NIH Public Access
Author Manuscript
Tuberculosis (Edinb). Author manuscript; available in PMC 2011 July 1.

Published in final edited form as:
Tuberculosis (Edinb). 2010 July ; 90(4): 245–251. doi:10.1016/j.tube.2010.04.002.

Solution structure of Rv2377c - founding member of the MbtH-like protein family

Garry W. Buchko[a,*], **Chang-Yub Kim**[b], **Thomas C. Terwilliger**[b], and **Peter J. Myler**[c,d]

Garry W. Buchko: garry.buchko@pnl.gov; Chang-Yub Kim: cykim@lanl.gov; Thomas C. Terwilliger: terwilliger@lanl.gov; Peter J. Myler: peter.myler@sbri.org

[a]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA

[b]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

[c]Seattle Biomedical Research Institute, 307 Westlake Avenue N, Seattle, WA 98109-5219, USA

[d]Department of Medical Education and Biomedical Informatics & Department of Global Health, University of Washington, Seattle, WA 98195, USA

Summary

The *Mycobacterium tuberculosis* protein Rv2377c (71 residues, MW = 8.4 kDa) has been characterized using nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy. Rv2377c was the first identified member of the MbtH-like family of proteins. MbtH-like proteins have been implicated in siderophore biosynthesis, however, their precise biochemical function remain unknown. Size exclusion chromatography and NMR spectroscopy show that Rv2377c is a monomer in solution. Circular dichroism spectroscopy indicates that Rv2377c unfolds upon heating and will reversibly fold into its native conformation upon cooling. Using NMR-based methods the solution structure of Rv2377c was determined and some of the dynamic properties of the protein studied. The protein contains a three-strand, anti-parallel β-sheet (β3:β1:β2) nestled against one C-terminal α-helix (S44-N55). Weak or absent amide cross peaks in the ^{1}H-^{15}N HSQC spectrum for many of the β1 and β2 residues suggest intermediate motion on the ms to μs timescale at the β1:β2 interface. Amide cross peaks in the ^{1}H-^{15}N HSQC spectrum are absent for all but one residue at the C-terminus (W56 - D71), a region that includes a highly conserved sequence WXDXR, suggesting this region is intrinsically disordered. The latter observation differs with the crystal structure of another MbtH-like protein, PA2412 from *Pseudomonas aeruginosa*, where a second ordered α-helix was observed at the extreme C-terminus.

Keywords

tuberculosis; siderophore assembly; mycobactin; circular dichroism; structural genomics; protein dynamics

[*]Corresponding Author: Dr. Garry W. Buchko, Biological Sciences Division, Pacific Northwest National Laboratory, P.O. Box 999, Mail Stop K8-98, Richland, WA 99352, garry.buchko@pnl.gov, Phone: 509-371-6543, Fax: 509-376-2303.

Funding: Please see Acknowledgments.

Competing interests: None declared.

Ethical approval: Not required.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

1. Introduction

Mycobacterium tuberculosis, the aetiological agent responsible for the chronic infectious disease tuberculosis (TB), annually kills 1.6 million people and newly infects another 8-9 million.[1,2] These levels of TB infection show no signs of decreasing in the near future due to the emergence of multi-drug and extremely drug-resistant *M. tuberculosis* strains along with the prevalence of high human immunodeficiency virus infections.[3-5] Conventional treatment of TB is a short course of multiple drugs over a long period of time, typically isoniazid, rifampin, pyrazinamide, and ethambutol for two months followed by isoniazid and rifampin for four months.[6] While such a therapy is often successful, incomplete implementation contributes to the increasing levels of multi-drug (to at least isoniazid and rifampin) and extreme-drug (to isoniazid and rifampin plus three or more second-line agents) resistant *M. tuberculosis* strains.[7] Consequently, it is of fundamental importance to develop a new generation of anti-TB agents.

One target of novel anti-TB drugs focuses on the disruption of the biochemical pathways responsible for the sequestration of iron by *M. tuberculosis*.[8,9] Iron is an essential catalytic and structural component of many proteins[10] and one of the defense mechanisms of infected eukaryotic hosts is to dramatically restrict its availability to the bacteria.[11] To overcome iron deficiencies in the host, *M. tuberculosis* and other bacteria have evolved sophisticated iron-acquisition systems that make use of small molecules, called siderophores, that are excreted into extracellular space, bind available iron, and then reinternalized with the iron into the bacteria via specific cell surface receptors.[12,13] *M. tuberculosis* produces two types of siderophores, called mycobactins, that are based on a 2-hydroxyphenyloxazoline-ring system and differ by the nature of a single fatty acyl side chain linked to the ε-amino group of N-hydroxy lysine.[14-16] These two siderophores are biosynthesized through the action of a number of proteins located in a cluster of genes involved in nonribosomal peptide and polyketide synthesis.[17] One of these genes encodes a small, 71-residue protein, MbtH,[18] that is the founding member of the MbtH-like (*Mycobacterium tuberculosis*) family of proteins present in numerous non-ribosomal protein synthetase clusters responsible for siderophore and antibiotic peptide synthesis. The function of the MbtH-like proteins remain unknown with only one structure, PA2412 from *Pseudomonas aeruginos*,[19] determined for the 150-plus members of this family. To obtain further insights into the biological function of the MbtH-like family of proteins that may enable the conception and development of new therapies and strategies to treat and control TB, the solution structure for the first member of the MbtH-like protein family, Rv2377c, has been determined by NMR-based methods. The dynamics of Rv2377c is explored, the solution structure compared to the crystal and solution structures of PA2412, and the thermostablity of Rv2377c assayed using circular dichroism spectroscopy.

2. Materials and Methods

2.1 Cloning, expression, and purification

The DNA coding sequence for the *Rv2377c* gene was cloned into a modified pET28b vector (Novagen, Madison, WI) that provided an N-terminal 6-histidine tag (MGSSHHHHHHSSGLVPRGSH-) upstream of the NdeI site. This recombinant vector was then transfected into the host *Escherichia coli* bacterial strain BL21PRO (Clontech, Palo Alto, CA). Uniformly ^{15}N- and ^{15}N-, ^{13}C-labeled Rv2377c was obtained following established protocols that have been previously described in detail[20] using the antibiotics kanamycin and spectinomycin during cell growth in minimal medium, isopropyl β-D-1-thiogalactopyranoside induction at 298 K, initial purification of the French pressed cell lysate on a Ni-NTA affinity column (Qiagen, Valencia, CA), thrombin cleavage of the N-terminal polyhistidine tag, and final purification on a Superdex75 HiLoad 10/30 column (GE

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Healthcare, Piscataway, NJ) that simultaneously exchanged Rv2377c into NMR buffer (300 mM NaCl, 20 mM TrisHCl, 1.0 mM dithiothreitol, pH 7.1). The final concentrations of the NMR samples were between 1 and 2 mM.

