

Harriet A. Watkins* and
Edward N. Baker

Maurice Wilkins Centre for Molecular
Biodiscovery and School of Biological Sciences,
3A Symonds Street, Private Bag 92019,
Auckland, New Zealand

Correspondence e-mail:
h.watkins@auckland.ac.nz

Received 12 June 2008

Accepted 8 July 2008

Cloning, expression, purification and preliminary crystallographic analysis of the RNase HI domain of the *Mycobacterium tuberculosis* protein Rv2228c as a maltose-binding protein fusion

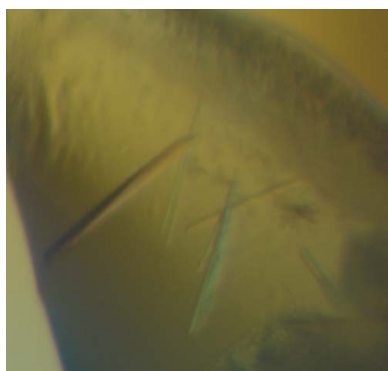
The predicted ribonuclease (RNase) HI domain of the open reading frame Rv2228c from *Mycobacterium tuberculosis* has been cloned as a hexahistidine fusion and a maltose-binding protein (MBP) fusion. Expression was only observed for the MBP-fusion protein, which was purified using amylose affinity chromatography and gel filtration. The RNase HI domain could be cleaved from the MBP-fusion protein by factor Xa digestion, but was very unstable. In contrast, the fusion protein was stable, could be obtained in high yield and gave crystals which diffracted to 2.25 Å resolution. The crystals belong to space group $P2_1$ and have unit-cell parameters $a = 73.63$, $b = 101.38$, $c = 76.09$ Å, $\beta = 109.0^\circ$. Two fusion-protein molecules of 57 417 Da were present in each asymmetric unit.

1. Introduction

The bacterium *Mycobacterium tuberculosis* is the causative agent of the disease tuberculosis, which kills 2–3 million people worldwide every year. One third of the world's population have latent infection and 10% of these will develop the active form of the disease. The evolution of multidrug-resistant strains and the increase in HIV-related immunocompromisation has led to a serious re-emergence of the disease. The sequencing and annotation of the *M. tuberculosis* genome (Cole *et al.*, 1998) has enabled proteins with suitable activities to be selected as possible novel drug targets.

Rv2228c is annotated as a bifunctional two-domain protein comprising an N-terminal domain with RNase HI activity and a C-terminal domain with α -ribazole phosphatase (CobC) activity (<http://www.doe-mbi.ucla.edu/TB/>). The RNases HI are endonucleases which digest the RNA strand of an RNA–DNA duplex. Digestion results in the production of 3'-hydroxyl and 5'-phosphate termini and is dependent on divalent metal cations such as Mg^{2+} and Mn^{2+} (Wu *et al.*, 2001). The human RNase H1 enzyme has positional but not sequence specificity for its substrate conferred by its RNA-binding domain (Wu *et al.*, 2001; Lindahl & Lindahl, 1984) and cleaves 7–12 nucleotides away from the 3'-DNA/5'-RNA terminus (Lima *et al.*, 2004); however, the *Escherichia coli* form cleaves only four nucleotides away from the terminus (Crooke *et al.*, 1995). The human form of RNase H1 has been shown to be involved in the removal of Okasaki fragments during the maturation of mitochondrial DNA and thus the development of multicellular organisms (Cerritelli *et al.*, 2003; Dasgupta *et al.*, 1987). RNase HI enzymes have also been shown to have a role in bacterial replication, including the processing of RNA primers for the initiation of ColE1-type plasmid DNA replication (Dasgupta *et al.*, 1987; Itoh & Tomizawa, 1980) and the suppression of replication other than at the OriC origin of replication (Horiuchi *et al.*, 1984; Kogoma *et al.*, 1985; Ogawa *et al.*, 1984).

The three-dimensional structures of a number of type 1 RNases H have now been determined, including those from *E. coli* (Yang *et al.*, 1990; Katayanagi *et al.*, 1990), human (Nowotny *et al.*, 2007), *Bacillus halodurans* (Nowotny *et al.*, 2005), *Thermus thermophilus* (Ishikawa *et al.*, 1993), *Sulfolobus tokodaii* (You *et al.*, 2007), the reverse transcriptase domain of the HIV virus (Davies *et al.*, 1991) and that of the Moloney murine leukaemia virus (Lim *et al.*, 2006). Analysis of the *M. tuberculosis* genome has not shown any other open reading frame



corresponding to an RNase HI or its gene *rnhA*. The RNase HII gene, *rnhB*, has been identified as corresponding to open reading frame Rv2902c and no *rnhC* gene corresponding to an RNase HIII is present. It can therefore be inferred that the Rv2228c N-terminal domain is the RNase HI in *M. tuberculosis*.

The fusion of protein tags to a gene of interest is used in many areas such as immunodetection, the study of protein structure, protein folding and protein–protein interactions (Smyth *et al.*, 2003; Beckwith, 2000). Such fusions have been found to be particularly useful in structural biology, where large amounts of protein are required for crystallization screening and approximately 50% of recombinant proteins expressed in *E. coli* are found to be insoluble or in the form of inclusion bodies. This is a major bottleneck in structural genomics pipelines and often results in the recovery of inactive or insoluble proteins (Christendat *et al.*, 2000). These fusions are termed affinity tags because of their usefulness in affinity chromatography.

The most commonly used tags are the hexahistidine tag (Bornhorst & Falke, 2000), glutathione *S*-transferase (GST) from *Schistosoma japonicum* (Smith, 2000), thioredoxin (LaVallie *et al.*, 2000) and maltose-binding protein (MBP; Sachdev & Chirgwin, 2000). These are generally used to purify the tagged protein from crude cell lysate. The larger affinity tags can also be used to aid in the soluble expression of the recombinant protein of interest in *E. coli* (Smith, 2000; Sachdev & Chirgwin, 2000). As an affinity tag, MBP has been shown to interact with unfolded proteins and to have chaperone-like properties (Richarme & Caldas, 1997). Its fusion to many small proteins has in some cases led to the solution of their structures as fusion proteins (Center *et al.*, 1998; Liu *et al.*, 2001; Rivas *et al.*, 2005).

Here, we report the successful use of an MBP-fusion construct in obtaining a stable form of the RNase HI domain of Rv2228c as a first step towards discovery of the functional significance of the two seemingly functionally unrelated domains of this protein and of routes to their possible use as drug targets. The high-quality crystals obtained of this fusion protein will facilitate determination of its structure by X-ray crystallography.

2. Materials and methods

2.1. Cloning and expression

The gene product corresponding to the RNase HI domain (residues 1–140) of Rv2228c was amplified from genomic DNA from the H37Rv strain of *M. tuberculosis* using the polymerase chain reaction. It was cloned using the Gateway system (Invitrogen) into the His₆-tagged vector pDEST17 with gene-specific primers for the sense strand of bases 1–15 and antisense strand of bases 385–390. The amplified gene was also cloned into pMAL-c2 (New England Biolabs) containing an MBP tag followed by a factor Xa cleavage site by restriction enzyme-based cloning. The sense primer 5-GTGAA-AGTTGTCATCGAA and antisense primer 5-AATCGGCTGGC-GGCGGATTAGAAGCTTCGGG corresponding to bases 1–18 and 404–420 of the gene sequence at the *Xmn*I and *Hind*III sites, respectively, were used for insertion into the vector. The resultant MBP fusion was expressed in BL21 (DE3) pRIL cells in Luria–Bertani (LB) medium at 310 K until the OD_{600nm} reached 0.6. The expression culture was then induced to a final isopropyl β -D-thiogalactopyranoside (IPTG) concentration of 0.3 mM, transferred to 291 K and incubated overnight with shaking. Expression trials of the His₆-tagged clone were carried out using autoinduction (Studier, 2005) and IPTG-based induction at both 310 and 291 K.

2.2. Purification

The cells from the expression cultures were harvested by centrifugation at 4600g for 15 min and resuspended in buffer A (50 mM Tris–HCl pH 7.5, 200 mM NaCl). The resuspended cells were lysed using a cell disruptor (Constant Cell Disruption Systems) at 124 MPa and incubated for 30 min at 277 K with DNase I. The soluble proteins were then separated by further centrifugation at 30 000g for 45 min. The solubly expressed MBP fusion was purified using amylose affinity chromatography (New England Biolabs) and eluted in 50 mM Tris–HCl pH 7.5 and 200 mM NaCl containing 10 mM maltose. The fusion protein was further purified by Superdex 200 size-exclusion chromatography in 50 mM Tris–HCl pH 7.5 and 200 mM NaCl. Removal of the MBP tag for generation of the RNase HI domain alone was carried out by digestion with factor Xa (Pierce) overnight at 277 K with gentle rocking. The cleaved RNase HI domain was purified by gel filtration as previously described with the addition of 0.01% dodecyl maltoside to the buffer. The dodecyl maltoside was subsequently removed by the use of BioBeads (BioRad; Rigaud *et al.*, 1998). The pure fusion protein was concentrated to 75 mg ml^{−1} and the RNase HI domain alone was concentrated to 2 mg ml^{−1}.

2.3. Crystallization

Screening for crystallization conditions was carried out using Honeybee (Cartesian Dispensing Systems) nanolitre robot technology in a sitting-drop format with 100 nl protein solution mixed in a 1:1 ratio with reservoir solution in Intelliplates (Art Robbins Instruments). Initial screens included Crystal Screens 1 and 2 (Hampton Research), a systematic PEG/pH screen (Kingston *et al.*, 1994), the Footprint Screen (Stura *et al.*, 1992) and the PEG/Ion Screen (Hampton Research). A well solution volume of 75 μ l was used and trays were incubated at 291 K. Crystals were grown in 0.2 M diammonium tartrate with 20% PEG 2000 and were mounted directly from the sitting drop and cryoprotected with 20% ethylene glycol before being flash-cooled in liquid nitrogen.

3. Results and discussion

Soluble overexpression with large yields was obtained for the MBP–RNase HI domain fusion protein after induction at 291 K (Fig. 1). This soluble protein was purified from crude cell lysate using amylose affinity chromatography and was further purified by gel filtration.

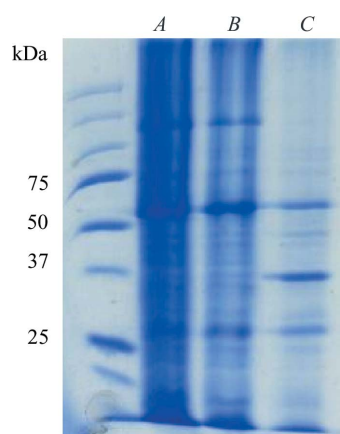


Figure 1
SDS–PAGE gel showing total cell extract (lane A), soluble fraction (lane B) and insoluble fraction (lane C) from IPTG-based induction of the RNase HI–MBP fusion protein in pMAL–C2.

Table 1

Crystal data and data-collection statistics.

Values in parentheses are for the outer shell.

Space group	$P2_1$
Unit-cell parameters (\AA , $^\circ$)	$a = 73.63$, $b = 101.38$, $c = 76.09$, $\alpha = 90$, $\beta = 109.0$, $\gamma = 90$
Wavelength (\AA)	0.983970
Resolution range (\AA)	41.45–2.25 (2.37–2.25)
Completeness (%)	99.99 (100)
Observed reflections	187273 (27345)
Unique reflections	50220 (7326)
$\langle I/\sigma(I) \rangle$	4.4 (1.3)
Multiplicity	3.7 (3.7)
R_{merge}^\dagger	0.121 (0.462)

$$\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

Pure fusion protein was concentrated to 75 mg ml^{-1} and placed into crystallization trials. Expression testing of the His₆-tagged RNase HI domain showed no overexpression in either LB or autoinduction media or with a change in temperature from 310 to 291 K. In order to generate the RNase HI domain alone, the pure fusion protein was incubated with factor Xa and after incubation at 277 K the separated RNase HI domain protein was purified by a further gel-filtration step. In this case separation was only achieved in the presence of 0.1% dodecyl maltoside, which was subsequently removed by incubation with BioBeads. Cleavage of the maltose-binding protein moiety proceeded to 100% digestion. Equal quantities of MBP and RNase HI domain were not obtained, however, as the RNase HI domain portion was very unstable and subject to degradation before purification (Fig. 2). The small amount of domain obtained was concentrated to saturation at 2 mg ml^{-1} and placed in crystallization trials.

Crystals of the 58 kDa fusion protein were obtained in 0.2 M diammonium tartrate with 20% PEG 2000 and were soaked directly from the Intelliplate wells in cryoprotectant and then flash-cooled in liquid nitrogen prior to mounting on the beam (Fig. 3). These crystals took approximately six weeks to form and reached dimensions of $0.4 \times 0.1 \times 0.05 \text{ }\mu\text{m}$. No crystals were obtained of the RNase HI domain alone.

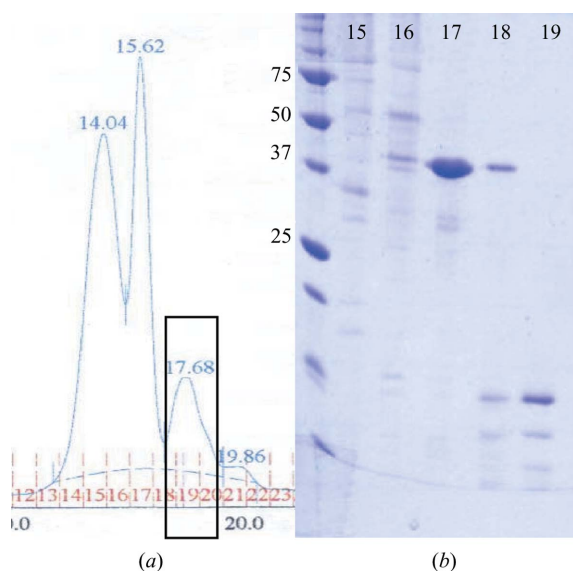


Figure 2

Separation of the isolated RNase HI domain. The boxed region in (a) shows the peak corresponding to fractions 18 and 19. These show the separation and degradation of the RNase HI domain from the cleaved maltose-binding protein (fraction 17).

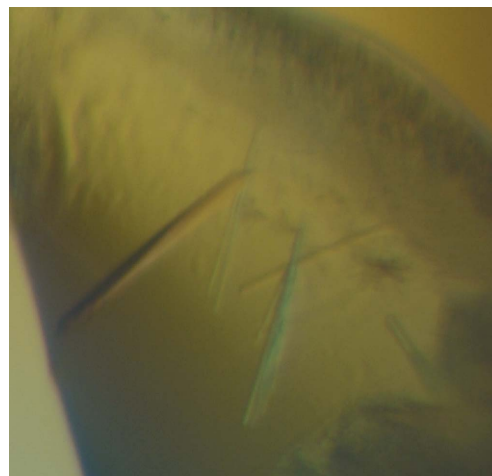


Figure 3

Crystals of the RNase HI domain-MBP fusion protein formed in 20% PEG 2000 and 0.2 M diammonium tartrate.

Data collection was carried out at 113 K at a wavelength of 0.98397 \AA on beamline 9-2 at the Stanford Synchrotron Radiation Laboratory, California, USA. Data to 2.25 \AA resolution were recorded and were processed using *MOSFLM* (Leslie, 1999) and *SCALA* (Collaborative Computational Project, Number 4, 1994), giving the statistics shown in Table 1. The crystals were monoclinic, space group $P2_1$, with unit-cell parameters $a = 73.63$, $b = 101.38$, $c = 76.09 \text{ \AA}$, $\beta = 109.0^\circ$. Assuming the presence of two fusion protein molecules of 58 kDa in each asymmetric unit, the Matthews coefficient V_M is $2.31 \text{ \AA}^3 \text{ Da}^{-1}$ and the solvent content is 47.8%. The crystals are of sufficient quality to enable us to undertake the structure solution of the MBP-RNase HI domain fusion protein by molecular replacement.

The lack of expression of the RNase HI domain as a His₆-tagged protein can be explained by its obvious instability in isolation, as illustrated by its rapid degradation upon cleavage from the MBP moiety. The failure of small proteins to crystallize other than as MBP-fusion proteins has previously been illustrated by the human T-cell leukaemia virus type 1 envelope protein gp21 (Center *et al.*, 1998), the *Staphylococcus* accessory regulator R from *S. aureus* (Liu *et al.*, 2001) and the MATA1 homeodomain protein from *Saccharomyces cerevisiae* (Ke & Wolberger, 2003). However, the crystallization of these proteins also required the optimization of the MBP-protein linker length before crystals of adequate quality were obtained. In the case of the *M. tuberculosis* Rv2228c RNase HI domain, adequate stability for crystallization and detectable expression is only conferred upon its fusion to the MBP tag. This may also imply a role for the functionally seemingly unrelated C-terminal CobC domain and the position of Rv2228c in the putative CobC operon. The high quality of the crystals obtained will enable us to determine the structure of the RNase HI domain of this protein in order to confirm its annotation and direct further studies into the full-length protein and the functional significance of the fusion of two domains predicted to have seemingly unrelated functions.

The authors gratefully acknowledge research support from the Health Research Council of New Zealand, the kind gift of the MBP-RNase HI expression plasmid from Michelle Baker and Valerie Mizrahi, and Stephanie Dawes for many helpful discussions (MRC/NHLS/WITS Molecular Mycobacteriology Research Unit and DST/

NRF Centre of Excellence for Biomedical Tuberculosis Research, School of Pathology, University of the Witwatersrand and the National Health Laboratory Service, Johannesburg, South Africa).

References

- Beckwith, J. (2000). *Methods Enzymol.* **326**, 3–7.
- Bornhorst, J. A. & Falke, J. J. (2000). *Methods Enzymol.* **326**, 245–254.
- Center, R. J., Kobe, B., Wilson, K. A., Teh, T., Howlett, G. J., Kemp, B. E. & Pountourios, P. (1998). *Protein Sci.* **7**, 1612–1619.
- Cerritelli, S. M., Frolova, E. G., Feng, C., Grinberg, A., Love, P. E. & Crouch, R. J. (2003). *Mol. Cell.* **11**, 807–815.
- Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Gerstein, M., Arrowsmith, C. H. & Edwards, A. M. (2000). *Prog. Biophys. Mol. Biol.* **73**, 339–345.
- Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Crooke, S. T., Lemonidis, K. M., Neilson, L., Griffey, R., Lesnik, E. A. & Monia, B. P. (1995). *Biochem. J.* **312**, 599–608.
- Dasgupta, S., Masukata, H. & Tomizawa, J. (1987). *Cell.* **51**, 1113–1122.
- Davies, J. F. II, Hostomska, Z., Hostomsky, Z., Jordan, S. R. & Matthews, D. A. (1991). *Science*, **252**, 88–95.
- Horiuchi, T., Maki, H. & Sekiguchi, M. (1984). *Mol. Gen. Genet.* **195**, 17–22.
- Ishikawa, K., Okumura, M., Katayanagi, K., Kimura, S., Kanaya, S., Nakamura, H. & Morikawa, K. (1993). *J. Mol. Biol.* **230**, 529–542.
- Itoh, T. & Tomizawa, J. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 2450–2454.
- Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T. & Morikawa, K. (1990). *Nature (London)*, **347**, 306–309.
- Ke, A. & Wolberger, C. (2003). *Protein Sci.* **12**, 306–312.
- Kingston, R. L., Baker, H. M. & Baker, E. N. (1994). *Acta Cryst.* **D50**, 429–440.
- Kogoma, T., Subia, N. L. & von Meyenburg, K. (1985). *Mol. Gen. Genet.* **200**, 103–109.
- LaVallie, E. R., Lu, Z., Diblasio-Smith, E. A., