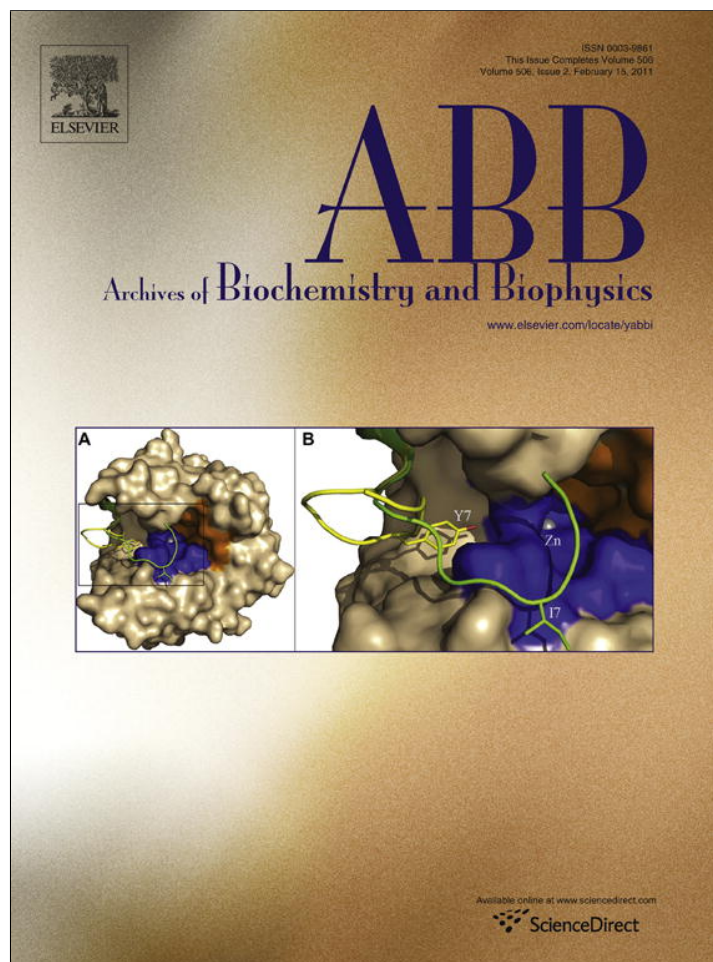


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

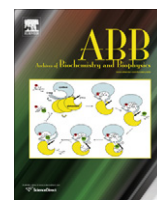
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## Archives of Biochemistry and Biophysics

journal homepage: [www.elsevier.com/locate/yabbi](http://www.elsevier.com/locate/yabbi)

## Inaugural structure from the DUF3349 superfamily of proteins, *Mycobacterium tuberculosis* Rv0543c

Garry W. Buchko<sup>a,\*</sup>, Isabelle Phan<sup>b</sup>, Peter J. Myler<sup>b,c,d</sup>, Thomas C. Terwilliger<sup>e</sup>, Chang-Yub Kim<sup>e,\*</sup>

<sup>a</sup>Biological Sciences Division and Seattle Structural Genomics Center for Infectious Disease, Pacific Northwest National Laboratory, Richland, WA 99352, USA

<sup>b</sup>Seattle Biomedical Research Institute and Seattle Structural Genomics Center for Infectious Disease, 307 Westlake Avenue N, Seattle, WA 98109-5219, USA

<sup>c</sup>Department of Medical Education and Biomedical Informatics, University of Washington, Seattle, WA 98195, USA

<sup>d</sup>Department of Global Health, University of Washington, Seattle, WA 98195, USA

<sup>e</sup>Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

### ARTICLE INFO

#### Article history:

Received 27 September 2010  
and in revised form 1 December 2010  
Available online 6 December 2010

#### Keywords:

Tuberculosis  
Circular dichroism  
Structural genomics  
Protein dynamics

### ABSTRACT

The first structure for a member of the DUF3349 (PF11829) family of proteins, Rv0543c from *Mycobacterium tuberculosis*, has been determined using NMR-based methods and some of its biophysical properties characterized. Rv0543c is a 100 residue, 11.3 kDa protein that both size exclusion chromatography and NMR spectroscopy show to be a monomer in solution. The structure of the protein consists of a bundle of five  $\alpha$ -helices,  $\alpha$ 1 (M1 – Y16),  $\alpha$ 2 (P21 – C33),  $\alpha$ 3 (S37 – G52),  $\alpha$ 4 (G58 – H65) and  $\alpha$ 5 (S72 – G87), held together by a largely conserved group of hydrophobic amino acid side chains. Heteronuclear steady-state  $\{^1\text{H}\}$ - $\{^{15}\text{N}\}$  NOE,  $T_1$ , and  $T_2$  values are similar through-out the sequence indicating that the backbones of the five helices are in a single motional regime. The thermal stability of Rv0543c, characterized by circular dichroism spectroscopy, indicates that Rv0543c irreversibly unfolds upon heating with an estimated melting temperature of 62.5 °C. While the biological function of Rv0543c is still unknown, the presence of DUF3349 proteins predominately in *Mycobacterium* and *Rhodococcus* bacterial species suggests that Rv0543c may have a biological function unique to these bacteria, and consequently, may prove to be an attractive drug target to combat tuberculosis.

© 2010 Elsevier Inc. All rights reserved.

### Introduction

Approximately 1.6 million people die annually from infection with *Mycobacterium tuberculosis*, the aetiological agent responsible for the chronic infectious disease tuberculosis (TB)<sup>1</sup> [1]. Via inhalation of aerosolized bacilli as droplet nuclei exhaled from coughing infected individuals [2], almost 10 million people become newly infected annually [1]. Indeed, it is estimated that one-third of the human population would have a positive skin test for the disease [3]. While TB is held in check by effective public health care systems in the Western world, it is endemic in 22 “high-burden” nations [4]. Due to the evolution of multi-drug and extremely drug-resistant *M. tuberculosis* strains along with the prevalence of high human immunodeficiency virus infections [5–7], there is a fear that the industrialized world may lose its ability to control this disease [4].

\* Corresponding authors.

E-mail addresses: [garry.buchko@pnl.gov](mailto:garry.buchko@pnl.gov) (G.W. Buchko), [cykim@lanl.gov](mailto:cykim@lanl.gov) (C.-Y. Kim).

<sup>1</sup> Abbreviations used: TB, tuberculosis; WHO, World Health Organization; DOTS, Directly Observed Treatment and Short-course drug therapy; RSCB, Research Collaboratory for Structural Bioinformatics; BMRB, Biological Magnetic Resonance Data Bank.

Consequently, it is of fundamental importance to develop a new generation of intervention strategies against TB [8].