2.2 Optical spectroscopy

Circular dichroism data were collected on a Aviv Model 410 spectropolarimeter (Lakewood, NJ) calibrated with an aqueous solution of ammonium d-(+)camphorsulfonate. Measurements were obtained on a Rv2377c sample (0.06 mM) in NMR buffer and in a quartz cell of 0.1 cm path length. A thermal denaturation curve for Rv2377c was obtained by recording and plotting the ellipticity at 210 nm in 2.0°C intervals from 10 to 80°C. Wavelength scans for Rv2377c were recorded between 200 and 260 nm at 25°C, 80°C, and 25°C (post-heating) in 0.5 nm increments. Each 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. The wavelength spectra were processed by subtracting a blank spectrum from the protein spectrum and then automatically line smoothing the data using Aviv software. A rough estimation of the elements of secondary structure in the protein at 25°C, presented in Table 2, was obtained by deconvoluting the spectrum using Aviv software.

2.3 NMR spectroscopy and resonance assignments

All NMR experiments were collected on 1 - 2 mM samples at 25°C using Varian 900-, 750-, 600- and 500-Inova spectrometers equipped with triple resonance probes and pulse field gradients. The data were processed with Felix2007 (Felix NMR, Inc., San Diego, CA) and analyzed with Sparky (v3.115). All chemical shifts were referenced to DSS (DSS = 0 ppm) using indirect methods.[21]

The ^{1}H, ^{13}C, and ^{15}N chemical shifts of the backbone and side chain resonances were obtained from standard two-dimensional ^{1}H-^{15}N and ^{1}H-^{13}C HSQC spectra and three-dimensional HNCACB, CBCA(CO)NH, HNCACB, HNCO, HCCH-TOCSY, HBHA(CO)NH, HCC-TOCSY-NNH and CC-TOCSY-NNH spectra. Distance restraints were obtained from a suite of three-dimensional, ^{13}C- and ^{15}N-edited NOESY-HSQC experiments using a mixing time of 80 ms. Deuterium-exchange studies were performed by lyophilizing a NMR sample and re-dissolving in 99.8% D_2O (the ^{1}H-^{15}N HSQC spectrum of a lyophilized sample re-dissolved into 10% D_2O/90% H_2O was essentially identical to the ^{1}H-^{15}N HSQC spectrum obtained prior to lyophilization). Two-dimensional ^{1}H-^{15}N HSQC spectra were recorded 0.2, 0.5 and 1.0 hour after the exchange. Steady-state $\{^{1}H\}$-^{15}N heteronuclear NOE values (NOE = I_{sat}/I_{unsat}) were measured from the ratios of ^{1}H-^{15}N HSQC cross peak volumes in spectra recorded in the presence (I_{sat}) and absence (I_{unsat}) of three seconds of proton presaturation prior to the ^{15}N excitation pulse.[22] An overall rotational correlation time (τ_c) for Rv2377c was rapidly estimated from backbone amide ^{15}N $T_{1\rho}/T_1$ ratios measured using a modified ^{1}H-^{15}N HSQC experiment to record ^{15}N-edited one-dimensional spectra.[23] All pulse programs were from the Varian BioPack suite.

2.4 Structure calculations

Structure calculations were performed iteratively using CYANA (v 2.1)[24] and the ^{1}H, ^{13}C, and ^{15}N chemical shift assignments and peak-picked NOESY data as experimental inputs. A total of 66 dihedral angle restraints for Phi (Φ) and Psi (Ψ) were introduced on the basis of the elements of secondary structure identified in the early structural ensembles and TALOS calculations.[25] Twenty-six 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, for residues in β-sheet regions, the observation of slowly exchanging amides in the deuterium exchange experiment. Alpha-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

helical hydrogen bond networks generally exchange before hydrogens in β-strand hydrogen bond networks[26] and, for the small protein Rv2377c, had all exchanged before the first ^{1}H-^{15}N HSQC spectrum could be acquired (< 12 minutes).

The final ensemble of 20 CYANA derived structures were then refined with explicit water[27] using force constants of 500 and 700 kcal for the NOE and dihedral restraints, respectively. The upper boundary of the CYANA distance restraints was increased by 10% and the lower bound was set to the vdw limit for the water refinement calculations. This final ensemble was then used to calculate a mean structure and average RMSD values to the mean structure using the structured regions of the protein (F13 - N55). Structural quality was assessed using the Protein Structure Validation Suite (PSVS, v1.3).[28] The structural statistics are summarized in Table 1.

2.5 Protein structure accession numbers

The atomic coordinates for the ensemble of 20 lowest energy structures for *M. tuberculosis* Rv2377c have been deposited in the RCSB under PDB code 2KHR. The chemical shifts assignments have been deposited with the BMRB under accession number 16253. 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 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.

3. Results and Discussion

3.1 Solution structure for Rv2377c

The elution time of Rv2377c on a size exclusion column was consistent with a monomeric 8,383 Da protein (data not shown). The experimentally estimated rotational correlation time of 6.7 ± 0.2 ns (20°C) was also more consistent with a monomeric instead of a dimeric protein.[29] The line widths and chemical shift dispersion of the ^{1}H-^{15}N HSQC spectrum for Rv2377c, shown in Figure 1, was also characteristic of a folded, monomeric protein of molecular weight in the 8.4 kDa range. Forty-nine amide resonances were observed in the ^{1}H-^{15}N HSQC spectrum and 48 of these were unambiguously assigned. On the basis of these amide assignments and extensive assignment of the $^{13}C^{\alpha}$ and side chain proton and carbon chemical shifts (BMRB ID 16253), an ensemble of structures were calculated (Figure 2A) that satisfied all the available experimental NMR data (NOEs, chemical shifts, deuterium exchange experiments, and TALOS calculations).

As summarized in Table 1, a total of 602 interproton distance restraints, 26 hydrogen bond restraints, and 66 dihedral angle restraints were used in the final structure calculations. Each member of the final ensemble of 20 calculated structures agreed well with the experimental data with no upper limit violation greater than 0.05 Å and no torsion angle violation greater than three degrees. The quality of the structure ensemble was also shown to be good using the PSVS validation software package.[28] The Ramachandran statistics for all the residues in the ensemble were overwhelmingly in acceptable space (87% of the (φ,ψ) pairs for Rv2377c were found in the most favored regions and 12% within additionally allowed regions) and all the structure quality Z-scores were acceptable (greater than minus five).

The final set of 20 calculated structures in the ensemble converge well, as shown mathematically by the statistics in Table 1 and visually by the superposition in Figure 2A. The RMSD of the structured core region (F13 - N55) in the ensemble to the mean structure is 0.62 Å for the backbone atoms (N-C^{α}-C=O) and 1.10 Å for all heavy atoms. Both the N- and C-terminal regions, G1*- A12 and W56 - D71, respectively, are disordered, and therefore, most of the residues at the termini are not shown in the ensemble in Figure 2A.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Note that there is evidence for intermediate motion (ms to μs) at the interface between β1 and β2 (Section 3.3), and consequently, due to dynamic averaging of NOEs that bias shorter distances, the ensemble of calculated structures may also be slightly biased in this region. A cartoon representation of the secondary structure of Rv2377c is shown in Figure 2B. The central core of the protein contains a three-strand, anti-parallel β-sheet composed of residues F13-N18 (β1), S23-P26 (β2), and W35-H39 (β3). The third β-strand contains a one residue β-bulge (V38) and overall the β-sheet contains a conspicuous β-sheet twist[30] as illustrated in the ensemble and the single structure closest to the average in Figure 2A and 2C, respectively. Completing the structure is a single α-helix (S44 - N55) nestled on one side of the β-sheet that is held in place by an hydrophobic core of side chains: V15 (β1), L25 (β2), V39 (β3), Y51 (α1) and V52 (α1).

3.2 Optical spectroscopy

Circular dichroism spectroscopy was used to characterize the secondary structure of Rv2377c over a range of temperatures.[31] Figure 3A shows the CD spectra for Rv2377c at two temperatures. The solid 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 ~225 and ~207 nm and projected maximum at a wavelength < 200 nm is characteristic of a structured protein with a mixture of β-sheet, α-helical, and random coil content.[31,32] Such a mixture of secondary structure is corroborated by the deconvolution of the CD data and the calculated NMR-based structure, as shown in Table 2. The red dotted line in Figure 3A is the CD spectrum of the same sample at 80°C. Relative to the spectrum collected at 25°C, the minimum at ~225 nm has disappeared and the minimum originally at ~207 nm has increased and shifted ~ 2 nm to shorter wavelength. These observations at 80°C indicate that Rv2377c is more unstructured at elevated temperatures. However, the absence of an extrapolated negative minimum at ~198 nm and a positive maximum at ~218 nm indicates that Rv2377c is not entirely random coil at 80°C and still possesses spectral properties consistent with primarily β-sheet structure (single minimum). When the sample is cooled back to 25°C, the CD spectrum, indicated by the dashed blue line in Figure 3A, is very similar to the original CD spectrum collected prior to heating (solid black line). This indicates that the temperature-induced effects to the structure of Rv2377c are largely reversible.

To assay the thermal stability of Rv2377c, the ellipticity at 220 nm was measured as a function of temperature between 5 and 80°C. Typically, a phase transition can be detected when a structured protein becomes denatured by monitoring the increase in the ellipticity at 220 nm with increasing temperature.[33,34] As shown in Figure 3B, a gradual increase in ellipticity at 220 nm is observed beginning at ~20°C with a tailing-off starting at ~70°C. There isn't a well-defined inflection point, but, the midpoint of the plot is around 55°C. Because the CD spectra in Figure 3A suggest that at 80°C the protein is not entirely unstructured and will refold upon cooling back to 25°C, the end point of the temperature study in Figure 3B likely does not represent a fully unstructured protein. Because β-strands are typically more robust than α-helices[26] the structure remaining at high temperature may be primarily β-strands as suggested by the CD spectrum at 80°C in Figure 3A.

3.3 Dynamics

Protein dynamics often plays a role in the binding and catalysis properties of proteins at, and around, active sites.[35] Information regarding the dynamics along a protein's backbone can be extracted from the properties of the amide resonances. For example, in the absence of intermolecular associations, weak or absent amide cross peaks in the ^{1}H-^{15}N HSQC spectrum often identify regions that are not rigidly structured in solution and qualitatively implies that these residues are undergoing conformational exchange on an intermediate

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

timescale (millisecond to microsecond) that broadens their resonances beyond detection.[35-37] The plot in Figure 4A indicate the residues with missing (solid purple column) or weak (purple asterisk) amide cross peaks in the ^{1}H-^{15}N HSQC spectrum of Rv2377c in Figure 1. Aside from K62, amide cross peaks are absent from W56 to D71 suggesting that the C-terminal region is unstructured in solution. Indeed, the adjacent α-helix is also somewhat dynamic (breathes) as the amide resonances in α1 all exchanged with D_2O 12 minutes into the deuterium exchange experiment. The other region with a significant clustering of weak or missing amide resonances is β2 (S23 - P26) where the amide resonances were absent. While none of the β2 amide resonances were visible, most of the side chain carbon and proton resonances for β2 were assigned. The $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts in β2 were characteristic of a β-strand and ^{1}H-^{1}H NOEs were observed between side chains of β1 and β2. The resonances of the amides in β1 facing β2, F14, L16, and N18, were also weak. Collectively, this data indicates that the S23 - P26 sequence does form a β-strand and interacts with β1, however, there is motion on the intermediate time scale (ms to μs) between the two strands.

Backbone motion on the fast time scale (picosecond) may be identified by regions with small, or negative, heteronuclear steady-state $\{^{1}H\}$-^{15}N NOE values.[38] Figure 4B is a plot of the backbone $\{^{1}H\}$-^{15}N heteronuclear NOE values measured for Rv2377c. Due to spectral overlap, and weak or missing amide resonances, the only region with continuous data is between D31 - W56. The heteronuclear steady-state $\{^{1}H\}$-^{15}N NOE values in this region are all 0.7 or greater indicating an overall structure that is rigid and an absence of residues undergoing motion on a picosecond timescale.

3.4 An intrinsically disordered region containing a highly conserved sequence

The MbtH-like family of proteins contains over 150 members and they all share highly conserved regions as illustrated for the two MbtH-like proteins with solved structures in Figure 5.[19] Universally conserved are three tryptophan residues, W25, W35, and W56. The relative orientation of these side chain are illustrated for Rv2377c in Figure 2B. The side chain indole rings of W25 and W35 sit parallel to each other at a distance of ~9 Å, but, are not immediately over top each other. This is in slight contrast to the corresponding indole rings in both the crystal and NMR structure of PA2412 from *P. aeruginosa* that sit more directly over top each other at a closer distance of ~7 and ~8 Å, respectively. One possible explanation for the difference between the crystal structure of PA2412 and solution structure of Rv2377c is that F_o - F_c differential electron density was observed in the tryptophan pocket suggesting a small molecule of unknown identity was bound to the protein in the crystal structure.[19] The third tryptophan, W56, is the first residue in the C-terminal highly conserved sequence WXDXR that is in a disordered region of the protein. Another highly conserved region, SXWP, is part of β2, a structured region with motion in the intermediate time scale. The two other highly conserved regions, (V/L)XN and PXGW, form β1 and the loop plus first residue of β3, respectively. Hence, of the four highly conserved regions of Rv2377c, two are associated with mobile elements and two are more rigid.

The last highly conserved region in MbtH-like proteins, WXDXR, is in a disordered region of the protein. Intrinsically disordered proteins represent a unique functional and structural category of proteins[39, 40] implicated in several vital biological functions.[41] As the name suggests, these regions lack a rigid, well-ordered structure and exist as a highly dynamic ensemble in solution. Intrinsically disordered proteins or regions of proteins are often involved in binding interactions with other proteins, nucleic acids, or small molecules, undergoing a disorder to order transition upon ligand binding.[42] Relationships between amino acid sequence and intrinsic disorder have been identified, and consequently, it is possible to predict the tendency of a polypeptide to be disordered (and ordered) from the primary amino acid sequence using programs such as PONDR (www.pondr.com).[43, 44] The

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

top of Figure 4 is a graphical output of a PONDR prediction for Rv2377c plotted over the experimentally determined secondary structure. The PONDR score for residues A12 - D50 are all below 0.2, values highly predictive of an ordered structure. These residues embrace all the ordered regions of the protein determined from the NMR-based structure, including β1 and β2 residues with absent or weak amide resonances in the ^{1}H-^{15}N HSQC spectrum. Residues at both the N- and C-termini have PONDR scores above 0.5, values predictive of disordered structure. These termini are disordered in the ensemble of calculated NMR-based structures for Rv2377c. While it is not unusual to observed disordered regions at the termini of proteins, the presence of a highly conserved region near the C-terminal of Rv2377c suggests this may be an intrinsically disordered region of the protein still in search of its binding partner.

Further evidence that the highly conserved WXDXR sequence is intrinsically disordered is from the solution structure determined for PA2412. Figure 6 compares the XRD- and NMR-determined structures for PA2412 with the NMR-determined structure for Rv2377c. The core of each structure is similar, a three-strand anti-parallel β-sheet with a C-terminal α-helix. However, in the crystal structure of PA2412, a short, well-ordered second α-helix was also observed at the extreme end of the C-terminal. While a subset of the NMR-determined structures for PA2412 also contain a second α-helix, Figure 6 clearly indicates that this region exists as a highly dynamic ensemble in solution, as observed for Rv2377c. The dynamic C-terminal region of PA2412 also contains the highly conserved WXDXR sequence. This means that each of these two independently determined solution structures of proteins in the MbtH-like family show the highly conserved WXDXR is disordered in solution. The second ordered helix observed in the PA2412 crystal structure may be a state stabilized by crystal packing interactions, or indeed, be due to the binding of an unknown small ligand suggested by the presence of unidentifiable electron density in the crystal structure.[19]

3.5 Insights into biological function

To identify a possible biochemical function for Rv2377c based on its structure, the Protein Data Bank was searched for structures with similarities to Rv2377c using the DALI search engine.[45] Only two proteins with DALI Z-scores greater than 3.0 were identified and these were the crystal (4.6) and NMR (3.9) structures determined for the *P. aeruginosa* protein PA2412. The confinement of the protein fold to the MbtH-like family of proteins may indicate that these proteins have a unique biological function.

While the exact biological function for MbtH-like family of proteins is unknown, there is substantial evidence that this protein plays a role in helping *M. tuberculosis* survive low-iron conditions. DNA microarray studies of *M. tuberculosis* response to low-iron and high-iron conditions show elevated levels of mRNA encoding Rv2377c under low-iron conditions.[46] Under conditions more relevant to biological infection, inside macrophage, elevated levels of mRNA for MbtH and other genes in the nonribosomal protein synthetase clusters responsible for siderophore and antibiotic peptide synthesis were found to be intraphagosomally induced using microarray experiments.[47] Some bacteria contain more than one non-ribosomal protein synthetase cluster each with their own or multiples copies of *mbtH*-like genes. It has been shown that the *mbtH*-like genes in *Streptomyces coelicolor* are able to cross-talk between biosynthetic pathways and can functionally replace each other.[48,49] On the other hand, the entire genome of *P. aeruginosa* contains only one MbtH-like gene, *PA2412*. Studies with a deletion mutant of PA2412 demonstrated that the MbtH-like protein was necessary for the normal production or secretion of pyoveridine, a siderophore found in *P. aeruginosa*.[19] Because the addition of pyoveridine to the PA2412 mutant cell line restored normal growth it was concluded that the MbtH-like protein was likely involved in the maturation, stabilization, or export of the siderophore. The function of PA2412 likely

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

does not involve direct interaction with pyoveridine because MbtH-like proteins are found in gene clusters of various species that generate diverse types of siderophores. Hence, Drake *et al.*[19] concluded that the conserved regions of the MbtH-like proteins may interact with the conserved regions of the other proteins in their gene clusters.

4. Conclusions

Rv2377c, the founding member of the MbtH-like family of proteins, is a small, 71-residue polypeptide that folds into a three-strand anti-parallel β-sheet with a C-terminal α-helix nestled onto one side. The structure is robust as CD spectroscopy shows it will refold after heating to 80°C. This core structure is similar to that observed in the XRD- and NMR-determined structure for the MbtH-like protein from *P. aeruginosa*, PA2412. A DALI search indicates that this fold is unique to the MbtH-like family of proteins, suggesting these proteins may also have a unique function. Further characterization of the dynamics of Rv2377c indicated that the interface between β1 and β2 displays features characteristic of motion on the ms to μs time scale and the C-terminal is highly disordered. Because these regions also contain residues highly conserved in all proteins in the MbtH-like family, the dynamics of these regions may be necessary for the hypothesized interactions with other proteins in their gene cluster. Indeed, intrinsically disordered regions are associated with binding to multiple partners[50] or functional diversity,[51] and hence, perhaps it is even possible that Rv2377c binds to more than one specific substrate and has more than one biological function. The observation that *mtbH*-like genes are able to cross-talk between biosynthetic pathways and functionally replace each other supports multiple binding partners and functional diversity.[48,49] Experimental evidence suggest the MbtH-like proteins play a biochemical role in virulence as various microarray studies of the *M. tuberculosis* response to iron deprivation show that Rv2377c levels are upregulated. Hence, the MbtH-like family of proteins may become a target for a new generation of anti-TB agents. Further studies are necessary to uncover the biochemical function of Rv2377c with an emphasis directed at determining if it interacts with any of the other proteins within their gene cluster responsible for siderophore and antibiotic peptide synthesis. When a function for Rv2377c is found, the structure presented here will assist the molecular understanding of its function and potentially speed up the conception and development of new therapies to treat and control the spread of tuberculosis around the world.

Acknowledgments

The structure of Rv2377c was a community request made to the Seattle Structural Genomics Center for Infectious Disease (SSGCID). The research was funded by NIAID 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). PNNL is operated for the U.S. Department of Energy by Battelle.

References

1. Global tuberculosis control: surveillance, planning, financing. Geneva: World Health Organization; 2007.
2. Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN. Molecular epidemiology of tuberculosis: current insights. Clin Micro Rev 2006;19:658–85. [PubMed: 17041139]
3. Mitchison DA. Drug resistance in tuberculosis. Eur Res J 2005;25:376–9.
4. Basu S, Friedland GH, Medlock J, Andrews JR, Shah NS, Gandhi NR, Moll A, Moodley P, Sturm AW, Galvani AP. Averting epidemics of extremely drug-resistant tuberculosis. Proc Natl Acad Sci USA 2009;106:7672–7. [PubMed: 19365076]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

5. Kamholz SL. Drug resistant tuberculosis. J Assoc Acad Minor Phys 2002;13:53–6. [PubMed: 12362568]
6. Jindani A, Nunn AJ, Enarson DA. Two 8-month regimens of chemotheraphy for treatment of newly diagnosed pulmonary tuberculosis: international multicentre randomised trial. Lancet 2004;364:1244–51. [PubMed: 15464185]
7. Zignol M, Hosseini MS, A W, Weezenbeek CL, Nunn P, Watt CJ, Williams BG, Dye C. Global incidence of multi-drug resistant tuberculosis. J Infect Dis 2006;194:479–85. [PubMed: 16845631]
8. Somu RV, Boshoff H, Qiao C, Bennett EM, Barry CE III, Aldrich CC. Rationally designed nucleoside antibiotics that inhibit siderophore biosynthesis of *Mycobacteria tuberculosis*. J Med Chem 2006;49:31–4. [PubMed: 16392788]
9. Miethke M, Bissseret P, Beckering CL, Vignard D, Eustache J, Marahiel MA. Inhibition of acyl acid adenylation domains involved in bacterial siderophore synthesis. FEBS J 2006;273:409–19. [PubMed: 16403027]
10. Ratledge C. Iron, mycobacteria and tuberculosis. Tuberculosis 2004;84:110–30. [PubMed: 14670352]
11. Kostoghiorghes GJ, Weinberg ED. Iron: Mammalian defense systems, mechanisms of diseases, and chelation therapy approaches. Blood Rev 1995;9:33–45. [PubMed: 7795423]
12. Ratledge C, Dover LG. Iron metabolism in pathogenic bacteria. Annu Rev Microbiol 2000;54:881–941. [PubMed: 11018148]
13. Faraldo-Gómez JD, Sansom MSP. Acquisition of siderophres in Gram-negative bacteria. Nat Rev Mole Cell Biol 2003;4:105–16.
14. Snow GA. Mycobactins: iron-chelating growth factors from mycobacteria. Bacteriol Rev 1970;34:99–125. [PubMed: 4918634]
15. De Voss JJ, Rutter K, Schroeder BG, Barry CE III. Iron acquisition and metabolism by Mycobacteria. J Bact 1999;181:4443–51. [PubMed: 10419938]
16. De Voss JJ, Rutter K, Schroeder BG, Su H, Zhu Y, Barry CE III. The silicylate-derived mycobactin siderophores of *Mycobacterium tuberculosis* are essential for growth in macrophage. Proc Natl Acad Sci USA 2000;97:1252–7. [PubMed: 10655517]
17. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE III, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barell BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998;393:537–44. [PubMed: 9634230]
18. Quadri LEN, Sello J, Keating TA, Weinreb PH, Walsh CT. Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem Biol 1998;5:631–45. [PubMed: 9831524]
19. Drake EJ, Cao J, Qu J, Shah MB, Straubinger RM, Gulick AM. The 1.8 Å crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of Pseudomonas aeruginosa. J Biol Chem 2007;282:20425–34. [PubMed: 17502378]
20. Buchko GW, Kim CY, Terwilliger TC, Kennedy MA. Solution structure of the conserved hypothetical protein Rv2302 from *Mycobacterium tuberculosis*. J Bact 2006;188:5993–6001. [PubMed: 16885468]
21. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD. ^{1}H, ^{13}C and ^{15}N Chemical shift referencing in biomolecular NMR. J Biomol NMR 1995;6:135–40. [PubMed: 8589602]
22. Farrow NA, Muhandiram DR, Singer AU, Pascal SM, Kay LE, Gish G, Shoelson SE, Pawson T, Forman-Kay JD. Backbone dynamics of a free and a phosphopeptide-complexed Src homlogy 2 domain studied by ^{15}N NMR relaxation. Biochemistry 1994;33:5984–6003. [PubMed: 7514039]
23. Szyperski T, Yeh D, Sukumaran D, Moseley H, Montelione G. Reduced-dimensionality NMR spectroscopy for high-throughput protein resonance assignment. Proc Natl Acad Sci USA 2002;99:8009–14. [PubMed: 12060747]
24. Guntert P. Automated NMR structure calculation with CYANA. Meth Mol Biol 2004;278:353–78.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

25. Cornilescu G, Delagio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 1999;13:289–301. [PubMed: 10212987]
26. Creighton, TE. Proteins: structure and molecular properties. New York: H.W. Freeman & Company; 1993.
27. Linge JP, Nilges M. Influence of non-bonded parameters on the quality of NMR structures: A new force field for NMR structure calculation. J Biomol NMR 1999;13:51–9. [PubMed: 10905826]
28. Bhattacharya A, Tejero R, Montelione G. Evaluating protein structures determined by structural genomics consortia. Proteins 2007;66:778–95. [PubMed: 17186527]
29. Palczewska M, Groves P, Ambrus A, Kaeta A, Kover K, Batta G, Kuznicki J. Structural and biochemical characterization of neuronal calretinin domain I-II (residues 1-100). Eur J Biochem 2001;268:6229–37. [PubMed: 11733019]
30. Chothia C. Conformation of twisted β-pleated sheets in proteins. J Mol Biol 1973;75:295–302. [PubMed: 4728692]
31. Woody, RW. Studies of theoretical circular dichroism of polypeptides: Contributions of β-turns. New York: John Wiley & Sons; 1974.
32. Holzwarth GM, Doty P. The ultraviolet circular dichroism of polypeptides. J Amer Chem Soc 1965;87:218–28. [PubMed: 14228459]
33. Buchko GW, Hess NJ, Bandaru V, Wallace SS, Kennedy MA. Spectroscopic studies of zinc(II)- and colbalt(II)-associated *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg): extended X-ray absorption fine structure evidence for a metal-binding domain. Biochemistry 2000;40:12441–9. [PubMed: 11015225]
34. Karantzeni I, Ruiz C, Liu CC, LiCata VJ. Comparative thermal denaturation of *Thermus aquaticus* and *Escherichia coli* type 1 DNA polymerases. Biochem J 2003;374:785–92. [PubMed: 12786603]
35. Ishima R, Torchia DA. Protein dynamics from NMR. Nat Struct Biol 2000;7:740–3. [PubMed: 10966641]
36. Millet O, Loria JP, Kroenke CD, Pons M, Palmer AG III. The static magnetic field dependence of chemical shift exchange linebroadening defines the NMR chemical shift time scale. J Amer Chem Soc 2000;122:2867–77.
37. Buchko GW, McAteer K, Wallace SS, Kennedy MA. Solution-state NMR investigation of DNA binding interactions in *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg): a dynamic description of the DNA/protein interface. DNA Repair 2005;4:327–39. [PubMed: 15661656]
38. Mandel AM, Akke M, Palmer AG III. Backbone dynamics of *Escherichia coli* ribonuclease H1 correlations with structure and function in an active enzyme. J Mol Biol 1995;246:144–63. [PubMed: 7531772]
39. Tompa P. Instrinsically unstructured proteins. Trends Biochem Sci 2002;27:527–33. [PubMed: 12368089]
40. Fink AL. Natively unfolded proteins. Curr Opin Struct Biol 2005;15:35–41. [PubMed: 15718131]
41. Dunker AK, Brown CJ, Lawsor JD, Iahoucheva LM, Obradovic Z. Intrinsic disorder and protein function. Biochemistry 2002;41:6573–82. [PubMed: 12022860]
42. Uversky VN. Intrinsically disordered proteins and their environement: Effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. Protein J 2009;28:305–25. [PubMed: 19768526]
43. Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting protein disorder for N-, C- and internal regions. Genome Info 1999;10:30–40.
44. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42:38–48. [PubMed: 11093259]
45. Holm L, Sander C. Touring protein fold space with Dali/FSSP. Nucleic Acids Res 1998;26:316–9. [PubMed: 9399863]
46. Rodriguez GM, Voskuil MI, Gold B, Schoolnik GK, Smith I. IdeR, an esential gene in *Mycobacterium tuberculosis*: Role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect Immun 2002;70:3371–81. [PubMed: 12065475]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

47. Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK. Transcriptional adaption of *Mycobacterium tuberculosis* within macrophages: Insights into the phagosomal environment. J Exp Med 2003;198:693–704. [PubMed: 12953091]

48. Wolpert M, Gust B, Kammerer B, Heide L. Effects of deletions of *mbtH*-like genes on clorobiocin biosynthesis in *Streptomyces coelicolor*. Microbiology 2007;153:1413–23. [PubMed: 17464055]

49. Lautru S, Oves-Costales D, Pernodet JL, Challis GL. MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. Microbiology 2007;153:1405–12. [PubMed: 17464054]

50. Haynes C, Oldfield E, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iahoucheva LM. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2006;2:e100. [PubMed: 16884331]

51. Xie H, Vucetic S, Iahoucheva LM, Oldfield E, Dunker AK, Uversky VN, Obradovic Z. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res 2007;6:188–1898.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

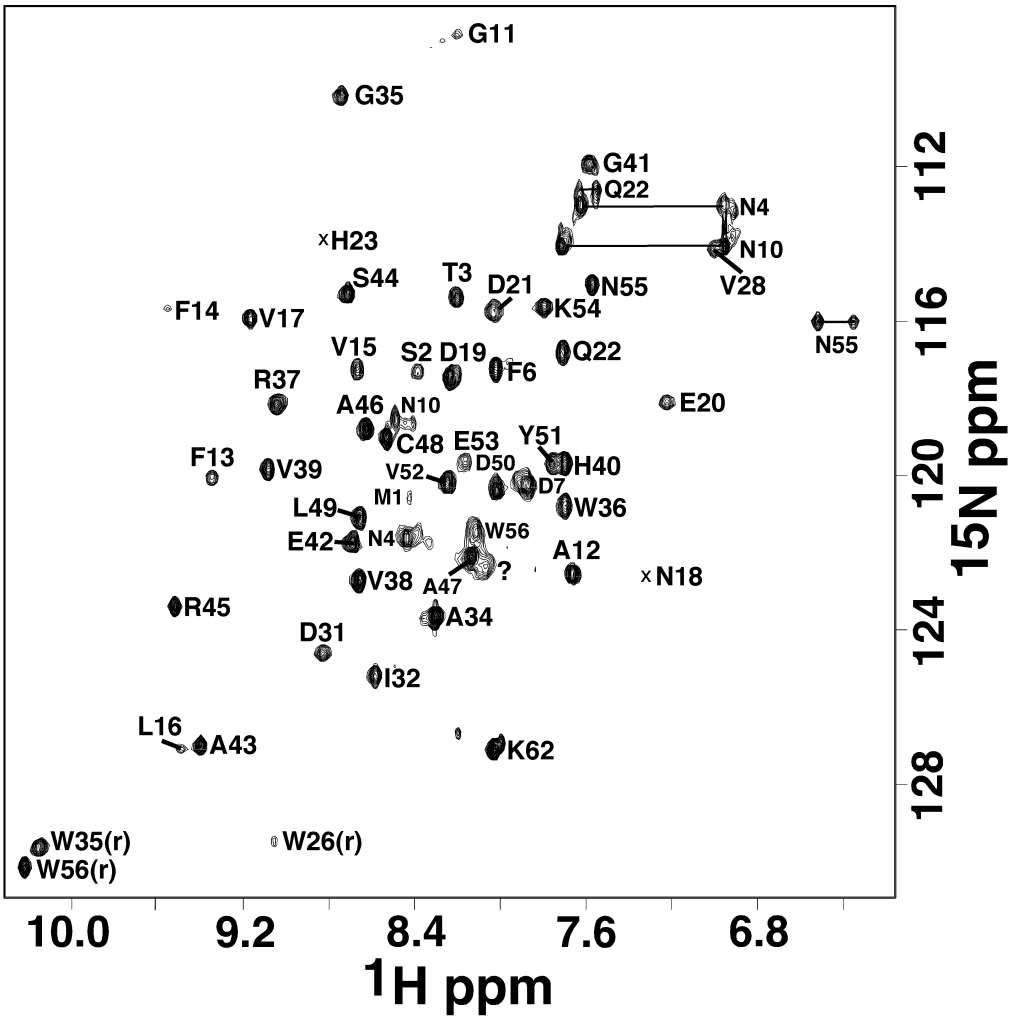

Figure 1.

Assigned ^{1}H-^{15}N HSQC spectrum of double-labeled Rv2377c collected at 25°C in NMR buffer (300 mM NaCl, 20 mM TrisHCl, 1.0 mM DTT, pH 7.1) at a ^{1}H resonance frequency of 900 MHz. Side chain resonances are indicated by horizontal solid lines and the single significant unassigned amide cross peak with a question mark.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

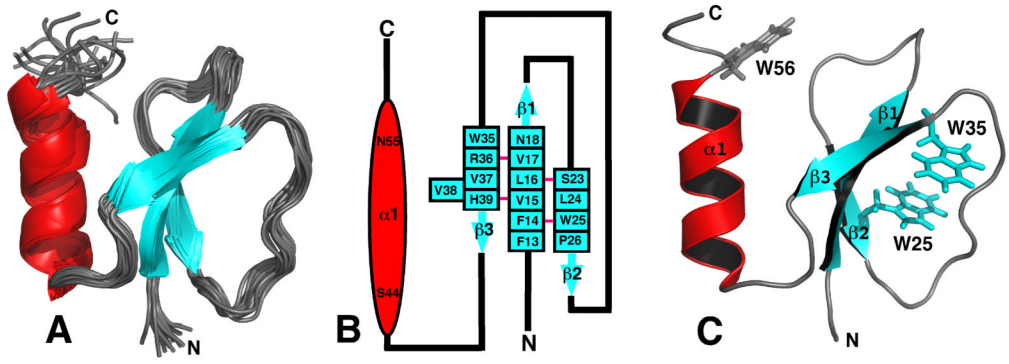

Figure 2.

A) Superposition of the ensemble of 20 lowest energy structures calculated for Rv2377c. Because both the N- and C-terminal regions are disordered, G1* - A12 and W56 - D71, respectively, most of these regions have been removed from the Figure for clarity. B) Secondary structure diagram of Rv2377c. The α-helix is drawn as a red ovals and the β-strands as solid blue arrows with the residue number at the beginning and the end of each structural element shown. The β-sheet contains a bulge at residue V38. A solid pink line between the β-strand residues indicates dual hydrogen bonds between two residues in an antiparallel β-sheet. C) Ribbon representation of the structure in the ensemble closest to the average structure for Rv2377c highlighting the position of the three exclusively conserved tryptophan residues in the MbtH-like family of proteins. While the side chains of W25 and W35 superimpose well in the ensemble, W56 assumes many positions. The β-strands are colored blue and the α-helix is colored red.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

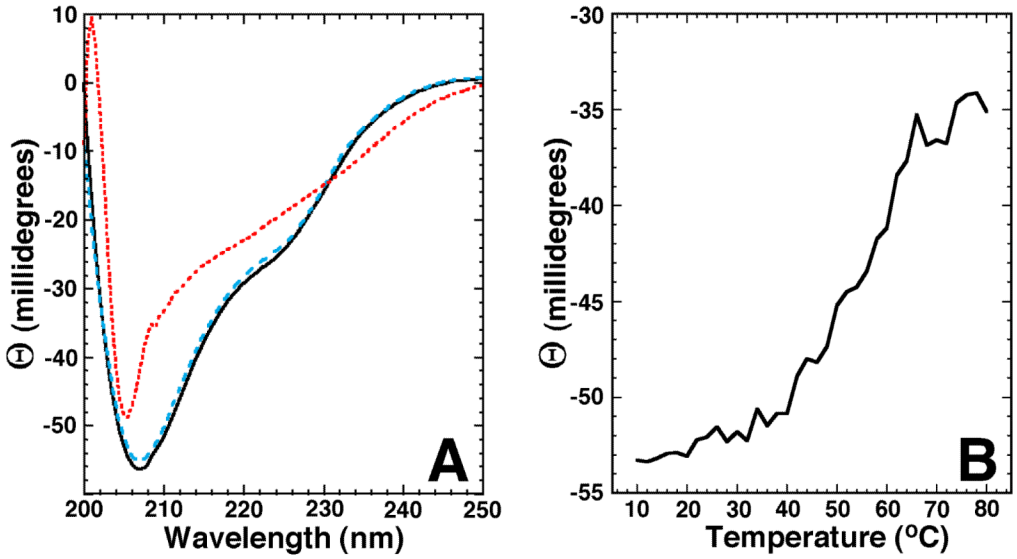

Figure 3.

A) Circular dichroism spectra of a single sample of Rv2377c (0.06 mM) in NMR buffer collected at 25°C (black solid line), 80°C (red dotted line), and 25°C post-heating (blue dashed line). B) The CD thermal melt for Rv2377c obtained by measuring the ellipticity at 210 nm in 2.0°C intervals between 10 and 80°C.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

