Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Cloning, expression, purification and crystallization of a transcriptional regulatory protein (Rv3291c) from *Mycobacterium tuberculosis* H37Rv

Tripti Shrivastava, Sandeep Kumar and Ravishankar Ramachandran*

Molecular and Structural Biology Division, Central Drug Research Institute, PO Box 173, Chattar Manzil Palace, Mahatma Gandhi Marg, Lucknow 226001 (U.P.), India

Correspondence e-mail: ravishankarr@indiatimes.com

Received 30 April 2004
Accepted 5 July 2004

Rv3291c, the translational product of the *Mycobacterium tuberculosis Rv3291c* gene, is an 18 kDa protein. It is a putative transcriptional regulatory protein belonging to the leucine-responsive regulatory protein/asparagine synthase C (Lrp/AsnC) family, which are proteins that have been identified in archaea and bacteria. Rv3291c probably plays a significant role during the persistent/latent phase of *M. tuberculosis*, as supported by its up-regulation several-fold during this stage. Orthorhombic crystals of recombinant Rv3291c have been grown from trisodium citrate dihydrate-buffered solutions containing monoammonium dihydrogen phosphate. Diffraction data extending to 2.7 Å have been collected from a single crystal with unit-cell parameters $a = 99.6$, $b = 100.7$, $c = 100.6$ Å. Assuming an octamer in the asymmetric unit results in a Matthews coefficient (V_M) of 1.75 Å^3 Da^{-1}, corresponding to a solvent content of about 30%.

1. Introduction

Mycobacterium tuberculosis exhibits significant changes in gene expression during the latent/persistent stage of infection. Proteomic analysis of the nutrient-starved/latent-phase bacteria suggested a slowing down of the transcriptional apparatus, energy metabolism, lipid biosynthesis and cell division in addition to induction of stringent response and several other genes that may play a role in long-term survival within the host (Betts *et al.*, 2002; Stewart *et al.*, 2003). *Rv3291c* is homologous to transcriptional regulators of the Lrp/AsnC family. It lies within the genomic region upstream of *sigF* involved in mycobacterial stress response (Wu *et al.*, 1997) and is up-regulated 15-fold during persistence compared with active infection (Betts *et al.*, 2002).

Lrp/AsnC-family regulatory proteins are present in bacteria and archaea (Brinkman *et al.*, 2003; Kyrfrides & Ouzounis, 1995) and regulate amino-acid metabolism and related processes. Lrp and AsnC-type proteins have been found to be evolutionarily related (Friedberg *et al.*, 2001; Ouhammouch *et al.*, 2003) (Fig. 1*a*). They are also known as feast/famine regulatory proteins (FFRPs) to summarize the general function of Lrp (Calvo & Matthews, 1994; Koike *et al.*, 2004). Many

```
E.coli Lrp      MVDSKKRPGKDLDRIDRNILNELQKDGRISNVELSKRVGLSPTPCLERVRRLE-RQGFIQ 59
Rv3291c         -------MNEALDDIDRILVRELAADGRATLSELATRAGLSVSAVQSRVRRLE-SRGVVQ 52
E.coli AsnC     ------MENYLIDNLDRGILEALMGNARTAYAELAKQFGVSPGTIHVRVEKMKKQAGIIT 54
Pf LrpA         ----------MIDERDKIILEILEKDARTPFTEIAKKLGISETAVRKRVKALE-EKGIIE 49

E.coli Lrp      GYTALLN--PHYLDASLLVFVEITLNRGAPDVFEQFNTAVQKLEEIQECHLVSGDFDYLL 117
Rv3291c         GYSARIN--PEAVGHLLSAFVAITPLD--PSQPDDAPARLEHIEEVESCYSVAGEESYVL 108
E.coli AsnC     G--ARIDVSPKQLGYDVGCFIGIILKS--AKDYPSALAKLESLDEVTEAYYTTGHYSIFI 110
Pf LrpA         GYTIKININPKKLGYSLVTITGVDTKP--EKLF-EVAEKLKEYDFVKELYLSSGDH-MIM 105

E.coli Lrp      KTRVPDMSAYRKLLGETLLRLPGVNDTRTYVVMEEVKQSNRLVIKKTR 165
Rv3291c         LVRVASARALEDLLQR--IRTTANVRTRSTIILNTFYSDRQHIP---- 150
E.coli AsnC     KVMCRSIDALQHVLINKKIQTIDEIQSTETLIVLQNPIMRTIKP---- 154
Pf LrpA         AVWAKDGEDLAEIISN-KIGKIEGVTKVCPAIILEKLK---------- 142
```

