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

Cloning, expression, purification, crystallization and preliminary X-ray crystallographic analysis of isocitrate dehydrogenase 2 (Rv0066c) from *Mycobacterium tuberculosis*

Georgios N. Hatzopoulos,[a] Georgia Kefala[b] and Jochen Mueller-Dieckmann[a]*

[a]EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, and [b]Structural Biology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

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

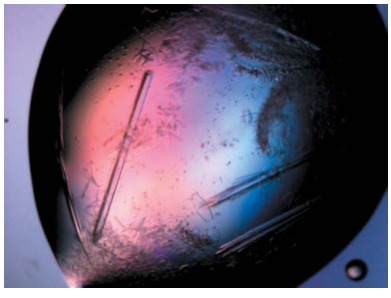
Received 5 August 2008
Accepted 29 October 2008

Isocitrate dehydrogenase 2 (Icd-2, Rv0066c) from *Mycobacterium tuberculosis* 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 and reduced to 3.25 Å resolution in space group *C*2. Preliminary diffraction data analysis suggests a complex packing with at least six molecules in the asymmetric unit.

1. Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (*Mtb*), remains one of the most common and deadly infectious diseases of our time and is responsible for approximately two million deaths every year worldwide. According to the World Health Organization, one third of the world's population is infected with TB and 10% of those infected (200 million individuals) will develop an active form of the disease during their lifetime (World Health Organization, 2007). TB can be controlled; the treatment, however, is prolonged and requires the combination of several drugs. Those complications, combined with severe side effects, often result in patient noncompliance and/or premature therapeutic termination. Both factors contribute to the increasing emergence of multidrug-resistant strains of *Mtb*.

A hallmark of *Mtb* is its ability to evade the immune system by hiding inside macrophages and granulomas. Infection with the pathogen can persist and last a lifetime without any obvious signs of illness (Helmuth, 2000). Intracellular pathogens, including *Mtb*, release proteins *in vitro* and into their phagosomes within macrophages. The use of extracellular proteins for protective immunity, induced by vaccination (Horwitz *et al.*, 1995), or as immunogenic markers for serodiagnosis (Banerjee *et al.*, 2004) has been demonstrated. Among the *Mtb* proteins which become accessible to proteolytic processing and subsequent presentation as major histocompatibility complex-bound fragments is isocitrate dehydrogenase (Icd; EC 1.1.1.42). In *Mtb* Icd is expressed in two isoforms, Icd-1 and Icd-2. Both enzymes catalyze the $NADP^+$-dependent oxidative decarboxylation of 2*R*,3*S*-isocitrate to 2-oxoglutarate and CO_2 in the Krebs cycle (Chen & Gadal, 1990). Icd-1, with a molecular weight of 40–50 kDa, is found in a wide variety of prokaryotes and eukaryotes, where it forms homodimers. Structural analysis revealed that residues from both subunits of Icd-1 contribute to the active site, which is located at the dimer interface (Hurley *et al.*, 1989). The molecular weight of Icd-2, which is found in a few species of bacteria, is 55–80 kDa, almost twice that of Icd-1. The structural analyses of Icd-2 from *Azotobacter vinelandii* in complex with isocitrate (Yasutake *et al.*, 2002) and of substrate-free Icd-2 from *Corynebacterium glutamicum* (Imabayashi *et al.*, 2006) showed that the enzyme consists of two domains and that the folding topologies of Icd-1 and Icd-2 are related. However, there is contradictory evidence in the literature



regarding the oligomeric state of Icd-2. Both crystallographic structures support the general view of a monomeric state of Icd-2. Biochemical analyses of Icd-2 from *Mtb*, on the other hand, indicate a dimeric or trimeric state of the enzyme (Banerjee *et al.*, 2005). Our studies point to a trimeric state of Icd-2 from *Mtb* in solution which is in accordance with these findings.

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 *Rv0066c* 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 25 PCR cycles with an annealing temperature of 333 K.

The sequences of the forward and reverse primers were 5′-CACCATGAGCGCCGAACAGCCGACCATCATTTACAC-3′ and 5′-AAAACTCGAGTTATCAGCCTTGGACAGCCTCCAGCG-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 *Rv0066c* 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-1-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.

