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

Cloning, expression, purification, crystallization and preliminary X-ray crystallographic analysis of initiation factor 1 from *Mycobacterium tuberculosis*

Georgios N. Hatzopoulos and Jochen Mueller-Dieckmann*

EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany

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

Received 19 December 2006
Accepted 26 January 2007

Initiation factor 1 (IF-1; Rv3462c) from *Mycobacterium tuberculosis*, a component of the 30S initiation complex, was cloned and heterologously expressed in *Escherichia coli*. The protein was purified by affinity and size-exclusion chromatography and crystallized. A complete data set has been collected to high resolution. The crystals belonged to space group $P2_12_12$, with two molecules per asymmetric unit which are related by translational symmetry.

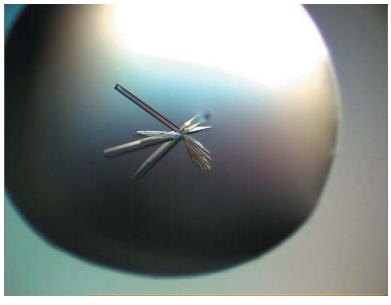
1. Introduction

Tuberculosis (TB) is a worldwide pandemic and remains one of the most common and deadly infectious diseases of our time. According to the World Health Organization, one third of the world's population is infected with TB and 10% (200 million individuals) of those infected will develop an active form of the disease during their lifetime. The highest rates of TB, which kills approximately two million people each year, are found in sub-Saharan Africa and Asia, but as a consequence of increased global mobility the disease is spreading to all countries. The disease is curable; however, treatment is prolonged and requires a combination of several drugs. At the same time, multidrug-resistant strains of *Mycobacterium tuberculosis* (Mtb), the bacillus that causes TB, have been detected in virtually all of the 109 countries surveyed in 2006. It is estimated that 450 000 new cases of multidrug-resistant TB occur every year (World Health Organization, 2006).

The best-characterized and most widely used strain of Mtb is H37Rv. Its genome comprises 4 411 529 base pairs which encode around 4000 genes (Cole *et al.*, 1998). Combined efforts by several organizations are presently under way to determine the three-dimensional structures of a large number of Mtb proteins; for example, the TB Structural Genomics consortium, a worldwide structural genomics collaboration (http://www.doe-mbi.ucla.edu/TB; Terwilliger *et al.*, 2003). These efforts lay the foundation for functional characterization of the gene products of Mtb in order to improve our knowledge of this deadly organism.

One partner of the TB Structural Genomics consortium is the German X-MTB consortium (http://www.xmtb.org), which focuses on a subset of validated gene products. Targets of this subset were identified by comparing the expression and transcription profiles of Mtb during different life cycles or under different growth conditions. These differences often indicate that the corresponding proteins are involved in and are important for the persistence or pathogenicity of Mtb. Knowledge of the three-dimensional structures of these targets provides information on their molecular function and supports structure-based efforts to design novel compounds to manipulate their cellular activity.

Rv3462c from Mtb encodes initiation factor 1 (IF-1), which is a component of the 30S initiation complex. Formation of the 30S complex precedes the assembly of the elongation-competent 70S ribosome and is the rate-limiting and most highly regulated step of



protein biosynthesis in prokaryotes. The complex consists of the 30S ribosome, fMet-tRNA, mRNA and the three initiation factors IF-1, IF-2 and IF-3 (Gualerzi & Pon, 1990). The three initiation factors control a rate-limiting rearrangement within this complex promoting the correct codon–anticodon formation at the ribosomal P-site.

The different initiation factors are structurally unrelated and IF-1 is the smallest of the three factors, with a molecular weight of 8.5 kDa (73 amino-acid residues) in Mtb.

2. Experimental methods

2.1. Cloning

Genomic DNA from the H37Rv strain of Mtb was used as a template for the polymerase chain reaction (PCR). Amplification of *Rv3462c* was carried out using the touchdown PCR reaction in order to avoid spurious priming (Don *et al.*, 1991). The protocol consisted of five cycles of annealing, starting at 345 K. The temperature was lowered by 2 K after each round of annealing to a final temperature of 335 K. The reaction was finished with another 35 PCR cycles with an annealing temperature of 337 K. The sequences of the forward and reverse primers were 5′-CACCATGGCCAAGAAGGACGGTGC-CATAGAGGTCGAGGG-3′ and 5′-CTACTTGTACCGGTACAC-GATGCGGCCCCGG-3′, respectively.

