Chapter 3: Biophysical and Biochemical features of *Mtb*-CRP

A part of the work presented in this Chapter has been published as:

**Available online at www.sciencedirect.com**


****

Novel biochemical properties of a CRP/FNR family transcription factor from *Mycobacterium tuberculosis*

Yusuf Akhtar\(^a\), Smanla Tundup\(^b\), Seyed E. Hasnain\(^{a,b,c,d,*}\)

\(^{a}\)Laboratory of Molecular and Cellular Biology, CDRI, Hyderabad 500076, India  
\(^{b}\)Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore 560064, India  
\(^{c}\)University of Hyderabad, Hyderabad 500046, India  
\(^{d}\)Institute of Life Sciences, Hyderabad 500046, India
3.1 Introduction

CRP/FNR (cAMP receptor protein/fumarate and nitrate reductase regulator) is one of the members of a family of transcriptional regulators. With over 370 family members, these DNA-binding proteins predominantly function as positive transcriptional regulators and are known to be associated with defense against oxygen stress and starvation, and at the same time respond to other environmental signals. The distinctive features of CRP/FNR superfamily proteins include the presence of a nucleotide-binding domain and a helix-turn-helix motif containing DNA-binding domain at the N- and C-terminal, respectively. The prototype cAMP-binding domain (Schultz et al., 1991) is a versatile structure that has evolved to accommodate different functional specificities in response to a broad range of signals (Green et al., 2001; Körner et al., 2003). *Mtb* H37Rv ORF Rv3676 codes for a putative CRP/FNR protein (Cole et al., 1998), which is required for virulence in mice and controls transcription of specific genes (Rickman et al., 2005). Recently, Bai et al. reported the characterization of *Mtb*-CRP (Rv3676) protein using computational and experimental methods (Bai et al., 2005). Phylogenetically, *Mtb* Rv3676 is nearest to the CooA branch represented by the CO sensor protein of *Rhodospirillum rubrum* (Körner et al., 2003). However, the relative positions of the regulatory and the DNA-binding domain are strikingly different in that the recognition helix of CooA is
rotated 180° away from the position occupied in CRP-cAMP. Further differences between CooA and CRP include an extended N-terminus providing a ligand to the heme of the opposite subunit, an 11-amino-acid extension in the regulatory domain (positions 72–82) to accommodate the heme and a different composition of the hinge region toward the C-terminus, which causes the displacement of the DNA-binding domain (Lanzilotta et al., 2000). Sequence alignment suggests that these 11 residues are not fully conserved in Rv3676.

*Escherichia coli* CRP and FNR regulate transcription globally in response to glucose starvation and anaerobic conditions, respectively (Kolb et al., 1993). *E. coli* FNR is structurally related to CRP except for the presence of four conserved cysteine residues at the N-terminal extension, which form part of an iron–sulfur cluster and a redox-sensing domain of FNR. This iron–sulfur cluster is absent in Rv3676 similar to other members of the same family from other systems such as *Pseudomonas stutzeri* (Vollack et al., 1999) and *Bradyrhizobium japonicum* (Mesa et al., 2003). Although these proteins do not have an iron–sulfur cluster, they are the regulators of oxygen tension *sensu stricto*. The earlier report by Bai et al. (2005) focused on CRP regulon prediction and the experimental validation of the same and provided the first direct evidence for cAMP binding to a transcription factor in *Mtb*, thereby suggesting a role for cAMP-mediated signal transduction in this
bacterium. In this Chapter, the purification and comprehensive characterization of a CRP/FNR regulator from *Mtb* in terms of oligomeric state, cAMP and DNA binding has been described. These results point to some new unusual properties of Rv3676 protein, which could have physiological relevance.

### 3.2 Materials and Methods

#### 3.2.1 Bacterial strains and plasmids

*E. coli* DH5α and *E. coli* BL21DE3 bacterial strains were used for cloning and expression purposes, respectively. DNA manipulations were carried out in pET23a plasmid vector using standard techniques. Integrity of the plasmid constructs was confirmed by DNA sequencing.

