Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

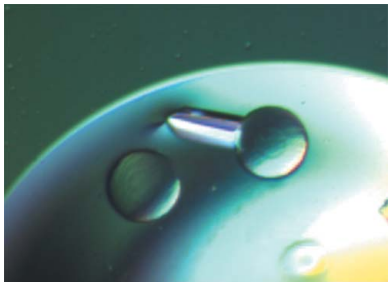
Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of Rv2827c from *Mycobacterium tuberculosis*

Robert Janowski,[a] Ali Nasser Eddine,[b] Stefan H. E. Kaufmann[b] and Manfred S. Weiss[a]*

[a]EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, and [b]Max-Planck-Institute for Infection Biology, Schumannstrasse 21/22, D-10117 Berlin, Germany

Correspondence e-mail: msweiss@embl-hamburg.de

Received 10 April 2006
Accepted 23 June 2006



The hypothetical protein Rv2827c from *Mycobacterium tuberculosis* was cloned and heterologously expressed in *Escherichia coli*. It was purified using affinity and size-exclusion chromatographic techniques and then crystallized. Preliminary X-ray diffraction data analysis suggests the presence of two translationally related molecules in the asymmetric unit of the orthorhombic crystals.

1. Introduction

At the beginning of the 21st century, tuberculosis remains one of the major infectious diseases globally. According to a recent WHO report, about one third of the world's population is expected to be infected with *Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis. In the year 2004, approximately nine million new tuberculosis cases occurred worldwide and approximately two million people died from the disease (World Health Organization, 2006).

The best-characterized and most widely used laboratory strain of the acid-fast bacterium Mtb is H37Rv. Its genome encodes approximately 4000 genes (Cole *et al.*, 1998; Camus *et al.*, 2002), many of which have no assigned functions (Kantardjieff & Rupp, 2004). Combined efforts by several groups are presently under way aimed at determining a large number of Mtb protein structures, for instance the TB Structural Genomics consortium, a worldwide structural genomics project (http://www.doe-mbi.ucla.edu/TB; Terwilliger *et al.*, 2003). These structures will provide the basis for the identification of novel candidate drug targets.

Information about gene functions can also be obtained by genome-wide expression profiling. The identification of single genes or whole gene networks which are differentially expressed under different conditions can often provide clues to the role of these genes, for instance in the persistence and pathogenicity of Mtb (Rachman, Strong, Schaible *et al.*, 2006; Rachman, Strong, Ulrichs *et al.*, 2006). Currently, only limited information on the three-dimensional architecture and structural features of the protein products of such genes is available. Therefore, knowledge of the three-dimensional structures of these proteins is needed because this may well provide a valuable basis for a better understanding of the biology of Mtb as well as for structure-based design of novel intervention strategies against tuberculosis. Many of these proteins are currently under investigation by the German-based X-MTB consortium (http://www.xmtb.org).

Rv2827c has been annotated as a hypothetical protein with unknown function (Cole *et al.*, 1998). Mtb mutants lacking a functional copy of the Rv2827c gene fail to grow *in vitro* (Sassetti *et al.*, 2003; Lamichhane *et al.*, 2003), which demonstrates that Rv2827c is indispensable for the growth of Mtb. Upon infection of macrophages with Mtb, the gene product of Rv2827c is upregulated approximately threefold compared with cultures of Mtb grown *in vitro*. IFN-γ activation of the infected macrophages leads to a further threefold increase (Rachman, Strong, Schaible *et al.*, 2006). It is therefore tempting to assume a critical role of Rv2827c in Mtb survival within macrophages, thus rendering Rv2827c an interesting target for future intervention strategies.

In this report, we describe the cloning, purification, crystallization and initial X-ray analysis of Rv2827c, which consists of 295 amino-

Table 1
Data-collection and processing statistics.

Values in parentheses correspond to the highest resolution shell.

No. of crystals	1
Beamline	X13
Wavelength (Å)	0.8031
Temperature (K)	100
Crystal-to-detector distance (mm)	180
Rotation range per image (°)	1.0
Total rotation range (°)	156
Space group	$P2_12_12$
Unit-cell parameters (Å)	a = 87.42, b = 180.65, c = 35.11
Mosaicity (°)	0.60
Resolution limits (Å)	50.0–1.93 (2.00–1.93)
Total No. of reflections	268981
Unique reflections	43027
Redundancy	6.3
$I/\sigma(I)$	23.3 (3.7)
Completeness (%)	99.9 (99.8)
$R_{\rm merge}$ (%)	7.6 (49.8)
$R_{\rm r.i.m.}$ (%)	8.3 (54.8)
$R_{\rm p.i.m.}$ (%)	3.3 (22.5)
Overall B factor from Wilson plot (Å^2)	24.2
Optical resolution (Å)	1.55

