ligand (L) and $\Delta F_{\text{max}}$ is the maximal change in signal. The corrected data were fitted to the following equations using Prism 4.0 (Graph Pad Software, Inc.).

$$\Delta F = \Delta F_{\text{max}}[L] / K_d + [L]$$  \hspace{1cm} (3)

$$\Delta F = \Delta F_{\text{max}} - K_d\Delta F / [L]$$  \hspace{1cm} (4)

The binding stoichiometry of nucleotides and C-terminal domain was determined by plotting the titration data as a mass action plot according to the following equation:

$$r /[L]_{\text{free}} = n / K_d - r / K_d$$  \hspace{1cm} (5)

The fluorescently labeled ATP analog, N-methylantraniloyl-ATP (MANT-ATP) (Molecular Probes) was used to qualitatively substantiate the binding of the nucleotide to the proteins. All spectra were corrected for the inner filter and dilution effects. Nucleotide binding to CT-Rv3868 was followed by the changes in MANT-ATP emission at 450 nm, (excitation and emission wavelengths). 1 μM MANT-ATP was titrated with increased concentrations of the protein in the experiments.

2.2.11 Fluorescence quenching and calculation of Stern-Volmer constant

Fluorescence quenching of tryptophan in the presence of increasing concentrations of acrylamide were monitored by following the emission at 340 nm after excitation at 285 nm. Samples were prepared in buffer consisting of 50 mM Tris, pH 7.5, 50 mM NaCl and 5 mM MgCl₂. Aliquots from a 2 M acrylamide stock solution were consecutively added in 5 mM steps to 1 ml reaction mixture. Experiments were performed in triplets and corrected for dilution effects. Quenching data were plotted as the ratio of fluorescence in absence of quencher ($F_0$) to the intensity in the presence of quencher (F) against quencher concentration. The resulting data were fit against dynamic parameters according to the Stern-Volmer equation (Lakowicz, 1983),
\[ F_0 / F = 1 + (K_{sv} \times [Q]) \]  

\( K_{sv} \) is the *Stern-Volmer* constant for quenching, given by the slope when data are plotted as \( F_0 / F \) versus \([Q]\) where the latter parameter is the concentration of the quencher.

### 2.2.12 Thermal and chemical 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. The experiments were carried out both in the presence and absence of 0.5 mM ATP. Samples were heated at a constant rate of 1°C/min in a 2 mm cell. For chemical denaturation, samples were incubated in increasing concentrations of guanidinium chloride at 25°C for 3-4 hours before the measurements. The folded fraction of protein at any temperature was determined as follows:

\[ ([\theta]^{\text{obs}} - [\theta]^{\text{den}})/([\theta]^{\text{nat}} - [\theta]^{\text{den}}) \]  

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

### 2.2.13 ANS binding

Titrations were performed to estimate the binding affinities of the proteins to 8-anilino-1-naphthalenesulfonic acid (ANS). Incremental amounts of ANS were added to a series of otherwise identical solutions of protein in buffer (50 mM Tris pH 7.5, 50 mM NaCl and 5 mM MgCl2). The excitation wavelengths were set to 370 and the emission was measured from 410 nm to 600 nm respectively. For each measurement, the fluorescence intensity was corrected by subtracting the fluorescence of the sample containing only ANS. The data were plotted against the total concentration of ANS. The apparent \( K_d \) was estimated by fitting the data to equation 8:

\[ F = F_{\text{max}}[\text{ANS}] / (K_d + [\text{ANS}]) \]
Where $F$ is the corrected fluorescence intensity, $F_{\text{max}}$ is the fluorescence intensity upon saturation of the ANS binding sites, $[\text{ANS}]$ is the total concentration of ANS and $K_d$ is the apparent dissociation constant.

### 2.2.14 Glutaraldehyde cross-linking

The cross-linking of protein samples was carried out in the presence of 1\% glutaraldehyde of CT-Rv3868 at a concentration of 0.2 mg/ml was used in the experiments. The molecular mass of the cross-linked products were determined by 12\% SDS-PAGE.

### 2.2.15 Molecular modeling

Sequence analysis led to the identification of putative Walker A and B motifs at the C-terminal end. A model of the putative ATP binding site was generated by comparative modeling approaches (Marti-Renom et al., 2000) by MODELLER corresponding to residues 331-481 of CT-Rv3868. The NT-Rv3868 domain model (residues 18-250) was generated using the fold prediction method implemented in the PHYRE server (http://www.sbg.bio.ic.ac.uk/phyre). The initial models were minimized using the DISCOVER module implemented in Insight II (Accelrys).

#### 2.2.15.1 MODELLER 6v2

MODELLER is widely used for homology or comparative modeling of protein three-dimensional structures (Marti-Renom et al., 2000). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints (Fiser et al., 2000; Sali and Blundell, 1993), and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc.. In obtaining a 3D model, Modeller tries to optimize the molecular probability density
function with the variable target function procedure in cartesian space using methods of conjugate gradient and molecular dynamics with simulated annealing.

2.2.16 NTP docking

*In silico* screening methods support the decision-making process in drug discovery by the evaluation of large virtual libraries / databases, which further helps in zeroing in on a smaller number of potential hits. Automated docking procedures coupled with biochemical assays have been successfully applied to database screening, *de novo* design and the analysis of binding modes of individual molecules.

Docking programs used nowadays are designed to carry out both ‘docking’ as well as ‘scoring’ tasks. The distinction between docking and scoring defines also the two major technical challenges faced by docking programs to predict the binding mode of a molecule correctly (herewith also referred to as ‘pose prediction’, where ‘pose’ refers to the orientation and conformation of a molecule at the receptor binding site and to predict the binding affinity of compounds (or to produce a relative rank-ordering for a number of compounds) in a reliable manner (Stahl and Rarey, 2001).

Basically, ligand-docking studies can be carried out using three distinct algorithms:

1. Lamarckian Genetic Algorithm (LGA)
2. Genetic Algorithm (GA)
3. Simulated Annealing (SA)

LGA is a more robust algorithm and generally provides better-docked results than the other two. It is a hybrid search technique that implements an adaptive global optimizer with local search. The local search method is based on the optimization algorithm of Solis and Wets (SW), which has the advantage that it does not require gradient information in order to proceed. Genetic Algorithm (GA) uses ideas of natural genetics and biological evolution. In this case of molecular docking, the particular arrangement of a ligand and a protein can be defined by a set of values describing transition, orientation and conformation of the ligand with respect to the protein. These are ligand state variables and in the GA each state variable corresponds to a genotype.
Whereas atomic co-ordinates correspond to the phenotype. Due to the well-known fact that amino acid sequence homology at a given level leads to similar 3D structure of proteins, several databases are interrelating the databases of sequences and structures. However, the term homology, a fundamental concept in bioinformatics, is often used incorrectly. Sequences are homologous if they are related by divergence from a common ancestor (as a first consequence, the search for homology in the sequence database is used to determine indications for function of proteins).

Conversely, analogy relates to the acquisition of common structural or functional features via convergent evolution from unrelated ancestors. Homology is not a measure of similarity, but rather an absolute statement that sequences have a divergent rather than a convergent relationship. Among homologous sequences we can distinguish orthologs (proteins having the same function in different species) and paralogs (proteins performing different but related functions within one organism).

The model building of a target sequence based on the comparison with the data extracted from homologous sequences with known structures (parents or templates) is termed comparative modelling. This can also be extended to homologs with low percentage of identity. Several software tools / Servers are available for homology modeling, e.g. CASP, Swiss-model server, Modeller, Wloop, Homology (InsightII), UCLA-DOE etc to name a few.

All current comparative modelling methods consist of four sequential steps:
1. fold assignment and template selection;
2. template-target alignment;
3. model building;
4. model evaluation.

2.2.16.1 Autodock 3.0.5

The program Autodock (Morris et al., 1998) was developed at The Scripps Research Institute and provides an automated procedure for predicting the interaction of ligands with biomacromolecular targets using LGA. Docking procedure can be divided into 4 steps:
1. Preparing Macromolecule
2. Preparing Ligand
3. Generating Grid Maps
4. Running Autodock
Step 1: Macromolecule preparation involves removal of all non-protein atoms and addition of all polar hydrogens. This is followed by addition of Kollman united atom charges and solvation parameters defined for each residue in Addsol.

Step 2: Ligand preparation involves addition of all hydrogens followed by Gasteiger charge calculation. Non-polar hydrogens and their corresponding charges are merged with the parent heavy atom. Rotatable bonds in the ligand are identified with Autotors and appropriate records are written to an output file.

Step 3: Autodock requires pre-calculation of grid maps one for each atom type present in the ligand being docked. This helps to make the docking calculations extremely fast. These maps are calculated by AutoGrid. A grid map consists of a three dimensional lattice of regularly spaced points, surrounding and centered on some region of interest. Figure 2.1 illustrates the main features of a grid map. The user must specify an even number of grid points in each dimension. This is because AutoGrid adds a central point and Autodock requires an odd number of grid points. The probe’s energy at each grid point is determined by the set of parameters supplied for that particular atom type and is the summation of non-bonded cutoff radius over the entire macromolecule. Parameters for Autogrid can be defined using mkgpf 3 script.

Figure 2.1: Features of a grid map
Step 4: Docking through Autodock can be done using any one of the several methods available. Parameters are different for each docking algorithm. These parameters can be defined using mkdpf script (Autodock manual).

![Figure 2.2: Binding of enzyme and inhibitor](image)

Autodock 3.0.5 implements a new scoring function that is based on the principles of QSAR. Figure 2.2 shows the thermodynamic cycle for the binding of an enzyme ‘E’ and an inhibitor ‘I’ in both the solvated phase and *in vacuo*. The solvent molecules indicated by filled circles tend to be ordered around the larger molecules, but when E and I bind, several solvent molecules are liberated and become disordered.