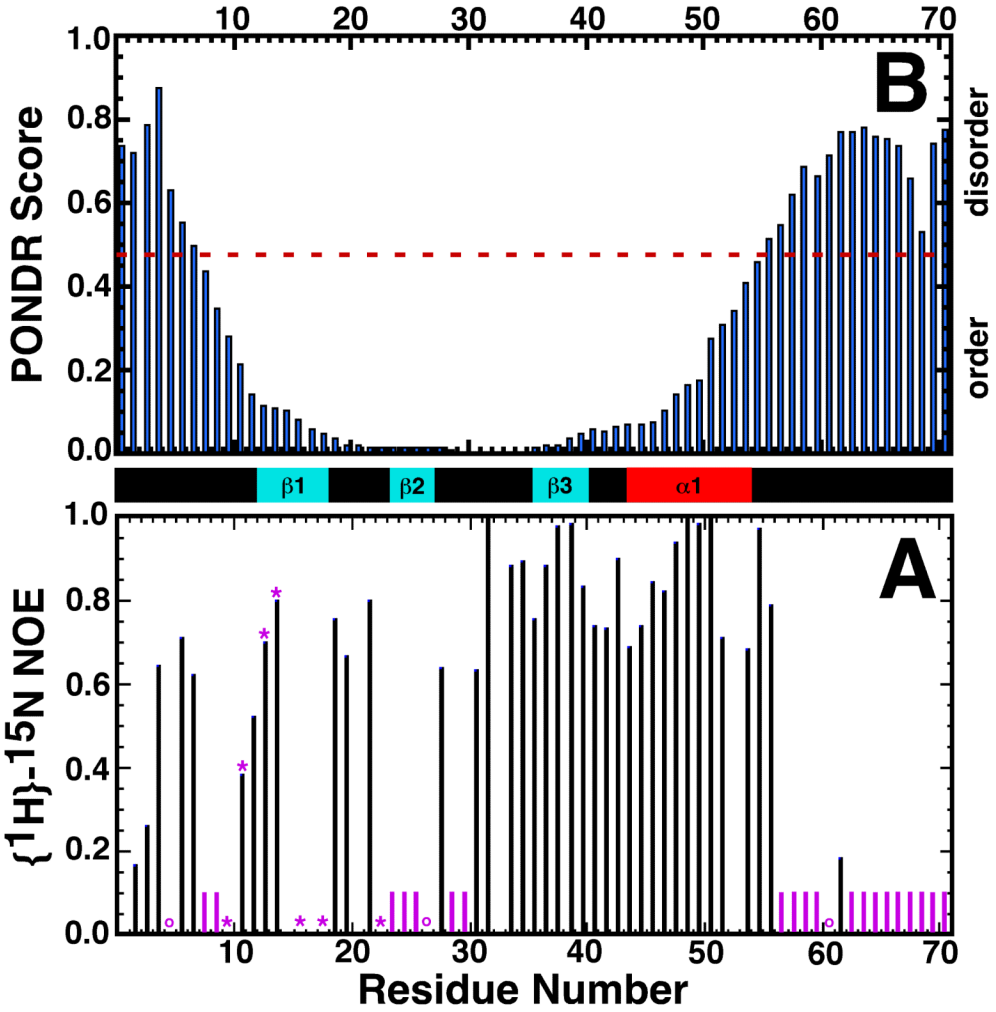

Figure 4.

A) Backbone {^{1}H}-^{15}N heteronuclear NOE values for Rv2377c. The position of proline residues in the sequence are indicated by purple circles. Amide cross peaks that were absent or very weak in the original ^{1}H-^{15}N HSQC spectrum are indicated by purple bars and asterisks, respectively. B) Graphical output of a PONDR prediction on Rv2377c using the VL-XT algorithm. Consecutive values above and below 0.5 predict disordered and ordered regions, respectively, within the protein. The experimentally observed elements of secondary structure are shown between the two plots: cyan = β-strand, red = α-helix.

Figure 5.

Alignment of the amino acid sequences for the three determined structures of proteins from the MbtH-like family: *M. tuberculosis* Rv2377c (2KHR); *P. aeruginosa* PA2412 (2PST (XRD)[19] and 2GPF (NMR)). The sequences are colored coded according to the observed regions of secondary structure: blue = β-strand; red = α-helix. The residues conserved in >95% of the family members are highlighted with grey shaded boxes.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

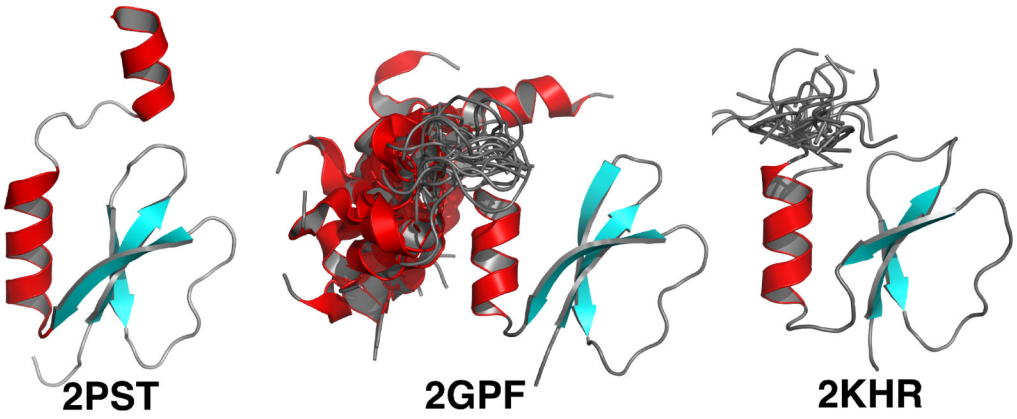

Figure 6.
Comparison of the structure of *P. aeruginosa* PA2412 determined by XRD (2PST) and NMR (2GPF) methods with the NMR determined structure of Rv2377c (2KHR). The structure in the NMR ensemble closest to the average structure is shown. The position of the second α-helix in 2GPF, relative to the structured core, is shown for each of the 20 structures in the PDB deposited ensemble. Similarly, a short part of the C-terminal region for the PDB deposited ensemble for 2KHR is also shown. In both cases the extreme N- and C-terminal regions for 2GPF and 2KHR have been removed for clarity, but, do include the highly conserved region, WXDXR. The β-strands are colored blue and the α-helices are colored red.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1
Summary of the structural statistics for Rv2377c[a]

Restraints for Structure Calculations	
Total NOEs	602
Intraresidue NOEs	142
Sequential (i, i + 1) NOEs	196
Medium-range (i, i + j; 1 < j ≤ 4) NOEs	77
Long-range (i, i + j; j > 4) NOEs	187
Phi (Φ) angle restraints	33
Psi (Ψ) angle restraints	33
Hydrogen bond restraints	26
Structure Calculations	
Number of structures calculated	100
Number of structures used in ensemble	20
Structures with Restraint Violations	
Distance Restraint Violations > 0.05Å	0
Dihedral Restraint Violation > 2°	1
Dihedral Restraint Violation > 3°	0
RMSD to Mean (Å) Ordered Residues: 13-55	
Backbone N-C^{α}-C=O Atoms	0.62±0.15
Heavy Atoms	1.10±0.20
All Atoms	1.25±0.19
Ramachandran Plots Ordered Residues	
Most favored regions	87.3%
Additionally allowed regions	12.4%
Generously favored regions	0.1%
Disallowed	0.1%
Global Quality Scores - Ordered Residues[b]	**Z-score (Raw)**
Procheck (all)	-3.02 (-0.51)
Procheck (Φ, Ψ)	-1.53 (-0.47)
MolProbity clash score	-1.26 (16.22)

[a] All statistics are for the 20-structure ensemble deposited in the Protein Data Bank (2KHR) using the residues of the structured central core (13-55).

[b] Calculated for the ordered residue, 13-55, using the PSVS program.[28]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Elements of secondary structure in Rv2377c as determined from the NMR structure and estimated from the deconvolution of the CD spectrum

Method	α-helix	β-strand	turn	random coil
NMR	16%	20%	22%	42%
CD	29%	20%	18%	33%