The current TB control strategy employed by the World Health Organization (WHO) is based on Directly Observed Treatment and Short-course drug therapy (DOTS) [9]. In addition to saving the lives of infected individuals, early diagnosis and treatment of the disease interrupts transmission. A major focus in the search for new anti-TB drug therapies is a molecular understanding of the *M. tuberculosis* gene products and the interaction of various metabolic pathways in the microenvironment in the host that allow *M. tuberculosis* to survive [4]. One highly conserved gene product present in *M. tuberculosis* and other *Mycobacterium* and *Rhodococcus* species is a small protein of 100 amino acid residues, Rv0543c. This protein contains a “Domain of Unknown Function” that falls into the DUF3349 (PF11829) superfamily of proteins. To date there is no known function for this protein and a structure has not been solved for any of the DUF3349 family members. To obtain further insights into the biological function of Rv0543c and the other proteins in the DUF3349 superfamily, the solution structure for Rv0543c was determined using NMR-based methods, its thermostability assayed using circular dichroism spectroscopy, and the dynamics of the protein explored by NMR spectroscopy.

## Materials and methods

### Cloning, expression, and purification

The DNA coding sequence for the *M. tuberculosis* Rv0543c gene was cloned into a modified pET28b vector (Novagen, Madison, WI) upstream of the NdeI site such that the expressed protein contained an N-terminal poly-histidine tag (MGSSHHHHHSSGLVPR GSH-). This recombinant vector was then transfected into the host *Escherichia coli* bacterial strain BL21PRO (Clontech, Palo Alto, CA). Uniformly  $^{15}\text{N}$ - and  $^{15}\text{N}$ -,  $^{13}\text{C}$ -labeled Rv0543c was obtained by growing the transformed cells (37 °C) in minimal medium (Miller) containing  $^{15}\text{NH}_4\text{Cl}$  (1 mg/mL) and  $\text{D-}[^{13}\text{C}_6]\text{glucose}$  (2.0 mg/mL) supplemented with  $\text{Fe}_2\text{Cl}_3$  (50  $\mu\text{g}/\text{mL}$ ) and the antibiotics kanamycin (34  $\mu\text{g}/\text{mL}$ ) and spectinomycin (100  $\mu\text{g}/\text{mL}$ ). Once the cells reached an  $\text{OD}_{600}$  reading of  $\sim 0.8$ , the cells were cooled to 25 °C and protein expression induced with isopropyl  $\beta$ -D-1-thiogalactopyranoside (0.026  $\mu\text{g}/\text{mL}$ ). Cells were harvested 4–6 h later and frozen at  $-80$  °C. Following established protocols [10,11], the frozen pellet was later thawed, the cells passed three times through a French press and the cell lysate applied to a Ni-NTA affinity column (QIAGEN, Valencia, CA). The fraction containing target protein was exchanged into thrombin cleavage buffer by dialysis (150 mM NaCl, 20 mM Tris-HCl, pH 8.4), concentrated to  $\sim 2$  mL, and the N-terminal poly-histidine tag removed by overnight incubation with thrombin at room temperature. Using a flow rate of 1.0 mL/min, the reaction solution was then loaded on a Superdex75 HiLoad 10/30 column (GE Healthcare, Piscataway, NJ) to simultaneously purify the protein and exchange it into NMR buffer (300 mM NaCl, 20 mM Tris-HCl, 1.0 mM dithiothreitol, pH 7.1). The band containing Rv0543c (retention time = 81 min) was collected and the volume reduced (Amicon Centriprep-10) to generate NMR samples in the 1–2 mM range (Lowry analysis). The SDS-PAGE analysis of the final NMR samples showed the protein to be  $> \sim 95\%$  pure.

### Optical spectroscopy

Circular dichroism data was collected on an Aviv Model 410 spectropolarimeter (Lakewood, NJ) calibrated with an aqueous solution of ammonium  $\text{D-}(+)\text{camphorsulfonate}$  using a 0.06 mM Rv0543c sample in NMR buffer and in a quartz cell of 0.1 cm path length. A thermal denaturation curve was obtained by recording and plotting the ellipticity at 220 nm in 2.0 °C intervals from 10 to 80 °C. A melting temperature,  $T_m$ , was estimated from the thermal denaturation curve by taking a first derivative of the plot using the Aviv software [12]. Steady-state wavelength spectra for Rv0543c were recorded in 0.5 nm increments between 200 and 260 nm at 25 °C, 80 °C, and 25 °C (post-heating). Each reported steady-state wavelength spectrum was the result of averaging two consecutive scans with a bandwidth of 1.0 nm and a time constant of 1.0 s. These spectra were processed by subtracting a blank spectrum from the protein spectrum and then automatically line smoothing the data using Aviv software.

### NMR spectroscopy and resonance assignments

Varian 750- and 600-Inova spectrometers equipped with triple resonance probes and pulse field gradients were used to collect the NMR data required for resonance assignments and structure determination. The NMR data, collected on 1–2 mM samples at 20 °C, were processed with Felix2007 (Felix NMR, Inc., San Diego, CA) and analyzed with Sparky (v3.115) [13]. Chemical shifts were referenced to DSS (DSS = 0 ppm) using indirect methods [14].

The  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  chemical shifts of the backbone and side chain resonances were obtained from standard two-dimensional

$^1\text{H}$ - $^{15}\text{N}$  HSQC,  $^1\text{H}$ - $^{13}\text{C}$  HSQC, HBCBCGCDHD, and HBCBCGDCHE experiments and three-dimensional HNCACB, CBCA(CO)NH, HNCO, HCC-TOCSY-NNH and CC-TOCSY-NNH experiments using Varian Protein-pack pulse programs. Distance restraints were obtained from a suite of three-dimensional,  $^{13}\text{C}$ - and  $^{15}\text{N}$ -edited NOESY-HSQC experiments using a mixing time of 80 ms. Deuterium-exchange studies were performed by lyophilizing an  $^{15}\text{N}$ -labeled NMR sample, re-dissolving in 99.8%  $\text{D}_2\text{O}$ , and immediately collecting  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra 10, 20, and 60 min after the exchange. Steady-state  $\{^1\text{H}\}$ - $^{15}\text{N}$  heteronuclear NOE values ( $\text{NOE} = I_{\text{sat}}/I_{\text{unsat}}$ ) were measured in triplicate from the ratios of  $^1\text{H}$ - $^{15}\text{N}$  HSQC cross peak volumes in spectra recorded in the presence ( $I_{\text{sat}}$ ) and absence ( $I_{\text{unsat}}$ ) of three seconds of proton presaturation prior to the  $^{15}\text{N}$  excitation pulse [15–17]. Nitrogen-15  $T_1$  and  $T_2$  values were measured using previously described experiments [15] with seven different time delays (s): 0.0, 0.1, 0.2, 0.3, 0.4, 0.6, 0.8 ( $T_1$ ); 0.01, 0.03, 0.05, 0.07, 0.09, 0.11, and 0.13, ( $T_2$ ). Values for  $T_1$  and  $T_2$  and an estimate of their associated errors were obtained using the rate analysis function in Sparky (v3.115) [13] by fitting the measured peak heights to a two-parameter exponential function,  $I(t) = I_0 \exp(-t/T_{1,2})$ . An overall rotational correlation time ( $\tau_c$ ) for Rv0543c was estimated from backbone amide  $^{15}\text{N}$   $T_1/T_{1\rho}$  ratios measured using a modified  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiment to record  $^{15}\text{N}$ -edited one-dimensional spectra [18,19] and from the average of all the individual amide nitrogen-15  $T_1/T_2$  ratios. The steady-state  $\{^1\text{H}\}$ - $^{15}\text{N}$  heteronuclear NOE,  $T_1$ , and  $T_2$  data were collected at a  $^1\text{H}$  resonance frequency of 600 MHz (20 °C).

