Structural analysis reveals DNA binding properties of Rv2827c, a hypothetical protein from *Mycobacterium tuberculosis*

Robert Janowski · Santosh Panjikar · Ali Nasser Eddine · Stefan H. E. Kaufmann · Manfred S. Weiss

Received: 31 October 2008 / Accepted: 14 January 2009 / Published online: 31 January 2009
© Springer Science+Business Media B.V. 2009

**Abstract** Tuberculosis (TB) is a major global health threat caused by *Mycobacterium tuberculosis* (Mtb). It is further fueled by the HIV pandemic and by increasing incidences of multidrug resistant Mtb-strains. Rv2827c, a hypothetical protein from Mtb, has been implicated in the survival of Mtb in the macrophages of the host. The three-dimensional structure of Rv2827c has been determined by the three-wavelength anomalous diffraction technique using bromide-derivatized crystals and refined to a resolution of 1.93 Å. The asymmetric unit of the orthorhombic crystals contains two independent protein molecules related by a non-crystallographic translation. The tertiary structure of Rv2827c comprises two domains: an N-terminal domain displaying a winged helix topology and a C-terminal domain, which appears to constitute a new and unique fold. Based on structural homology considerations and additional biochemical evidence, it could be established that Rv2827c is a DNA-binding protein. Once the understanding of the structure-function relationship of Rv2827c extends to the function of Rv2827c in vivo, new clues for the rational design of novel intervention strategies may be obtained.

**Keywords** X-ray crystallography · DNA binding · *Mycobacterium tuberculosis* · Hypothetical protein · Rv2827c · Winged helix domain

**Abbreviations**

- HTH Helix-turn-helix
- Mtb *Mycobacterium tuberculosis*
- PDB Protein Data Bank
- TB Tuberculosis
- WH Winged helix
- wHTH Winged helix-turn-helix

**Introduction**

Globally, tuberculosis (TB) represents a major health threat with 9 million cases and close to 2 million deaths annually [80]. Although drugs to treat TB are available, the complex and long-lasting treatment scheme with three to four drugs over a period of 6–9 months leads to poor patient compliance. In several regions, notably in sub-Saharan Africa, the re-emergence of TB was further fueled by the pandemic with human immunodeficiency virus (HIV) responsible for acquired immunodeficiency syndrome (AIDS) [40]. As a corollary, incidences of multidrug-resistant (MDR) and even extensively drug-resistant (XDR)-TB have increased profoundly. MDR-TB is resistant to treatment with first-line drugs, whereas XDR-TB is virtually untreatable with the currently available drugs. Drug development for TB has been largely neglected in the last decades and hence appropriate intervention methods to efficiently combat the re-emergent threat of TB are urgently required. The etiologic agent of TB, *Mycobacterium tuberculosis* (Mtb), is a
bacterium capable of persisting in resting macrophages. After adequate activation by cytokines, notably interferon-gamma (IFN-γ), macrophages acquire increased antibacterial capacities. These activated macrophages can control Mtb growth though they fail to achieve sterile eradication [84]. Consequently, gene products essential for Mtb growth or persistence in resting or activated macrophages, respectively, represent potential targets for novel drugs.

Genome-wide expression profiling of microbial pathogens has become a useful tool to identify single genes and gene networks that are differentially expressed under different conditions. In numerous cases, these analyses have provided clues to gene functions and persistence. Among the close to 4,000 genes encoded by the H37Rv strain of Mtb [15, 21], a group of proteins has been identified in microarray experiments as being differentially expressed and therefore considered potentially important for persistence and pathogenicity of Mtb [66, 67]. Currently, only limited information on the three-dimensional (3-D) architecture and the structural features of these proteins is available. It is well conceivable that the understanding of the 3-D structures of these proteins will provide a valuable basis for a better understanding of pathogenesis and persistence of Mtb and for structure-based design of novel intervention strategies against tuberculosis.

Based on its amino acid sequence, Rv2827c has been annotated as a hypothetical protein with unknown function [21]. The protein is composed of 295 amino acid residues with a molecular weight of 32.3 kDa and an isoelectric point of 9.3. According to Sassetti et al. [70] and Lamichhane et al. [44], this protein is critical for Mtb replication. Mtb mutants lacking a functional copy of the rv2827c gene fail to grow in vitro. The protein Rv2827c is approximately threefold upregulated upon infection of macrophages with Mtb in comparison to in vitro grown cultures of Mtb [66, 67]. IFN-γ activation of macrophages leads to a further threefold increase as compared to the situation in resting macrophages. Based on these observations we proposed that Rv2827c plays a critical role in Mtb survival in macrophages and hence represents a potential target for future intervention strategies.

The crystallization and preliminary diffraction experiments for Rv2827c have recently been reported [38]. Here, we describe the X-ray structure of Rv2827c, solved by the MAD method [36] utilizing bromide-derivatized crystals [23]. The structure of Rv2827c and its potential function as a DNA binding protein will be described and discussed in detail. The elucidation of the structure—function relationship for Rv2827c may provide a blueprint for the rational design of a drug candidate targeted at this molecule.

Materials and methods

Cloning, expression, purification and crystallization

The cloning, expression, purification, crystallization and preliminary X-ray diffraction experiments for the native crystals of Rv2827c have been described previously [38]. Briefly: recombinant, full-length Rv2827c, with three additional N-terminal residues (Gly-2, Ala-1 and Met0) introduced for cloning purposes, was crystallized from 2 M sodium formate and 100 mM sodium acetate (pH 4.7) supplemented with the 3-(1-pyridino)-1-propane sulfonate (aka non-detergent sulfo-betaine 201) or 6-aminocaproic acid at 277 K. Single crystals grew out of the surface of spherulites within several weeks. The crystals are orthorhombic, space group P2₁2₁2 with unit-cell parameters \( a = 87.42 \ \text{Å}, b = 180.65 \ \text{Å} \) and \( c = 35.11 \ \text{Å} \) and diffract X-rays to a resolution of better than 2.0 Å.