This is an entropic effect and is the basis of the hydrophobic effect. The solvent ordering around E and I, both bound and unbound, is strongly influenced by the hydrogen bonding between these molecules. These hydrogen bonds between solvent and E, and solvent and I, contribute enthalpic stabilization. According to Hess’s law of heat summation, the change in free energy between two states will be the same, no matter what the path. So, the free energy of binding in solvent $\Delta G_{binding,\text{solution}}$ can be calculated by the following equation:

$$
\Delta G_{binding,\text{solution}} = \Delta G_{binding,\text{vacuo}} + \Delta G_{solvation}(EI) - \Delta G_{solvation}(E+I)
$$

$\Delta G_{binding,\text{vacuo}}$ is calculated during docking runs.
The program was used in the *in silico* docking studies involving NTPs with the CT-Rv3868. The volume chosen for the grid maps was 80 x 80 x 80 grid points with a spacing of 0.375 Å. The Lamarckian Genetic Algorithm was used for the calculations. Docked complexes were visualized using InsightII (Accelrys) and PYMOL (DeLano, 2002).

### 2.2.17 Generation of hexameric model

First model structures of NT-Rv3868 and CT-Rv3868 were joined close together, based on surface area and charge in Insight II. The oligomer of full length was modeled by superposing the C-terminal model structure onto the hexameric D2 domain of NSF (PDB code, 1NSF) (Richard C. Yu *et al.*, 1998) by PYMOL.

### 2.2.18 NMR spectroscopy

For the NMR experiments $^{15}$N-labeled CFP10 protein in 20 mm sodium phosphate (pH 6.5), 50 mm NaCl, 0.1% sodium azide and 5% (v/v) $^2$H$_2$O was purified as reported earlier (Meher *et al.*, 2006). The spectra were recorded on a Varian 600-MHz instrument equipped with triple nuclei inverse probe, at 30 °C. 2D $^{15}$N-$^1$H-HSQC spectra were recorded for the ($^{15}$N-labeled CFP-10) as reported earlier as also for the ($^{15}$N-labeled CFP-10-unlabeled Rv3868) protein. The HSQC spectrum for each experiment was acquired with 1024 and 128 complex points in the $^1$H and $^{15}$N dimensions, respectively.

### 2.2.19 Chaperone activity assays

The assays were carried out using procedures similar to those described earlier (Ferreira *et al.*, 2006) at 43°C using Hen egg white lysozyme (*Sigma*) and porcine mitochondrial citrate synthase (*Sigma*) as test substrates.
2.3 Results and discussion

2.3.1 Sequence and phylogenetic analysis of Rv3868

The protein has been classified as a conserved hypothetical protein in the databases such as Tuberculist (http://genolist.pasteur.fr/Tuberculist/) and TB structural genomics consortium (http://www.doe-mbi.ucla.edu/TB/). Our sequence analysis and the construction of a phylogenetic tree using the neighbour-joining method supports that Rv3868 is a novel member of the CbxX/CfqX sub-family of AAA-ATPases (Neuwald et al., 1999) also known as CbbX (Fig. 2.3). The sister group of CbbX proteins are sporulation factors and the related proteins in mycobacteria have apparently acquired alternative functions (Maier et al., 2000). The protein itself consists of 573 amino acids with Mr ~63 kDa. No crystal structure or solution structure was available for Rv3868 or related protein from mycobacterium sp. The genome wide BLAST analysis shows that the N-terminal of Rv3868 has alanine rich motifs but exhibits no sequence homology to other genes. The C-terminal domain contains the Walker motifs characteristic of the ATPase domain in AAA-ATPases. Based on sequence alignments and phylogenetic analysis the specific walker motif’s residues in Rv3868 could be demarcated, which are found to be conserved (Fig. 2.4) in other hypothetical proteins also. Rv3868 was also analyzed for prediction of functionally linked protein in string database. The application of this tool for the prediction of co-occurrence of genes in related species, proteins exhibiting similar phylogenetic profiles can be predicted to be functionally linked i.e. they occur as a structural complex or are involved in a common pathway. The phylogenetic analysis reveals that Rv3868 is closely related to Rv0282 which is localized to downstream of Rv0287 and Rv0288. Rv Rv0287 and Rv0288 are members of CFP10 family and their complex stability is closely related to CFP10/ESAT-6 complex (Lightbody et al., 2008). This suggests that Rv3868 and related proteins in mycobacteria regulating essential function in secretion of virulent components.
Figure 2.3: Evolutionary analysis of Rv3868
Phylogenetic tree calculated for different AAA-ATPase protein families with the major proteins highlighted in boxes:
A, CbbxX family; B, Periplasmic chaperones; C&D, Cdc/p97; E, Metallopeptidases; F, DNA helicases
The Swiss-Prot No. is given in brackets. Bootstrap values (%) are indicated for the major branch points in the tree. To generate the family relationships shown in the tree, the protein sequences were aligned initially and then bootstrapped 1,000 times using the PAM 250 amino acid comparison table.
Figure 2.4: Phylogeny analysis of Rv3868 in *mycobacterium tuberculosis*

(A) Multiple sequence alignment of amino acid sequence of Rv3868 of *M. tuberculosis* and respective homologs. The alignment was carried out using ClustalX.

(B) Phylogenetic analysis of Rv3868 in *M. tuberculosis*. 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.
2.3.2 Cloning, expression and purification of Rv3868 and CT-Rv3868

The coding sequence for this protein was amplified from genomic DNA of *M. tuberculosis* by PCR producing a 1722 bp amplified product. *NdeI* and *HindIII* sites were designed in the primers to facilitate cloning in pET 23a expression vector. Since *NdeI* is a poor cutter, leading to failure in ligation of Rv3868 PCR product to pET23a expression vector, it was first cloned in pGEM-T Easy vector. The vector was supplied linearized with single 3’-T overhangs for TA cloning. These single 3’-T overhangs at the insertion site greatly improve the efficiency of ligation of a PCR product into the plasmids by preventing recircularization of the vector and providing a compatible overhang for PCR products generated by certain thermostable polymerases. Since Pfx DNA polymerase is unable to add ‘A’ overhang, ‘A’ tail was generated in the PCR product according to the said protocol of Promega. The orientation of the gene in pGEM-T was confirmed by restriction enzyme digestion with *HindIII*. There are two possibilities regarding the orientation of the cloned gene as elucidated in the schematic diagram (Fig. 2.5). If the gene is present in orientation II, restriction digestion with *HindIII* will give a larger fragment of and a smaller fragment of which could not be seen on the gel. While in orientation I, restriction digestion with *NdeI* will give only one fragment so in orientation I. When double digested with *NdeI* and *HindIII*, an insert of 1722 bp (i.e., Rv3868 gene) will be obtained, in case of orientation II possibility of extra base pairs will reside, so clones have been selected having orientation I for further cloning in pET23a expression vector. After sub cloning in pET23a expression vector a clone was sequenced and the result confirmed the presence of Rv3868 gene in-frame with C-terminal His-tag in pET23a multiple cloning site. The *CT-Rv3868* is 1347 bp long and was amplified by PCR by using sense and antisense primer set for C-terminal his-tagged protein. The PCR product corresponds to 1347 bp was eluted using PCR gel extraction kit and cloned into T7 RNA polymerase-based bacterial expression vectors, pET23a between *BamHI* and *HindIII*. Clones were confirmed through restriction digestion and PCR. The expression plasmids were introduced into *E.coli* C-41(DE3), a strain that contains T7 RNA polymerase gene under the control of *lac-UV5* promoter. Rv3868 was optimized for growth and induction at 37°C for 4 hrs by inducing with 0.3 mM IPTG (Section). Rv3868 was optimally soluble (50 – 60%) and over-expressed under these conditions and a
prominent ~ 63 kDa polypeptide was detected in case of Rv3868 by SDS-PAGE in whole cell extracts of IPTG induced bacteria (Fig. 2.5). These polypeptides were not present in uninduced samples or when bacteria containing the pET23a alone were induced with IPTG.

Figure 2.5: Cloning strategy and over-expression of Rv3868
Schematic presentation of steps involved in the cloning, and expression of Rv3868 protein. The coding region for Rv3868 gene was amplified from genomic DNA of M.tuberculosis using Taq DNA polymerase. “A” tailing procedure was followed for cloning in pGEM-T Easy vector.
2.3.3 The hypothetical protein ORF Rv3868 of *M. tuberculosis* encodes an ATPase

Purified recombinant Rv3868 was free of any contamination according to SDS-PAGE (Fig. 2.6A). However our attempts failed to detect ATPase activity on pmol quantity from colorimetric assay so far purified Rv3868 was assayed for ATPase activity using [$\gamma^{-32}$ P] ATP as a substrate the reaction product free Pi was separated by TLC and quantified by auto- radiography. The release of free phosphate increases linearly over time (Fig. 2.6B); while a plot of ATP hydrolysis as a function of the protein concentration is in Figure 2.6C. A $K_m$ of $0.8 \pm 0.1 \mu M$ and $V_{max}$ of $139 \pm 8.8 \text{ fmol/min}$ respectively for the ATPase activity of Rv3868 was derived from a Michaelis-Menten plot (Fig. 2.6D) following a non-linear regression analysis using Prism 4.0 (Graph Pad Software, Inc.). There is no co-cooperativity in ATPase activity of Rv3868 which is stated by hill coefficient ($n = 0.976$).