### Structure calculations

Structure calculations were performed iteratively using CYANA (v 2.1) [20], the  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  chemical shift assignments, and the peak-picked NOESY data as initial experimental inputs. Sixty-six dihedral angle restraints for both Phi ( $\Phi$ ) and Psi ( $\Psi$ ) were introduced on the basis of the elements of secondary structure identified in the early structural ensembles and TALOS calculations [21]. Near the end of the iterative process 64 hydrogen bond restraints (1.8–2.0 Å and 2.7–3.0 Å for the NH–O and N–O distances, respectively) were introduced into the structure calculations on the basis of proximity in early structure calculations and the observation of slowly exchanging amides in the deuterium exchange experiment.

The final ensemble of 20 CYANA derived structures were then refined with explicit water [22] with CNS (version 1.1) using force constants of 500 and 2000 kcal for the NOE and dihedral restraints, respectively. For the water refinement calculations the upper boundary of the CYANA distance restraints was increased by 1% and the lower bound was set to the vdW limit. The 15 water refined structures with the fewest and smallest unfavorable contacts (best clash scores) were then used to calculate a mean structure and average RMSD values to the mean structure using the structured regions of the protein (residue M1 – G87). Structural quality was assessed using the Protein Structure Validation Suite (PSVS, v1.3) [23]. The structural statistics are summarized in Table 1.

### Protein structure accession numbers

The atomic coordinates for the final ensemble of 15 structures for *M. tuberculosis* Rv0543c have been deposited in the Research Collaboratory for Structural Bioinformatics (RSCB) under PDB code 2KVC. The chemical shifts assignments have been deposited with the Biological Magnetic Resonance Data Bank (BMRB) under Accession Number 16774. Note that the non-native three residues at the N-terminus, GSH-, are numbered sequentially starting with G1 in the RSCB and BMRB depositions. However, here the three

**Table 1**  
Summary of the structural statistics for Rv0543c<sup>a</sup>.

<i>Restraints for structure calculations</i>	
Total NOEs	1118
Intraresidue NOEs	364
Sequential ( <i>i, i + 1</i> ) NOEs	304
Medium-range ( <i>i, i + j; 1 &lt; j ≤ 4</i> ) NOEs	209
Long-range ( <i>i, i + j; j &gt; 4</i> ) NOEs	241
Phi ( $\Phi$ ) angle restraints	66
Psi ( $\Psi$ ) angle restraints	66
Hydrogen bond restraints <sup>b</sup>	64
<i>Structure calculations</i>	
Number of structures calculated	100
Number of structures used in ensemble	15
<i>Structures with restraint violations</i>	
Distance restraint violations >0.04 Å	0
Dihedral restraint violation >1°	0
<i>RMSD to mean (Å)</i>	
Ordered residues: 1–87	
Backbone N–C $\alpha$ –C=O atoms	0.64 ± 0.08
Heavy atoms	1.14 ± 0.09
All atoms	1.31 ± 0.05
<i>Ramachandran plots summary</i>	
Ordered residues from Procheck <sup>c</sup>	
Most favored regions	92.3%
Additionally allowed regions	7.5%
Generously favored regions	0.2%
Disallowed	0.0%
<i>Global quality scores – Ordered residues<sup>c</sup></i>	
	Z-score (Raw)
Procheck (all)	–0.41 (–0.07)
Procheck ( $\Phi, \Psi$ )	0.90 (0.14)
MolProbity clash score	–2.70 (24.60)

<sup>a</sup> All statistics are for the 15-structure ensemble deposited in the Protein Data Bank (2KVC) using the residues containing the structured central core (1–87).

<sup>b</sup> Obtained from 32 amide cross peaks still present 60 min after the start of the deuterium exchange experiment: I8–Y16, G19, V28–C33, L36, V41–A45, E47–M49, V76, E77, E79–L83.

<sup>c</sup> Calculated for the ordered residue, 1–87, using the PSVS program.

non-native residues are numbered sequentially with an asterisk (G1\* – H3\*) and the first native residue, M4 in the RSCB and BMRB depositions, is labeled as M1.

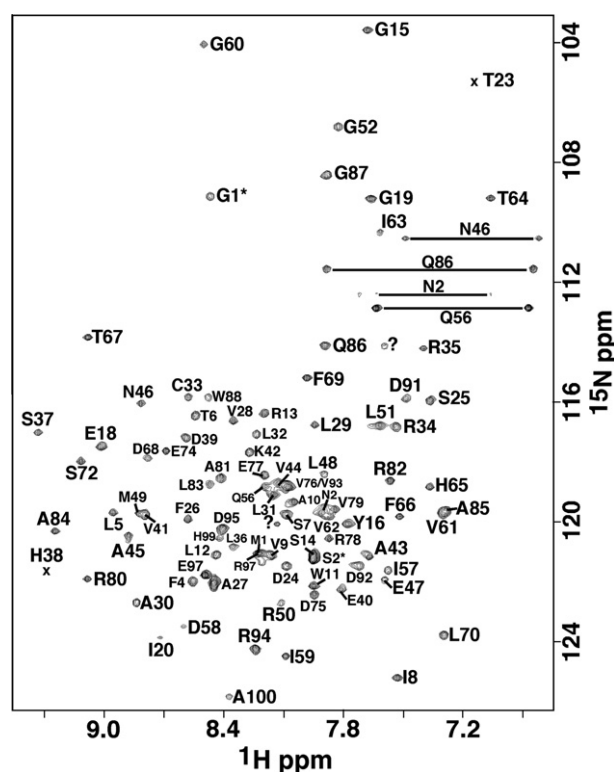
### Bioinformatics analyses

A search for structures similar to Rv0543c in the RSCB Protein Data Bank was performed using the DALI server ([http://ekhidna.biocenter.helsinki.fi/dali\\_server/](http://ekhidna.biocenter.helsinki.fi/dali_server/)) [24]. The amino acid conservation pattern on the three-dimensional structure of Rv0543c was calculated using the ConSurf server (<http://consurf.tau.ac.il/overview.html>) [25]. Using the proteins listed in the DUF3349 (PF11829) superfamily, a multiple sequence alignment was performed with 10 closely related protein sequences from genera other than *Mycobacterium* and *Rhodococcus* using the program ClustalW2 [26]. Figures of structures were generated using PyMOL [27].