Data collection and processing

Due to the lack of a suitable search model for molecular replacement, three-wavelength MAD data were collected from a crystal soaked with 0.3 M NaBr (in crystallization buffer) for 2 days and additionally for 20 min in 0.5 M NaBr (in crystallization buffer) immediately before the X-ray experiment. A crystal with dimensions 200 \( \times \) 150 \( \times \) 150 \( \mu \text{m} \) was mounted in a nylon fiber loop, cryo-protected for 10 s in reservoir solution containing 15% (v/v) MPD and 0.5 M NaBr and flash-cooled to 100 K in a nitrogen gas stream. Diffraction data were then collected on the EMBL beamline BW7A (DESY, Hamburg, Germany) using a MARCCD detector. For all three wavelengths, 360° of data were collected to 2.6 Å resolution (Table 1). Data were indexed and integrated using DENZO [59] and scaled using SCALEPACK [59]. The redundancy-independent merging R-factor \( R_{\text{merge}} \) as well as the precision-indicating merging R-factor \( R_{\text{merge}} \) [79] were calculated using the program RMERGE (available from http://www.embl-hamburg.de/~msweiss/projects/msw_qual.html or from MSW upon request). Intensities were converted to structure-factor amplitudes using the program TRUNCATE [17, 26] and the optical resolution was calculated using the program SFCHECK [78].

Structure determination and refinement

The structure of Rv2827c was solved using the three-wavelength MAD protocol of Auto-Rickshaw; the EMBL-Hamburg automated crystal structure determination platform [63]. The input diffraction data were uploaded to the Auto-Rickshaw server and then prepared and converted for
TABLE 1 Data collection and processing statistics

| Number of crystals | 1 | 1 |
|--------------------|---|---|
| Beamline           | X13 | BW7A |
| Wavelength [Å]     | 0.8031 | 0.9195 (peak) | 0.9203 (inflection) | 0.9095 (remote) |
| Temperature [K]    | 100 | 100 |
| Crystal-to-detector distance [mm] | 180 | 210 |
| Rotation range per image [°] | 1.0 | 1.0 |
| Total rotation range [°] | 156 | 360 |
| Space group        | P2₁2₁2 | P2₁2₁2 |
| Unit cell parameters [Å] | a = 87.42 | a = 87.48 | a = 87.47 | a = 88.58 |
|                    | b = 180.65 | b = 180.79 | b = 180.69 | b = 182.95 |
|                    | c = 35.11 | c = 35.43 | c = 35.39 | c = 35.82 |
| Mosaicity [°]      | 0.60 | 1.00 | 1.00 | 0.95 |
| Resolution limits [Å] | 50.0–1.93 | 99–2.60 |
| Total number of reflections | 268,981 | 257,694 | 257,209 | 258,683 |
| Unique reflections | 43,027 | 18,227 | 18,143 | 18,166 |
| Redundancy         | 6.3 | 14.1 | 14.2 | 14.2 |
| I/σ(I)             | 23.3 (3.7) | 30.3 (5.8) | 32.8 (6.9) | 27.9 (5.7) |
| Completeness [%]   | 99.9 (99.8) | 99.9 (99.2) | 99.9 (99.7) | 99.9 (100.0) |
| Rmerge [%]         | 7.6 (49.8) | 10.2 (64.6) | 9.2 (55.8) | 10.8 (63.0) |
| Rfree [%]          | 8.3 (54.8) | 10.7 (76.7) | 9.4 (59.5) | 11.3 (66.8) |
| Rp.i.m. [%]        | 3.3 (22.5) | 2.8 (20.8) | 2.5 (16.2) | 3.0 (17.6) |
| R anom [%]         | – | 3.2 (16.2) | 2.2 (13.3) | 3.3 (15.5) |
| Overall B factor from Wilson plot [Å²] | 24.2 | 41.9 | 47.1 | 48.8 |
| Optical resolution [Å] | 1.55 | 1.88 | 1.89 | 1.91 |

Values in parentheses correspond to the highest resolution shell.

\[ R_{\text{merge}} = \frac{\sum_{i=1}^{N} |I_{i}(hkl)| - \langle |I(hkl)| \rangle}{\sum_{i=1}^{N} |I_{i}(hkl)|} - \left( \frac{1}{N}\right)^{1/2} \sum_{i=1}^{N} |I_{i}(hkl)| / \sum_{i=1}^{N} I_{i}(hkl), \]

(1)

\[ R_{\text{free}} = \frac{\sum_{i=1}^{N} |I_{i}(hkl)| - \langle |I(hkl)| \rangle}{\sum_{i=1}^{N} |I_{i}(hkl)|} - \left( \frac{1}{N}\right)^{1/2} \sum_{i=1}^{N} |I_{i}(hkl)| / \sum_{i=1}^{N} I_{i}(hkl), \]

(2)

\[ R_{\text{p.i.m.}} = \frac{\sum_{i=1}^{N} |I_{i}(hkl)| - \langle |I(hkl)| \rangle}{\sum_{i=1}^{N} |I_{i}(hkl)|} - \left( \frac{1}{N}\right)^{1/2} \sum_{i=1}^{N} |I_{i}(hkl)| / \sum_{i=1}^{N} I_{i}(hkl), \]

(3)

\[ R_{\text{anom}} = \frac{\sum_{i=1}^{N} |I_{i}(hkl)| - \langle |I(hkl)| \rangle}{\sum_{i=1}^{N} |I_{i}(hkl)|} - \left( \frac{1}{N}\right)^{1/2} \sum_{i=1}^{N} |I_{i}(hkl)| / \sum_{i=1}^{N} I_{i}(hkl), \]