acid residues and has a molecular weight of 32.3 kDa. A sequence-similarity search using commonly used programs such as *BLAST* and *PSI-BLAST* (Altschul *et al.*, 1997) did not identify homologues of Rv2827c except for two hypothetical proteins with 100% amino-acid identity to Rv2827c encoded by ORFs in the genomes of *M. bovis* (Mb2851c) and *M. microti* (not yet annotated). To date, no tertiary structure has been reported in the PDB (Berman *et al.*, 2000) for any protein similar to Rv2827c. Also, Rv2827c has not been classified into any superfamily in the COG (Tatusov *et al.*, 2001) or Pfam databases (Bateman *et al.*, 2000). The only available information concerning the potential function of Rv2827c originates from the function-prediction program *ProKnow* (Pal & Eisenberg, 2005), which predicts that Rv2827c may have transferase, ATP-binding, lyase, kinase, tRNA-ligase, hydrolase or signal transduction activity, with an identical but rather low probability for each prediction. Thus, the predicted information is neither helpful nor reliable and consequently the elucidation of the three-dimensional structure of the protein Rv2827c may provide relevant additional information pointing towards its function.

2. Experimental methods

2.1. Cloning

The target sequence was amplified using genomic DNA from the H37Rv strain of Mtb as a template. The polymerase chain reaction was carried out using KOD Hot Start DNA polymerase (Novagen). It consisted of 30 cycles of denaturation (45 s), annealing (45 s) and extension (240 s) at temperatures of 368, 339 and 345 K, respectively. The sequence of the forward primer used was 5′-AAAA**CC-ATGG**TGGTGAGCCCAGCCGGCGCCGATC-3′ and that of the reverse primer was 5′-AAAA**CTCGAG**TTACGCCTTGCCGATCA-CGCGCAGCAGCGG-3′. The amplified fragment containing the 5′-*Nco*I and 3′-*Xho*I restriction sites (marked in bold in the primer sequences) was digested and ligated to the pETM-11 expression vector (EMBL), which was digested with the same restriction enzymes. The Rv2827c gene starts with the GTG codon for residue Val1 (Cole *et al.*, 1998). The pETM-11 vector adds an N-terminal His_6 tag and the recombinant tobacco etch virus protease (TEV) cleavage site to the expressed recombinant protein. Two additional residues, Gly and Ala, are left on the N-terminus of the protein after the removal of the His_6 tag. In addition, the expressed protein contains a Met residue left over from the *Nco*I restriction site. Thus, the expressed and TEV-cleaved protein contains the three-residue N-terminal extension Gly-Ala-Met. The construct was sequenced to confirm the cloning of the Rv2827c gene sequence in the frame.

2.2. Expression and purification

The recombinant plasmid was used to transform *Escherichia coli* Rosetta(DE3)pLysS cells (Novagen). Cells from an overnight 5 ml pre-culture were grown in 1 l LB broth medium containing chloramphenicol (34 µg ml^{-1}) and kanamycin (30 µg ml^{-1}) at 310 K and 210 rev min^{-1}. The culture was induced with 1 m*M* isopropyl β-D-thiogalactopyranoside (IPTG) at an OD_{600} of approximately 0.6 at 293 K. After induction, the culture was incubated for about 20 h at 293 K and 210 rev min^{-1} and then harvested. The cells were frozen and stored at 253 K until further processing. The cell pellet was dissolved in buffer *A* [50 m*M* Tris pH 7.3, 150 m*M* KCl, 5 m*M* 2-mercaptoethanol and one Complete Mini EDTA-free Protease Inhibitor Cocktail tablet (Roche) per 25 ml] and then lysed by sonication four times for 4 min in 0.3 s pulses at 277 K. For each gram of cell pellet, 10 ml buffer *A* was used. The cell debris was pelleted by centrifugation for 60 min at 277 K and 20 000 rev min^{-1}. The crude lysate was filtered through a 0.2 µm membrane and was loaded onto a 5 ml Hi-Trap Chelating HP (Amersham Pharmacia Biotech) column charged and equilibrated with Ni^{2+} and buffer *A*, respectively. In order to remove unbound proteins, the column was first washed with seven column volumes of buffer *A* and then with five column volumes of buffer *B* (50 m*M* Tris pH 7.3, 150 m*M* KCl, 50 m*M* imidazole, 5 m*M* 2-mercaptoethanol). The protein was eluted by running a linear gradient from 50 to 750 m*M* imidazole (in buffer *B*). Attempts to remove imidazole from the protein solution or to replace it by increasing salt concentration were unsuccessful. Therefore, TEV cleavage as well as all further steps was performed in the presence of imidazole. For TEV cleavage, the major peak fractions were pooled, TEV was added at a ratio of 1:50 and the mixture was left overnight at 277 K. The protein was subsequently purified by gel filtration (Superdex 200, 16/60, Amersham Pharmacia Biotech) using buffer *C* (50 m*M* Tris pH 7.3, 150 m*M* KCl, 350 m*M* imidazole, 2 m*M* DTT) for both equilibration and elution. The peak fractions were pooled and concentrated to 14 mg ml^{-1} for crystallization. The purity of the prepared sample was judged using 10% SDS–PAGE stained with Coomassie Brilliant Blue. The oligomeric state of the protein was assessed by gel filtration and dynamic light scattering (DLS).