#### 3.2.2 Cloning, expression and purification of recombinant *Mtb* Rv3676

*Mtb* ORF Rv3676 was PCR amplified from *Mtb* H37Rv genomic DNA using forward (GGATAT**CATATG**GTGGACGAGATCCTGGCCAGGG) and reverse (CG**CTCGAG**CCTCGCTCGGCGGGCCAGTC) primers with restriction enzyme sites for cloning (shown in bold). The amplicon was cloned into the corresponding sites of pET23a, and recombinant Rv3676
protein was purified as a 6XHis-tagged fusion protein from *E. coli* BL21 (DE3) cells.

![Image](image_url)

**Figure 3.1:** Cloning and expression of *Rv3676* (*Mtb*-CRP). (a) Identification of the positive clone of *Rv3676* in pET23a vector by restriction digestion with *Nde*I and *Xho*I. Lane 1, 1Kb marker; lane 2, lane 3 and lane 4 shows release of *Rv3676* insert cloned in pET23a after. (b) Integrity of positive clones was verified by sequencing and the plasmid carrying the insert was transformed into BL21 (DE3) cells and cultures were grown in 200 ml LB supplemented with 30 µg/ml kanamycin. The culture was induced at an OD600 of 0.4 with 0.4 mM IPTG at 300 K and 200 rev per min to allow protein expression. Protein expression was monitored by SDS–PAGE. Lanes marked “un-induced” and “induced” show expression of *Rv3676* without and with IPTG. Lane marked “M” shows protein molecular size markers.

Protein concentration was estimated using the dye-binding method (Bradford, 1976). To determine suitable storage conditions, aliquots of recombinant *Rv3676* (r*Rv3676*) were dialyzed in different buffers, namely phosphate-buffered saline (PBS), 10 mM Tris and 10 mM HEPES. Storage temperature was also optimized, and the conditions under which r*Rv3676* was most stable were selected.
3.2.3 Analytical size exclusion chromatography

The oligomeric state of native recombinant protein was determined by analytical size exclusion chromatography using a Superose 6 fast protein liquid chromatographic column (BIORAD) at room temperature with PBS as running buffer. A standard curve was prepared according to the instruction manual using standard molecular weight markers. The void volume was determined using Blue Dextran 2000. The elution parameter $K_{av}$ was calculated as follows: $K_{av} = (V_c - V_0)/V_s$, where $V_c$ is the elution volume for the protein, $V_0$ the column void volume, and $V_s$ the total stationary phase volume. $K_{av}$ was plotted against log molecular weight.

3.2.4 Spectral analyses

To detect the presence, if any, of any associated co-factor, absorption was measured between 200 and 800 nm using a Perkin-Elmer spectrophotometer. Fluorescence spectrometric measurements and ligand-binding assays were carried out using a Perkin-Elmer LS50B luminescence spectrometer and a sample volume of 200 ml with 0.3 cm path length. Tryptophan fluorescence was measured at an excitation wavelength of 295 nm. The slit widths for excitation and emission were 10 and 20 nm, respectively. Emission spectra were recorded between 310 and 500 nm. All spectra measurements were corrected by subtracting
the corresponding buffer backgrounds. Increasing concentrations of urea (1–8 M) and a constant concentration (3 µM) of recombinant protein was used to study the denaturation kinetics of the protein. At 8 M urea the protein was fully unfolded, and the spectrum of fully unfolded protein was further compared with that of 6 µM free tryptophan. The circular dichroism (CD) spectra of recombinant native protein and liganded protein, incubated with different concentrations of cAMP (6–16 µM), were recorded using a JASCO CD spectrometer (Model J-810) between 200 and 250 nm in steps of 0.5 nm with four accumulations in each step. The spectral baseline was corrected by subtracting the respective blanks. Molar ellipticity, expressed in millidegrees, was plotted as a function of wavelength. The secondary structure content of the protein was calculated by using k2d software (www.embl-heidelberg.de/~andrade/k2d/). For CD and fluorimetric spectral analysis, 5 and 3 mM recombinant protein was used, respectively.