(4)

use in Auto-Rickshaw using programs of the CCP4 suite [17]. F_A values were calculated using the program SHELXC [74]. Based on an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 3.4 Å. Twenty-three bromide positions were located using the program SHELXD [71]. The correct hand for the substructure was determined using the programs ABS [34] and SHELXE [73]. The occupancy of all substructure atoms was refined using the program MLPHARE [17]. The initial phases were improved using density modification and phase extension to 2.60 Å resolution using the program DM [22]. Approximately 50% of the model was built automatically using the program ARP/wARP [52, 65]. The missing parts of the model were then added by assembling the intermediate models generated from ARP/wARP. As soon as the model was 80% complete, refinement was continued against the 1.93 Å resolution native dataset. Refinement was performed in REFMAC5 [54] using the maximum likelihood target function including TLS parameters [82]. For TLS refinement, four TLS groups were used per protein chain (Gly-2-Ile80, Pro83-Asp94, Gly98-Thr253 and Val255-Gly293). In between refinement cycles, the structure was rebuilt manually in COOT [24]. The final model is characterized by R and R free factors of 18.3% and 22.4%, respectively (Table 2). Structural superpositions and searches were carried out using the program ALJIN [20] and the SSM server (http://www.ebi.ac.uk/msd-srv/ssm, [42]). Electrostatic potential analysis was performed using the GRASP [55] and PYMOL (www.pymol.org) programs. The stereochemistry of the final model was analyzed using PROCHECK [45]. The refined structure and corresponding structure-factor amplitudes have been deposited with the PDB under the accession code 1ZEL.
Table 2 Refinement statistics

| Parameter                          | Value         |
|-----------------------------------|---------------|
| Resolution limits [Å]             | 3.00–1.93     |
| Data cutoff [Fo/σ(F)]             | 0.0           |
| Completeness [%]                  | 99.9          |
| Total No. of reflections          | 41,592        |
| No. of reflections in working set | 40,251        |
| No. of reflections in test set    | 1,341         |
| R [%]^a                          | 18.3          |
| Rfree [%]^b                       | 22.4          |
| No. of amino acid residues        | 588           |
| No. of protein atoms              | 4,620         |
| No. of sodium ions                | 2             |
| No. of acetate atoms              | 8             |
| No. of MPD atoms                  | 8             |
| No. of formate atoms              | 45            |
| No. of solvent atoms              | 340           |
| R.m.s.d. bond lengths [Å]         | 0.012         |
| R.m.s.d. bond angles [°]          | 1.38          |
| Ramachandran plot [%]: most favored/additionally allowed region | 92/8 |

* R = Σhkl | F o,d - | F c,d | VΣhkl|F o,d for all reflections, where F o and F c are observed and calculated structure factors, respectively
* Rfree was calculated against 3.2% (1,341) of all reflections, randomly excluded from the refinement

Domain division

For the definition of domains in the 3-D structure of Rv2827c, various approaches were used. First, the structure was analyzed using the program 123D+ (http://123d.ncifcrf.gov/123D+).html, [1]), which combines sequence profiles, secondary structure prediction, and contact capacity potentials to thread a protein sequence through the set of structures. Secondly, the vector alignment search tool VAST Search [31, 48] was employed using services of the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml). VAST defines domains based on structure similarity searches. Thirdly, automatic domain definition was attempted by analyzing accessible surface areas (ASA) on a residue by residue basis. A high value of ASA (>100 Å²) for a few adjacent residues indicates that a particular sequence fragment is well exposed and may thus constitute a linker or a hinge between domains. This calculation was carried out using the PISA server [43]. A further attempt on domain definition was made using a normal mode calculation [76] using the elNemo server (http://www.igs.cnrs-mrs.fr/elnemo/index.html). Finally, a TLS group analysis was performed using the TLS Motion Determination server (TLSMD, http://skuld.bmsc.washington.edu/~tlsmd/). For this analysis, the protein was divided into two to six TLS groups.

Initial DNA binding test

Without any information about the potential DNA sequence being recognized by Rv2827c, initial DNA binding tests were performed with four dsDNA fragments (Table 3), taken from the structures of various DNA binding proteins containing a wHTH motif: 5'-GGTCTCA GAACC-3' (PDB code 3HTS, [46]), 5'-CTATGAGTCGTTG-3' (PDB code 1RH6, [69]), 5'-AAAAAGGGGAAGT GG-3' (PDB code 1PUE, [41]), and 5'-GAGAAGTGA AGACTTTTCACCTTCTC-3' (PDB code 1IF1, [25]). The oligonucleotides were obtained from MWG Biotech and dissolved in a buffer composed of 10 mM Tris pH 8.5, 50 mM NaCl, 1 mM EDTA. In order to obtain a double stranded form, the palindromic fragments 5'-GGTCTCA GAACC-3' and 5'-GAGAAGTGAAGACTTTCACCTT CTC-3' were annealed at 95°C for 10 min and left for slow cooling. The other two fragments 5'-CTATGAGTCGTTG-3' and 5'-AAAAAGGGGAAGTGGG-3' were mixed with their complementary fragments 5'-CAACAGACTAC ATAG-3' and 5'-CCCACTTCCCTTCTTT-3', respectively, and prior to annealing them as described above. Equimolar amounts of Rv2827c (in 50 mM Tris pH 7.3, 150 mM KCl, 350 mM imidazole, 2 mM DTT) and the corresponding dsDNA fragments were mixed and dialyzed overnight against a buffer composed of 50 mM Tris (pH 7.3), 150 mM KCl and 2 mM DTT to slowly remove the imidazole present in the protein sample [38].

PCR-assisted binding site selection method

The polymerase chain reaction (PCR) assisted binding site selection method used here was described earlier by Nørby et al. [56]. The following ssDNA fragment was designed and obtained from MWG Biotech: 5'-CAATCCATGGC GACTCTGCATCGCG(N)30GTGTCACCGGCATGA-3' (PDB code 1RH6, [69]), 5'-AAAAGGGGAAGTGGG-3' and 5'-GAGAAGTGA AGACTTTTCACCTTCTC-3'. It contains 30 nucleotides with a random sequence flanked on both sides by a 25-base long conserved fragment with recognition sites for NcoI and Xhol at the 5' and 3' ends, respectively. The recognition sites (underlined) were necessary for subcloning purposes.