2.3. Crystallization

Recombinant Rv2827c from Mtb was crystallized in 24-well plates using the hanging-drop vapour-diffusion method. The initial crystallization screening was performed using the protein in buffer *C* (14 mg ml^{-1}) and commercially available sparse-matrix screens (Jancarik & Kim, 1991) from Hampton Research. Protein solution (1.5 µl) was mixed with reservoir solution (1.5 µl) and equilibrated over a 1 ml reservoir. In condition No. 34 of the Crystal Screen, as well as several other conditions containing high salt (1.5–2.0 *M*), Rv2827c forms spherulites at 292 K. Optimization of these conditions at room temperature was not successful. However, single crystals of Rv2827c could be grown from 2 *M* sodium formate and 100 m*M* sodium acetate pH 4.7 at 277 K. To improve the quality of the crystals, Additive Screens (Hampton Research) were used. The best results were obtained in the presence of 0.2 *M* nondetergent sulfobetaine 201 (NDSB) or 3%(*v*/*v*) 6-aminocaproic acid. In these cases, the drop

Table 2
Average $I/\sigma(I)$ values for the even and odd reflections of classes (h00), (0k0) and (00l).

Resolution limits (Å)	(h00), h even	(h00), h odd	(0k0), k even	(0k0), k odd	(00l), l even	(00l), l odd
50.0–6.0	25.8	0.6	30.2	1.9	34.1	15.0
6.0–3.0	26.2	−3.1	25.4	0.5	22.2	9.2
3.0–1.93	2.9	−0.1	7.8	−0.7	7.8	7.4
50.0–1.93	16.8	−1.0	20.2	0.4	18.4	10.3

was composed of 1.2 µl protein solution, 0.3 µl additive solution and 1.5 µl reservoir solution.

2.4. Diffraction data collection and processing

A crystal with dimensions 300 × 150 × 150 µm was separated from the spherulite from which it grew, mounted in a nylon-fibre loop and flash-cooled to 100 K in a nitrogen-gas stream. For cryoprotection, the reservoir solution was supplemented with 15%(*v*/*v*) MPD and the crystal was soaked in this solution for 10 s. Diffraction data were collected on beamline X13, EMBL Hamburg, Germany using a MAR CCD detector. The data were indexed and integrated using *DENZO* (Otwinowski & Minor, 1997) and scaled using *SCALEPACK* (Otwinowski & Minor, 1997). The redundancy-independent merging R factor $R_{\rm r.i.m.}$ as well as the precision-indicating merging R factor $R_{\rm p.i.m.}$ (Weiss, 2001) 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* (French & Wilson, 1978; Collaborative Computational Project, Number 4, 1994). Table 1 summarizes the data-collection and processing statistics. The optical resolution was calculated using the program *SFCHECK* (Vaguine *et al.*, 1999) and the self-rotation function was calculated using the program *MOLREP* (Collaborative Computational Project, Number 4, 1994; Vagin & Teplyakov, 1997).

3. Results and discussion

Expression of Rv2827c in *E. coli* Rosetta(DE3)pLysS cells resulted in approximately 80% of the protein in the soluble fraction and 20% in inclusion bodies. After the two-step chromatographic procedure, the final yield of pure protein was approximately 5 mg from a 1 l culture. The purity of the sample prepared for crystallization was at least 95% as estimated by SDS–gel electrophoresis. From the gel-filtration column, the protein eluted with an apparent molecular weight of approximately 32 kDa, which is consistent with a monomeric state. This result was confirmed by DLS.