In order to check the specificity of the enzyme for ATP, the GTPase activity of the protein was also checked using [$\gamma^{-32}$ P] GTP. The GTPase activity of the enzyme was approximately 20% of the corresponding ATPase activity under similar assay conditions (Fig. 2.6E). It has been observed in some AAA-ATPases that the ATPase activity is stimulated by casein, DNA or other specific factors (Barnett *et al.*, 2000; Purkey and Ebisuzaki, 1977). Consequently activity assays were carried out in the presence of casein (Fig. 2.6F). The ATPase activity is not stimulated in the presence of casein. The activity is however much reduced in the presence of 0.5 M NaCl and abolished in the presence of EDTA (Fig. 2.6F). It is possible that NaCl could disrupt the oligomeric associations in the enzyme and these were investigated subsequently. EDTA chelates out the Mg$^{2+}$ ions which are necessary for the ATPase activity. The characterization thus reveals a specific ATPase activity for the oligomeric form of the full-length enzyme. There is some possibilities of reduced activity since structure-function relationship have not been demonstrated for Rv3868 and to gain further insights into regulation of ATP hydrolysis, domain organization of Rv3868 is need to postulate.
2.3.4 Effect of temperature and pH on activity

The activity profile of Rv3868 was checked in the pH range of 4.5 to 9.5. The optimal pH for MtuLigA was 8.0 with greater than 20% activity observed at 7.5 (Fig. 2.7A). Activity declined progressively before pH 7.0 while at higher pH activity of Rv3868 did not decline. The Rv3868 ATPase activity was also checked at different temperature (Fig. 2.7B). At 30°C maximum activity was observed while at higher temperature activity was declined.

2.3.5 Molecular weight and subunit characterization

The molecular mass of the purified recombinant Rv3868 was determined under non-dissociating conditions by size exclusion chromatography. Gel filtration of the recombinant protein on the Superdex S-200 column calibrated with the various molecular weight standards showed a single peak with a retention volume of 11.9 ml (Fig. 2.8A). When the elution volumes of the marker proteins were plotted as a function of log of molecular mass, Rv3868 was found to have a molecular mass of about 390 kDa. The calculated molecular mass (from primary amino acid sequence) of 63 kDa for the Rv3868 corroborated that Rv3868 is composed of six monomer. The results of the subunit mass (as determined by SDS-PAGE) along with SEC studies demonstrate that Rv3868 exists as a Hexamer under physiological conditions.

2.3.5.1 Effect of ATP on oligomerization

Size exclusion chromatography experiments were carried out in the presence of ATP also. The results show that the full-length protein is a hexamer (Mr ~390kDa) both in the presence and absence of ATP and does not exhibit further self-association in presence of ATP. This contrasts to the observation in many other AAA-ATPases which exhibit ATP-dependent self-association (Krzewska et al., 2001; Zolkiewski et al., 1999)
Figure 2.6: Purification and functional characterization of Rv3868
(A) SDS-PAGE analysis of the purified protein. Lanes 1-5 represent molecular weight markers, supernatant of induced culture lysate, flow-through and protein eluted in 300 mM imidazole from the Ni-NTA column respectively. (B) The protein was incubated...
with [\gamma^{32}\text{P}] ATP for various time intervals to observe the ATPase activity. The amount of ATP hydrolyzed at each time point is shown as a percentage of the original [\gamma^{32}\text{P}] ATP before incubation at 30°C. Each point is the average of the values obtained from three independent experiments. The insets represent the autoradiography profiles taken from a TLC. (C) Plot of ATP hydrolysis with respect to protein concentration (D) Michaelis-Menten plot of ATP hydrolysis by Rv3868. (E) Relative NTP hydrolysis (arbitrary units) when [\gamma^{32}\text{P}] ATP and [\gamma^{32}\text{P}] GTP were used as substrates of Rv3868. (F) Effects of NaCl, casein and EDTA on ATPase activity. Inset: lanes 1-5 represent control experiments in the presence of 0.5M NaCl, 5\mu g, 10\mu g casein and 10 mM EDTA respectively.

Figure 2.7: ATPase activity of Rv3868 under various conditions
(A) Influence of pH on Rv3868 ATPase activity (n=3)
(B) TLC profile of ATPase activity determined at different temperatures (n=3). In TLC, C indicates control without Rv3868 and other numbers indicate specific temperatures.
2.3.5.2 Effect of concentration and salt on Rv3868 oligomerization

Analytical gel filtration experiments show that Rv3868 predominantly exists as a hexamer at protein concentrations up to ~3mg/ml. At higher concentrations the hexameric population reduces and the protein forms higher order oligomer, both in the presence and absence of ATP (Fig. 2.8 B and C). Dynamic light scattering experiments further support that the higher oligomeric state is a multiple of hexamer. Increasing the salt concentration inhibited the oligomeric association. NaCl concentrations up to 100 mM did not disrupt the oligomer; but concentrations 0.75 M and above resulted in mainly a monomer, in the presence and absence of ATP, in the size exclusion chromatography experiments (Fig. 2.8 D). At a salt concentration of 0.5 M the protein is predominantly dimeric. Similarly, a monomeric form was mainly observed in the presence of 1 M GdmCl, a known disrupter of ionic interactions in proteins. The quaternary structure was therefore deduced as being stabilized by ionic interactions. The ATPase activity is also abolished at GdmCl concentrations greater than 0.5 M as observed with NaCl earlier. The ATP binding site in AAA-ATPases is made up of residues from neighboring subunits. The loss in activity at high salt concentrations can therefore be rationalized by the corresponding reduction in oligomerization.

2.3.6 Dynamic light scattering studies

Dynamic light scattering is a technique for measuring the size of molecules and nanoparticles. DLS measures the time dependent fluctuations in the scattering intensity to determine the translational diffusion coefficient (DT), and subsequently the hydrodynamic radius (RH). In Rv3868 formation of higher order oligomerization was observed in concentration dependent manner. To discriminate between aggregates or higher order oligomer deconvolution of gel filtration elution peaks and light scattering was done in different concentration of protein (for a solution containing ~6 μM to 10μM Rv3868 in 25 mM Tris, pH 7.5 & 50mM NaCl). It is clear from volume and size distribution of light scattering experiments that Rv3868 in not in aggregate form (polydispersity) and multimeric of hexamer (Fig. 2.9C). Polydispersity (Pd) less than 20% of the sample indicates the absence of non-specific aggregates.
The autocorrelation function, shown in Fig. 2.9D is as determined by the method of cumulants, with random residuals, indicating that it is monodisperse. It also shows a bimodal size distribution for the sample suggesting two populations of different oligomeric states. The apparent hydrodynamic radius extracted from the autocorrelation function centers on 23.16 nm.
Figure 2.8: Analytical gel-filtration of Rv3868 under various conditions
(A) Size exclusion chromatography profile of Rv3868 (~4 μM) on a Superdex 200HR column in 25 mM Tris, 50 mM NaCl (pH 7.0) and at 25°C. Inset represents the calibration curve of the column. (B) Hexamer (dotted line) to higher order oligomer (Black) equilibrium of Rv3868 under concentration dependent manner. (C) Rv3868 at concentrations of ~6 μM (Blue), 7-15 μM (Green) and >15 μM (Black) was incubated for 1 h at 25 °C before centrifugation and applied onto a Superdex 200HR column. The hexameric form is represented by I while II refers to a dodecameric form and a high order oligomeric form (multiple of hexamer) is represented by III. (D) Size exclusion chromatography profile of Rv3868 (~3 μM) in 25 mM Tris buffer. The protein was incubated in the respective buffers for 1 h. The gray and black lines represent the elution profile in the presence of 500 mM and 750 mM NaCl respectively in the same buffer.
Figure 2.9: Dynamic light scattering experiments involving Rv3868
Size distribution based on (A) Intensity and (B) Volume studied by DLS
(C) Analysis of molecular weight and monodispersity of higher order oligomers (peak 2) and aggregates (peak 1).
(D) A plot of the correlation coefficient with respect to time (μs).
2.3.7 Identification, purification, and subunit characterization of a compact N-terminal Domain of Rv3868

The factors determining the vulnerability of a protein for proteolysis by protease depends on the conformational parameters such as accessibility, segmental motion, and protrusions. For this reason limited proteolysis has been effectively used to monitor structural domains in protein (Hubbard, 1998). Domain organization calculated by Fold index (Fig. 2.10A and 2.10B) (Prilusky et al., 2005) of Rv3868 deduce that protein is composed of two domain. To find out the stable domain of Rv3868 trypsin digestion has been performed. Figure 2.11A summarizes the SDS-PAGE profile of the protein fragments obtained on limited proteolysis of the recombinant Rv3868 with trypsin. Almost similar pattern of proteolysis, depending on the protein to protease ratio used, was observed. At a protein to protease ratio of 50:1 and 100:1 two protein fragments, namely Fragment I and II, corresponding to molecular mass of approximately 30 and 20 kDa, respectively were obtained. However, when the proteolysis was carried out for a different span, Fragment II was proteolysed. The former fragment could be purified from the reaction mixture by size exclusion chromatography as a monomer (Fig. 2.11 B) and was identified as having a molecular weight of 29.9 kDa from ESI-MS (Fig. 2.11C). The monomer was stable even at high protein concentrations and did not exhibit any self-association as in the case of the full-length protein. This fragment corresponds to about half the molecular weight of the protein. Peptide sequencing established that this fragment/domain occurs in the N-terminus of the protein with starting sequence TDRLA. Sequence analysis (Fig. 2.12) had shown that only the N-terminal stretch contains tryptophan any. Subsequent activity assays also revealed that the fragment, as expected does not possess any ATPase activity observed in the full-length enzyme.
Figure 2.10: Domain organization of Rv3868

(A) An estimate of the folded regions in Rv3868 calculated using the program Fold Index. Predicted folded regions are shown in green while unstructured regions are in red and clearly show the demarcation of the protein sequence into two distinct domains. (B) A schematic depiction of the predicted domain boundaries and motifs of Rv3868
Chapter 2

Figure 2.11: Limited proteolysis of Rv3868; purification of NT-Rv3868.

(A) SDS-PAGE analysis of Rv3868 digested with trypsin at a protease to protein ratio of 1:100 (w/w) and 1:50 (w/w). Lanes 2-5 as also lanes 6-9 represent the reactions after 5, 10, 15 and 30 min respectively.

(B) Size exclusion chromatography profile of Rv3868 after trypsin digestion for 30 min (NT-Rv3868). The protein was loaded onto column in 25 mM Tris, 50 mM NaCl (pH 7.0) and at 25°C. The inset in panel D depicts the SDS-PAGE analysis of the purified N-terminal domain. Lanes 1-4 represent molecular weight markers, reaction after 0 min, after 30 min and purified protein following size exclusion chromatography respectively.

(C) ESI-MS of NT-Rv3868.
Figure 2.12: Amino acid sequence of Rv3868
N- and C- terminal domains are depicted in green and gray respectively. Tryphotophan residues and those of the Walker motifs are depicted in red and blue respectively.
2.3.8 Identification and characterization of the ATP-binding domain (CT-Rv3868)

Fold index (Prilusky et al., 2005) (Fig. 2.10A) results for Rv3868 suggested that residues between 330-481 in the C-terminus contain the Walker motifs/ATP binding site and should encode for an ~18 kDa fragment. This fragment was cloned and purified for further studies. (Fig. 2.13A & B) Size exclusion chromatography and glutaraldehyde cross-linking experiments show that this CT-Rv3868 associates predominantly as a dimer in the absence of ATP and forms higher order oligomers in the presence of the nucleotide (Fig. 2.14A, B&D). This suggests that this domain is largely responsible for oligomerization. Purified CT-Rv3868 was fractionated on Superdex S-200 column and the fractions were examined for total protein content and enzymatic activity (Fig. 2.14C). Total protein determination of the gel filtration fractions revealed that CT-Rv3868 eluted in three forms and predominant form was dimeric conformation (form III) and other two forms (form I and form II) appeared to contain larger oligomers or aggregates. It is obvious from above discussion that form I which has much higher ATPase activity is associated with the larger oligomeric fraction than with the dimeric fraction.

C-terminal domain hydrolyzes ATP in time dependent manner, with free phosphate increasing linearly in the initial time (Fig. 2.15A). The reaction is dependent on the concentration of ATP and the enzyme activity is saturated when ATP is greater than 1mM (Fig. 2.15D). The kinetics were modeled using Enzyme kinetics template from software prism 3.0. In presence of 5mM magnesium, $K_m$ 73.39±20μM, $V_{max}$ 141.2±12 nmol/min/mg and hill coefficient 1.4 for CT-Rv3868. (Fig. 2.15D & E)

2.3.8.1 C-Terminal Domain of Rv3868 displays cooperative activity

To establish whether the concentration of CT-Rv3868 influences its specific activity, ATP hydrolysis was examined at different enzyme concentrations under saturating substrate conditions (1 mM ATP). If enzyme concentration does not affect activity, then a plot of ATP hydrolysis versus enzyme concentration should demonstrate a linear relationship; however, the plot of ATP hydrolysis versus enzyme concentration (Fig. 2.15C) for C-terminal domain revealed a concave mode of behavior, indicating
that the activity is concentration dependent. At low enzyme concentrations, the specific activity of C-terminal domain was not maximal under saturating substrate conditions. As the enzyme concentration increased, so did the specific activity until the maximal enzyme activation was reached near 50nmol/min/mg. Concentration-dependent activity may indicate cooperative association of C-terminal domain. This behavior has been identified for a number of NTPase that have subsequently been functionally characterized as oligomers (Pozidis et al., 2003).

Figure 2.13: Cloning and purification of CT-Rv3868
(A) Cloning of CT-Rv3868 as his-tagged protein. Lane 1, 1 kb DNA ladder (Amersham); Lane 2, and 3, Restriction digestion of CT-Rv3868 clone in pET23a.
(B) SDS-PAGE analysis: Lanes 1-5 represent molecular weight markers, supernatant of induced culture lysate, flow-through after loading on the NI-NTA column, after a washing step using 80 mM imidazole and released protein after treating the column to a 300 mM imidazole elution step, respectively.
Figure 2.14: ATP dependent self association, comparative relative activity of CT-Rv3868

(A) SEC profile of C-terminal domain on a Superdex 200HR column at pH 7.0 and 25°C inset in this panel cross-linking of the protein eluted as major peak showing dimeric fraction of protein (B) CT-Rv3868 elutes in three forms in concentration dependent manner. (C) Relative ATPase activity of different forms. (D) Size exclusion chromatography profiles of CT-Rv3868 in the absence (black line) and presence of ATP (gray line). The column was equilibrated with 25 mM Tris, 50 mM NaCl, pH 7.0, (+/-2 mM ATP) and at 25°C.
Figure 2.15: ATPase activity of CT-Rv3868

(A) Time dependence of ATPase activity of CT-Rv3868. (B) ATP hydrolysis at various concentrations of CT-Rv3868 (■) and NT-Rv3868 (▲). (C) Concentration dependence of ATP hydrolysis was analyzed using a plot involving specific activity (nmol min\(^{-1}\) mg\(^{-1}\)) versus CT-Rv3868 concentration (µg); inset represents time dependence of ATPase activity of CT-Rv3868. The reactions were carried out at 30°C for the indicated time periods (min). (D) Michaelis-Menten plot of ATP hydrolysis by CT-Rv3868. The ATPase assays were carried out using 1.0 µg protein and at the indicated concentrations of ATP. (E) Figure depicts the co-operativity of the CT-Rv3868 ATPase activity. The steady state activity (V\(_0\)) was transformed to fractional rates (v/V\(_{\text{max}}\) - v) where V\(_{\text{max}}\) is the Michaelis-Menten constant, so as to fit to the Hill-like equation. The gradient of the plot gives the degree of cooperativity.
2.3.8.2 Heat induced ATPase activity of the CT-Rv3868

Previous studies have suggested an involvement of Rv3868 in stress tolerance and sequence similarity between CT-Rv3868 and D1 domain of VCP, in vitro ATPase activity profile of CT-Rv3868 was analyzed under various physiological conditions reflecting environmental stresses. Interestingly the activity increased at elevated pH and temperature peaked between 50-55°C (Fig. 2.16A).

The ATPase activity reached a maximum at an ATP concentration of 1mM. It is reasonable to speculate that the ATPase activity can be modulated by the level of ATP within a normal physiological concentration of ATP (0.5-4mM) to determine the nucleotide specificity, NTP hydrolysis assays were carried out with ATP, GTP, UTP, CTP and ADP at a concentration of 1mM. High specificity of CT-Rv3868 for ATP was observed (Fig. 2.16B). CT-Rv3868 did not have hydrolysis activity towards ADP and other nucleotide except GTP (30% relative activity compared to ATP). The activity was found to be maximal between 7.5-8.5 pH and at 5mM Mg²⁺ (Fig. 2.16 C, D). Titration experiment further determine the optimum concentration for Mg²⁺ to be 5mM, notably at concentration greater than 10mM, the ATPase activity is markedly suppressed. This probably reflects a general inhibitory effect of the high salt concentration rather than a Mg²⁺ Specific phenomenon. More than 4mM EDTA inhibit the ATPase activity of CT-Rv3868 (Fig. 2.16E). The presence of higher concentration of NaCl and KCl also suppress the ATPase activity (Fig. 2.16 F).
Figure 2.16: ATPase activity of CT-Rv3868 under various conditions
(A) Effect of temperature on the ATPase activity of CT-Rv3868. (B) Optimal substrate for hydrolysis. NTPs were added to a final concentration was 0.8mM. (n=4). (C) Influence of pH on Rv3868 ATPase activity (n=4). (D) Effect of MgCl₂ on ATPase activity of CT-Rv3868 (n=2). (E) Effect of EDTA on ATPase activity of CT-Rv3868 (n=2). (F) Effect of NaCl and KCl on ATPase activity of CT-Rv3868(n=2)
2.3.9 *In silico* modeling studies and analysis of the nucleotide binding site

A modeling study of the tertiary structure of the protein was carried out to gain insights into the structure-function relationship. The individual domains were modeled separately in the first instance. The C-terminal domain was modeled using comparative methods (Fig. 2.17 A) implemented in the *Modeller* program with high confidence because of the larger number of available structures containing the AAA-ATPase domain in the PDB. The N-terminal domain was modeled based on fold prediction methods because of the low sequence homology to other proteins. The predicted fold is mostly helical and agrees with the far UV CD spectra analysis delineated in the next section.

The size exclusion chromatography and DLS results on the full-length protein suggested that the oligomeric state is a hexamer while the work involving the individual domains supports that the ATP-binding domain is the one involved in oligomerisation. Accordingly the hexameric association crystal structure of the D2 domain of the AAA-ATPase N-ethylmaleimide sensitive factor (Fig. 2.17B) (R. C. Yu *et al.*, 1998). The features of the ATP binding site were subsequently analyzed.

Different nucleotides including ATP, ADP, GTP, CTP and UTP were docked into the binding site of Rv3868 using the Lamarckian genetic algorithm implemented in the Autodock program to gain insights into their interactions with the protein (Fig 2.18). The scoring functions suggest that the ATP should have the best affinity followed by GTP and subsequently by other nucleotides (Table 2.2). The docking results agree with the protein characterization where it was shown that the GTPase activity of the protein is only about 1/3rd of that observed with ATP.

A close-up of the nucleotide binding site is in Figure 2.19. The ATP moiety interacts with several residues of the Walker motifs and is in a similar orientation as observed in other AAA-ATPases eg. N-ethylmaleimide sensitive factor complexed with ATP (R. C. Yu *et al.*, 1998). The (426-436) loop in the C-terminal domain of a neighboring subunit comes spatially close to the ATP binding site. Very interestingly it was observed that Arg-429 of this loop is in a spatially advantageous position to function as a ‘sensor arginine’ reported in AAA-ATPases (Fig. 2.19) (Ogura *et al.*, 2004).
The arginine in these ATPases helps distinguish between ATP and ADP bound forms of the enzyme and results in corresponding conformational changes in the respective proteins (Ogura et al., 2004).

Table 2.2: *In silico* docking energy and relative activity (%) of different nucleotides for CT-Rv3868

<table>
<thead>
<tr>
<th>Activity</th>
<th>Free Energy (Kcal/mol)</th>
<th>Relative Activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>-9.06</td>
<td>100</td>
</tr>
<tr>
<td>GTP</td>
<td>-6.8</td>
<td>35</td>
</tr>
<tr>
<td>UTP</td>
<td>-5.61</td>
<td>N.D</td>
</tr>
<tr>
<td>CTP</td>
<td>-5.92</td>
<td>N.D</td>
</tr>
<tr>
<td>ADP</td>
<td>-7.52</td>
<td>5</td>
</tr>
</tbody>
</table>
Figure 2.17: In silico modeling of Rv3868

(A) Cartoon representation of the structure of the N- and C-terminal domains depicted in yellow and violet respectively. The docked ATP moiety (green) is shown in stick representation. Dotted lines indicate the linker region between two domains. (B) Hexameric association in Rv3868. The nucleotide binding sites occur at the inter-subunit interfaces as in other AAA-ATPases. The arrows indicate movement of N-terminal domain predicted by the dynamic quenching and other experiments. The ATP-binding site is marked by a box.
Figure 2.18: Predicted docking mode of different NTPs
Molecular docking with ATP (orange), GTP (blue); UTP (green) and CTP (red) to CT-Rv3868 was carried out using AUTODOCK
Figure 2.19: Close-up of the ATP binding site in Rv3868.
The residues corresponding to the Walker A motifs are indicated by cyan space-filled models. Arg 429, the predicted ‘sensor arginine’ from the modeling studies, is depicted as a yellow stick and is from a neighboring subunit. The predominantly negatively charged electrostatic potential surface around the adenosine nucleoside moiety of ATP indicates that the environment around the moiety is mainly polar.
2.3.10 Mutational analysis and identification of Arg429 as a sensor-arginine

Four mutants of CT-Rv3868 viz. P336A, T338A, K340A and R429A have been generated based on the modeling studies to probe for the roles of the residues in ATP binding and hydrolysis. The first three mutants correspond to those residues that belong to the same subunit in the nucleotide binding site while the Arg residue is from a symmetry related subunit of the oligomer. Thr338 and Lys340 lie close to the γ-phosphate in the docked complex. Arg429 was chosen to examine its role as a probable sensor-arginine while the Pro residue was mutated to check for possible structural effects on the binding site architecture.

Table 2.3 lists the various parameters of the respective mutants. The wild-type protein has a catalytic efficiency of about 577 as suggested by the $K_{cat}/K_m$ ratio. The Hill coefficient of 1.4 is indicative of the positive co-operativity in CT-Rv3868. The P336A mutant does not seem to distort the binding site architecture; the catalytic efficiency as also the $V_{max}$ is only marginally reduced in the mutant (Fig. 2.20). The Thr338 and Lys340 residues apparently perform different roles in the hydrolysis. Thr338 contributes to the binding and its mutation leads to ~7-fold decrease in the binding of the substrate as suggested by the $K_m$ values. The catalytic efficiency also is reduced ~10-fold. The reduction in the affinity of the substrate in this mutant is also supported by the positive change in the free energy. The K340A mutant does not affect the binding of the substrate and the $K_m$ is relatively unaffected. But there is an approximately 35% reduction in the catalytic efficiency as also the $V_{max}$ indicating that this residue has a rigorous role in the catalytic action as opposed to stabilizing the substrate. In the three mutations detailed above the co-operativity is relatively unaffected.

The R429A mutant’s also exhibits a large increase in the $K_m$ and a drastic reduction in the catalytic efficiency (Fig. 2.21). The positive free energy change also indicates a loss in the binding of the substrate. These parameters are similar to those seen in the T338A mutant. An important difference is that while the Thr mutation did not lead to loss in co-operativity, the R429A mutation almost abolishes the co-operativity as seen by the Hill co-efficient of 1.15 (Fig. 2.22 A). Binding and release of the nucleotide in AAA-ATPases are known to lead to changes in the conformation of the oligomer and gives rise to co-operative effects.
The above conclusions are supported by ATP-dependent self-association experiments. CT-Rv3868 as mentioned earlier exhibits ATP-dependent self-association and also shows co-operativity. All mutants except the Arg429 mutant exhibit ATP-dependent self-association and also the co-operativity is relatively unaffected. The Arg mutant loses the ATP dependent self-association and the co-operativity is also abolished (Fig. 2.22 A & B). Obviously the conformational adjustments necessary for the binding of the nucleotide are precluded in the mutant. These properties are consistent with the in silico prediction of Arg429 as a sensor-arginine.

Table 2.3: Effect of mutations in the ATP binding site in CT-Rv3868.

All Kcat and Km values are expressed as a mean ± S.E. for n=3.

<table>
<thead>
<tr>
<th></th>
<th>Vmax</th>
<th>Kcat</th>
<th>Km</th>
<th>Kcat/Km</th>
<th>n*</th>
<th>ΔGATP**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(nmol/min/mg)</td>
<td>(min⁻¹)</td>
<td>(µM)</td>
<td>(M⁻¹S⁻¹)</td>
<td></td>
<td>(kcal/mol)</td>
</tr>
<tr>
<td>CT-Rv3868</td>
<td>141.2±12</td>
<td>2.541±0.23</td>
<td>73.39±20</td>
<td>577</td>
<td>1.40</td>
<td>0.0</td>
</tr>
<tr>
<td>P336A</td>
<td>130.9±12</td>
<td>2.357±0.23</td>
<td>79.97±21</td>
<td>491</td>
<td>1.37</td>
<td>0.096</td>
</tr>
<tr>
<td>T338A</td>
<td>87.33±30</td>
<td>1.572±0.5</td>
<td>524.3±70</td>
<td>49</td>
<td>1.34</td>
<td>1.47</td>
</tr>
<tr>
<td>K340A</td>
<td>117.2±15</td>
<td>2.110±0.28</td>
<td>92.29±30</td>
<td>380</td>
<td>1.31</td>
<td>0.256</td>
</tr>
<tr>
<td>R429A</td>
<td>106.4±25</td>
<td>1.914±0.5</td>
<td>439.8±60</td>
<td>72</td>
<td>1.15</td>
<td>1.244</td>
</tr>
</tbody>
</table>

* n  Hill coefficient  
**ΔG = -RTln [(kcat/Km)mut]/[(kcat/Km)wt]
Figure 2.20: Effect of mutations on kinetics of CT-Rv3868 ATPase activity

Figure 2.21: Relative Catalytic efficiency of CT-Rv3968 and its mutants
Figure 2.22: Comparative cooperativity and ATP-dependent self association

(A) Relative hill plot if CT-Rv3868 and its mutants. (B) ATP-induced alterations in the molecular dimension of CT-Rv3868<sup>R429A</sup> (upper panel) and CT-Rv3868 (lower panel). Size-exclusion chromatographic profiles for proteins (black) and on incubation with 2mM ATP (grey) on a Superdex-200 column at pH 7.0 and 25 °C. The columns were run with the same concentration of ATP in which the protein sample was incubated. The samples were incubated for 1 hr in ATP before column chromatography.
2.3.11 Nucleotide binding affinity and binding site environment

The NTPase activity assays of the ATP-binding domain delineated earlier clearly showed that the high order oligomer form I exhibited the highest activity. The affinity of ATP and other nucleotides to all three forms was determined to gain insights into oligomerization and binding.

Exploiting the intrinsic fluorescence of Trp and Tyr residues of proteins has for a long time been relied upon as an efficient technique for studying protein ligand complexes. Several groups have studied nucleotide binding by monitoring changes in the intrinsic fluorescence of the protein. The fluorescence spectrum of Rv3868 is shown in Fig. 2.23A. The emission wavelength of the tryptophan fluorescence of full length was observed at 340nm, the solvent exposed tryptophan residues in folded protein show fluorescence emission maximum at 330-340 nm. Hence, in the Rv3868 protein the tryptophan molecule is exposed to ionic environment. According to the primary amino acid sequence, Rv3868 has eight tryptophan residues in N-terminal domain while no tryptophan is present in C-terminal domain (330-481)(Fig. 2.23 A&B). Emission maximum in NT-Rv3868 was also observed at 340nm while in CT-Rv3868 emission wavelength of the tyrosine fluorescence was observed at 304nm in Form I and Form II while in Form III the emission maximum was observed at 340nm. It reflects that form III exists in higher order(Fig. 2.23C). Intrinsic fluorescence of Trp residues of N-terminal domain was explored to monitor domain movement after binding of nucleotide to C-terminal part. Small aliquots of ATP added to the full length Rv3868 solution resulted in a decrease in fluorescence emission intensity with a $\Delta F_{\text{max}}$ of ~ 30 %. This quenching is saturable with $K_d$ of 1.2 mM. Upon addition of ADP and AMP-PNP, no quenching was observed and $K_d$ value was very high (Fig. 2.24). In full length in absence of MgCl$_2$ the no quenching was observed after addition of ATP. To rule out of the possibility of any artifactual measurement in fluorescence quenching upon nucleotide binding, nucleotide binding in N-terminal domain was also monitored that is known not to change in conformation after binding nucleotides. As expected, the N terminal domain showed very poor quenching of intrinsic Trp fluorescence ($\Delta F_{\text{max}}$ ~10 %).