Table 3 DNA sequences tested for DNA binding

| DNA sequence 5'–3' | G + C [%] | Stabilizing |
|--------------------|-----------|------------|
| GGTTCCTAGAACCC     | 50        | Y          |
| CTATGAGTCGTTG       | 40        | Y          |
| AAAAGGGGAAGTGGG     | 50        | Y          |
| GAGAAGTGAAGACTTTCACCTTC-3' | 38   | N          |
The primers 5′-CAATCCATGCGACTCTGCATCCGC-3′ (forward) containing a recognition site for NcoI, and 5′-TGGTTCAGAGTCATGCCGGTGACAC-3′ (reverse) with an XhoI recognition site were designed for PCR amplification. The ssDNA fragment was converted to dsDNA by means of DNA polymerase I Klenow fragment and the reverse primer. For the protein-DNA binding test, 2 μl of 10 mg/ml Rv2827c were spotted onto 1 cm² of nitrocellulose membrane and air dried. The filter was blocked by washing it for 30 min at 4°C in a buffer composed of 50 mM Tris pH 7.3, 40 mM KCl, 3 mM MgCl₂, 2 mM DTT and 0.5% (w/v) carnation milk. The membrane was exposed overnight at 4°C to 200 μl of the binding buffer (50 mM Tris pH 7.3, 40 mM KCl, 3 mM MgCl₂, 2 mM DTT) supplemented with 10 pmol of the random dsDNA described above. Afterward, the membrane was extensively washed in binding buffer first with 0.5% (w/v) carnation milk and then without. As a control, the membrane without Rv2827c was exposed to the mixture of random dsDNA. To dissociate DNA bound to the protein (and to the membrane in the control experiment), the membrane was washed in 0.5 M KCl and to check for the presence of DNA in the eluted samples and to amplify it, PCR was performed using the primers described above. Four different samples were investigated: (1) the eluent from the control membrane washed in washing (low salt) buffer and (2) in 0.5 M KCl, (3) the eluent from the membrane with immobilized Rv2827c washed in low salt buffer (as additional control) and (4) in 0.5 M KCl. The last sample (4) was then used for the next round of the experiment. The whole procedure was repeated seven times. After each cycle of binding, PCR was performed to analyze the quality and quantity of binding. After the final cycle, the amplified DNA fragments were digested with NcoI and XhoI and subcloned into the corresponding sites of pETM-11 vector containing a kanamycine resistance. TOP10 cells (Invitrogen) were transformed with the recombinant plasmid. The presence of the inserted DNA fragment was verified by PCR. Seventeen randomly selected clones were sent for oligonucleotide sequencing (Table 4).

Modeling the complex of Rv2827c with dsDNA

In order to construct a model of the Rv2827c/dsDNA-complex, the N-terminal domain of Rv2827c was superimposed onto the winged helix domain of the interferon regulatory factor 3 (IRF-3, PDB code 1T2K, [64]) complexed with a 31-mer DNA fragment. Then, the C-terminal domain (residues Pro83-Ala295) was manually rotated and translated toward the DNA using the molecular graphics program COOT [24]. By repeating this operation with C-terminal fragment Asp94-Ala295 the fit of Rv2827c onto the DNA can be markedly improved. Further minor adjustments in the loop between Leu125 and Val139 followed by geometry idealisation resulted in the final model.

Results and discussion

Analysis of the primary structure of Rv2827c

A sequence similarity search using the programs BLAST and PSI-BLAST [2] provided only limited information.
about homologues of Rv2827c in other organisms. The sole exception is the hypothetical protein Mb2851c from M. bovis [29] with 100% sequence identity to Rv2827c. Further hits, albeit at a much lower confidence level, include phosphoribosylaminomimidazole carboxylase from Azospirillum brasilense [16] (EMBL/GenBank/DDBJ databases), with 26% identity and 53% similarity for a 255 amino acid overlap, a hypothetical protein from Pseudomonas sp. WBC-3 [47] with 28% identity and 50% similarity in a 264 aa overlap and the hypothetical protein SCO15 from Streptomyces coelicolor [6] with 29% identity and 46% similarity in a 256 aa overlap. Searches with shorter sequence fragments revealed that the first 133 amino acids of Rv2827c exhibit 25% identity and 38% similarity to seryl-tRNA synthetase from Methanopyrus kandleri [75]. The short fragment Asp133-Leu184 displays some evidence for a leucine-zipper motif (LX_6) and shares 52% identity and 69% similarity with the putative integral membrane protein from S. avermitilis [57]. The sequence region Ala172-Glu254 appears to be homologous to a putative serine/threonine protein kinase from S. coelicolor [6] with 34% identity and 45% similarity. Rv2827c has not been associated with any superfamily in the COG [77] or Pfam databases [4]. The only information about its potential function comes from the function prediction program ProKnow [62], which predicts that Rv2827c may have ATP binding activity or may participate in pantothenate biosynthesis as well as protein amino acid phosphorylation. However, the evidence ranks [62], are low and comparable for all three functions. Therefore, the predicted information is not really reliable. To date, no tertiary structure has been reported to the Protein Data Bank (PDB) [7] for any protein similar to Rv2827c. The highest score from sequence-based search against the PDB identifies 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from Escherichia coli (PDB code: 1RU1, [8]) with 29% identity for the 120-residue long fragment between Leu175 and the C-terminus of Rv2827c as a potential homologue.

The overall fold of Rv2827c and its topology

The protein structure consists of 15 α-helices and 12 β-strands and belongs to the class 3, α/β family according to the CATH protein structure classification [58] (Fig. 1a, b). The N-terminal part starts with a short β-strand (β1) composed of Ala-1-Ser3. Even though the residues Ala-1 and Met0 are a cloning artifact (together with Gly-2 which is visible only in chain B), it can be expected that in the native protein, which starts with Val1, the structure is the same, albeit shorter by three amino acids. Three main independent molecules in the asymmetric unit is 0.32 Å for 276 pairs of superimposed Cα atoms, which is not significantly higher than the overall coordinate error of the structure.

Quality of the structure

The final refined model of Rv2827c consists of amino acids Ala-1 to Ala295 in chain A and Gly-2 to Ala295 in chain B with the exception of the solvent-exposed loop Arg267-Arg269 in both chains. It also includes 340 water molecules, two sodium ions, one MPD molecule, 15 formate ions and two acetate ions. Based on the refinement statistics and stereochemical parameters (Table 2) the quality of the model is high. About 92% of all residues are in the core region of the Ramachandran plot and no residues appeared in the generously allowed or unfavorable area. The root-mean-square deviation (r.m.s.d.) value between the two
chain–main chain hydrogen bonds connect the fragment Val1-Ser3 to strand β5 (Ala78-Ser81) in an antiparallel fashion. The residues Ala-1 and Met0 strengthen this antiparallel β-sheet with an additional main chain–main chain hydrogen bond which connects Ala-1 to Asp82 located at the C-terminus of the strand β5. After leaving the first β-strand, the polypeptide chain continues as a loop and then enters the three-helix bundle, which is built up of helices z1 (Ala15-Arg26), z2 (Lys32-Ala42) and z3 (Pro48-Ile58). Helices z2 and z3 form a helix-turn-helix motif followed by a β-hairpin consisting of the strands β3 (Leu61-Leu64) and β4 (Thr69-Ile73), which, together with the short strand β2 (Val29-Thr31) form a three-stranded antiparallel β-sheet. Following another loop is the short strand β5 (Ala78-Ser81), which interacts with β1 in an antiparallel fashion. The entire fragment between z1 and β4 exhibits the canonical winged helix (WH) fold making Rv2827c a member of the WH or the winged helix-turn-helix (wWH) family. The canonical WH motif consists of two wings (W1 and W2) which are extended loop structures, three z helices (H1, H2, and H3) and three β-strands (S1, S2, and S3), arranged in the order H1-S1-H2-S2-W1-S3-W2 [19, 68]. In Rv2827c, the order is as follows: z1-β2-z2-β3-β-W1-β4-W2 where W1 and W2 correspond to the fragments Pro65-Gly68 and Pro74-Glu77, respectively.

Residues Tyr84-Asp94 form the helical fragment z4 occurring between the N- and C-terminal parts of Rv2827c. Helix z4 leads to a four-stranded mixed β-sheet exhibiting the topology β9-β6-β7-β8. Further z-helical fragments occur in the loops connecting the β-strands: Helix z5 (Gly103-Leu110) between strands β6 (Met100-Ala102) and β7 (Ile121-Leu125), helix z6 (Asp133-Ser137) between β7 and β8 (Val139-Val142), and the two helices z7 (Asp150-Leu154) and z8 (Arg157-Arg164) before strand β9 (Pro176-Leu178). After leaving the β-sheet, the protein chain enters the z-helical region composed of helices z9 (Gly179-Arg190), x10 (Pro196-Val201) and x11 (His203-Asp210). These three helices are arranged in a circular structure separating the β-sheet and the three-helix bundle consisting of the almost parallel helices z12 (Ser212-Ser221), z13 (Pro224-Gly238), and z14 (Glu240-Ala249). The last part of Rv2827c structure consists of the three-stranded mixed β-sheet with the topology β10-β12-β11, which comprises the sequence segments Val258-Thr262, Gln279-Glu283 and Ser272-Ala275, respectively. The polypeptide chain is terminated with the helical segment z15 (Leu284-Ala295) that points into the solvent.

**Division of the structure into domains**

The division of the Rv2827c structure into domains turned out to be somewhat ambiguous (Fig. 2). The program 123D+ [1] divides Rv2827c into three domains: Gly-2-Ala97, Gly98-Leu183 and Leu184-Ala295. While the predicted border between the first two domains appears to be reasonable, the one between domains 2 and 3 occurs exactly in the middle of helix x9 (Gly179-Arg190) and it is buried in the protein core. The vector alignment search tool VAST Search [31, 48] predicts four domains: Gly-2-Gly98, Phe99-Ala152, Leu175-Val255 and Met256-Ala295, where the first predicted domain is identical to the one predicted by the program 123D+. This can be explained that for this part of the structure many related structures are available. The fragment Phe99-Ala152 was assigned as the next domain although it does not show any similarity to any known 3-D structures. The sequence Leu153-Gly174 was left out of the domain prediction by VAST and the next segment Leu175-Val255 was recognized as a separate domain. The last sequence stretch (Met256-Ala295) was again recognized as a separate domain. This fragment corresponds to the three-stranded β-sheet terminated by an x-helix. The assignment of this stretch to a domain is probably sensible, since it interacts with the rest of the molecule mostly by means of hydrophobic contacts. The difficulties of VAST in assigning domains in the C-terminal part of Rv2827c are a consequence of the lack of structural homologues for this region. This in turn corroborates the notion that the C-terminal part of Rv2827c constitutes a novel fold. By examining the solvent-accessible surface area (ASA) on a residue-by-residue basis using the PISA server [43] two potential linker regions were defined: Arg93-Asn96 and Lys250-Val255, thus dividing Rv2827c into three domains. The calculation of the normal modes using the ElNemo server [76] for Rv2827c divides the structure into two domains: Gly-2 to Ser81 and Asp82 to Ala295. The definition of the first domain contradicts to some extent the definitions of the programs 123D+ and VAST in that it assigns the helical fragment z4 (Tyr84-Asp94) to the C-terminal domain of Rv2827c rather than the N-terminal. The segment Tyr84-Asp94 exhibits strong interactions with both N- and C-terminal parts of the protein. The residues Tyr84, Leu87 and Trp90 are surrounded by the hydrophobic side chains of Ala97, Phe99, Leu101, Ile121, Ile123, Leu125, Leu131, Leu135 and Val139 from the C-terminal part of Rv2827c. Furthermore, the segment participates in the hydrogen bonds Tyr84-OH...His109-NE2, Arg88-NH1...Thr173-O and Arg88-NH2...Thr173-OG1. On the other side of z4, Leu85 and Ala92 interact with the hydrophobic side chains of Val11, Ile80, Val28 and Val29 of the N-terminal part, Ser89-OG forms a hydrogen bond with Val66-N and the side chains of Arg93 and Glu33 are connected by a salt bridge. Based on these considerations the segment Tyr84-Asp94 should probably be considered part of the C-terminal domain rather than the N-terminal. Similarly, the
The division of the C-terminal part into further TLS groups was unclear and did not yield any indication about potential further domain division, similar to the other programs mentioned above. Rv2827c is therefore most sensibly divided into two structural domains.

The N-terminal domain of Rv2827c exhibits DNA binding features

Table 5 shows the top scored structures related to the N-terminal domain fragment Thr13-Ile73 of Rv2827c identified by the secondary structure matching (SSM) server. Most of these proteins interact with DNA. The structure superposition shown in Fig. 3 clearly indicates the structural similarity between the N-terminal domain of Rv2827c and other proteins belonging to the WH or wHTH family. To date, many X-ray structures of WH proteins have been determined. Whereas the first structures were those of eukaryotic proteins, such as the hepatocyte nuclear factor 3 (HNF-3) [19] and histone H5 [68], it was soon recognized that the WH motif is common to eukaryotes and prokaryotes [9]. The most prominent feature of the WH motif is the presence of the two helices H2 (the stabilization helix) and H3 (the recognition helix) and the turn between them, the length of which is variable [27]. In the majority of WH protein structures H3 is responsible for the interaction with DNA. One exception to this rule is the human regulatory factor X1 where the W1 face is responsible for DNA binding [28]. In WH proteins, the H3 helix typically lies in the major groove and makes most of the sequence-specific contacts with nucleic acids via a number of hydrogen bonds and hydrophobic interactions [10]. A superposition of the N-terminal domain of Rv2827c with structures of protein-DNA complexes such as heat shock transcription factor (PDB code 3HTS, [46]), serum response factor (PDB code 1K6O, [51]), transcription factor Pu.1 (PDB code 1PUE, [41]), or interferon regulatory factor (PDB code 1IF1, [25]) indicates that the sequence fragment 48-ProAspSerAlaIleArgGluLeuArgArgIle-58 of Rv2827c (z3) might be responsible for the interaction with DNA. The presence of three Arg residues in the fragment further supports this hypothesis. An analysis of the electrostatic surface potential (Fig. 4) reveals that one side of the protein is negatively charged and includes a metal binding site whereas the other side shows a continuous path of positive potential extending along the whole molecule including the

![Springer]
potential nucleic acid binding motif in the N-terminal domain containing the helix α3. The second and third helices of the WH unit form a HTH variant motif containing a longer turn than the corresponding turn in canonical HTH proteins [10]. The HTH motif has been found in many DNA binding proteins that regulate gene expression and also in proteins involved in DNA repair and replication, as well as in RNA metabolism. It consists of two helices connected by the turn, in which Gly is usually found in its first [10] or second position [37]. The length of the turn connecting two helices of the typical HTH motif is 3 or 4 residues. In Rv2827c the HTH motif is built up of 27 residues, characterized by helices α2 (Lys32-Ala42) and α3 (Pro48-Ile58) linked by a five-residue turn which contains two Gly residues (Gly43 and Gly45). Although the turn contains two extra residues, its conformation resembles more closely the typical HTH motif than that found in WH proteins. In contrast, the angle between the two helices is approximately 100° although for a typical HTH motif it is usually about 120° [11]. For WH proteins, the angle between the two helices ranges from 100° in biotin operator repressor protein BirA [81] to 150° in transcription factor DP2 [85].

The C-terminal domain constitutes a novel fold

A structural search using the entire C-terminal domain did not yield any similarity hits to other known structures. Only when smaller segments were used, similar motifs were found in other protein structures. For the fragment Ile80-Pro145, the highest scoring structure identified by SSM is the
three-helix bundle (α12, α13, and α14) several other structures were identified. Among them are the hypothetical protein ST1625p from the hyperthermophilic archaeon Sulfolobus tokodaii (PDB code 1WY6, [83]) and human PEX5 (PDB code 1FCH, [30]). For the fragment Leu175-Ala295, no other similar structure could be identified, indicating that this fragment is unique, despite a sequence identity of 29% to 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from E. coli (PDB code 1RU1, [8]).

The sodium binding site

The C-terminal domain of Rv2827c subunit harbors a metal binding site formed by two Asp residues, one water molecule and the side chain of an Arg residue. In the refined structure the metal ion has been identified as Na⁺, although both Na⁺ and K⁺ were present in the crystallization solution at concentrations of 1.0 M and 0.075 M, respectively. The coordination number for the metal ion is six with the Me…X distances in chains A and B being 2.25 and 2.28 Å (Asp206-OD1), 2.55 and 2.44 Å (Asp206-O), 2.26 and 2.38 Å (Asp210-OD1), 2.46 and 2.54 Å (Asp210-OD2), 2.43, 2.27 Å (water-O) and 3.10 and 3.28 Å (Arg164-CZ). If the two carboxylate oxygens of Asp210 are counted as two ligands, the coordination geometry of the metal ion constitutes a distorted octahedron. This, together with the observed distances favors the presence of Na⁺ over K⁺ [35]. An analysis of the bond-valence parameters [12–14, 53] confirms this observation. The values of this parameter for Na⁺ are 1.16 and 1.31 (for the site in chains A and B, respectively), which is close to the expected value for this ion, while for K⁺ the values are 2.82 and 2.99. An usual feature of this site is that the guanidinium side chain of Arg164 is placed such that is

![Fig. 4](https://www.pymol.org) Surface presentation of M. tuberculosis Rv2827c, showing the electrostatic potential. The figure was prepared using PYMOL (www.pymol.org). The distribution of the electrostatic surface potential indicates that one side of the protein is negatively charged (red); this area includes the metal binding site. On the opposite side there is a continuous positively charged patch extending for the entire length of the molecule (blue), which includes the potential nucleic acid binding motif in the N-terminal domain containing the α3 helix.

![Fig. 5](https://www.pymol.org) Stereo view of the metal binding site in the C-terminal domain of Rv2827c. The unusual coordination of the metal ion by an Arg residue via a π-interaction is shown in green. The distances between the metal ion and the ligands are given in Å.
makes a π-contact with the metal ion while its hydrogen atoms are involved in interactions with Leu110-O, Asp206-O and a water molecule (Fig. 5). In the crystal structure of the GABA(A) repressor-associated protein (PDB code 1KJT, [5]) Na⁺ is also coordinated by an Arg side chain but in this case, the NH1-atom of Arg65 is directed toward the Na⁺ so that a π-contact as the one found in Rv2827c can not form. The role of the metal binding site in Rv2827c is unknown. It is located on the opposite side of the protein with respect to the potential nucleic acid binding site, in the region with mostly negatively charged character. It may thus just serve the purpose of structurally stabilizing the protein.

DNA binding properties of Rv2827c

Initial hints toward DNA binding properties of Rv2827c were obtained from experiments, which revealed that in the presence of DNA Rv2827c becomes more stable in solution and remains soluble even at low imidazole concentrations. Without DNA, Rv2827c is only stable at high concentrations of imidazole [38]. Of the four randomly chosen DNA sequences tried (Table 3), three were able to stabilize Rv2827c. Only the 26-bp fragment taken from the DNA complex of interferon regulatory factor 1 (PDB code 1IF1, [25]) did not stabilize Rv2827c, which precipitated during the dialysis. Neither the sequences nor the G+C content of the nucleotides used in the experiment provide an explanation for this phenomenon. A DNA binding test using the PCR-assisted binding site selection method [56] clearly demonstrated that Rv2827c possesses dsDNA binding capability. However, the DNA sequences derived from the 17 randomly picked clones did not reveal any sequence preference (Table 4). Based on this experiment we can only conclude that Rv2827c binds DNA in a non-specific manner.

Model of the complex of Rv2827c with DNA

A superposition of the N-terminal domain of Rv2827c and the WH domain of the interferon regulatory factor 3 (IRF-3, PDB code 1T2K, [64]) complexed with a 31-mer DNA fragment results in the model shown in Fig. 6a. The DNA fragment fits well to the positively charged N-terminal domain, and it continues alongside Rv2827c next to the positively charged side of the C-terminal domain (Fig. 4). If the C-terminal domain is slightly rotated relative to the N-terminal domain by adjusting the orientation of the C-terminal domain, then fit can be markedly improved (Fig. 6b). An ASA calculation on the Rv2827c/DNA-complex model reveals that DNA binding buries 1,330 Å² of surface area of Rv2827c, which amounts to 8.7% of its total surface. This lends further support to the hypothesis that Rv2827c is a DNA binding protein. In an anomalous difference Fourier map calculated based on the Br peak data set, 23 bromide binding sites can be identified. All but one bromide binding sites are conserved and occur in both independent molecules in the asymmetric unit. Four out of eleven sites in chain A and four out of twelve sites in chain B are located in the protein-DNA interface. This indicates a clear preference for negatively charged bromide ions to bind to the potential positively charged DNA binding surface further supporting the DNA binding hypothesis and the model of the Rv2827c/DNA complex.

Conclusions

In this article we present the three-dimensional structure of the hypothetical protein Rv2827c from Mtb determined at 1.93 Å resolution using the three-wavelength anomalous diffraction method. Rv2827c consists of two structural domains. The structure of the C-terminal domain of Rv2827c constitutes a novel fold whereas the structure of the N-terminal domain of Rv2827c exhibits a winged helix topology. A structural Na⁺ binding site was identified with an unusual coordination of the metal ion by a guanidinium side chain of Arg. It could also be shown that the presence of oligonucleotides can stabilize Rv2827c and prevent spontaneous precipitation in imidazole-free buffer. Furthermore, by PCR-assisted binding site selection it was demonstrated that Rv2827c indeed binds dsDNA. The analysis of the charge of the accessible surface area of Rv2827c suggests that both domains are involved in the interaction with DNA or RNA.
References

1. Alexandrov NN, Nussinov R, Zimmer RM (1995) Fast protein fold recognition via sequence to structure alignment and contact capacity potentials. In: Hunter L, Klein TE (eds) Pacific symposium on biocomputing ’96, World Scientific Publishing Co., Singapore, pp 53–72.

2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. doi:10.1093/nar/25.17.3389

3. Arunkumar AI, Klimovich V, Jiang X, Ott RD, Mizoue L, Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Hopwood DA (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417:141–146. doi:10.1038/52634

4. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242. doi:10.1093/nar/28.1.235

5. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242. doi:10.1093/nar/28.1.235

6. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Hopwood DA (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417:141–146. doi:10.1038/417141a

7. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242. doi:10.1093/nar/28.1.235

8. Blaszczyszyn J, Li Y, Wu Y, Shi G, Ji X, Yan H (2004) Essential roles of a dynamic loop in the catalysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. Biochemistry 43:1469–1477. doi:10.1021/bi0306531

9. Brennan RG (1993) The winged-helix DNA-binding motif: another helix-turn-helix takeoff. Cell 74:773–776. doi:10.1016/0092-8674(93)90456-Z

10. Brennan RG, Matthews BW (1989) The helix-turn-helix DNA binding motif. J Biol Chem 264:1903–1906

11. Brennan RG, Takeda Y, Kim J, Anderson WF, Matthews BW (1986) Crystallization of a complex of cro repressor with a 17 base-pair operator. J Mol Biol 188:115–118. doi:10.1016/0022-2836(86)90488-2

12. Brese NE, O’Keefe M (1991) Bond-Valence parameters for solids. Acta Crystallogr B 47:192–197. doi:10.1107/S0108768190011041

13. Brown ID (1977) Predicting bond lengths in inorganic crystals. Acta Crystallogr B 33:1305–1310. doi:10.1107/S0567740777005998

14. Brown ID, Altermatt D (1985) Bond-valence parameters obtained from a systematic analysis of the inorganic crystal structure database. Acta Crystallogr B 41:244–247. doi:10.1107/S0108768185002063

15. Camus J-C, Pryor MJ, Medigue C, Cole ST (2002) Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv. Microbiology 148:2967–2973

16. Carreno-Lopez R, Sanchez-Villa A, Camargo-Diaz N, Elmerich C, Baca BE (2006) Characterization of chsA, a new gene controlling the chemotactic response, in Azospirillum brasilense Sp7. EMBL/GenBank/DDBJ

17. CCP4, Collaborative Computational Project, Number 4 (1994) Acta Crystallogr. D50:760–763

18. Chatake T, Mizuno N, Voordouw G, Higuchi Y, Arai S, Tanaka I, Niimura N (2003) Crystallization and preliminary neuron analysis of the dissipatory sulfite reductase D (DsrD) protein from the sulfate-reducing bacterium Desulfovibrio vulgaris. Acta Crystallogr D Biol Crystallogr 59:2306–2309. doi:10.1107/S0907444903020596

19. Clark KL, Halay ED, Lai E, Burley SK (1993) Co-crystal structure of the HNF-3/fork head DNA recognition motif resembles histone H5. Nature 364:412–420. doi:10.1038/364420