The growth of single crystals was always preceded by the appearance of solid spheroidal objects, *i.e.* spherulites. 1 d after the crystallization experiment was set up, spherulites appeared in the crystallization drops. The spherulites could be stained with Methylene Blue, indicating that they consist of protein material. Within the next few days, small needles grew from the surface of the spherulites. The needles then took another 2–5 weeks to reach a sufficient size. Fig. 1 shows a crystal of Rv2827c from Mtb grown on the surface of a spherulite. From the same crystal, an X-ray diffraction data set of about sixfold redundancy was collected to 1.93 Å resolution. The data-processing and scaling statistics (Table 1) indicate that the data set is of good quality. The $I/\sigma(I)$ value of 3.7 in the outer resolution bin seems to suggest that the crystals may even diffract a little further than 1.9 Å. This notion is corroborated by the value for the optical resolution of 1.55 Å (Table 1). However, owing to the appearance of a large number of overlapping reflections at high resolution, the maximum resolution was limited to 1.93 Å. The crystals belong to the orthorhombic crystal system, space group $P2_12_12$, with unit-cell parameters a = 87.42, b = 180.65, c = 35.11 Å. Based on the molecular weight of the protein (32.6 kDa, including the three extra amino-acid residues) and the volume of the asymmetric unit, Matthews parameters (Matthews, 1968) of 4.3, 2.1 and 1.4 Å^3 Da^{-1} for one, two and three molecules of Rv2827c in the asymmetric unit can be calculated. This suggests the presence of two molecules per asymmetric unit, which corresponds to a solvent content of 42%. The self-rotation function does not show any distinct peak on the $\kappa = 180^\circ$ section, but the native Patterson map exhibits a strong peak at fractional coordinates (x, y, z) = (0.50, 0.28, 0.50), which indicates that the two independent molecules in the asymmetric unit are related by translational symmetry. Initially, the presence of translational symmetry made the space-group assignment ambiguous, since the translational symmetry should render both the odd (h00) reflections as well as the odd (00l) reflections weak. However, a closer look at the relative $I/\sigma(I)$ values of the respective reflection classes (Table 2) revealed that the systematic extinctions are clear along the h and the k directions and that along the l direction (at least at higher resolution) the even and odd reflections exhibit nearly the same $I/\sigma(I)$ values. Thus, $P2_12_12$ is the most likely space group.

No homologous structure for Rv2827c is currently available from the PDB (Berman *et al.*, 2000). Therefore, structure determination will be attempted using isomorphous replacement and anomalous scattering methods.

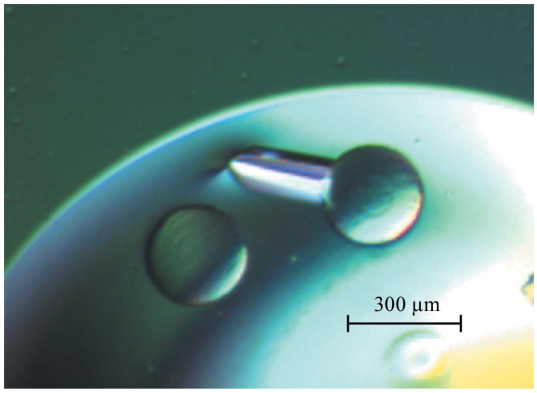


Figure 1
A single crystal of Rv2827c from *M. tuberculosis* grown from a spherulite.

4. Conclusions

Rv2827c from Mtb has been cloned and heterologously expressed in *E. coli*. Moreover, it has been purified to homogeneity using affinity and size-exclusion chromatographic techniques. Single crystals of the protein could only be grown from spherulites which had formed prior to crystal formation in the crystallization drop. Preliminary diffraction data analysis suggests the presence of two molecules in the asymmetric unit of the crystals, which are related by translational symmetry.

We would like to thank Dr L. Jeanne Perry (UCLA) for providing genomic Mtb H37Rv DNA and the X-Mtb consortium (http://www.xmtb.org) for funding through BMBF/PTJ grant No. BIO/0312992A.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. L. (2000). *Nucleic Acids Res.* **28**, 263–266.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Camus, J.-C., Pryor, M. J., Médigue, C. & Cole, S. T. (2002). *Microbiology*, **148**, 2967–2973.

Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

French, G. S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–525.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Kantardjieff, K. & Rupp, B. (2004). *Curr. Pharm. Des.* **10**, 3195–3211.

Lamichhane, G., Zignol, M., Blades, N. J., Geiman, D. E., Dougherty, A., Grosset, J., Broman, K. W. & Bishai, W. R. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 7213–7218.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pal, D. & Eisenberg, D. (2005). *Structure*, **13**, 121–130.

Rachman, H., Strong, M., Schaible, U., Schuchhardt, J., Hagens, K., Mollenkopf, H., Eisenberg, D. & Kaufmann, S. H. E. (2006). *Microbes Infect.* **8**, 747–757.

Rachman, H., Strong, M., Ulrichs, T., Grode, L., Schuchhardt, J., Mollenkopf, H., Kosmiadi, G. A., Eisenberg, D. & Kaufmann, S. H. E. (2006). *Infect. Immun.* **74**, 1233–1242.

Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. & Koonin, E. V. (2001). *Nucleic Acids Res.* **29**, 22–28.

Terwilliger, T. C. *et al.* (2003). *Tuberculosis*, **83**, 223–249.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.

Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.

World Health Organization (2006). *Global Tuberculosis Control: Surveillance, Planning, Financing.* WHO/HTM/TB/2006.362. Geneva: World Health Organization.