(*a*)

```
Upstream Rv3291c    GGCGACCGCAGCGAGTTCGGCCGGGTTGCCCAC 60
E.coli Lrp          --------CAGCAAATTTTGCTC---------- 15
```

(*b*)

Figure 1
(*a*) Multiple sequence alignment of Rv3291c with *E. coli* Lrp and AsnC and *P. furiosus* Lrp. Sequences were aligned using *CLUSTAL_W* (Thompson *et al.*, 1994). The percentage sequence identities between Rv3291c and *E. coli* Lrp and AsnC are 30.3 and 25.3%, respectively. Conserved residues are marked in bold, while those conserved between Rv3291c and *E. coli* Lrp are marked in grey. (*b*) Sequence alignment of the binding site of *E. coli* Lrp with the region upstream of *Rv3291c*. Identical bases are marked in bold.

Lrp/AsnC proteins can also repress their own expression (Brinkman *et al.*, 2003; Friedberg *et al.*, 2001; Ouhammouch *et al.*, 2003). They exhibit a range of oligomeric states including dimers, tetramers, octamers and hexadecamers (Brinkman *et al.*, 2003; Leonard *et al.*, 2001). AsnC regulators are involved in asparagine biosynthesis and particularly in the regulation of asparagine synthetase A in *Escherichia coli*. Asparagine is involved in the regulation of the *asnA* gene, while the autoregulation of *asnC* is itself asparagine-independent. Lrp proteins are implicated in globally regulating various operons and leucine biosynthesis (Willins *et al.*, 1991). The DNA-binding helix–turn–helix (HTH) domain present in transcriptional regulators of the AsnC/Lrp family usually occurs in the N-terminal part and consists of about 60 amino acids. The C-terminal part can contain an effector-binding domain and/or an oligomerization domain (Brinkman *et al.*, 2003). In Lrp, the middle part mediates transcriptional activation (Brinkman *et al.*, 2003; Platko & Calvo, 1993; Wang *et al.*, 1994). Lrp is a global regulator of *E. coli* and is involved in amino-acid metabolism and pili synthesis by affecting transcription of at least 10% of its genes. Most Lrp homologues appear to be specific regulators of amino-acid metabolism (Brinkman *et al.*, 2003). Various amino acids act as specific effectors and can either activate or repress transcription of metabolic enzymes. A striking feature of the *lrp* regulon is the variety of ways that leucine and Lrp interact in order to regulate gene expression (Newman *et al.*, 1992). In some cases activation requires leucine, while in others the activation is negated by leucine; sometimes the activation is independent of leucine (Wang *et al.*, 1994). Analogously, in operons that are negatively regulated by Lrp, the same three categories have been observed: leucine negates the effect, leucine is required for the effect and leucine has no effect. The molecular mechanisms underlying these six different patterns of regulation involving Lrp and leucine are only partially understood (Wang *et al.*, 1994).