3.2.5 Electrophoretic mobility shift assay (EMSA)

Gel retardation assays were carried out as described earlier (Prakash et al., 2005). Complementary synthetic oligodeoxyribonucleotides corresponding to the CRP/FNR-binding site (AATGTGATCTAGGTCACGTG) present upstream of Rv1552 (frdA) were end labeled with [g-32P]ATP using T4 polynucleotide kinase. One
nanogram labeled oligonucleotide was incubated with 3 mg recombinant protein in binding buffer (10 mM Tris–HCl, 50 mM NaCl, 50 mM MgCl2, 1 mg BSA, 1 mg poly dI:dC, 1 mM EDTA, 1 mM DTT, 1 mM PMSF and 10% glycerol) in 20 ml reaction volume, incubated for 30 min at room temperature and fractionated on a 5% polyacrylamide gel in TBE. After electrophoresis at 200 V at 4º C, the gel was dried and analyzed by autoradiography. To check for the specificity of the complex, unlabeled homologous oligonucleotide or an oligonucleotide carrying a specific mutation (mut) critical for binding (AATTTGATCTAGGTCACGTG, shown as underlined) was used in competition assays.

3.3 Results

3.3.1 Purified rRv3676 exists in dimeric state

*Mtb* rRv3676 was purified as a 6XHis-tagged protein using affinity chromatography as described in Materials and Methods. Purified rRv3676 was stable at 4º C in PBS while at lower temperatures and in other buffers it formed insoluble aggregates.
Figure 3.2: Recombinant Rv3676 protein exists as a dimer: (a) Coomassie-stained polyacrylamide gel showing the ORF Rv3676-encoded protein of *Mycobacterium tuberculosis* purified from *E. coli*. M represents the protein molecular size marker (Medium range, Genei, India.), E1–E6 show successive TALON column eluted fractions of the recombinant protein. (b) The purified protein was pooled and fractionated on a Superose 6 FPLC column, resulting in a single peak. The calculated molecular mass of the recombinant protein was ~53 kDa corresponding to a dimeric state.

The purified recombinant protein was fractionated by electrophoresis on a 10% polyacrylamide gel and stained with Coomassie Brilliant Blue (Figure 3.2a, inset). Gel filtration analysis was carried out to determine the oligomeric nature, if any, of the rRv3676 protein. rRv3676 exists as a pure dimer of ~53 kDa, as evident from analytical size exclusion chromatography (Figure 3.2).
3.3.2 Purified rRv3676 has no associated cation co-factors

In most oxygen tension-sensing proteins belonging to the same family, transition metals like Fe or Ni are associated with the protein to sense the fluctuations of oxygen availability via redox mechanisms (Körner et al., 2003).

Therefore, the absorption spectrum of purified rRv3676 was scanned to check for the presence, if any, of a metal ion co-factor. Spectral analysis revealed two peaks, one at 295 nm (corresponds to tryptophan) and the other at 280 nm (corresponds to tyrosine and phenylalanine). The fact that no other peak was observed clearly indicates that rRv3676 does not have any other
associated metal ion co-factor. This rather unexpected finding suggests that *Mtb* Rv3676 apparently uses some other mechanism(s) to sense effector signals.

### 3.3.3 cAMP binds to purified rRv3676 in a concentration dependent manner

Results of protein family searches revealed the presence of a putative cAMP-binding domain at the N-terminal end of rRv3676 protein, thereby raising a strong probability that cAMP may be acting as an effector of Rv3676 protein. Therefore purified rRv3676 was subjected to CD analysis in the presence and absence of cAMP as ligand.