## Results and discussion

### Solution structure for Rv0543c

The elution time of Rv0543c on a size exclusion column, 81 min, and estimations of the rotational correlation time at 20 °C,  $9.3 \pm 0.3$  ns (modified  $^1\text{H}$ – $^{15}\text{N}$  HSQC experiment) and  $8.2 \pm 2.3$  ns ( $T_1/T_2$  ratios), were consistent with a monomeric 11.3 kDa protein [28]. The wide chemical shift dispersion in the nitrogen and especially the proton dimension of the  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectrum of Rv0543c, shown in Fig. 1, indicates the protein adopts a stable

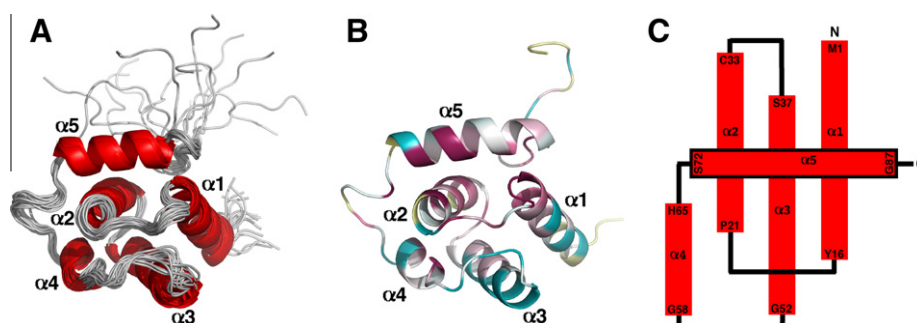


**Fig. 1.** Assigned  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectrum of double-labeled Rv0543c collected at 20 °C in NMR buffer (300 mM NaCl, 20 mM TrisHCl, 1.0 mM DTT, pH 7.1) at a  $^1\text{H}$  resonance frequency of 750 MHz. Side chain resonances are indicated by solid horizontal lines and the two unassigned cross peak are labeled with a question mark. Not shown are the ring amide resonances for W11 (9.481 and 131.7 ppm) and W88 (10.35 and 130.1 ppm).

folded conformation in solution. All but 6 of the 96 expected amide resonances were unambiguously assigned in the  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectrum of Rv0543c. Cross peak amides were not identified for D53–D55, V76, L90, and E98. On the basis of the amide assignments and extensive assignment of the  $^{13}\text{C}\alpha$  and side chain proton and carbon chemical shifts (BMRB ID 16774), an ensemble of structures was calculated (Fig. 2A) that satisfied all the available experimental NMR data (NOEs, chemical shifts, deuterium exchange experiments, and TALOS calculations).

As summarized in Table 1, the final set of structure calculations included 1118 interproton distance restraints (NOE data), 64 hydrogen bond restraints (deuterium exchange data), and 132 dihedral angle restraints (TALOS calculations). While the final set of 20 CYANA derived structures contained no unfavorable steric clashes, refinement with explicit water introduced some unfavorable steric clashes into the structures despite extensive efforts adjusting the refinement parameters. In the end the 15 structures with the best clash scores following water refinement were chosen for the ensemble submitted to the RCSB PDB. Each member of this final ensemble agreed well with the experimental data with no upper limit violation greater than 0.04 Å and no torsion angle violation greater than 1°. The quality of the structure ensemble was also shown to be good using the PSVS validation software package [23]. The Ramachandran statistics for all the residues in the ensemble were overwhelmingly in acceptable space (92.3% of the ( $\Phi, \psi$ ) pairs for Rv0543c were found in the most favored regions and 7.5% within additionally allowed regions) and all the structure quality Z-scores were acceptable (greater than –5) including the final MolProbity clash score of –2.70.

The final set of 15 calculated structures in the ensemble converged well, as shown mathematically by the statistics in Table 1



**Fig. 2.** (A) Superposition of the final ensemble of 15 structures calculated for Rv0543c. (B) Ribbon representation of the structure in the ensemble closest to the average structure for Rv0543c color-coded using ConSurf. (C) Secondary structure diagram of Rv0543c with the  $\alpha$ -helices drawn as red rectangles with the residue number at the beginning and the end of each helix shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

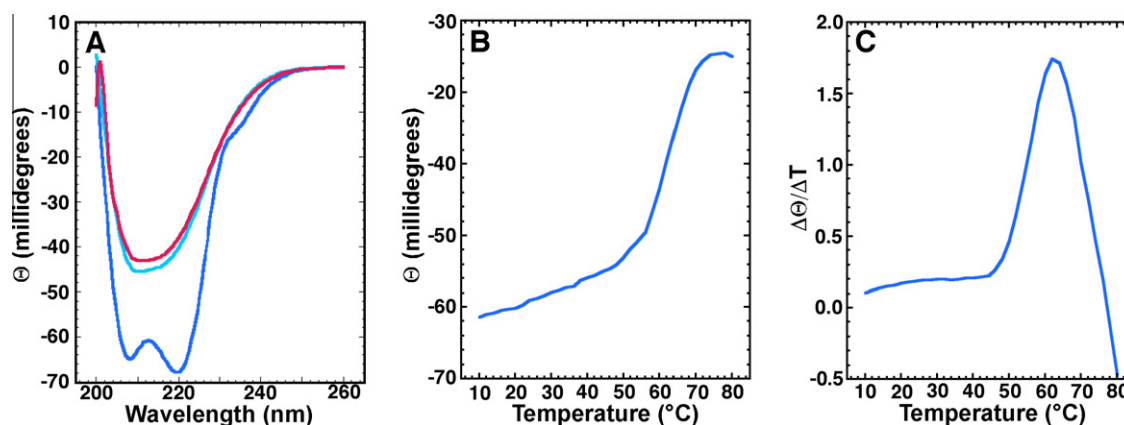
and visually by the superposition of the ordered residues in Fig. 2A. The RMSD of the structured core of ordered residues (M1 – G87) in the ensemble to the mean structure is 0.64 Å for the backbone atoms (N–C $\alpha$ –C=O) and 1.14 Å for all heavy atoms. The C-terminal region, W91–A100, is unstructured and disordered, and this is evident in Fig. 2A. A cartoon representation of a single structure is shown in Fig. 2B along with a schematic secondary structure diagram of Rv0543c in Fig. 2C. The protein consists of a bundle of five  $\alpha$ -helices,  $\alpha$ 1 (M1 – Y16),  $\alpha$ 2 (P21 – C33),  $\alpha$ 3 (S37 – G52),  $\alpha$ 4 (G58 – H65) and  $\alpha$ 5 (S72 – G87), held in place by an hydrophobic core of largely conserved amino acid side chains.

#### Optical spectroscopy

Circular dichroism spectroscopy was used to characterize the structure and stability of Rv0543c over a range of temperatures [29]. Fig. 3A shows the steady-state wavelength spectra for Rv0543c at two temperatures. The solid blue line is the CD spectrum of a fresh, ~0.06 mM sample collected at 25 °C in the same buffer used to collect NMR data for the structure calculations. The double minimum at ~220 and ~208 nm and projected maximum at a wavelength <200 nm is characteristic of a structured protein with significant  $\alpha$ -helical content [29,30]. The red line in Fig. 3A is the CD spectrum of the same sample at 80 °C. Relative to the spectrum collected at 25 °C, the double minimum has disappeared and is replaced by a single, more positive minimum at ~212 nm. These observations at 80 °C indicate that Rv0543c is more unstructured at elevated temperatures, losing most of its  $\alpha$ -helical character. However, the absence of an extrapolated

negative minimum at ~198 nm and a positive maximum at ~218 nm indicates that Rv0543c is not entirely random coil at 80 °C and still possesses spectral properties consistent with primarily  $\beta$ -sheet structure (single minimum). When the sample is cooled back to 25 °C, the CD spectrum, indicated by the cyan line in Fig. 3A, is almost identical to the CD spectrum collected at 80 °C (red line). This indicates that the temperature-induced effects to the structure of Rv0543c are irreversible.

To assay the thermal stability of Rv0543c, the ellipticity at 220 nm was measured as a function of temperature between 10 and 80 °C. Typically, a phase transition can be detected when a structured protein becomes denatured by monitoring the increase in the ellipticity at a specific wavelength with increasing temperature [10,11,31,32]. As shown in Fig. 3B, a gradual increase in ellipticity at 220 nm is observed up to ~56 °C followed by a more rapid increase in ellipticity that tails-off and plateaus at ~74 °C. Because the steady-state wavelength spectra in Fig. 3A suggest that the protein is not entirely unstructured at 80 °C, the end point of the temperature study in Fig. 3B likely does not represent a fully unstructured protein but one that now contains significant  $\beta$ -strand character. Beta-strands are typically more robust than  $\alpha$ -helices [33] and the structure that forms at high temperature does not unfold into its native structure upon cooling to 25 °C. While the irreversible nature of the transition means the CD data may not be analyzed thermodynamically [32], a quantitative estimation of the  $T_m$  for this transition may be obtained by assuming a two-state model and taking a first derivative of the curve in Fig. 3B [12]. The maximum of this first derivative, shown in Fig. 3C, is 62.5 °C.



**Fig. 3.** (A) Circular dichroism steady-state wavelength spectra of a single sample of Rv0543c (0.06 mM) in NMR buffer collected at 25 °C (blue), 80 °C (red), and 25 °C post-heating (cyan). (B) The CD thermal melt for Rv0543c obtained by measuring the ellipticity at 220 nm in 2.0 °C intervals between 10 and 80 °C. (C) The first derivative of the curve in Fig. 3B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

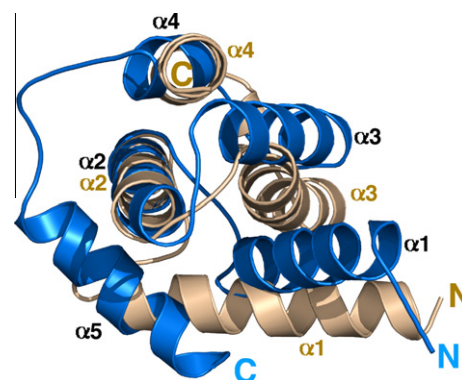
### Backbone dynamics

The solution structure calculations for Rv0543c show that the protein consists of five  $\alpha$ -helices arranged in an ordered bundle. Because protein dynamics often plays a role in the binding and catalysis properties of proteins at, and around, active sites [34], motion unique to any one of the five  $\alpha$ -helices may have some mechanistic consequences. Information regarding the dynamics along a protein's backbone can be extracted from the relaxation properties of the amide resonances obtained from heteronuclear steady-state  $\{^1\text{H}\}$ - $^{15}\text{N}$  NOE,  $T_1$ , and  $T_2$  measurements [35,36] because internal motion affects the rate excited nuclei exchange energy with their local environment and relax. Fig. 4 is a plot of the backbone  $\{^1\text{H}\}$ - $^{15}\text{N}$  heteronuclear NOE,  $T_1$ , and  $T_2$  values measured for Rv0543c. The average value for the heteronuclear steady-state  $\{^1\text{H}\}$ - $^{15}\text{N}$  NOE,  $T_1$ , and  $T_2$  is  $0.77 \pm 0.11$ ,  $595 \pm 108$  ms, and  $77 \pm 21$  ms, respectively. There is no evidence that the motion of any of the helices are significantly different from each other as most of the relaxation values are relatively uniform through-out the sequence. These observations suggest that the protein has restricted flexibility on the sub-ns timescale [35] with the only significant exception towards the end of the C-terminus in an unstructured region of the protein where more rapid motion may be expected.

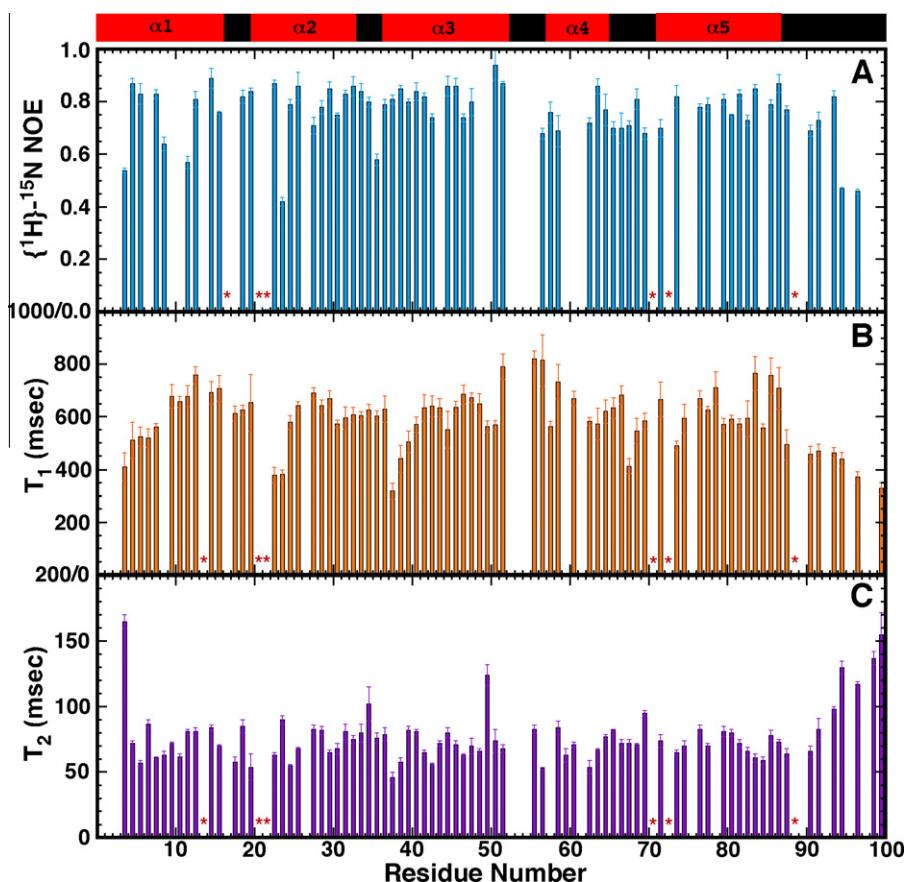
### Insights into biological function

To identify a possible biochemical function for Rv0543c based on its structure, the RCSB Protein Data Bank was searched for