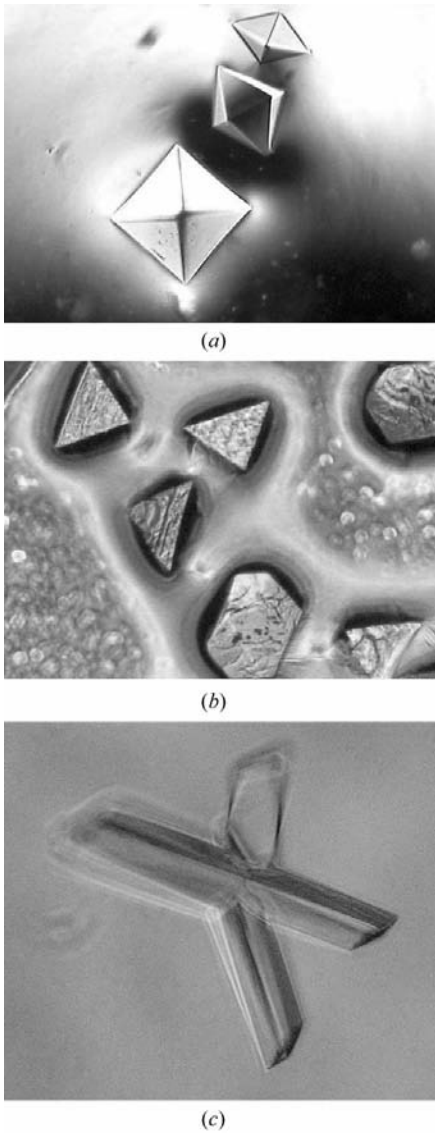

Figure 2
Crystals of Rv3291c grown in (*a*) 1 *M* monoammonium dihydrogen phosphate, 0.1 *M* trisodium citrate dihydrate pH 5.35, (*b*) 0.05 *M* monopotassium dihydrogen phosphate, 18%(*w*/*v*) PEG 8000 and (*c*) 1.3 *M* lithium sulfate monohydrate, 0.1 *M* HEPES pH 7.5.

Crystal structures of the protein are available only from two archaeal sources (Leonard *et al.*, 2001; Koike *et al.*, 2004). No structure of the protein is available from eubacterial sources nor are structures available of complexes with DNA. Owing to its implied importance in the persistent stage of *M. tuberculosis* infection, we have purified and crystallized the protein in order to understand the molecular mechanisms underlying its regulatory activity.

2. Protein expression, purification and mass spectrometry

Rv3291c is a 150-residue polypeptide. Rv3291c with a C-terminal His_6 tag was expressed from a lactose-inducible promoter in *E. coli* strain BL21 (DE3). The gene was amplified from *M. tuberculosis* genomic DNA using the forward primer 5′-GAGAATT**CCATGG**ACGAGGCGCTCGAC-3′ and the reverse primer 5′-CAAT**AAGCTT**TGGTATATGCTGCCTATCGCTG-3′ containing *Nco*I and *Hin*dIII sites (in bold), respectively. The PCR product was digested with *Nco*I and *Hin*dIII and ligated to pET21d (Novagen), which was also digested with the same enzymes. The resulting construct was transformed into BL21 (DE3).

The transformed cells were grown at 310 K in LB medium supplemented with ampicillin to an A_{600} of 0.6. Expression of recombinant Rv3291c-His_6 was induced by addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 m*M* for 8 h. The cells were harvested by centrifugation at 15 000*g* for 15 min and resuspended in ice-cold 50 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl (buffer *A*) and 10 m*M* imidazole with 12% glycerol. The cells were frozen, thawed and lyzed by sonication at 20% output power, 45% pulsar duty cycle for a pulse time of 8 min giving a 15 s pulse using a Sonicator (Ultrasonic processor XL, Heat Systems Inc). The lysate was centrifuged at 27 000*g* for 30 min and the supernatant was loaded onto an Ni-chelating column equilibrated in buffer *A* containing 10 m*M* imidazole. A 0.1–1 *M* imidazole gradient in buffer *A* was applied to the column. Protein eluted at around 400 m*M* imidazole. Purified fractions were precipitated using ammonium sulfate and the solution was centrifuged at 27 000*g* for 15 min using a Sorvall Super T-21 (Kendro). The pellet was dissolved in minimum volume of buffer *B* (50 m*M* Tris–HCl pH 8.0 100 m*M* NaCl and 2 m*M* EDTA) and loaded onto a Superdex-200 HR 10/30 (Amersham) gel-filtration column pre-equilibrated with buffer *B*. The protein yield was around 10 mg from 500 ml of culture. Rv3291c was concentrated to 10–12 mg ml^{-1} in 50 m*M* Tris–HCl pH 8.0, 50 m*M* NaCl and 2 m*M* EDTA using a 10 kDa cutoff centricon (Amicon). Protein concentrations were determined by the Bradford method (Bradford, 1976) using bovine serum albumin as a standard. The protein remained stable at 277 K without degradation for several weeks. The purity of the protein was confirmed using SDS–PAGE.