![Figure 3.4: Purified rRv3676 protein displays cAMP-binding activity as evident from circular dichroism (CD) spectral analysis (see Materials and Methods). Binding of cAMP to rRv3676 is evident from a change in secondary structure of the native protein upon interaction with cAMP.](http://www.bork.embl-heidelberg.de/andrade/k2d)

The change in secondary structure was calculated using k2d software (http://www.bork.embl-heidelberg.de/andrade/k2d) using a
method developed earlier (Yang et al., 1986). A comparison of CD spectra of cAMP-free and cAMP-bound rRv3676 provides evidence of binding (change in secondary structure of purified rRv3676 upon interaction with cAMP).

![Figure 3.5: Denaturation of recombinant Rv3676 in the presence of urea. Recombinant Rv3676 is completely denatured in the presence of 8 M urea as evident from maximum fluorescence.](image)

This change in secondary structure clearly appears to be a function of increasing concentration of cAMP (Figure 3.4). That cAMP indeed causes concentration-dependent conformational alterations within rRv3676 was further confirmed by tryptophan fluorescence spectrometry.
Figure 3.6: Comparison of relative fluorescence intensities contributed by free tryptophan residues and tryptophan residues present in rRv3676 protein. Fluorescence intensities obtained from 6µM of free tryptophan and 3µM rRv3676 (native and unfolded) were compared. Each molecule of rRv3676 contains two tryptophan residues. These comparisons show that at 8 M urea protein is completely unfolded and both the tryptophan residues are fully exposed to the solvent and contribute to equivalent amount of fluorescence intensity as free tryptophan.

The two tryptophan residues (Trp112 and Trp203) present in Rv3676 protein were used as probe to study conformational changes in the protein in solution upon urea-induced denaturation. Purified rRv3676 unfolds completely in the presence of 8 M urea without any further increase in fluorescence (Figure 3.5), indicating the presence of fully unfolded protein molecules. The maximum wavelength of absorbance of denatured rRv3676 is approximately 360 nm, which is equal to the maximum wavelength of absorbance of 6 mM free tryptophan (Figure 3.6). As protein unfolds (relaxed) tryptophan residues are exposed to the solvent, resulting in an increase in relative
fluorescence. Therefore, fluorescence method was used to assay whether increasing concentrations of cAMP have any effect on the unfolding of rRv3676 protein.

PBS was used as solvent, which has a physiological pH and an ionic strength similar to the intracellular milieu of the bacilli. Physiological cAMP levels are in the range of 0–10 mM. At lower concentrations (6–10 mM) the binding shows positive cooperativity, and at 10 mM cAMP the protein is in the most open conformation. This is evident from the increase in tryptophan fluorescence (Figure 3.7). With further increase of cAMP (12–16 mM), the relative tryptophan

![Figure 3.7: Fluorescence emission spectra of rRv3676 as a function of cAMP concentration (6–16 mM). The fluorescence maximum of the protein increases steadily up to 10 mM cAMP and later drops as a function of increasing cAMP concentration (12–16 mM).]
fluorescence decreases, suggesting that the protein is getting compacted. This protein compaction could be a reflection of a feedback regulation.

3.3.4 Purified rRv3676 binds *in vitro* to the CRP/FNR cognate nucleotide sequence motif present upstream of *Rv1552*

Having shown that Rv3676 is a likely member of the CRP/FNR family of DNA-binding proteins, purified rRv3676 was tested whether it indeed displays such an activity.

<table>
<thead>
<tr>
<th>Purified rRv3676</th>
<th>-</th>
<th>+</th>
<th>+</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competitor DNA (50X)</td>
<td>-</td>
<td>-</td>
<td><em>Rv1552</em></td>
<td>mut</td>
</tr>
<tr>
<td>Probe <em>Rv1552</em></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

*Figure 3.8: Recombinant Rv3676 binds to the CRP/FNR-binding element present upstream of *Rv1552* (*frdA*). Specificity of binding was confirmed by homologous competition with a 50-fold excess of unlabeled ligand (lane 3). The complex is unaffected when a mutant oligonucleotide (mut) carrying an alteration in the binding sequence is used in the competition assay (lane 4).*

