Acta Crystallographica Section F
Structural Biology Communications
ISSN 2053-230X

Structure of the carboxy-terminal domain of *Mycobacterium tuberculosis* CarD protein: an essential rRNA transcriptional regulator

Shanti P. Gangwar, Sita R. Meena and Ajay K. Saxena*

Structural Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110 067, India

Correspondence e-mail: ajaysaxena@mail.jnu.ac.in

Received 15 October 2013
Accepted 21 December 2013

PDB reference: CarD$_{83-161}$, 4kmc

The CarD protein is highly expressed in mycobacterial strains under basal conditions and is transcriptionally induced during multiple types of genotoxic stress and starvation. The CarD protein binds the β subunit of RNA polymerase and influences gene expression. The disruption of interactions between CarD and the β subunit of RNA polymerase has a significant effect on mycobacterial survival, resistance to stress and pathogenesis. To understand the structure of CarD and its interaction with the β subunit of RNA polymerase, *Mycobacterium tuberculosis* CarD (*Mtb*CarD) and the *Thermus aquaticus* RNA polymerase β subunit were recombinantly expressed and purified. Secondary-structure analysis using circular-dichroism spectroscopy indicated that *Mtb*CarD contains ~60% α-helix, ~7% β-sheet and ~33% random-coil structure. The C-terminal domain of *Mtb*CarD (CarD$_{83-161}$) was crystallized and its X-ray structure was determined at 2.1 Å resolution. CarD$_{83-161}$ forms a distorted Y-shaped structure containing bundles of three helices connected by a loop. The residues forming the distorted Y-shaped structure are highly conserved in CarD sequences from other mycobacterial species. Comparison of the CarD$_{83-161}$ structure with the recently determined full-length *M. tuberculosis* and *T. thermophilus* CarD crystal structures revealed structural differences in residues 141–161 of the C-terminal domain of the CarD$_{83-161}$ structure. The structural changes in the CarD$_{83-161}$ structure occurred owing to proteolysis and crystallization artifacts.

1. Introduction

According to World Health Organization statistics, 30% of the world population is infected by *Mycobacterium tuberculosis* (*Mtb*) and it causes 1.4 million deaths per year (World Health Organization, 2011). There is an essential requirement to develop drugs and vaccines against multidrug-resistant, extensively drug-resistant and totally drug-resistant strains of mycobacteria. During a study of the stress response of mycobacteria, it was found that the *carD* gene was transcriptionally induced under multiple types of stress (Stallings *et al.*, 2009). On transient knockout of the CarD gene in mycobacteria, the cells become sensitive to killing by starvation, reactive oxygen species and ciprofloxacin stresses (Stallings *et al.*, 2009). *M. tuberculosis* bacilli with a depleted *carD* gene did not replicate and persist in mice, and a global change in the transcriptional profiles of mycobacteria was observed in a microanalysis experiment (Stallings *et al.*, 2009).

Mycobacterial CarD protein is required for the efficient control of rRNA transcription by (p)ppGpp (Stallings *et al.*, 2009). CarD is a functional homologue of DksA and interacts directly with the RNA polymerase (RNAP) β subunit (Stallings *et al.*, 2009). A study of mycobacterial strains with attenuated CarD interactions with the RNAP β subunit indicated that this interaction is required for resistance to oxidative stress and for viability and is not required for fluoroquinolone resistance (Weiss *et al.*, 2012). A complete phenotype was not observed on weakening the interaction of *Mtb*CarD with the *Mtb*RNAP β subunit (Weiss *et al.*, 2012).

*Mtb*CarD contains 162 residues with a molecular mass of ~17.9 kDa. The N-terminal region of *Mtb*CarD contains a TRCF-RID module and interacts with the N-terminus of the *Mtb*RNAP β subunit. Immunoprecipitation of CarD from *M. smegmatis* cell lysate indicated that the α, β and β' domains of the mycobacterial RNA

polymerase bind directly to CarD. The C-terminal region of *Mtb*CarD contains a leucine zipper-like fold and is essential for mycobacterial viability, as shown in a gene-deletion experiment (Stallings *et al.*, 2009). Recently, the crystal structure of full-length *M. tuberculosis* CarD in complex with RNAP β lobes (Gulten & Sacchettini, 2013; PDB entry 4kbm) and the crystal structure of *Thermus thermophilus* CarD (Srivastava *et al.*, 2013; PDB entry 4l5g) have been determined.

Mycobacterial CarD binds directly to RNA polymerase and influences gene expression (Stallings *et al.*, 2009). To understand the *Mtb*CarD structure and its interaction with the *T. aquaticus* RNAP (*Taq*RNAP) β subunit, we expressed and purified the *Mtb*CarD and *Taq*RNAP β subunit proteins. We performed CD spectroscopy to identify the secondary-structure content of *Mtb*CarD. The crystal structure of the $CarD_{83–161}$ domain was analyzed and compared with recently determined X-ray structures of *M. tuberculosis* and *T. thermophilus* CarD proteins.

2. Materials and methods

2.1. *Mtb*CarD expression and purification

The gene encoding residues Met1–Ser162 of CarD (accession No. NP_218100) was amplified using genomic DNA of *M. tuberculosis* H37Rv strain by polymerase chain reaction (PCR) and inserted into pET-28a(+) expression vector (Novagen). The forward primer 5′-GATCCATATGATGATTTTCAAGGTCGGA-3′ and the reverse primer 5′-CATGAAGCTTTCAAGACGCGGCGGCTAA-3′ were used in the polymerase chain reaction. The amplified PCR product was digested and ligated into pET-28a(+) expression vector using *Nde*I and *Hin*dIII restriction sites. The CarD plasmid was transformed into *Escherichia coli* BL21(DE3) cells for protein expression.

The *E. coli* cells were grown in 2 l Luria–Bertani medium supplemented with 50 µg ml^{-1} kanamycin at 37°C until the OD_{600} reached 0.6–0.7. The culture was induced with 250 µ*M* IPTG at 37°C and grown for a further 3 h. The cells were harvested by centrifugation at 10 000*g* and resuspended in 100 ml lysis buffer consisting of 20 m*M* HEPES pH 7.0, 100 m*M* NaCl, 1 m*M* benzamidine–HCl, 0.1% Triton X-100, 5% glycerol, 1 m*M* phenylmethylsulfonyl fluoride, 0.2 mg ml^{-1} lysozyme. The cells were kept on ice for a further 1 h and disrupted by sonication. The cell lysate was centrifuged at 16 000*g* for 20 min and the supernatant was collected.

For Ni–NTA affinity chromatography, the supernatant was mixed with pre-equilibrated Ni–NTA resin and incubated for 2 h at 4°C using a 360° rocker. The resin was washed with buffer consisting of 20 m*M* HEPES pH 7.0, 500 m*M* NaCl, 1 m*M* benzamidine–HCl, 5% glycerol, 1 m*M* phenylmethylsulfonyl fluoride, 40 m*M* imidazole. The *Mtb*CarD protein was eluted with buffer consisting of 20 m*M* HEPES pH 7.0, 150 m*M* NaCl, 1 m*M* benzamidine–HCl, 5% glycerol, 1 m*M* phenylmethylsulfonyl fluoride, 250 m*M* imidazole. The eluted *Mtb*CarD fractions were pooled, concentrated and loaded onto a gel-filtration column (HiLoad 16/60 Superdex 75 pg, GE Healthcare). The column was pre-equilibrated with buffer consisting of 20 m*M* HEPES pH 7.0, 150 m*M* NaCl. The peak fractions were pooled and concentrated to 10–12 mg ml^{-1} using a centrifugal filter device (Millipore, USA). The protein was stored at 4°C. The final recombinant *Mtb*CarD contained 183 residues with a molecular mass of 20.2 kDa [an extra 21 residues at the N-terminus (MGSSHHHHHHSSGLVPRGSHM containing a thrombin cleavage site) and 162 residues of CarD]. The purity of the recombinant *Mtb*CarD was analyzed by SDS–PAGE and mass spectrometry. The protein concentration was estimated using the Bradford technique (Bradford, 1976).

2.2. *Taq*RNAP β subunit expression and purification

The pET-21c plasmid containing the *Taq*RNAP β subunit gene was transformed into *E. coli* BL21(DE3) cells for protein expression. A single colony was inoculated in 5 l Luria–Bertani medium supplemented with 100 µg ml^{-1} ampicillin. The culture was grown at 37°C until the OD_{600} reached 0.6–0.7. The culture was induced with 1 m*M* IPTG at 37°C and grown for a further 3 h. The cells were harvested by centrifugation at 10 000*g* and the pellet was resuspended in 100 ml lysis buffer consisting of 20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 1 m*M* benzamidine–HCl, 0.1% Triton X-100, 10% glycerol, 1 m*M* phenylmethylsulfonyl fluoride, 0.5 mg ml^{-1} lysozyme. The cells were kept on ice for a further 1 h and disrupted by sonication. The cell lysate was centrifuged at 16 000*g* for 20 min and the supernatant was collected.

For Ni–NTA affinity chromatography, the supernatant was mixed with pre-equilibrated Ni–NTA resin and incubated for 2 h at 4°C using a 360° rocker. The mixture was loaded into an empty column and the resin was washed with buffer consisting of 20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 1 m*M* benzamidine–HCl, 10% glycerol, 1 m*M* phenylmethylsulfonyl fluoride, 45 m*M* imidazole. The protein was eluted in buffer consisting of 20 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl, 1 m*M* benzamidine–HCl, 10% glycerol, 1 m*M* phenylmethylsulfonyl fluoride, 250 m*M* imidazole. The eluted protein fractions were pooled, concentrated and loaded onto a gel-filtration column (HiLoad 16/60 Superdex 200 pg, GE Healthcare). The column was pre-equilibrated with buffer consisting of 20 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl, 10% glycerol, 5 m*M* β-mercaptoethanol. The peak fractions were pooled and concentrated to 5 mg ml^{-1} using a centrifugal filter device (Millipore, USA). The protein was stored at 4°C. The purity of the *Taq*RNAP β subunit was analyzed by SDS–PAGE and mass spectrometry. The protein concentration was estimated using the Bradford method (Bradford, 1976).

2.3. Circular-dichroism (CD) analysis

CD measurements were recorded using a Chirascan CD spectropolarimeter (Applied Photophysics) with a water bath to maintain a constant temperature. The *Mtb*CarD was diluted to 0.1 mg ml^{-1} in 10 m*M* Tris–HCl buffer pH 8.0 and loaded into a 0.1 cm quartz cuvette. The blank for all experiments was 10 m*M* Tris–HCl buffer pH 8.0. The final spectrum was the average of three sequential scans. The CD data were converted to mean residue ellipticities (in deg cm^2 dmol^{-1}). The *DichroWeb* server (Whitmore & Wallace, 2004) was used to estimate the amount of secondary structure from the CD spectra and estimated ~60% α-helix, ~7% β-sheet and ~33% random-coil structure.

2.4. Crystallization and heavy-atom derivatization

Selenomethionine-substituted *Mtb*CarD protein was prepared using the following protocol. Complete selenomethionine medium was prepared by mixing the base medium with nutrient mixture from Molecular Dimensions (MD 12-501). The base medium was autoclaved and filter-sterilized nutrient mixture was added to the base medium to prepare complete selenomethionine medium. 40 mg seleno-DL-methionine from Sigma–Aldrich was added to 1 l selenomethionine medium. The primary culture was grown overnight at 37°C and the obtained cell pellet was resuspended in complete selenomethionine medium. It was centrifuged again and resuspended in complete selenomethionine medium. This culture was used as a starting culture for large-scale purification of selenomethionine-derivatized *Mtb*CarD. The expression and purification procedure

Table 1
Intensity data-collection and refinement statistics for $CarD_{83–161}$.

Values in parentheses are for the highest resolution shell.

Data set	Native	Iodide-soaked
Space group	$P4_32_12$	$P4_32_12$
Temperature (K)	100	100
Wavelength (Å)	0.97625	1.5418
X-ray source	BM14, ESRF	AIRF, JNU
Resolution (Å)	50.00–2.15 (2.19–2.15)	60.93–2.75 (2.90–2.75)
Unit-cell parameters (Å)	$a = b = 60.95$, $c = 59.47$	$a = b = 60.93$, $c = 59.37$
Observed reflections	179423	53716
Unique reflections	6559 (334)	3111 (349)
Completeness (%)	99.9 (100)	96.8 (77.8)
R_{merge}† (%)	8.0 (50.3)	13.6 (48.4)
Average $I/\sigma(I)$	10.3 (7.9)	19.5 (6.3)
Multiplicity	27.4 (28.6)	17.3 (17.2)
Wilson B factor (Å^2)	43.81	
Molecules in asymmetric unit	1	1
V_M (Å^3 Da^{-1})	2.87	2.87
Solvent content (%)	57.2	57.2
Refinement		
Resolution (Å)	30.47–2.14	
R_{work}/R_{free}‡	0.21/0.25	
Protein atoms	616	
Water atoms	15	
R.m.s.d., bonds (Å)	0.007	
R.m.s.d., angles (°)	0.866	
Average B factor (Å^2)		
Protein	49.70	
Water	50.40	
Ramachandran plot (%)		
Favoured	100	
Allowed	0	
Generously allowed	0	
Forbidden	0	
PDB code	4kmc	

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith intensity measurement of reflection hkl and $\langle I(hkl)\rangle$ is the average intensity of that reflection. ‡ $R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$.

used for selenomethionine-derivatized *Mtb*CarD was similar to that described for native *Mtb*CarD.

Both native and selenomethionine-derivative *Mtb*CarD proteins were used for crystallization experiments. Initial crystallization conditions were obtained using Crystal Screen and Crystal Screen 2 from Hampton Research and the JCSG-*plus* screen from Molecular Dimensions. The sitting-drop vapour-diffusion technique was used in initial crystallization trials at 4°C. In each trial, 1 µl protein solution was mixed with 1 µl reservoir solution and equilibrated against 100 µl reservoir solution. Needle-shaped crystals of *Mtb*CarD and selenomethionine-derivatized *Mtb*CarD appeared after 4–5 d in several crystallization conditions. Further optimization of the crystallization conditions yielded the best diffracting crystals of *Mtb*CarD using a reservoir consisting of 0.2 *M* $MgSO_4$, 16%(*w*/*v*) PEG 4000, 30%(*v*/*v*) ethylene glycol.