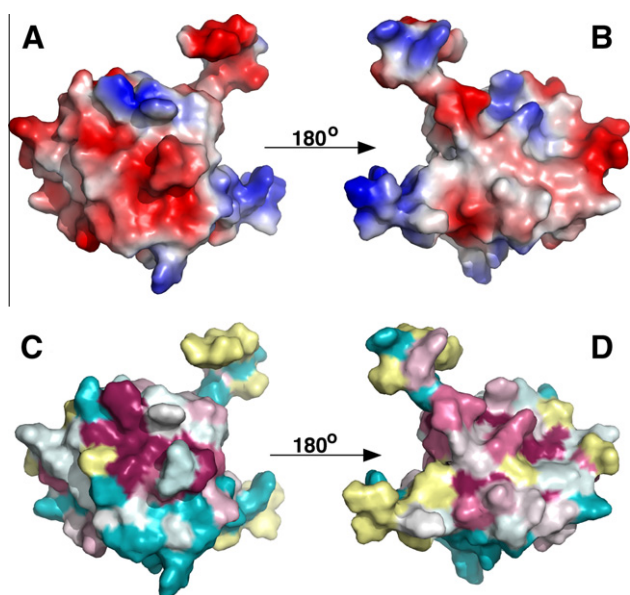
structures with similarities to Rv0543c using the DALI search engine [24]. The best Z-scores identified by the search were only slightly above 4.0, indicating a low structural similarity to any other protein in the RCSB-PDB. The structures with the best Z-scores were all DNA binding proteins, Cre recombinases. Indeed, the 30 structures with the highest Z-scores were Cre recombinases (1CRX-A,  $Z = 4.4$ ; 1X00,  $Z = 4.1$ ) bound to DNA [37,38]. Fig. 5 shows an overlay of the structure of Rv0543c with the PDB structure that resulted in the best RMSD (5.9 Å), 1NZB-E [39], using the web pro-



**Fig. 5.** Superposition of the structure of Rv0543c (M1 – G90, blue) on the structure of *E. coli* phage P1 Cre recombinase 1NZB-E (S38 – G128, wheat) [39] using the program SuperPose [40]. The five  $\alpha$ -helices for Rv0543c are labeled in black following the scheme in Fig. 2 and the four  $\alpha$ -helices shown for Cre recombinase are labeled sequentially in wheat. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** (A) Backbone  $\{^1\text{H}\}$ - $^{15}\text{N}$  heteronuclear NOE (blue), (B)  $T_1$  (orange), and (C)  $T_2$  (purple) values for Rv0543c collected at a  $^1\text{H}$  resonance frequency of 600 MHz (20 °C). Proline residues in the sequence are indicated by red asterisks. The location of experimentally observed  $\alpha$ -helices in the structure is shown in the schematic above the graph. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 6.** (A and B) PyMOL generated maps [27] of the electrostatic potentials at the solvent-accessible surface of Rv0543c. The orientation in (A) is similar to the orientation shown in Fig. 2B with (B) a 180° rotation about the vertical axis. (C and D) ConSurf generated maps [25] of the conserved residues on the solvent-accessible surface of Rv0543c. The orientation of (C and D) is identical to (A and B), respectively. Highly conserved residues are colored magenta and pink, poorly conserved residues are colored cyan, and residues with insufficient information are colored yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

gram SuperPose [40]. The overlaid region of 1NZB (S38 – G128) contains four  $\alpha$ -helices and Rv0543c (M1 – G90) contains five. It can be seen from Fig. 5 that one of the  $\alpha$ -helices, labeled  $\alpha$ 2 in both molecules, superimpose fairly well. Helices  $\alpha$ 1,  $\alpha$ 3, and  $\alpha$ 4 are only

crudely in similar space with  $\alpha$ 5 from Rv0543c overlaying in part with the C-terminal end of  $\alpha$ 1 from 1NZB.

If the biological role of Rv0543c involved binding DNA, then, as observed for the Cre recombinases and other DNA binding proteins, Rv0543c should also have a positively charged surface that interacts with the negatively charged phosphodiester backbone of DNA [37–39]. Fig. 6A and B illustrate the electrostatic surface potential on the solvent-accessible surface of Rv0543c. The most significant feature is a negatively charged surface (red) on one face (Fig. 6A) with only a few small pockets of positively charged surfaces (blue), observations opposite to what would be expected if Rv0543c was a DNA binding protein. Furthermore, ConSurf analysis [25] of the structure of Rv0543c shows that the solvent-accessible surface of the negatively charged areas is more conserved than for the positively charged areas (Fig. 6C and D), suggesting the negatively charged surface may have more functional significance than the positively charged pockets.

Fig. 7 contains a ClustalW2 [26] sequence alignment of 10 proteins in the DUF3349 superfamily that do not belong to the genera *Mycobacterium* or *Rhodococcus*. The residues identified as conserved by ClustalW2 and ConSurf are marked with notations on the bottom of the alignment, and in general, they identify similarly conserved regions of the protein. The ClustalW2 alignment identifies 11 identical amino acid residues in all 10 sequences; eight hydrophobic, one hydrophilic, one negatively charged, and one positively charged. Twelve conserved residues are identified by ClustalW2; 10 hydrophobic and two negatively charged. The ClustalW2 identification of only one conserved negatively charged residue in the DUF3349 family further rules out the possibility that these proteins are involved with a DNA-binding function. Instead, most of the amino acid conservation in the DUF3349 family involves hydrophobic residues and these residues are likely responsible for driving the five  $\alpha$ -helices to adopt their tertiary fold. This conversation of the interior of the protein is better shown in Fig. 2B, a ConSurf color-coded ribbon representation of the