MALDI–TOF spectra were generated using a Shimadzu QP-2000 (GC/MS) Micromass TofSpec 2E MALDI TOF MS instrument. Our results show that the protein has a weight of 17.984 kDa, which agrees within experimental error with the theoretically calculated value of 18.0 kDa.

3. Crystallization

Crystallization experiments were set up at 295 K using the hanging-drop vapour-diffusion method exploiting two different screening kits (Mazeed *et al.*, 2003; Jancarik & Kim, 1991; Cudney *et al.*, 1994). Crystals were obtained in different screening solutions. Large single crystals (1.4 × 0.7 ×

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell (2.8–2.7 Å).

Wavelength (Å)	1.5418
Resolution range (Å)	12.0–2.7
No. measured reflections	130282
No. unique reflections	28037 (2756)
Multiplicity	4.65 (4.63)
$I/\sigma(I)$	19.2 (2.6)
Completeness (%)	99.7 (99)
$R_{\rm merge}$ (%)	9.2 (44.6)
$R_{\rm r.i.m.}$† (%)	10.4 (50.3)

† $R_{\rm r.i.m.} = \sum_{hkl}[N/(N-1)]^{1/2}\sum_i I_i(hkl) - I(hkl)/\sum_{hkl}\sum_i I_i(hkl)$.

0.4 mm) of Rv3291c were grown over 2–3 d from a hanging drop consisting of 2 µl of 9–10 mg ml^{-1} protein and 1 µl of 0.1 *M* trisodium citrate dihydrate solution in the pH range 5.3–5.7 containing monoammonium dihydrogen phosphate. The size and diffraction quality of these crystals were improved by varying the concentration of the precipitant. Increasing the protein concentration led to clustering of crystals as well as reduction in size. Crystals were also obtained in conditions containing 0.1 *M* HEPES in the pH range 7.0–8.0 along with $LiSO_4$ and also in conditions containing 0.05–0.1 *M* monopotassium dihydrogen phosphate along with different concentrations of PEG 8000 (Fig. 2). The latter crystals did not diffract well.

4. Data collection and analysis

A single crystal obtained from the trisodium citrate dihydrate and monoammonium dihydrogen phosphate conditions was mounted in a capillary (Fig. 2*a*). Data extending to 2.7 Å were collected at room temperature using a MAR 345 area-detector system and a Rigaku RU300 rotating-anode generator operating at 44 kV and 74 mA. The crystal remained stable at room temperature throughout the data collection. The data were processed using the *DENZO*/*SCALEPACK* suite of programs (Otwinowski & Minor, 1997). Crystals are orthorhombic, with unit-cell parameters a = 99.6, b = 100.7, c = 100.6 Å and space group $P222_1$. The possibility of a higher symmetry space group was excluded by examining the calculated $R_{\rm sym}$ values between different pairs of symmetry-related structure-factor amplitudes. The data-collection statistics are summarized in Table 1. The crystal mosaicity refined to around 0.3 and an overall data completeness of 99.7% was obtained.

Size-exclusion chromatography experiments indicated that the protein exists as an octamer. Calculation of the Matthews coefficient (Matthews, 1968) suggests that the asymmetric unit can contain between four and eight subunits. Assuming that the asymmetric unit contains an octamer, the calculated Matthews coefficient is 1.75 Å^3 Da^{-1}, corresponding to a solvent content of about 30%. A tetramer in the asymmetric unit, on the other hand, corresponds to a calculated solvent content of about 65%. Efforts are being made to crystallize the protein–DNA complex and also to improve the crystals obtained under other conditions.

5. Discussion

In our search for a suitable model for use in molecular-replacement calculations, we found two structures, both of which belong to the archaeal Lrp-type proteins (Leonard *et al.*, 2001; Koike *et al.*, 2004), to be available. No structure of an AsnC-type protein is available to date. As mentioned earlier, Lrp- and AsnC-type proteins are evolutionarily related (Friedberg *et al.*, 2001; Ouhammouch *et al.*, 2003). In order to distinguish and identify the gene in question as Lrp or AsnC, we performed alignments and searches involving both the nucleotide and the protein sequences.