There are a total of four tyrosine moieties in the C-terminal domain and two of them are at a distance less than 4.5 Å from the bound nucleotide (Fig. 2.25). Small aliquots.
of nucleotides (ATP, ADP and AMP-PNP) added to ATPase domain resulted in Tyr quenching at 304 and 340nm for determination of binding constant and stoichiometry. A fixed amount of the C-terminal domain was titrated with increasing concentrations of the ATP (Fig. 2.27A). Binding isotherms were obtained by plotting fractional change in fluorescence $\Delta F/\Delta F_{\text{max}}$ as a function of ATP concentration (Fig. 2.26). Similar isotherms were generated for ATP in the absence of MgCl$_2$ also. Binding constants of different nucleotides were also calculated and it reveals that ATP has maximum binding affinity for CT-Rv3868 (Table 2.4). The affinity of ATP for CT-Rv3868 was found to be 0.27 ± 0.5 mM while it reduces ~4-fold in the R429A mutant to 1.08±0.2 mM (Fig. 2.26). ATP has the highest affinity for the high oligomer form II and the $K_d$ value for the binding is 0.27 mM. On the other hand, its affinity of ATP for the dimeric form III and form I oligomers are approximately half that observed in the case of Form II (Fig. 2.27 B, Table 2.5). The stoichiometry of binding of ATP as well as other for other nucleotides as derived from the scatchard plot (Fig. 2.27D) was 1.732 for the dimeric form III. This value is in line with the expected stoichiometry of 2 binding sites per dimer for ATP. In our experimental conditions the protein concentration was 1 μM and thus the $r$-value in equation 5 was same as $\Delta F/\Delta F_{\text{max}}$.

ATP binding co-operativity for the three oligomeric forms were derived from respective Hill plots and show that there is a high degree of co-operativity with a Hill co-efficient $>2.0$ for the three forms (Fig. 2.27C). This underscores the importance of nucleotide binding in oligomeric association. Overall, the results are in agreement with the modeling studies and the activity assays on the full-length protein where it was noted that oligomerization is necessary for activity.

To reconfirm intrinsic fluorescence data fluorescent analog of the nucleotide, MANT-ATP was used. The complex formation was studied by the changes in MANT-ATP fluorescence at the emission maximum of 445 nm. Binding to the protein caused decreased intensity of the MANT-ATP nucleotide emission and followed a hyperbolic curve, corresponding to the saturation of a single binding site (Fig. 2.28A). The dissociation constant of the complex was measured to be $K_d = 0.47 \mu$M (Fig. 2.28 B). The binding studies were performed in the presence of Tris buffer, and no inhibition effect was observed. The observed decrease of the MANT-ATP intensity upon binding to CT-Rv3868 suggests a polar surrounding of the bound nucleotide and static quenching of its emission. The MANT-ATP fluorescence is
known to be sensitive to the local environment, (Nowak et al., 2006; Toshiaki, 1983) and the ATP binding site in ATPase domain consists mainly of hydrophilic residues, which would be expected to cause the observed decrease.

Table 2.4: Affinity of different nucleotides for CT-Rv3868

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Binding constant (K_d) (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP(+MgCl_2)</td>
<td>0.4982±0.1230</td>
</tr>
<tr>
<td>ATP(-MgCl_2)</td>
<td>0.5907±0.05</td>
</tr>
<tr>
<td>ADP</td>
<td>0.6246±0.1363</td>
</tr>
<tr>
<td>AMP-PNP</td>
<td>0.6509±0.06</td>
</tr>
</tbody>
</table>

Table 2.5: ATP affinity for CT-Rv3868

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Form I</th>
<th>Form II</th>
<th>Form III</th>
</tr>
</thead>
<tbody>
<tr>
<td>K_d (ATP)(mM)</td>
<td>0.4982±0.1230</td>
<td>0.2728±0.03413</td>
<td>0.6001±0.05073</td>
</tr>
<tr>
<td>n (Cooperativity in ATP binding)</td>
<td>2.087 ± 0.3137</td>
<td>3.064 ± 0.4110</td>
<td>2.165 ± 0.3455</td>
</tr>
</tbody>
</table>
Figure 2.23: Fluorescence emission spectra

(A) Rv3868, (B) NT-Rv3868 & (C) Different forms of CT-Rv3868. All Trp residues are present only in NT-Rv3868 (and Rv3868). The samples were in 25mM Tris, pH 7.5, 50mM NaCl
Figure 2.24: Effect of nucleotides on tryptophan quenching of Rv3868
Emission spectra at varying ATP concentration measured at 25°C. ATP concentrations were varied (0.1 mM to 1.0 mM). The arrow indicates the quenching observed upon increasing the concentration of ATP. Steady-state binding of ATP (B) AMP-PNP (C) and ADP (D) to Rv3868.
Figure 2.25: Position of two tyrosine residues (439 & 466) in the close vicinity of the docked ATP

Figure 2.26: Steady-state binding of ATP to CT-Rv3868 (■) and CT-Rv3868R429A (▽). Data are plotted as the percentage of fluorescence change versus ATP concentration (mM).
Figure 2.27: Steady-state nucleotide binding of CT-Rv3868.

(A) Emission spectra of 1μM CT-Rv3868 (□) (Form III) and in presence of eg. 0.1 mM (○), 0.2 mM (△), 0.5 mM (■), 0.7 mM (▲), 1.0 mM (+) ATP respectively (B) Steady-state binding of ATP to Form I (●), Form II (■) and Form III (○). Data are plotted as the percentage of fluorescence change versus ATP concentration (mM). The Kₐ value for the binding of ATP to the different forms is given in Table 2.5 (C) cooperativity plots of ATP binding of Form I (●), Form II (■) and Form III (○). (D) The Scatchard plot of the titration of CT-Rv3868 (FormIII) and ATP according to Eq. (5) to determine the binding stoichiometry of the complex.
Figure 2.28: Quantitative study of the binding of MANT-ATP to CT-Rv3868 studied by fluorescence spectroscopy.

(A) Fluorescence emission spectra of 1.0μM CT-Rv3868 in the absence of MANT-ATP (---), 1μM MANT-ATP in the absence of protein (•) and in the presence of 0.5μM of CT-Rv3868 (Δ) and 1.0μM of CT-Rv3868 (○). (B) Steady-state binding of MANT-ATP to CT-Rv3868.
2.3.12 Structural features of Rv3868 and its variants

2.3.12.1 Circular Dichroism

Studies on the model polypeptides and proteins reveal that the α-helical and β-sheet proteins show characteristic far-UV CD spectra. The helical proteins have two minima at 222nm and 208nm and the sheet proteins have a single minimum at 216nm. Far UV spectra of the full length protein in the presence and absence of ATP clearly suggests that the protein in the absence of the nucleotide is predominantly helical while the addition of ATP or the analog AMP-PNP induces conformational changes and also in secondary structure content (Fig. 2.29 and Table 2.6). NT-Rv3868 and CT-Rv3868 also show the helical characteristic of secondary structure. In CT-Rv3868 after addition of ATP, no major change in secondary structure is obtained. It clear reflects that after addition of ATP in full length major conformational changes was occurred in NT-Rv3868.

Table 2.6: Secondary structure content of Rv3868 and variants

<table>
<thead>
<tr>
<th></th>
<th>α-helix (%)</th>
<th>β-sheet (%)</th>
<th>Random coil (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rv3868</td>
<td>54</td>
<td>13</td>
<td>33</td>
</tr>
<tr>
<td>Rv3868(+ATP)</td>
<td>13</td>
<td>69</td>
<td>17</td>
</tr>
<tr>
<td>Rv3868(+AMP-PNP)</td>
<td>59</td>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>CT-Rv3868</td>
<td>45</td>
<td>23</td>
<td>31</td>
</tr>
<tr>
<td>CT-Rv3868(+ATP)</td>
<td>46</td>
<td>16</td>
<td>38</td>
</tr>
<tr>
<td>CT-Rv3868(+AMP-PNP)</td>
<td>45</td>
<td>23</td>
<td>31</td>
</tr>
<tr>
<td>NT-Rv3868</td>
<td>87</td>
<td>-</td>
<td>13</td>
</tr>
</tbody>
</table>
Figure 2.29: Far-UV CD spectrums corresponding to:
(A) Rv3868; (B) Rv3868 (●), Rv3868+AMP-PNP (▲) & Rv3868+ATP (○); (C) NT-Rv3868; (D) CT-Rv3868
2.3.13 Denaturation studies on Rv3868

2.3.13.1 Thermal denaturation

The thermal denaturation profile of the full-length and individual domains of Rv3868 were studied by following the change in molar ellipticities at 222 nm as a function of the temperature in the presence and absence of 0.5 mM ATP respectively (Fig.2.30A). At ATP concentrations below this there was no significant structural change in the full-length protein. During thermal denaturation, Rv3868 apparently goes through two transitions. In the experiments without ATP, the mid-point of the first transition occurs at around 30°C while the other one occurs at around 48°C. Addition of ATP stabilized the second transition and the mid-point of the same was observed to be around 53°C. The first transition probably corresponds to the reduction in the oligomerisation while the second transition might reflect the loss of tertiary structure in each subunit. Going by the CD signals, about 50% of the structure is stabilized even at 90°C in the presence of ATP while in its absence; about 30% of the protein retains structure. This probably reflects stabilization of additional regions in the protein after ATP binding and also agrees with structural changes in the far UV CD experiments detailed earlier. Secondary structure of Rv3868 was calculated at different temperature and it also reflects the stabilization of Rv3868 at higher temperature.

To compare the thermal stability, Rv3868 and its variants thermal denaturation study was also carried out. Figure 2.30B shows comparative thermal denaturation profile proving that C-terminal domain is thermal stable domain and only 20% structure was lost at 100°C. In full length during thermal denaturation 0°C to 60°C, 60% structure was denatured and 60°C to 100°C no structure loss was observed due to thermal resistant C-terminal domain. While in N-terminal domain at 100°C, 90% structure was denatured (Fig. 2.30B). It reveals that N-terminal domain is protease resistant not thermal resistant while active C-terminal domain is providing thermal stability to Rv3868 and thermal stress induced ATPase activity. The melting temperature of all variants was calculated by fitting the values in two state equations and it reflects that some mutual interaction between NT-Rv3868 and CT-Rv3868 lowers the stability of Rv3868.
Figure 2.30: Thermal Denaturation of Rv3838 and its variants

(A) Mean residual ellipticity (%) at 222 nm was plotted at different temperatures. Rv3868 (0.25 μm) was in 25 mM Tris pH 7.5 without ATP (■) and in presence of 0.5 mM ATP (○) for the experiments. (B) Comparative denaturation profiles of Rv3868 (■), CT-rv3868 (▲) and NT-Rv3868 (●).
2.3.13.2 GdmCl and urea induced denaturation

The unfolding characteristic of Rv3868 was studied by monitoring the GdmCl and urea induced changes in the structural properties of Rv3868. Time-dependent changes in the structural parameters of Rv3868 at increasing GdmCl and urea concentrations (0.5, 1, and 4 M) were monitored to standardize the incubation time required to achieve equilibrium under these conditions. Under all the conditions studied, the changes occurred within a maximum of 3 h with no further alterations in the values obtained up to 12 hrs. These observations suggested that a minimum time of about 3 h is sufficient for achieving equilibrium under any of the denaturing conditions studied. Figure 2.31A summarizes the effect of increasing concentrations of GdmCl on the CD ellipticities at 222 nm of the Rv3868 protein. A biphasic loss of CD signal was observed between 0.0 and 6.0 M GdmCl. An initial sharp decrease of about 55% in the CD signal was observed between 0.0 M and 0.5 M GdmCl. This was followed by a further gradual sigmoidal decrease in CD signal from about 55 to 0% between 2 M and 5 M GdmCl. As the GdmCl concentration (0.5 M to 2 M) at which loss of 60% of secondary structure was observed, but two state model indicate a partially stabilization intermediate and probably this partially intermediate is monomer conformation of Rv3868. The stabilization of a partially unfolded intermediate of Rv3868 during GdmCl induced unfolding of protein was further supported by the tryptophan fluorescence studies (Fig.2.32A). To evaluate this partially stabilized conformation from 0.5 M to 2.0 M, gel filtration chromatography was run in 1.0 M GdmCl buffer and a hexamer to monomer equilibrium was observed in gel filtration profile (Fig.2.32B). The second transition (monomer conformation) was stabilized in range of 10 -20% after addition of ATP. It reflects that ATP is stabilizing monomer conformation and the results are in agreement with the thermal denaturation studies (Fig.2.31B).

Although urea and GdmCl are believed to have similar modes of action, GdmCl is a monovalent salt that has both ionic and chaotropic effects whereas urea has only chaotropic effects. Thus urea is an ideal control agent for distinguishing between the ionic and chaotropic effects of GdmCl. The urea-induced changes in the secondary structure of Rv3868 were studied by monitoring changes in CD ellipticity at 222 nm at increasing urea concentrations and are summarized in Figure 2.31. Unlike GdmCl denaturation in urea denaturation there are four transitions between 0.0 M to
1.5M, 1.5 M to 3.0M, 3.0M to 4.5 M and 4.5M to 6.5M. This suggests that the urea denaturation of Rv3868 is a multi-state process where the unfolding of the Rv3868 by GdmCl denaturation is a three state process. It also reflects that secondary structure of Rv3868 is stabilized by ionic interaction.

Figure 2.31: GdmCl induced structural changes
(A) GdmCl (□) and urea (○) induced changes in secondary structure of Rv3868 as monitored by following the changes in ellipticity at 222nm obtained from far-UV CD curves at increasing concentrations of GdmCl and urea (B) Plot of the fractional changes in the ellipticity at 222 nm of Rv3868 in the absence (■) and presence (●) of ATP after incubation with increasing molar concentrations of guanidinium chloride. \([\theta]_{\text{obs}}\) is the ellipticity for a particular sample, \([\theta]_{\text{nat}}\) is the ellipticity in the absence of guanidinium chloride, and \([\theta]_{\text{den}}\) is the ellipticity at 6 M guanidinium chloride.
Figure 2.32: GdmCl induced structural alteration in Rv3868
(A) Changes in tryptophan fluorescence emission wavelength maximum of Rv3868 on incubation with increasing concentrations of GdmCl.
(B) GdmCl-induced alterations in the molecular dimension of Rv3868. Size-exclusion chromatographic profiles for Rv3868 on incubation with 1.0M GdmCl on a Superdex 200 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 1 hr in GdmCl before column chromatography.
2.3.14 pH induced conformational change in Rv3868

To probe for pH-induced structural changes in Rv3868, the protein was exposed to different pH values ranging from 4 to 11. The effects of changing pH on the secondary structure content of the proteins were studied. Figure 2.33A summarizes the effects of the pH on the CD signal at 222nm for Rv3868. An approximately 40% decrease in the CD signal at 222 nm was observed between 7 and 11 pH units. However, decreasing the pH below 7 resulted in more loss of secondary structure (60%) as indicated Figure 2.33A. These results demonstrate that secondary structure content of Rv3868 is more resistant to change under alkaline pH conditions but are highly sensitive to acidic pH. At pH 6 and 7 emission maximum of Rv3868 is 338nm while form pH 7 to 11 emission maximum is increasing from 338nm to 343nm (Fig2.33B). These observations rationalized that at higher pH Rv3868 exits in open conformation and allows ATP to insert in active site which we previously described (Section 2.4.4) that at higher pH (8.0-9.0), ATPase activity of Rv3868 is maximum.

![Figure 2.33: pH-induced conformational change](image)

(A) changes in the secondary structure of full length as monitored by following changes in ellipticity at 222nm obtained form the far-UV CD curves at increasing pH. (B) Changes in tryptophan fluorescence emission maximum
2.3.15 N- and C-terminal domains are proximal and exhibit a relative conformational change on nucleotide binding

The full-length protein contains 8 tryptophan and these are situated only in the NT-Rv3868. Dynamic quenching study on the full length protein and the NT-Rv3868 using acrylamide as a quencher was carried out. This moiety, on account of its polar nature interacts with tryptophan residues which are exposed or partially buried and leads to a quenching of the fluorescence. This approach gives insights into relative conformational changes between the domains based on the quenching of the tryptophan fluorescence as also reported earlier (Georlette et al., 2004). Probing the individual accessibility of each Trp residue rigorously requires the determination of $k_q$, the bi-molecular rate constant: $k_q = K_{sv} \cdot \tau_0$, where $K_{sv}$ and $\tau_0$ are the Stern-Volmer constant and fluorescence life-time respectively. However, the presence of 8 Trp residues impeded the determination of $\tau_0$ for the individual residues. As is generally accepted, the conformational changes can alternatively be studied by comparing the Stern-Volmer constants rather than the bi-molecular rate constants. The Stern-Volmer plots for the NT-Rv3868 and full length Rv3868 in the presence and absence of ATP are in Figure 2.34A. The $K_{sv}$ for the N-terminal domain alone is $9.37 \pm 0.53$ M$^{-1}$. The $K_{sv}$ corresponding to the full length protein in the absence of ATP is $5.12 \pm 0.54$ M$^{-1}$ while it is $4.11 \pm 0.96$ M$^{-1}$ in its presence (Fig. 2.34). If the $K_{sv}$ values for the full length protein and the NT-Rv3868 alone were similar it would suggest that the two domains in the protein are not in close proximity as the accessibility of the individual Trp residues is relatively unaffected.

However the present experiments represent a direct evidence for the proximity of the two domains in the protein. A significant reduction was observed in the $K_{sv}$ value for the ATP-bound enzyme compared to the unbound form. This clearly suggests that the two domains move closer to each other from a relatively 'open' to a 'closed' conformation on addition of the nucleotide. From the above results, it is straightforward to visualize that the binding of nucleotide co-factor and its release should be accompanied by a concomitant change in the relative spatial dispositions of the N- and C-terminal domains suggesting an 'open-close' induced mechanism for the protein.
The above results were independently corroborated by following the intrinsic fluorescence of the Trp residues in the presence of ATP. Addition of the nucleotide led to a reduction in the observed Trp fluorescence in the full-length protein. On the other hand, addition of the nucleotide aliquots to the NT-Rv3868 alone leaves the observed fluorescence relatively undisturbed (Fig. 2.35). Since the Trp residues occur only in the N-terminal segment which has no nucleotide binding sites, the quenching can only be presumably due to the increased proximity of the two individual domains on addition of the nucleotide and corresponding reduction in the accessibility of surface exposed tryptophan residues.
Figure 2.34: Stern-Volmer plots of Rv3868 and its variants.
The quenching of intrinsic fluorescence of the proteins was followed after addition of increasing concentrations of acrylamide or ATP. (A) Fluorescence quenching observed in Rv3868 (●) Rv3868+ATP (▼) and NT-Rv3868 (▲) proteins on addition of acrylamide. The Stern-Volmer quenching constants ($K_{SV}$) corresponding to plot slopes.

Figure 2.35: (A) The $\Delta$ fluorescence intensity of Rv3868 (■), NT-Rv3868 (●) followed at 340 nm plotted as a function of ATP concentration. (B) Decrease in Trp fluorescence of NT-Rv3868 upon the addition of ATP and decrease in Trp fluorescence of Rv3868 upon the addition of different nucleotides.
2.3.16 Rv3868 does not interact with CFP-10

Previously, it was suggested that Rv3868 might interact with CFP-10 or ESAT-6 proteins (Brodin et al., 2006; Teutschbeina et al., 2006). Other groups have suggested a chaperone function for the protein in this context (Gao et al., 2004; McLaughlin et al., 2007). Since a predicted recognition motif is present in the C-terminal segment of CFP-10 (Champion et al., 2006), NMR studies were undertaken to identify possible interactions with the latter protein. 2D $^{15}$N- $^1$H HSQC spectrum serves as a fingerprint of the overall structure of a protein. The HSQC spectrum recorded with $^{15}$N-labelled CFP-10 at 30 °C is shown in Fig. 4A. The spectrum is characterized by sharp but narrowly dispersed peaks along the $^1$H dimension (within 7 – 8.5 ppm), which is consistent with CFP-10 being unstructured in its native form. CFP-10 by itself is unstructured as reported earlier and the spectra show no change in the presence of unlabeled Rv3868 both in the presence and absence of ATP (Fig.2.36). The experiments clearly rule out any interactions of CFP-10 and therefore the C-terminal recognition motif with Rv3868.

2.3.17 Rv3868 does not exhibit chaperone-like activity

Possible chaperone-like activities were also probed and ruled out using substrates like hen Egg white lysozyme and porcine citrate synthase where the possible disaggregation of the substrates in the presence of Rv3868 was monitored spectroscopically. To address these questions and test whether Rv3868 has a general chaperone activity a non-native substrate was used. A classical assay for general chaperone function involves following thermal aggregation of citrate synthase. In this assay aggregation is triggered by incubation at 45°C and the extent of aggregation is measured by light scattering. Over time aggregated particles grow in size and number, causing increased light scattering. The presence of a general molecular chaperone is expected to prevent thermal aggregation and thus to decrease light scattering. During incubation at 45°C thermal aggregation of citrate synthase was observed (Fig.2.37A). Aggregation was not inhibited with increasing concentrations of Rv3868 and in presence of ATP also. Similar observation was observed in DTT induced lysozyme
aggregation. In this context aggregation was also not inhibited with increasing concentrations of Rv3868 (Fig. 2.37B).

In another set of experiments the presence of hydrophobic patches on the surface of the protein was probed using ANS or BIS-ANS binding studies. It is known that substrate polypeptides bind to large hydrophobic patches on the substrate binding domains in related AAA-ATPases like HslU, ClpA and Hsp with chaperone/Protease like activities (Fu et al., 2005; Ramachandran et al., 2002). Since Rv3868 contains two domains, the N-terminal domain is expected to have hydrophobic patches to bind to substrates if it had a chaperone function. The C-terminal domain on the other hand will expectedly bind to ANS at least because of the ATP-binding site.

For Rv3868 and Ct-Rv3868, $K_d$ value for ANS binding is $27\mu M$ and $25\mu M$ respectively. The binding data reveal that the N-terminal domain has no hydrophobic patches and is compact while the full length and ATP binding domains have similar affinity for ANS (Fig. 2.38). Altogether these studies suggest that the enzyme is not likely to have a chaperone like function.
Figure 2.36: HSQC spectra of CFP-10

(A) $^{15}$N-1H HSQC spectra of $^{15}$N-labeled CFP-10 in Free State. (B) $^{15}$N-1H HSQC spectra of $^{15}$N-labeled CFP-10 in presence of unlabeled Rv3868. (C) $^{15}$N-labeled CFP-10+Rv3868+ATP.
Figure 2.37: Assessment of chaperoning ability of Rv3868.

(A) Aggregation of citrate synthase as a function of time: aggregation of citrate synthase at 43°C in the absence (○) and the presence of equimolar ratios of *M. tuberculosis* Rv3868 (□); and in presence of ATP (▲), measured as a function of light scattered at 465 nm. Increase in molar excess ratio of Rv3868 to two fold or 14-fold resulted in no suppression of aggregation of citrate synthase.

(B) Possible chaperone functions were probed by following the thermal aggregation of lysozyme in the presence and absence of Rv3868. The thermal aggregation of lysozyme is not inhibited by Rv3868 in a concentration-dependent way. The aggregation of lysozyme was followed by light scattering at 360 nm in a Perkin-Elmer lambda 25 spectrophotometer with a Peltier system. (A) Temperature (45°C) induced aggregation of lysozyme in the absence (▲) and at increasing molar ratios of lysozyme: Rv3868 of 1:5 (▼); 1:10 (▲), and 1:10+ATP (○) respectively were followed. In both experiments control involved following the scattering by Rv3868 (♦) alone.
Figure 2.38: Study of ANS binding to Rv3868 and its variants by fluorescence spectroscopy

The fluorescence intensity of ANS is enhanced upon binding to Rv3868. Its fluorescence is also observed to increase significantly in the presence of both Rv3868 (A) and CT-Rv3868 (B). The direction of the arrow represents the enhancement seen in the presence of ANS (C). The relative change (%) in fluorescence intensity at 470 nm plotted against the ANS concentration in experiments involving Rv3868 (▲) and CT-Rv3868 (●) respectively. (D) No enhancement is observed in fluorescence intensity of bound ANS (460nm) in NT-Rv3868.
2.4 Conclusion

The characterization of the hypothetical protein Rv3868 reveals that it is a novel member of the CbbX family of proteins which contain an AAA-ATPase domain. The protein was shown to be composed of two domains. The C-terminal ATP-binding domain is responsible for oligomerisation and contains the AAA family signature motifs. The ATPase activity of the domain in isolation is several folds higher than that of the full-length protein. The N-terminal contains a helical domain with no homology to other known proteins. The reduced ATPase activity of the full-length enzyme coupled with the proximity of the domain to the ATPase binding domain suggests that it has a role in modulating the activity. It is highly plausible that the activity is stimulated by interactions of the domain with other specific protein partners of the ESX-1 system; although the ATPase activity in the presence of casein or DNA did not show any increase in our experiments. An analogous situation has been noted in some AAA-ATPases earlier where interactions with specific substrates/partners stimulate the respective functions.

The N-terminal domain might also have a role in modulating the oligomerization observed in the isolated ATPase domain alone. The latter domain, in the presence of nucleotide, has the tendency to form very high oligomeric structures while the full-length protein is predominantly a hexamer. It is tempting to surmise that the proximity of the N-terminal to the ATP-binding domain modulates the oligomeric association by blocking interactions of specific regions of the latter domain involved in oligomerization.

The \textit{in silico} modeling, docking calculations and mutational analysis has helped rationalize the observed activities and also the affinity of the protein for different nucleotides. An exciting outcome of these studies is the identification of Arg-429 as a potential 'sensor arginine' (Ogura \textit{et al.}, 2004). This residue is known to play a special role by transducing the ATP hydrolysis/binding event into a mechanochemical outcome in AAA-ATPases. It comes close to the binding site of the neighboring subunit in the oligomer to form a part of the binding site. This residue is known to play a special role by transducing the ATP hydrolysis/binding event into a mechanochemical outcome in AAA-ATPases. However the catalytic functions in the
respective proteins are known to be different and they play a context-specific role in the ATPases.

Other groups have identified at least four substrates of the ESX-1 system viz. ESAT-6 CFP-10, EspA and EspB and it is known that disruption of the Rv3868 gene prevents secretion of the substrates although their expression is not impaired (Brodin et al., 2006; Mahairas et al., 1996). This has led to a suggestion that the protein either affects the translocation or stability of the exported substrates. The lack of a general chaperone activity in Rv3868 as also interactions with the unstructured CFP-10 substrate suggests that it probably has a role in the translocation of the substrates rather than their stability. This then brings us to the question as to which are the likely interacting partners of Rv3868? One possibility, based on earlier work (Brodin et al., 2006; Teutschbeina et al., 2006) against the backdrop of the current characterization could be Rv3873 a gating protein. The interactions of Rv3868 with the gating protein would specifically modulate the secretion of the virulence factors in agreement with the essential role of Rv3868 in secretion but would not affect the expression of these factors. However, more work is necessary to identify and characterize the interactions of Rv3868 with its interacting partners Many AAA-ATPases associate as an oligomer. The ATP binding site itself occurs between the subunits and the ATPase activity is associated with a specific conformational change which is translated to other parts of the protein to perform the respective functions. The characterization presents direct evidence that the two domains of Rv3868 are in close proximity in the presence of ATP and open up in its absence. This suggests that the ATPase action leads to a relative 'open-close' movement of the N- and C-terminal domains. The N-terminal domain can then transmit the ATPase activity induced conformational change to the other suggested interacting proteins downstream in the export pathway. The proposed mechanism and activity of Rv3868 substantiated by the present work is summarized in Figure 2.39.

This is the detailed functional and structural characterization of the hypothetical protein, Rv3868. A co-factor induced 'open-close' mechanism was identified through which the protein can interact with other proteins of the ESX-1 secretory pathway. The enzyme characterization, in-silico modeling, docking and mutational studies pave the way for utilizing rational structure based approaches to identify inhibitors with the potential of specifically targeting the critical ESX-1 pathway.
Figure 2.39: A mechanistic model for structure and function of Rv3868 and its variants

*(Top)* The addition of ATP increases the proximity of the N- and C-terminal domains. This open to close movement and vice-versa can conceivably give rise to functional interactions with other interacting protein partners downstream in the ESX-1 pathway.

*(Bottom)* Addition of salt disrupts the oligomeric interactions supporting that the inter-subunit association is stabilized by ionic interactions. Limited proteolysis breaks the protein into two domains. The N-terminal domain does not associate in the presence of nucleotide while the C-terminal domain exhibits ATP-dependent self association suggesting that it is primarily the oligomerization domain.
References:


Chapter 2

CFP-10 Form a Tight, 1:1 Complex and Characterization of the Structural Properties of ESAT-6, CFP-10, and the ESAT-6\(\cdot\)CFP-10 Complex. IMPLICATIONS FOR PATHOGENESIS AND VIRULENCE. *Journal of Biological Chemistry* **277**: 21598-21603.