		<b><math>\alpha</math>1</b>		<b><math>\alpha</math>2</b>		<b><math>\alpha</math>3</b>															
<i>Mycobacterium tuberculosis</i>	1	----	MNRFLTSIVAWLRAGY	PE	GIPPTBSFAVLALL	CRRLSHDEVKAVANELMR	----	50													
<i>Janibacter sp. HTCC2649</i>	110	----	RPTGERARPLERIVSWLR	EGYPOGVPPQ	DFPLLALLRRRLS	DDDEIKAVAKSLRKA	----	57													
<i>Nocardia farcinica</i>	1	----	MPPFLQAIIDWLRAGY	PEGVPSDY	IPLLAVLRRRLS	DDDEVAIASAVYGS	----	51													
<i>Frankia alni</i>	1	----	MLASIIQWLRAGY	PDGVP	RDYLPFALLRRRL	TGDEVTAIADALTTN	----	48													
<i>Tsakamurella paurometabola</i>	2	----	SSGQGFIAKIVD	WLRAGY	PEGVPSHDY	QPLLALLDRQLTKEEVKEVTRELIDDAHLT	----	58													
<i>Gordonia bronchialis</i>	1	----	MDRPQFLKRVVW	WLRAGYPNGV	PDYIPVALLRRRLS	DDDEVAIASAVYGS	----	57													
<i>Clavibacter michiganensis</i>	137	----	RATERITGLVSRIVAWLR	AGYPAGLPE	QDFVPLLALLRRRLS	DDDEVAIASAVYGS	----	193													
<i>Kineococcus radiotolerans</i>	2	----	TSTGSTNLATR	RVWLRAGY	PGGVPRQ	DYVVLGGLRRRLTDEEVHGIAAELAGLAGQ	----	59													
<i>Kocuria rhizophila</i>	86	----	PPROPGYVARVMNWL	RADYPNGV	PAQDYLPLVAVLRP	QLNDEEDMAAIAIQADAREQ	----	144													
<i>Corynebacterium glutamicum</i>	2	----	SHNDSPNFARRALN	WLROGYPT	GVPRHDT	TFALFYVLERELTEEDLNELAEILLIAGENN	----	60													
<i>Rothia mucilaginosa</i>	2	----	PKSMISRVI	DWLSVGY	PKVVPVQ	DRAAVMAILKMRLTDEQLQEVVROLMQSRAR	----	57													
ClustalW2																					
ConSurf																					
			<b><math>\alpha</math>4</b>		<b><math>\alpha</math>5</b>																
<i>Mycobacterium tuberculosis</i>	51	----	LGDFDQIDIGVVVITH	FTHDEL	SPEDVERVRARLAA	QGWP	LD	-----	VRDREESA	-----	100										
<i>Janibacter sp. HTCC2649</i>	58	----	VSPVTPADIGVEIT	RVTQEL	PSDEDLRVRARLE	KKGW	PLD	-----	FPERAQQG	-----	215										
<i>Nocardia farcinica</i>	52	----	GDRT	---	DIQVLI	TKVTNEMP	SEQDVARV	QERLAG	REWPE	SE	-----	92									
<i>Frankia alni</i>	49	----	GDSASALAIRTAIA	IVTHD	TPLGDI	ARVSARLA	AGGW	PLAS	-----	PSRPRPATET	-----	100									
<i>Tsakamurella paurometabola</i>	59	----	PEGVEPITRIDAG	VAISGHTRE	MPLEKDI	ARVRERL	ERKGV	PFDD	APLR	-----	ADEPPRE	-----	113								
<i>Gordonia bronchialis</i>	58	----	PE	---	RISKID	AGVKIT	ITHEL	PHEADI	ARVRKHL	EASGW	PFDD	PLPPY	PTS	DAHD	PD	TD	DP	-----	119		
<i>Clavibacter michiganensis</i>	194	----	GLPASRLDVA	DAIAG	VTSE	MPSSAD	VERVR	AYLE	AHGW	PEG	-----	FRI	-----	236							
<i>Kineococcus radiotolerans</i>	60	----	GEVITVAD	VERLVNE	ATLDE	PAAAD	VARV	SSHLA	AGGW	PLA	-----	HPQA	-----	105							
<i>Kocuria rhizophila</i>	145	----	GVAPSE	ADARA	AIQ	RATED	EP	PKPE	ELARI	RAML	SAHGW	PFEG	-----	187							
<i>Corynebacterium glutamicum</i>	61	----	GLHDNP	ITREK	IGKIL	ITHV	HSQ	PEDE	IDRI	OKKL	QAE	GF	PT	RN	-----	105					
<i>Rothia mucilaginosa</i>	58	----	EAYVSD	QRINE	YIRRV	VVDHT	PT	QIDR	VAKIL	GAHG	FLIA	NHVN	EDA	EP	TAGG	TE	YV	TY	DD	PT	123
ClustalW2																					
ConSurf																					

**Fig. 7.** Sequence alignment of Rv0543c with 10 other closely related sequences from genera other than *Mycobacterium* and *Rhodococcus* using the program ClustalW2 [26]. The proteins, in order are: Rv0543c, JNB\_16594, nfa52480, FRAAL1463, Gbro\_1053, CMS\_2894, Krad\_0190, KRH\_23040, NCg10440, ZP\_053675541.1. The residues are color coded as follows: cyan = hydrophobic and small (AVFPMILW), red = acidic (D and E), purple = basic (K and R), yellow = hydrophilic (STYHNCQG). Below the alignment is shown the position of residues that are highly conserved following ClustalW2 (\* = identical, : = conserved substitutions, and . = semi-conserved substitutions) and ConSurf [25] (scores of 9 = asterisk, 8 = caret) analysis of the sequences. The ClustalW2 analysis used only the amino acid sequence of Rv0543c and the 10 other proteins listed in Fig. 7 while the ConSurf analysis used the complete RCSB-PDB and SWISS-PROT databases. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

structure of Rv0543c, where a substantial number of the highly conserved residues (purple and pink) are either in the interior (much of  $\alpha 2$ ) or pointing towards the interior (C-terminal of  $\alpha 1$ , approximately half of  $\alpha 4$  and  $\alpha 5$ ) of the structure.

## Conclusions

Rv0543c is a monomeric bundle of five  $\alpha$ -helices rigidly held together by a largely conserved group of hydrophobic amino acid side chains. With a melting temperature of 62.5 °C, the protein is somewhat resistant to thermal denaturation. However, upon thermal denaturation the protein is unable to refold into its native conformation. A DALI search of the protein structures in the RCSB-PDB returned hits with low *Z*-scores ( $\leq 4.4$ ) and the hits with *Z*-scores above 4.0 were all Cre recombinases. However, both the electrostatic surface potential and ConSurf identified conserved surface on the solvent-accessible surface of Rv0543c are not consistent with a DNA-binding function for the protein. Consequently, the fold observed for Rv0543c may be unique for Rv0543c and the other members of the DUF3349 superfamily. Associated with this unique fold may be a unique function that still needs to be discovered, perhaps a function associated with the menaquinone-biosynthetic pathway since the Rv0543c gene is co-localized with genes from this pathway on the chromosome. When a function is found, the structure for Rv0543c presented here will assist the molecular understanding of its mechanism and potentially speed up the conception and development of new therapies to treat and control the spread of tuberculosis if the function turns out to be vital for *M. tuberculosis* survival in the microenvironment of the host.