The amplified fragment was cloned into the pET151/D-TOPO vector using the TOPO cloning kit (Invitrogen), which permits directional cloning into expression systems. The vector adds an N-terminal His_6 tag, a V5 epitope and a tobacco etch virus (TEV) protease cleavage site to the expressed recombinant protein. Correct cloning was verified by sequencing the DNA construct obtained by re-amplifying *Rv3462c* from the ligated pET151/D-TOPO vector using standard primers for T7 promoters.

2.2. Expression and purification

The recombinant plasmid was used to transform *Escherichia coli* BL21(DE3)-RP cells (Novagen). 1 l LB broth medium containing 50 µg ml^{-1} carbenicillin and 34 µg ml^{-1} chloramphenicol was inoculated with cells from a 20 ml overnight culture at 310 K and grown at 210 rev min^{-1} to an optical density of approximately 0.5 at 600 nm. The culture was cooled to 293 K before induction with 0.3 m*M* isopropyl *β*-D-thiogalactopyranoside (IPTG). After induction, the culture was incubated at 293 K and 210 rev min^{-1} for about 16 h. The cells were pelleted by centrifugation and stored at 253 K until further processing.

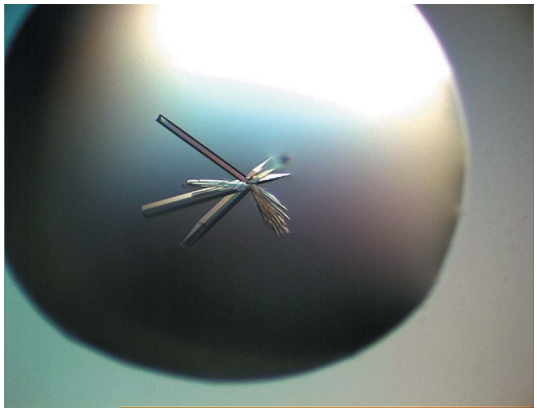


Figure 1
Cluster of crystals of IF-1 from *M. tuberculosis*.

Cell pellets were resuspended in 50 ml lysis buffer [50 m*M* MES pH 6.0, 200 m*M* NaCl, 10% glycerol, 3 m*M* 2-mercaptoethanol, 1 mg ml^{-1} DNase I and two Complete Mini EDTA-free Protease Inhibitor Cocktail tablets (Roche)] and then sonicated for 3 × 4 min in 0.3 s pulses at 277 K. The lysed cell suspension was cleared of debris by centrifugation for 60 min at 277 K and 5400*g*. The supernatant was filtered through a 0.2 µm membrane and loaded onto a 5 ml Hi-Trap Chelating HP column (Amersham Pharmacia Biotech) which had first been equilibrated against 100 m*M* $NiSO_4$ and then against buffer *A* (50 m*M* MES pH 6.0, 200 m*M* NaCl and 3 m*M* 2-mercaptoethanol). In order to remove unbound proteins, the column was washed with five column volumes of buffer *A* followed by five column volumes of high-salt buffer (buffer *A* containing 800 m*M* NaCl). The protein was eluted by running a linear gradient from 0 to 800 m*M* imidazole in buffer *A*. The pooled peak fractions were incubated with recombinant TEV protease in a 1:100 molar ratio for 12 h and left to dialyse against buffer *A* overnight. The protein solution was again loaded onto a Hi-Trap Chelating HP column in order to remove uncleaved protein molecules and the His_6 tag. Size-exclusion chromatography (HiLoad Superdex 75, 16/60, Amersham Pharmacia Biotech) was used as the final step of the purification procedure. The column was equilibrated with buffer *C* (50 m*M* sodium citrate pH 4.5, 50 m*M* NaCl, 5 m*M* EDTA, 5 m*M* DTT) and loaded with no more than 2 ml sample solution. The protein eluted as a single peak with an apparent molecular weight of about 9 kDa, which is consistent with the weight of recombinant Rv3462c (IF-1) in its monomeric state. Recombinant Rv3462c from this study carries six additional amino acids (GIDPFT) at the N-terminus from cloning. Peak fractions from the size-exclusion chromatography were analyzed with SDS–PAGE, pooled and concentrated to 10 mg ml^{-1}. Buffer *C* was also used as a storage buffer. The composition of buffer *C* was refined over several rounds of purification to prevent aggregation and precipitation of IF-1. The purity of the combined protein fractions was assessed with a 4–20% gradient SDS–PAGE stained with Coomassie Brilliant Blue. The monodispersity of the purified sample was also confirmed by dynamic light scattering (DLS).