*Mtb* ORF *Rv1552* was selected, which is putatively regulated by the CRP/FNR family of transcriptional regulators. Synthetic oligodeoxyribonucleotide corresponding to the CRP/FNR cognate DNA-binding element present in ORF *Rv1552* (*frdA*) was radiolabeled and used
as probe in EMSA using purified rRv3676. The results clearly show a shift in the mobility of the CRP/FNR probe upon incubation with rRv3676 protein (Figure 3.8, lane 2). The specificity of the binding is evident from the disappearance of the complex in a competition assay using a 50-fold excess of homologous unlabeled CRP/FNR oligonucleotides (Figure 3.8, lane 3). The specificity of this DNA–protein interaction is further evident from the absence of any competition when a mutated version of the oligonucleotide (mut in Figure 3.8, lane 4) is used. These results demonstrate that rRv3676 indeed specifically interacts with the CRP/ FNR cognate DNA sequence motifs.

3.4 Discussion

*Mtb* harbors a single member of the CRP/FNR superfamily, i.e. Rv3676. That this gene is important is evident from knockout studies. An Rv3676 knock-out strain is impaired in growth under *in vitro* conditions, in bone marrow-derived macrophages and also in an animal model (Rickman *et al.*, 2005). ORF Rv3676 was therefore selected for further analyses of its protein product. Purified rRv3676 exists in a single oligomeric state as a homodimer (Figure 3.2) and is active in terms of DNA binding. It is interesting to note that, despite an only weak DNA-binding activity, the interaction is very specific as could be seen from the inability of the mutant oligonucleotide to compete for binding. Most other
proteins of this family are active as dimers (Körner et al., 2003) but, unlike Rv3676, contain metal cations such as iron and nickel as co-factors. Interestingly, Rv3676 does not carry a metal-binding motif, and the absence of any metal co-factor is indeed evident from the spectral features of rRv3676. *Mtb* Rv3676 thus appears to be different from other oxygen-sensing proteins in terms of non-requirement of a metal cationic co-factor. To investigate the ability of rRv3676 to bind to its cognate DNA motif, EMSA was carried out using purified rRv3676 and a radiolabeled oligonucleotide carrying the CRP/FNR-binding site present upstream of the *frd* (*Rv1552*) gene encoding the fumarate reductase enzyme. This binding site was identified as a putative binding site in recent reports (Bai et al., 2005; Spreadbury et al., 2005). In *in silico* regulon prediction studies, this motif elicited the highest score (Chapter 2; Akhter et al., 2008), and therefore was selected for EMSA. It has been reported that Rv3676 senses oxygen (Bai et al., 2005; Spreadbury et al., 2005) indirectly by controlling the expression of genes such as *frd*. Fumarate serves as an alternative electron acceptor in the absence of oxygen, and this is mediated by a membrane-linked fumarate reductase enzyme complex (Lambden and Guest, 1976). The putative CRP/FNR binding site, present upstream of the *frd* operon, was recognized by rRv3676 protein as evident from EMSA.
The *Mtb* genome encodes as many as 15 adenylate cyclases, suggesting that cAMP may have an important role in mycobacteria. It has indeed been reported that cAMP can alter the gene expression profile of *Mtb* during anaerobic conditions (Gazdik and McDonough, 2005). The predicted cAMP-binding site in rRv3676 indeed shows binding to cAMP leading to conformational changes in the protein as evident from spectral analyses. The extent of change in secondary structure is maximal in the presence of 10 mM cAMP. The effect of cAMP binding on the DNA-binding efficiency of Rv3676 has already been reported earlier (Bai et al., 2005). cAMP, acting as effector, is known to modulate the regulation of a large number of target genes, and it is likely that Rv3676 is involved in this process. While these *in vitro* findings point to the importance of cAMP, it remains to be experimentally demonstrated whether cAMP is actually involved in regulating gene expression by recruiting Rv3676 protein. While the biophysical features of purified *Mtb* Rv3676 protein described here are physiologically relevant, experimental validation *in vivo* will be required to dissect the complete network of *Mtb* genes regulated by Rv3676 and cAMP.