## Acknowledgments

The structure of Rv0543c was a community request made to the Seattle Structural Genomics Center for Infectious Disease (SSGCID) and was given the internal identification code MytuD.17712.a. The research was funded by the National Institute of Allergy and Infectious Diseases under Federal Contract No. HHSN272200700057C and performed primarily at the W.R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by U.S. Department of Energy's Office of Biological and Environmental Research program located at Pacific Northwest National Laboratory (PNNL). Battelle operates PNNL for the U.S. Department of Energy.

## References

- [1] Global Tuberculosis Control, A Short Update to the 2009 Report, World Health Organization, Geneva, 2009.
- [2] G. Kaplan, F.A. Post, A.L. Moreira, H. Wainwright, B.N. Kreiswirth, M. Tanverdi, B. Mathema, S.V. Ramaswamy, G. Walther, L.M. Steyn, C.E. Barry III, L.G. Bekker, *Infect. Immun.* 71 (2003) 7099–7108.
- [3] D.A. Enarson, H. Hahn, in: S.H.E. Kaufmann (Ed.), *Mycobacteria and TB*, Karger, Basel, 2003, pp. 1–16.
- [4] D.G. Russell, C.E. Barry III, J.L. Flynn, *Science* 328 (2010) 852–856.
- [5] D.A. Mitchison, *Eur. Res. J.* 25 (2005) 376–379.
- [6] S. Basu, G.H. Friedland, J. Medlock, J.R. Andrews, N.S. Shah, N.R. Gandhi, A. Moll, P. Moodley, A.W. Sturm, A.P. Galvani, *Proc. Natl. Acad. Sci. USA* 106 (2009) 7672–7677.
- [7] S.L. Kamholz, *J. Assoc. Acad. Minor. Phys.* 13 (2002) 53–56.
- [8] J.C. Palomino, D.F. Ramos, P.A. da Silva, *Curr. Med. Chem.* 16 (2009) 1898–1904.
- [9] C. Dye, B.G. Williams, *Science* 328 (2010) 856–861.
- [10] G.W. Buchko, C.Y. Kim, T.C. Terwilliger, M.A. Kennedy, *J. Bact.* 188 (2006) 5993–6001.
- [11] G.W. Buchko, C.Y. Kim, T.C. Terwilliger, P.J. Myler, *Tuberculosis* 90 (2010) 245–251.
- [12] N.J. Greenfield, *Nat. Prot.* 6 (2006) 2527–2535.
- [13] T.D. Goddard, D.G. Kneller, Sparky 3, University of California, San Francisco.
- [14] D.S. Wishart, C.G. Bigam, J. Yao, F. Abildgaard, H.J. Dyson, E. Oldfield, J.L. Markley, B.D. Sykes, *J. Biol. NMR* 6 (1995) 135–140.
- [15] N.A. Farrow, D.R. Muhandiram, A.U. Singer, S.M. Pascal, L.E. Kay, G. Gish, S.E. Shoelson, T. Pawson, J.D. Forman-Kay, *Biochemistry* 33 (1994) 5984–6003.
- [16] N.J. Skelton, A.G. Palmer III, M. Akke, J. Kordel, M. Rance, W.J. Chazin, *J. Magn. Res. B* 102 (1993) 253–264.
- [17] G.W. Buchko, G.W. Daughdrill, R. De Lorimier, S. Rao, N.G. Isern, J. Lingbeck, J.-S. Taylor, M.S. Wold, L.D. Gochin, L.D. Spicer, D.F. Lowry, M.A. Kennedy, *Biochemistry* 38 (1999) 15116–15128.
- [18] T. Szyperki, D. Yeh, D. Sukumaran, H. Moseley, G. Montelione, *Proc. Natl. Acad. Sci. USA* 99 (2002) 8009–8014.
- [19] G.W. Buchko, B.J. Tarasevich, J. Bekhazi, M.L. Snead, W.J. Shaw, *Biochemistry* 47 (2008) 13215–13222.
- [20] P. Guntert, *Meth. Mol. Biol.* 278 (2004) 353–378.
- [21] G. Cornilescu, F. Delaglio, A. Bax, *J. Biol. NMR* 13 (1999) 289–301.
- [22] J.P. Linge, M. Nilges, *J. Biol. NMR* 13 (1999) 51–59.
- [23] A. Bhattacharya, R. Tejero, G. Montelione, *Proteins* 66 (2007) 778–795.
- [24] L. Holm, P. Rosenstrom, *Nucl. Acids Res.* 38 (2010) W545–W549.
- [25] M. Landau, I. Mayrose, Y. Rosengerg, F. Glaser, E. Martz, T. Pupko, N. Ben-Tal, *Nucl. Acids Res.* 33 (2005) W299–W302.
- [26] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, *Bioinformatics* 23 (2007) 2947–2948.
- [27] W.L. DeLano, The PyMOL Molecular Graphics System, DeLano Scientific LLC, Palo Alto, CA, USA, 2008.
- [28] M. Palczewska, P. Groves, A. Ambrus, A. Kaeta, K. Kover, G. Batta, J. Kuznicki, *Eur. J. Biochem.* 268 (2001) 6229–6237.
- [29] R.W. Woody, *Studies of Theoretical Circular Dichroism of Polypeptides: Contributions of  $\beta$ -Turns*, John Wiley & Sons, New York, 1974.
- [30] G.M. Holzwarth, P. Doty, *J. Amer. Chem. Soc.* 87 (1965) 218–228.
- [31] G.W. Buchko, N.J. Hess, V. Bandaru, S.S. Wallace, M.A. Kennedy, *Biochemistry* 40 (2000) 12441–12449.
- [32] I. Karantzeni, C. Ruiz, C.-C. Liu, V.J. LiCata, *Biochem. J.* 374 (2003) 785–792.
- [33] T.E. Creighton, *Proteins: Structure and Molecular Properties*, H.W. Freeman & Company, New York, 1993.
- [34] R. Ishima, D.A. Torchia, *Nat. Struct. Biol.* 7 (2000) 740–743.
- [35] D.I. Freedberg, R. Ishima, J. Jacob, Y.-X. Wang, I. Kustanovich, J.M. Louis, D.A. Torchia, *Protein Sci.* 11 (2001) 221–232.
- [36] E.E. Metcalfe, J. Zmoon, D.D. Thomas, G. Veglia, *Biophys. J.* 87 (2004) 1205–1214.
- [37] F. Guo, D.N. Gopaul, G.D. Van Duyne, *Nature* 389 (1997) 40–46.
- [38] K. Ghosh, C.K. Lau, F. Guo, A.M. Segall, G.D. Van Duyne, *J. Biol. Chem.* 280 (2005) 8290–8299.
- [39] E. Ennifar, J.E.W. Meyer, F. Buchholz, A.F. Stewart, D. Suck, *Nucl. Acids Res.* 31 (2003) 5449–5460.
- [40] R. Maiti, G.H. Domselaar, H. Zhang, D.S. Wishart, *Nucl. Acids Res.* 23 (2004) W590–W594.