2.3. Crystallization

Initial crystallization trials were performed as sitting-drop vapour-diffusion experiments at 292 K in 96-well trays at the EMBL Hamburg high-throughput crystallization facility (Mueller-Dieckmann, 2006). Over 1000 initial screening conditions were tested at a sample concentration of 10 mg ml^{-1}. Several crystallization droplets contained spherulites and the following four conditions yielded crystals: condition No. 48 from Hampton Research Crystal Screen 1 [0.1 *M* Tris–HCl pH 8.5, 2.0 *M* $(NH_4)H_2PO_4$], condition No. 13 from Jena Bioscience Screen (JBS) 2 (0.1 *M* HEPES pH 7.5, 22% PEG 4000, 0.1 *M* sodium acetate), condition No. 2 from JBS 5 (12% PEG 8000, 10% glycerol, 0.5 *M* KCl) and condition No. 14 from JBS 10 (0.1 *M* MES pH 6.5, 1.6 *M* $MgSO_4$). Optimization attempts using the conditions from Crystal Screen I, JBS 2 and JBS 10 at room temperature were unsuccessful. Well diffracting crystals were obtained in hanging-drop vapour-diffusion experiments by optimizing condition 2 from JBS 5. The best crystals grew in 12–16%(*w*/*v*) PEG 8000, 50–200 m*M* KCl and 8–16%(*v*/*v*) glycerol. Crystals grew rapidly (within ∼9 h) in 1 + 1 µl drops from single nucleation points and formed clusters of rod-like crystals (Fig. 1). Stored protein samples showed a considerable tendency to nucleate and form microcrystals when kept at 277 K. Once nucleation had occurred in a sample it was no longer possible to obtain crystals of good quality, even after attempts to remove nuclei either by dissolving them

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

No. of crystals	1
Beamline	BW7A, EMBL Hamburg
Wavelength (Å)	0.975
Temperature (K)	100
Crystal-to-detector distance (mm)	100
Rotation range per image (°)	0.4
Total rotation range (°)	192
Space group	$P2_12_12$
Unit-cell parameters (Å)	a = 73.95, b = 76.54, c = 28.03
Mosaicity (°)	0.70
Resolution limits (Å)	30.00–1.47 (1.50–1.47)
Total No. of reflections	197099
Unique reflections	27463 (1330)
Redundancy	7.2
$\langle I/\sigma(I)\rangle$	20.8 (5.6)
Completeness (%)	98.2 (98.2)
$R_{\rm merge}$† (%)	7.9 (35.5)
$R_{\rm r.i.m.}$‡ (%)	8.6 (38.8)
$R_{\rm p.i.m.}$§ (%)	3.2 (14.7)
Overall B factor from Wilson plot (Å^2)	17.8
Optical resolution¶ (Å)	1.25

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I_i(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm r.i.m.} = \sum_{hkl}[N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. § $R_{\rm p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ¶ Optical resolution = $[2(\sigma^2_{\rm Patt} - \sigma^2_{\rm sph})]^{1/2}$.

through dilution or through filtration. We suspect in both cases that it was not possible to completely remove all nuclei, which then interfered with subsequent crystallization attempts. It was therefore important to set up crystallization trials immediately after protein samples had been concentrated.

2.4. Collection and processing of diffraction data

A single crystal was separated from a cluster and mounted in a nylon-fibre loop. The crystal was then flash-cooled to 100 K in a nitrogen-gas stream. Prior to flash-freezing, the crystal was soaked for 10 s in reservoir solution supplemented with 25% glycerol as a cryoprotectant. Diffraction data were collected on beamline BW7A (EMBL Hamburg, c/o DESY, 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 merging R factors $R_{\rm p.i.m}$, which indicates the precision of the data set, and $R_{\rm r.i.m}$, which is redundancy-independent, were calculated using *RMERGE* (available at http://www.embl-hamburg.de/~msweiss/projects/msw_qual.html; Weiss, 2001). 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

Rv3462c (IF-1) from Mtb was expressed in BL21(DE3)-RP cells, resulting in close to 100% soluble protein. The recombinant protein was purified by affinity and size-exclusion chromatography and the final yield was 10 mg of pure protein per litre of LB medium. The sample was 99% pure as estimated by SDS–gel electrophoresis. The protein elutes from the size-exclusion chromatography column with an apparent molecular weight of 9 kDa. The experimental molecular weight of the recombinant Rv3462c of this study, as determined by MALDI–TOF mass spectrometry on a crystal of Rv3462c, is 9140.0 Da. This is in good agreement with the theoretical molecular weight of 9119.6 Da. The DLS analysis of the sample indicates that it is monodisperse.