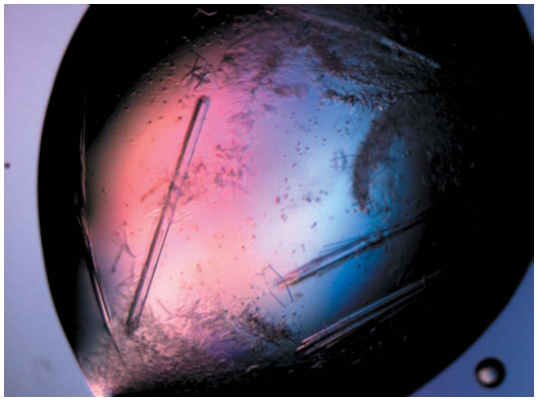


Figure 1
Crystals of Rv0066c from *M. tuberculosis*. Two different crystal forms appeared in the same droplets: plate-like and rod-shaped crystals. The rod-shaped crystals diffracted to about 3 Å resolution.

Cell pellets were resuspended in 50 ml lysis buffer [30 m*M* Tris–HCl pH 8.0, 250 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 from 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 been equilibrated first against 100 m*M* $NiSO_4$ and then against buffer *A* (30 m*M* Tris–HCl pH 8.0, 250 m*M* NaCl, 5% glycerol, 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 1:200 molar ratio for 12 h and left overnight to dialyse against buffer *A*. The protein solution was again loaded on a Hi-Trap Chelating HP column to remove uncleaved protein molecules and the His_6 tags. Size-exclusion chromatography (HiLoad Superdex 200, 16/60, Amersham Pharmacia Biotech) was used as the final step of the purification procedure. The column was equilibrated with buffer *C* (30 m*M* Tris–HCl pH 8.0, 50 m*M* NaCl, 5% glycerol, 1 m*M* DTT) and loaded with no more than 5 ml sample solution. The protein eluted with an apparent molecular weight of about 290 kDa, which corresponds to 3.5 times the molecular weight of Icd-2 (83 181 Da). The peak fractions were analyzed by SDS–PAGE, pooled and concentrated to 20 mg ml^{-1}, based on a specific absorbance coefficient $A_{280,1\,cm}(1\%) = 7.75$. Buffer *C* was also used as a storage buffer. 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 confirmed by dynamic light scattering (DLS; DynaPro 99 from Protein Solutions), indicating a molecular weight of 260 kDa.

2.3. Crystallization

Initial crystallization trials were performed in 96-well trays as sitting drops at 292 K in the EMBL Hamburg high-throughput crystallization facility (Mueller-Dieckmann, 2006). Over 1000 initial screening conditions were tested at sample concentrations of 20 and 40 mg ml^{-1} by equilibrating 600 nl droplets (300 nl sample plus 300 nl reservoir) against 50 µl reservoir solution. Several initial hits were obtained in conditions containing medium- to high-molecular-weight PEGs (PEG 3350 or PEG 8000), Mg^{2+} or Ca^{2+} and buffered at pH 5.5–8.5. All initial crystals were small and thin needles with poor or no diffraction. Extensive optimization using different molecular-weight PEG solutions led to crystals of a different plate-like morphology, which were grown in 3–7%(*w*/*v*) PEG 20 000, 0.1 *M* MES pH 6.0 and 0.05–0.2 *M* calcium acetate. These crystals diffracted X-ray radiation to no more than 6 Å resolution. Further optimization of the salt and the pH of the buffer led to a final crystallization condition which contained 3–5%(*w*/*v*) PEG 20 000, 0.1 *M* MES pH 5.7 and 0.2 *M* $CaCl_2$. Those conditions resulted in two different crystal forms which appeared in the drop at different time points during the equilibration process. Initially, after 3–4 d, small plate-like crystals appeared in the drop. After 7–10 d, larger crystals of similar morphology emerged. Finally, after 14–16 d, rod-shaped crystals with frayed ends decorated with multiple smaller rods materialized (Fig. 1). These crystals were fragile and diffracted to about 3 Å resolution.

2.4. Collection and processing of diffraction data

After careful removal of the additional crystals grown at the ends of larger rods, single crystals were transferred to reservoir solution supplemented with 30% ethylene glycol as a cryoprotectant for 10 s. Crystals were then flash-cooled to 100 K in liquid nitrogen. Diffraction data were collected on beamline ID23-1 (ESRF Grenoble, France) using a Quantum CCD detector. The data were indexed and integrated using *DENZO* and scaled using *SCALEPACK* (Otwinowski & Minor, 1997). The merging *R* factors $R_{\rm p.i.m.}$, indicating the precision of the data set, and $R_{\rm r.i.m.}$, which is redundancy-independent, were calculated using *RMERGE* (Weiss, 2001; available at http://www.embl-hamburg.de/~msweiss/projects/msw_qual.html). 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).