Other heavy-atom derivatives of *Mtb*CarD were prepared in which native *Mtb*CarD crystals were soaked in reservoir solution containing the following heavy-atom compounds: (i) 500 m*M* NaI for 2 min, (ii) 10 m*M* K_2PtCl_4 for 24 h and (iii) 5 m*M* $HgCl_2$ for 24 h.

2.5. Intensity data collection and X-ray structure solution

For intensity data collection, single crystals of *Mtb*CarD were cooled directly in liquid nitrogen, as 30% ethylene glycol acts as a good cryoprotectant. The native CarD crystal diffracted to 2.1 Å resolution and an X-ray intensity data set was collected on the BM14 beamline at the European Synchrotron Radiation Facility, Grenoble, France. The selenomethionine-derivative *Mtb*CarD crystal diffracted to 6.0 Å resolution and a single anomalous diffraction data set at the Se peak was collected at 6.0 Å resolution (data not shown).

A single anomalous diffraction data set was collected from an NaI-soaked *Mtb*CarD crystal using Cu *Kα* radiation (λ = 1.5418 Å) at the X-ray diffraction facility (AIRF) of Jawaharlal Nehru University (JNU), New Delhi, India. Merging of raw data sets was performed using *iMosflm* (Battye *et al.*, 2011). The scaling and processing of data sets were performed using *CCP*4 (Winn *et al.*, 2011). The intensity data-collection and processing statistics are given in Table 1. Iodide sites were obtained using single anomalous diffraction data collected from an NaI-soaked crystal using *SHELXD* from the *HKL2MAP* suite (Pape & Schneider, 2004). Initial phases were obtained using the *AutoSol* program and the model was built using *AutoBuild* from the *PHENIX* suite (Adams *et al.*, 2010). The *AutoBuild* program traced only the C-terminal residues Thr83–Ala161 of *Mtb*CarD, and the electron density for the N-terminal residues 1–82 was missing. The *Mtb*CarD crystals were dissolved and the protein was analyzed by SDS–PAGE. The protein showed a single band of ~9 kDa (the molecular mass of full-length CarD is ~20.2 kDa). To identify the cleavage site in the crystallized *Mtb*CarD protein, we dissolved the

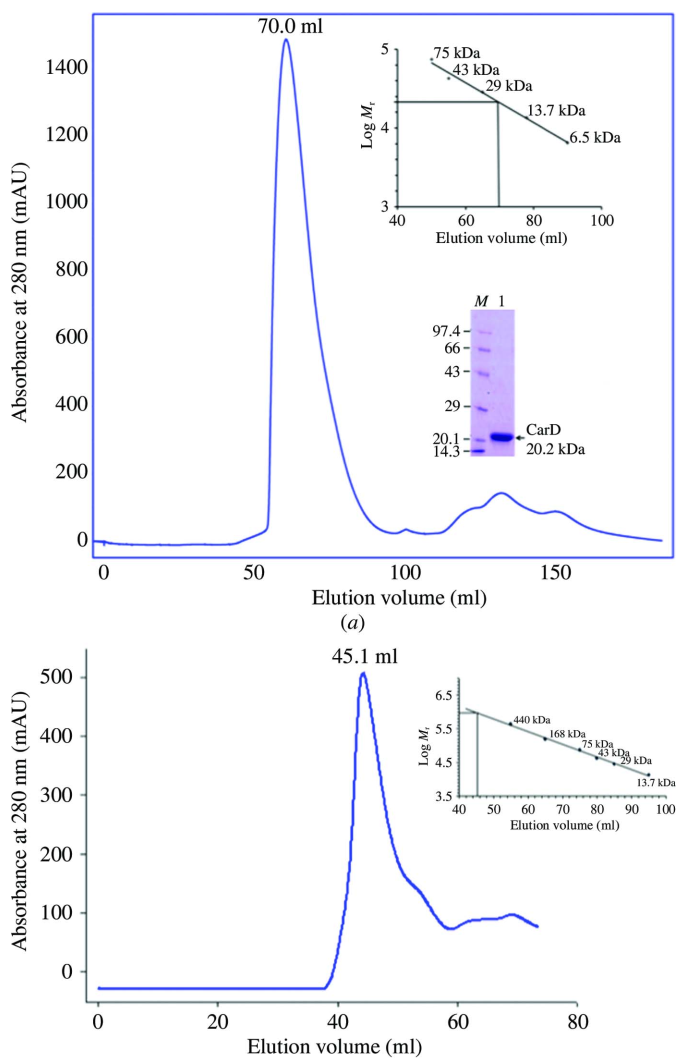

Figure 1
Size-exclusion chromatography of *Mtb*CarD and the *Taq*RNAP *β* subunit. (*a*) Size-exclusion chromatogram and SDS–PAGE of full-length *Mtb*CarD protein. *Mtb*CarD eluted as a monomer from a Superdex 75 (16/60) column. (*b*) Size-exclusion chromatogram of purified *Taq*RNAP *β* subunit. The *Taq*RNAP *β* subunit eluted as an aggregate from a Superdex 200 (16/60) column. The calculated molecular mass of each protein is denoted.

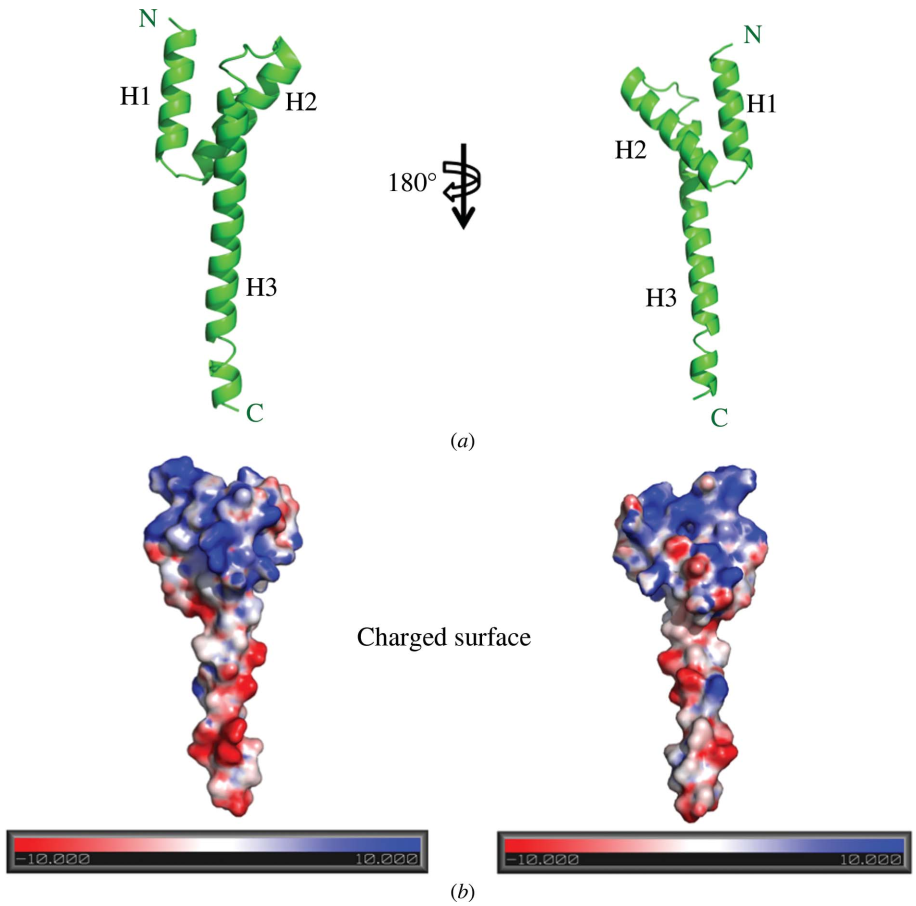

Figure 2

Crystal structure of the C-terminal domain of *Mtb*CarD ($CarD_{83–161}$). (*a*) $CarD_{83–161}$ adopts a distorted Y-shaped structure. Two views of a ribbon diagram of the $CarD_{83–161}$ crystal structure are shown (180° apart). (*b*) Two views of the molecular surface of $CarD_{83–161}$ coloured by electrostatic potential (180° apart). The surface potential ranges from $-10kT$ (red denotes negative charge) to $+10kT$ (blue denotes positive charge).

*Mtb*CarD crystals and performed N-terminal sequencing of the protein obtained from these crystals (Supplementary Table S1[1]). The sequencing result indicated the first seven residues to be APHTEEP, *e.g.* residues 76–82 of *Mtb*CarD. This indicates that residues 76–82 were disordered in our *Mtb*$CarD_{83–161}$ crystal structure and could not be traced. The $CarD_{83–161}$ structure was refined using *PHENIX* (Adams *et al.*, 2010) and the model was rebuilt using *Coot* (Emsley & Cowtan, 2004). Figures were generated by *PyMOL* (DeLano, 2002).

3. Results and discussion

3.1. *Mtb*CarD exists as a monomer in solution

Recombinant *Mtb*CarD (183 residues, molecular mass ∼20.2 kDa) and *Taq*RNAP β subunit (1119 residues, molecular mass ∼124.7 kDa) were expressed and purified as described in §2. The recombinant *Mtb*CarD eluted as a monomer from the size-exclusion column and showed a single peak on SDS–PAGE (Fig. 1). Single crystals of *Mtb*CarD were grown using the hanging-drop vapour-diffusion technique and diffracted to 2.1 Å resolution. However, the *Mtb*CarD protein degraded during crystallization and the X-ray intensity data yielded the structure of only the C-terminal residues Thr83–Ala161 of *Mtb*CarD. SDS–PAGE analysis of the protein obtained from the *Mtb*$CarD_{83–161}$ crystals indicated a molecular mass of ∼9 kDa. We dissolved the *Mtb*$CarD_{83–161}$ crystals and performed N-terminal sequencing of the obtained protein (Supplementary Table S1). This indicated that the protein is cleaved after residue 75 and that residues 76–162 of *Mtb*CarD are present in the crystallized protein. The recombinant *Taq*RNAP β subunit aggregated and eluted in the void volume of the Superdex 200 (16/60) column (Fig. 1).

3.2. Secondary-structure content of *Mtb*CarD

Far-UV CD spectroscopy was used to estimate the secondary-structure content of *Mtb*CarD. The CD data were deconvoluted using the *DichroWeb* server (Whitmore & Wallace, 2004) and the percentages of α-helix, β-sheet and random-coil structure were evaluated. The CD data predicted ∼60% α-helix, ∼7% β-sheet and ∼33% random-coil structure in full-length *Mtb*CarD. These values were close to the secondary-structure content observed in the crystal structure of *Mtb*CarD (Gulten & Sacchettini, 2013) and from various theoretical calculations on *Mtb*CarD (∼59% α-helix, ∼10.5% β-sheet and ∼30.5% random coil). The secondary-structure content of *Mtb*CarD obtained using *PSIPRED* (Jones, 1999) analysis is shown in Supplementary Fig. S1.

3.3. Crystal structure of $CarD_{83–161}$

We set up crystallization trials using full-length *Mtb*CarD protein. However, the protein degraded during crystallization and the X-ray intensity data obtained yielded the structure of only residues Thr83–

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: GX5219).

Ala161 of *Mtb*CarD ($CarD_{83–161}$). The rhombohedral-shaped crystals of $CarD_{83–161}$ grew in mother liquor consisting of 20%(*w*/*v*) PEG 4000, 0.2 *M* $MgSO_4$, 30% ethylene glycol to dimensions of approximately 0.5 × 0.3 × 0.1 mm. The phases for the $CarD_{83–161}$ structure were obtained by the single anomalous dispersion (SAD) technique using an iodide SAD data set collected at 2.8 Å resolution. The $CarD_{83–161}$ structure obtained from the iodide SAD data was placed in the native CarD data set collected at 2.1 Å resolution and refined to an *R* factor of 0.21 and an R_{free} of 0.25. All residues of $CarD_{83–161}$ fall in the favourable region of the Ramachandran plot.

$CarD_{83–161}$ forms a distorted Y-shaped structure containing bundles of three helices connected by short loops (Fig. 2). Two views of the $CarD_{83–161}$ electrostatic surface (180° apart) showing the distribution of positive (blue) and negative (red) surface charge are shown in Fig. 2. The protein carries a pronounced overall positive charge at the N-terminus (helices H1 and H2) and a negative charge in the H3 helix. In the $CarD_{83–161}$ structure, the first helix (H1) is connected to the second helix (H2) by a three-residue loop. The second helix (H2) is connected to the third helix (H3) by an eight-residue loop. The lengths of the various helices were observed to be 21 Å for the H1 helix, 24 Å for the H2 helix and 60 Å for the H3 helix. The following interactions were observed between the H1, H2 and H3 helices: (i) a salt bridge Lys95 $N^{\zeta}\cdots$Glu106 $O^{\varepsilon1}$ between the H1 and H2 helices, (ii) salt bridges Trp85 $N^{\varepsilon1}\cdots$Glu124 $O^{\varepsilon2}$ and Arg88 NH2$\cdots$Glu124 $O^{\varepsilon2}$ between the H1 and H3 helices and (iii) a salt bridge Asp115 $O^{\delta1}\cdots$Lys125 N^{ζ} between the H2 and H3 helices. These residues are highly conserved in the sequences of CarD from all mycobacterial species. These results indicate that the $CarD_{83–161}$ structure is unique and is a template for the structure of CarD from all mycobacterial species.

PISA server (Krissinel & Henrick, 2007) analysis of the $CarD_{83–161}$ interfaces indicated that the quaternary structure of $CarD_{83–161}$ has two macromolecular oligomeric states with a buried surface area of 1837.1 Å^2, a solvent-accessible surface area of 100 049.5 Å^2 and a dissociation barrier of 5.2 kcal mol^{-1}. Of the 79 residues of $CarD_{83–161}$, 76 residues constitute the total surface area of 5943.3 Å^2 with a free energy (ΔG) of −48.9 kcal mol^{-1}. The server identified four interfaces in the $CarD_{83–161}$ structure, with surface areas of 918.6 Å^2 for surface 1, 364.8 Å^2 for surface 2, 250.7 Å^2 for surface 3 and 98.0 Å^2 for surface 4.

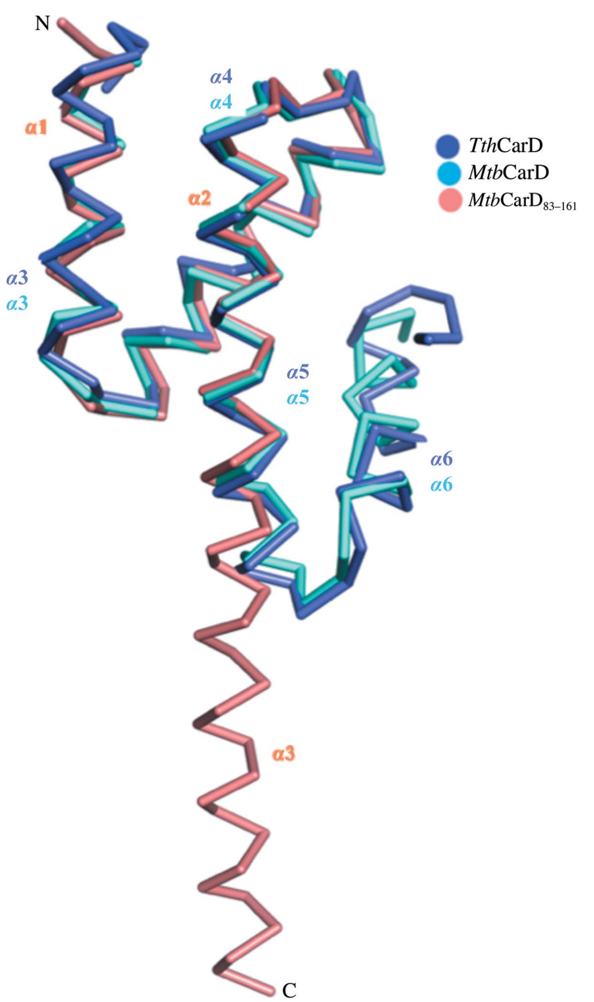

Figure 3
Superposition of *Mtb*$CarD_{83–161}$ with the crystal structures of the C-terminal domains of full-length *Mtb*CarD (Gulten & Sacchettini, 2013; PDB entry 4kbm) and *Tth*CarD (Srivastava *et al.*, 2013; PDB entry 4l5g). *Mtb*$CarD_{83–161}$ is shown in orange and is labelled as described in the text. The C-terminal domains of full-length *Mtb*CarD and *Tth*CarD are shown in cyan and blue, respectively, and all helices are labelled according to the corresponding publications.

3.4. Identification of the $CarD_{83–161}$ fold

Analysis of the $CarD_{83–161}$ sequence using the leucine-zipper domain server (Bornberg-Bauer *et al.*, 1998) indicated that $CarD_{83–161}$ does not contain a leucine-zipper domain as reported previously (Stallings *et al.*, 2009). The trimeric helix bundle formed by the $CarD_{83–161}$ structure did not superpose on any leucine-zipper domain structure available in the database. The *DALI* server (Holm & Rosenström, 2010) was used to obtain the homologous structure of $CarD_{83–161}$, which indicated a unique fold with no precedents.

A previous study (Westblade *et al.*, 2010) indicated that *Mtb*CarD is like a CdnL protein, which shares similarity to the N-terminal TRCF-RID-like domain and lacks an identifiable C-terminal DNA-binding domain. When the $CarD_{83–161}$ structure was aligned with the following HMG domain-containing proteins involved in DNA binding, (i) the three-dimensional structure of the human SRY-DNA complex solved by multidimensional heteronuclear-edited and filtered NMR (PDB entry 1hry; Werner *et al.*, 1995), (ii) the HMG domain structure (from mouse) complexed with DNA from NMR (PDB entry 2lef; Love *et al.*, 1995) and (iii) the NMR structure of rat HMG1 HMGA fragment (PDB entry 1aab; Hardman *et al.*, 1995), no sequence or structure homology was observed between $CarD_{83–161}$ and these proteins. These results indicate that $CarD_{83–161}$ contains a unique fold that was not observed in the existing structural database.

3.5. Comparison of the $CarD_{83–161}$ structure with *M. tuberculosis* and *T. thermophilus* full-length CarD structures

Recently, the crystal structures of full-length *M. tuberculosis* CarD in complex with RNAP β lobes (Gulten & Sacchettini, 2013; PDB entry 4kbm) and the crystal structure of *T. thermophilus* CarD (*Tth*CarD; Srivastava *et al.*, 2013; PDB entry 4l5g) have been determined. The structure of the C-terminal domain of *Mtb*CarD and *Tth*CarD consists of five α-helices, compared with the three α-helices observed in our *Mtb*$CarD_{83–161}$ structure. Ala122–Ala160 of our *Mtb*$CarD_{83–161}$ structure form a single α3 helix, while Ala122–Ala159 of full-length *Mtb*CarD and *Tth*CarD form a helix–turn–helix motif (Gulten & Sacchettini, 2013; Srivastava *et al.*, 2013; Fig. 3). In the *Mtb*CarD and *Tth*CarD structures, Ala122–Glu143 form helix α5 and Lys149–Ala159 form helix α6 and the helices are connected by a five-residue loop. Superposition of the *Mtb*$CarD_{83–161}$ structure with the *Tth*CarD structure indicated an r.m.s.d of 1.37 Å for 58 C^{α} atoms, while superposition with *Mtb*CarD gave an r.m.s.d. of 0.89 Å for 59 C^{α} atoms (Fig. 3).

The α5 helix of the C-terminal domain of *Mtb*CarD and *Tth*CarD superposes well with the α3 helix of *Mtb*$CarD_{83–161}$. However, a segment containing a loop and helix α6 (residues 143–161) of *Mtb*CarD and *Tth*CarD is oriented ~180° away from the α3 helix of *Mtb*$CarD_{83–161}$. In the full-length *Mtb*CarD and *Tth*CarD structures (Gulten & Sacchettini, 2013; Srivastava *et al.*, 2013), residues 71–75 of helix 1 interact with residues 153–157 of the C-terminal domain of CarD in a mostly hydrophobic manner. N-terminal sequencing of our crystallized protein indicated that residues 76–161 of CarD are present in our crystals. This means that the hydrophobic interactions between residues 71–75 and residues 153–157 are missing in our *Mtb*$CarD_{83–161}$ structure and this has caused the 143–161 helix of our structure to unfold. It indicates that the structural differences in our *Mtb*$CarD_{83–161}$ structure *versus* full-length *M. tuberculosis* and *T. thermophilus* CarD are owing to a combination of proteolysis and crystallization artifacts.

In the current study, we have expressed and purified full-length *Mtb*CarD and *Taq*RNAP β subunit. The *Mtb*CarD protein eluted as a monomer, while the *Taq*RNAP β subunit eluted as an aggregated protein from a gel-filtration column. We have determined the structure of the C-terminal domain of *Mtb*CarD ($CarD_{83–161}$). Comparison of our *Mtb*$CarD_{83–161}$ structure with recently determined full-length *M. tuberculosis* and *T. thermophilus* CarD crystal structures has revealed structural differences in the C-terminal region of our *Mtb*$CarD_{83–161}$ structure owing to proteolysis and crystallization artifacts.

This work was supported by a research grant from the Council of Scientific and Industrial Research (CSIR), New Delhi, India. Grants from UGC Networking, Capacity Buildup and a Department of Science and Technology (DST) purse grant are gratefully acknowledged. The authors declare that they have no competing interests. The authors thank Dr Sushma Nagpal for allowing them to use the BIAcore facility of the National Institutes of Immunology, New Delhi for the SPR experiment. The authors thank the staff of the BM14 beamline for their help in CarD intensity data collection and processing. We acknowledge the support of Dr Seth A. Darst from Rockefeller University, New York, USA in gifting us the plasmid containing the *Taq*RNAP β subunit gene.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Bornberg-Bauer, E., Rivals, E. & Vingron, M. (1998). *Nucleic Acids Res.* **26**, 2740–2746.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
DeLano, W. L. (2002). *PyMOL.* http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gulten, G. & Sacchettini, J. C. (2013). *Structure*, **21**, 1859–1869.
Hardman, C. H., Broadhurst, R. W., Raine, A. R., Grasser, K. D., Thomas, J. O. & Laue, E. D. (1995). *Biochemistry*, **34**, 16596–16607.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Jones, D. T. (1999). *J. Mol. Biol.* **292**, 195–202.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R. & Wright, P. E. (1995). *Nature (London)*, **376**, 791–795.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Srivastava, D. B., Leon, K., Osmundson, J., Garner, A. L., Weiss, L. A., Westblade, L. F., Glickman, M. S., Landick, R., Darst, S. A., Stallings, C. L. & Campbell, E. A. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 12619–12624.
Stallings, C. L., Stephanou, N. C., Chu, L., Hochschild, A., Nickels, B. E. & Glickman, M. S. (2009). *Cell*, **138**, 146–159.
Weiss, L. A., Harrison, P. G., Nickels, B. E., Glickman, M. S., Campbell, E. A., Darst, S. A. & Stallings, C. L. (2012). *J. Bacteriol.* **194**, 5621–5631.
Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. (1995). *Cell*, **81**, 705–714.
Westblade, L. F., Campbell, E. A., Pukhrambam, C., Padovan, J. C., Nickels, B. E., Lamour, V. & Darst, S. A. (2010). *Nucleic Acids Res.* **38**, 8357–8369.
Whitmore, L. & Wallace, B. A. (2004). *Nucleic Acids Res.* **32**, W668–W673.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
World Health Organization (2011). *Global Tuberculosis Report*. Geneva: World Health Organization. http://www.who.int/tb/publications/global_report/en/.