High-throughput screening of initial conditions resulted in several leads. The best diffracting crystals grow in 12–16%(*w*/*v*) PEG 8000, 50–200 m*M* KCl and 8–16%(*v*/*v*) glycerol as clusters of rod-like crystals which could be separated easily. The crystals belong to the orthorhombic crystal system, space group $P2_12_12$, with unit-cell parameters a = 73.95, b = 76.54, c = 28.03 Å. A complete X-ray diffraction data set of about sevenfold redundancy was collected to 1.47 Å resolution. Based on the molecular weight of the protein (9120 Da, including six additional N-terminal amino-acid residues as cloning artefacts) and the volume of the asymmetric unit, the Matthews parameters (Matthews, 1968) for one, two and three molecules in the asymmetric unit are 4.3, 2.2 and 1.4 Å^3 Da^{-1}, respectively. This suggests the presence of two molecules in the asymmetric unit, which corresponds to a solvent content of 43%. The self-rotation function did not reveal significant noncrystallographic peaks on the section corresponding to $\kappa = 180°$. A strong peak was found in a native Patterson map at fractional coordinates 0.41, 0.31, 0.50, indicating the presence of translational symmetry between the two independent molecules in the asymmetric unit (Fig. 2).

Sequence similarity searches using the *BlastP* algorithm against the Protein Data Bank (PDB; Berman *et al.*, 2002; http://www.rcsb.org) identified two similar structures. One is the NMR structure of IF-1 from *E. coli* (69% sequence identity with Rv3462c; PDB code 1ah9) and the other is the X-ray structure of IF-1 from *Thermus thermophilus* (65% sequence identity with Rv3462c; PDB code 1hr0) in complex with the 30S ribosomal subunit. Attempts to phase Rv3462c with molecular replacement (MR) using either structure as a search model in *MOLREP* and *CaspR* failed. The NMR structure of IF-1 from *E. coli* does not indicate highly flexible regions apart from the protein's termini. We believe that MR failed because of translational

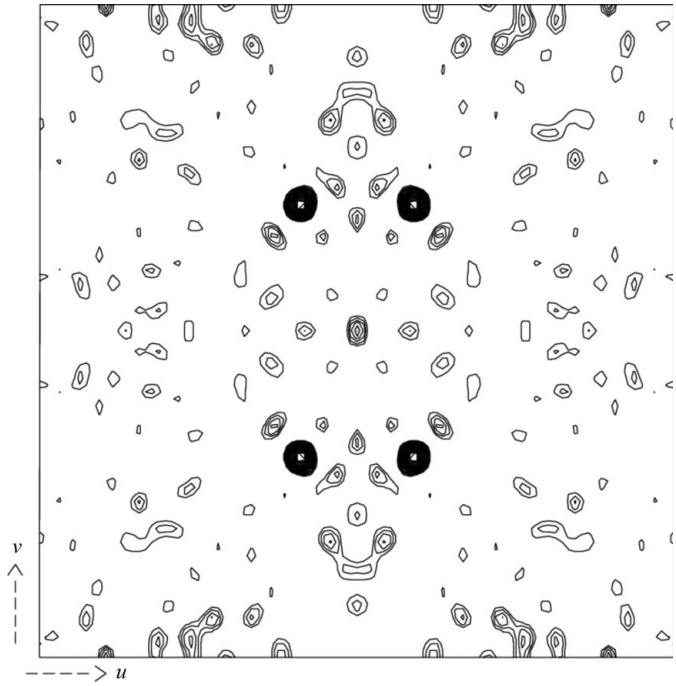


Figure 2
Section of the native Patterson map of Rv3462c at w = 0.5. The largest peak of 13.9σ corresponds to fractional coordinates u = 0.41, v = 0.31 and w = 0.5.

symmetry in the crystal. As a consequence, structure determination will continue using multiple isomorphous replacement and anomalous scattering methods.

References

Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002). *Acta Cryst.* D**58**, 899–907.

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

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

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. (1991). *Nucleic Acids Res.* **19**, 4008.

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

Gualerzi, C. O. & Pon, C. L. (1990). *Biochemistry*, **29**, 5881–5889.

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

Mueller-Dieckmann, J. (2006). *Acta Cryst.* D**62**, 1446–1452.

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

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). *2006 Tuberculosis Facts.* Geneva: World Health Organization. http://www.who.int/entity/tb/publications/2006/tb_factsheet_2006_1_en.pdf.