The sequence of Rv3291c is consistent with the known physico-chemical properties of other Lrp proteins, specifically *E. coli* Lrp. This comparison includes the size of the Lrp polypeptide, which ran as a 20–21.5 kDa protein in SDS–PAGE experiments compared with 17.9 kDa predicted from the nucleic acid sequences. This is analogous to the *E. coli* Lrp protein, which behaves like a 21.5 kDa protein compared with its actual weight of 18.8 kDa. Sequence alignment shows that Rv3291c is more homologous to Lrp proteins than to AsnC-type proteins (Fig. 1*a*). As expected, the main differences in the Lrp and AsnC sequences occur in the C-terminal region, which has been implicated in leucine-mediated regulation (Wang *et al.*, 1994).

We also examined the region upstream of *Rv3291c* in order to identify (if present) characteristic direct repeat sequences and binding sites of AsnC- and Lrp-type proteins. Three direct repeats of ten base pairs each in the upstream region have been reported to be the binding sites for AsnC-type regulators (Platko & Calvo, 1993). We were unable to identify such repeat sequences upstream of *Rv3291c*. We then examined the region upstream of *Rv3291c* in order to identify consensus sequences known to be binding sites of Lrp-like proteins (Cui *et al.*, 1995; Brinkman *et al.*, 2003). We found a sequence highly homologous to the binding sites of Lrp-type proteins (Fig. 1*b*). The above analysis suggests that Rv3291c belongs to the Lrp family. Attempts are therefore being made to solve the structure by molecular replacement using the archaeal Lrp structures as models. We are also searching for heavy-atom derivatives.

This work is supported by a grant from DBT. TS and SK are recipients of JRF and SRF from CSIR, respectively. This is manuscript number 6568 from CDRI.

References

Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. & Duncan, K. (2002). *Mol. Microbiol.* **43**, 717–731.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Brinkman, A. B., Ettema, T. J. G., de Vos, W. M. & van der Oost, J. (2003). *J. Mol. Microbiol.* **48**, 287–294.
Calvo, J. M. & Matthews, R. G. (1994). *Microbiol Rev.* **58**, 466–490.
Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* D**50**, 414–423.
Cui, Y., Wang, Q., Stormo, G. D. & Calvo, J. M. (1995). *J. Bacteriol.* **177**, 4872–4880.
Friedberg, D., Midkiff, M. & Calvo, J. M. (2001). *J. Bacteriol.* **183**, 4004–4011.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Koike, H., Ishijima, S. A., Clowney, L. & Suzuki, M. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 2840–2845.
Kyrfrides, N. C. & Ouzounis, C. A. (1995). *Trends Biochem. Sci.* **20**, 140–141.
Leonard, P. M., Smits, S. H.., Sedelnikova, S. E., Brinkman, A. B., De Vos, W. M., van der Oost, J., Rice, D. W. & Rafferty, J. B. (2001). *EMBO J.* **20**, 990–997.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mazeed, S., Ofek, G., Belachew, A., Huang, C. C., Zhou, T. & Kwong. P. D. (2003). *Structure*, **11**, 1061–1070.
Newman, E. B., D'Ari, R. & Lin, R. T. (1992). *Cell*, **68**, 617–619.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ouhammouch, M., Dewhourt, R. E., Hausner, W., Thomm, M. & Seiduschele, E. P. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 5097–5102.
Platko, J. V. & Calvo, J. M. (1993). *J. Bacteriol.* **175**, 1110–1117.
Stewart, G. R., Robertson, B. D. & Young, D. B. (2003). *Nature Rev. Microbiol.* **2**, 97–105.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Wang, Q., Jingcai, W. U., Friedberg, D., Platko, J. V. & Calvo, J. M. (1994). *J. Bacteriol.* **176**, 1831–1839.
Willins, D. A., Ryan, C. W., Platko, J. V. & Calvo, J. M. (1991). *J. Biol. Chem.* **266**, 10768–10774.
Wu, Q. L., Kong, D. Q., Lam, K. & Husson, R. N. (1997). *J. Bacteriol.* **179**, 2922–2929.