20. Cohen GE (1997) ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. J Appl Cryst 30:1160–1161. doi:10.1107/S0021889897006729

21. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D et al (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393:537–544. doi:10.1038/31159

22. Cowtan K (1994) Joint CCP4 and ESI-EACBM newsletter on protein crystallography 31:34–38

23. Dauter Z, Dauter M, Rajashankar KR (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. Acta Crystallogr D Biol Crystallogr 56:232–237. doi:10.1107/S0907449999016352

24. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126–2132. doi:10.1107/S0907444904019158

25. Escalante CR, Yie J, Thanos D, Aggarwal AK (1998) Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. Nature 391:103–106. doi:10.1038/34224

26. French GS, Wilson KS (1978) On the treatment of negative intensity observations. Acta Crystallogr A 34:517–525. doi:10.1107/S0567739478001114

27. Gajiwala KS, Burley SK (2000) Winged helix proteins. Curr Opin Struct Biol 10:110–116. doi:10.1016/S0959-440X(99)00573-7

28. Gajiwala KS, Chen H, Cornille F, Roques BP, Reith W, Mach B, Burley SK (2000) Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature 403:916–921. doi:10.1038/3502634

29. Garnier T, Eiglmeier K, Camus J-C, Medina N, Mansoor H, Pryor M, Duthoy S, Grondon S, Lacroix C, Monsempre C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrett BG, Parkhill J, Hopwood DA (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417:141–147. doi:10.1038/417141a

30. Garnier T, Eiglmeier K, Camus J-C, Medina N, Mansoor H, Pryor M, Duthoy S, Grondon S, Lacroix C, Monsempre C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrett BG, Parkhill J, Hopwood DA (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417:141–147. doi:10.1038/417141a

31. Gibrat JF, Madej T, Bryant SH (1996) Surprising similarities in protein structure comparison. Curr Opin Struct Biol 6:377–385. doi:10.1016/S0959-440X(96)80058-3

32. Ha SC, Lokanath NK, Van Quyen D, Wu CA, Lowenhaupt K, Rich A, Kim YG, Kim KK (2004) A poxvirus protein forms a complex with left-handed Z-DNA: crystal structure of a Yata-poxvirus Zalpha bound to DNA. Proc Natl Acad Sci USA 101:14367–14372. doi:10.1073/pnas.0405586101
Structure of Rv2827c from *M. tuberculosis*
of protein networks. Microbes Infect 8:747–757. doi:10.1016/j.micinf.2005.09.011

67. Rachman H, Strong T, Ulrichs T, Grode L, Schuchhardt J, Mollenkopf H, Kosmiadi GA, Eisenberg D, Kaufmann SHE (2006) Unique transcriptome signature of Mycobacterium tuberculosis in pulmonary tuberculosis. Infect Immun 74:1233–1242. doi:10.1128/IAI.74.2.1233-1242.2006

68. Ramakrishnan V, Finch JT, Graziano V, Lee PL, Sweet RM (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature 362:219–223. doi:10.1038/362219a0

69. Sam MD, Cascio D, Johnson RC, Clubb RT (2004) Crystal structure of the excisionase-DNA complex from bacteriophage lambda. J Mol Biol 338:229–240. doi:10.1016/j.jmb.2004.02.053

70. Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48:77–84. doi:10.1046/j.1365-2958.2003.03425.x

71. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 58:1772–1779. doi:10.1107/S0907444902011678

72. Schwartz T, Rould MA, Lowenhaupt K, Herbert A, Rich A (1999) Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. Science 284:1841–1845. doi:10.1126/science.284.5421.1841

73. Sheldrick GM (2002) Macromolecular phasing with SHELXE. Z Kristallogr 217:644–650. doi:10.1524/zkri.217.12.644.20062

74. Sheldrick GM, Hauptman HA, Weeks CM, Miller R, Usón I (2001) In: Rossmann MG, Arnold E (eds) International tables for macromolecular crystallography, vol F. Kluwer Academic Publishers, Dordrecht, pp 333–345 Chapter 16

75. Slesarev AI, Mezhevaya KV, Makarova KS, Polushin NN, Shcherbinina OV, Shakhova VV, Belova GI, Aravind L, Natale DA, Rogozin IB, Tatusov RL, Wolf YI, Malykh AG, Koonin EV, Kozyavkin SA (2002) The complete genome of hyperthermophile Methanopyrus kandleri AV19 and monophyly of archaeal methanogens. Proc Natl Acad Sci USA 99:4644–4649. doi:10.1073/pnas.032671499

76. Suhre K, Sanejouand YH (2004) ElNemo: a normal mode webserver for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res 32:W610–W614. doi:10.1093/nar/gkh368

77. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29:22–28. doi:10.1093/nar/29.1.22

78. Vaguine AA, Richelle J, Wodak SJ (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr D Biol Crystallogr 55:191–205. doi:10.1107/S0907444998006684

79. Weiss MS (2001) Global indicators of X-ray data quality. J Appl Crystal 34:130–135. doi:10.1107/S0021889800018227

80. WHO (2008) Global tuberculosis control: surveillance, planning, financing WHO, Geneva http://www.who.int/tb/publications/global_report/en/index.html

81. Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW (1992) Escherichia coli biotin synthetase/biorepressor crystal structure delineates the biotin- and DNA-binding domains. Proc Natl Acad Sci USA 89:9257–9261. doi:10.1073/pnas.89.19.9257

82. Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr D Biol Crystallogr 57:122–133. doi:10.1107/S0907444900014736

83. Yoneda K, Sakuraba H, Tsuge H, Katunuma N, Kuramitsu S, Kawabata T, Ohshima T (2005) The first crystal structure of an archaeal helical repeat protein. Acta Crystallogr F61:636–639

84. Young D, Stark J, Kirschner D (2008) Systems biology of persistent infection: tuberculosis as a case study. Nat Rev Microbiol 6:520–528. doi:10.1038/nrmicro1919

85. Zheng N, Fraenkel E, Pabo CO, Pavletich NP (1999) Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. Genes Dev 13:666–674. doi:10.1101/gad.13.6.666