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell (3.37–3.25 Å).

No. of crystals	1
Beamline	ID23, ESRF Grenoble
Wavelength (Å)	0.8726
Temperature (K)	100
Crystal-to-detector distance (mm)	373.6
Rotation range per image (°)	0.95
Total rotation range (°)	98.8
Space group	*C*2
Unit-cell parameters (Å, °)	a = 139.83, b = 182.46, c = 290.03, β = 103.32
Mosaicity (°)	0.80
Resolution limits (Å)	99–3.25 (3.37–3.25)
Total No. of reflections	203846
Unique reflections	105225
Redundancy	1.9
$I/\sigma(I)$	5.6 (1.5)
Completeness (%)	94.6 (94.2)
$R_{\rm merge}$ (%)	12.8 (53.5)
$R_{\rm r.i.m.}$ (%)	16.9 (70.7)
$R_{\rm p.i.m.}$ (%)	10.9 (45.7)
Overall *B* factor from Wilson plot (Å^2)	68.5
Optical resolution (Å)	2.28

3. Results and discussion

Expression of Rv0066c (Icd-2) from *Mtb* in BL21(DE3)RP cells resulted in close to 100% soluble protein. Recombinant protein was purified in a three-step procedure applying affinity and size-exclusion chromatography (SEC) to give a final yield of 120 mg of pure protein per litre of culture medium. The protein elutes from the Superdex 200 size-exclusion chromatography column with an apparent molecular weight of 290 kDa, corresponding to 3.5 times the molecular weight of the monomer. In order to further investigate the oligomeric state of Icd-2, the purified sample that was used for crystallization experiments was subjected to analytical size-exclusion chromatography (Superose 6 10/300 GL, Amersham Pharmacia Biotech). The sample eluted as a single peak with an apparent molecular weight of 250 kDa, which is consistent with a trimeric state of Icd-2. This result was confirmed by SLS data recorded during SEC, which indicated a molecular weight of 248 kDa.

The sample was 99% pure as estimated by SDS–PAGE. High-throughput screening of initial conditions resulted in several leads. Initial optimization experiments yielded thin plates which diffracted X-ray radiation to about 6 Å resolution. After extensive optimization of the crystallization conditions, crystals suitable for diffraction experiments were obtained in 3–5%(*w*/*v*) PEG 20 000, 0.1 *M* MES pH 5.7 and 0.2 *M* $CaCl_2$. Two clearly distinguishable crystal forms grew in the same droplet, but only one of them diffracted X-ray radiation to about 3 Å resolution.

The diffraction data of Rv0066c could be indexed in space groups *C*2, with unit-cell parameters a = 139.8, b = 182.5, c = 290.0 Å, β = 103.3°, and *F*222, with unit-cell parameters a = 142.5, b = 184.2, c = 568.3 Å. A doubling of c^* in the monoclinic setting links the different lattice types with respect to the required systematic extinctions. Converting the monoclinic into an orthogonal coordinate system is possible because $\arctan(c^*_{C2}/2a^*_{F222}) \simeq 14^\circ$. However, data reduction was only possible in the monoclinic setting. This was further corroborated by an analysis of unmerged data with *POINTLESS* (Evans, 2006), which precludes crystallographic symmetry elements with the exception of a twofold axis along [010].

The self-rotation function of Icd-2 (Fig. 2) indicates *D*6 symmetry with one dyad along the crystallographic *b* axis. However, sixfold symmetry is incompatible with the observed trimeric state in solution. The adoption of a higher symmetry arrangement on crystallization is conceivable but unlikely. Alternatively, there may be two trimers per

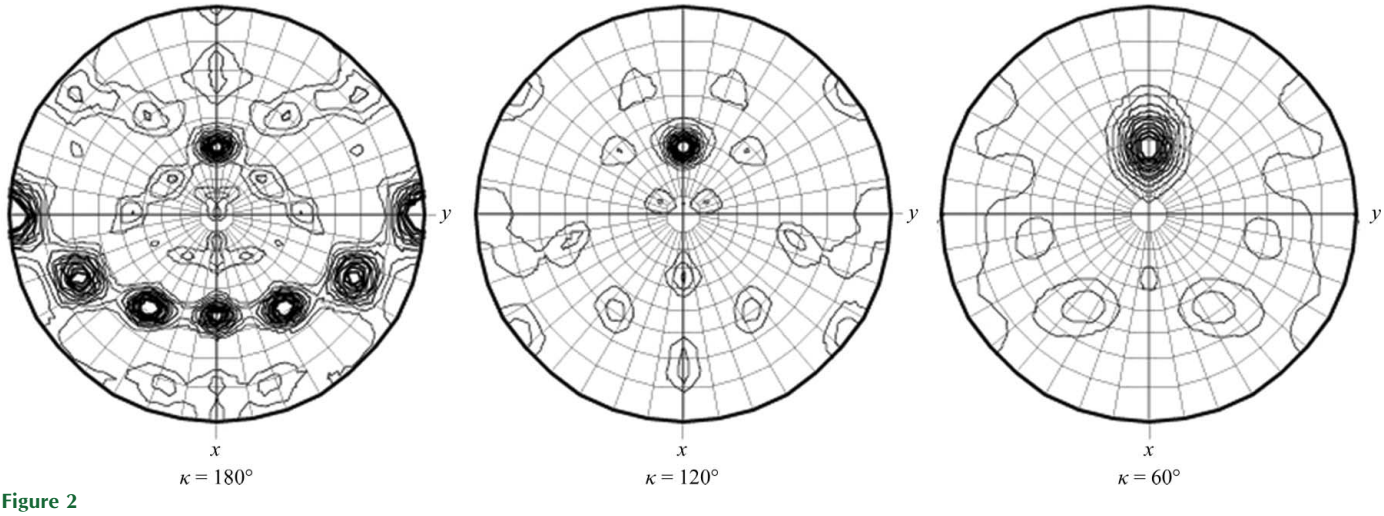


Figure 2
Self-rotation function for sections corresponding to (*a*) dyads, (*b*) threefold and (*c*) sixfold rotational symmetry based on data from Rv0066c.

asymmetric unit which are related by a twofold parallel to the observed noncrystallographic threefold (Fig. 2*b*). A definitive answer to the actual symmetry arrangement will have to await structure solution.

Both arrangements suggest the presence of six molecules of Rv0066c per asymmetric unit. This would correspond to a solvent content of 66%, based on a molecular weight of 83 181 Da and a unit-cell volume of 7 195 409 Å^3 (Matthews, 1968). While a high solvent content correlates with the observed weak diffraction of the crystals, the presence of additional molecules cannot be excluded. No peaks were found in a native Patterson map, excluding crystallographic symmetry.

The structures of prokaryotic Icd-2 from *A. vinelandii* (Yasutake *et al.*, 2002) and *C. glutamicum* (Imabayashi *et al.*, 2006) show 60% and 56% sequence identity, respectively, to Rv0066c. Therefore, structure solution will be attempted by molecular replacement.

References

Banerjee, S., Nandyala, A., Podili, R., Katoch, V. M. & Hasnain, S. E. (2005). *BMC Biochem.* **6**, 20.

Banerjee, S., Nandyala, A., Podili, R., Katoch, V. M., Murthy, K. J. R. & Hasnain, S. E. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 12652–12657.

Chen, R. D. & Gadal, P. (1990). *Plant Physiol. Biochem.* **28**, 411–427.

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.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

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

Helmuth, L. (2000). *Science*, **289**, 1123–1125.

Horwitz, A. M., Lee, B.-H. E., Dillon, B. J. & Harth, G. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 1530–1534.

Hurley, J. H., Thornsness, P. E., Ramalingam, V., Helmers, N. H., Koshland, D. E. & Stroud, R. M. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 8635–8639.

Imabayashi, F., Aich, S., Prasad, L. & Delbaere, L. T. J. (2006). *Proteins*, **63**, 100–112.

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.

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 (2007). *2007 Tuberculosis Facts.* Geneva, Switzerland: World Health Organization.

Yasutake, Y., Watanabe, S., Yao, M., Takada, Y., Fukunaga, N. & Tanaka, I. (2002). *Structure*, **10**, 1637–1648.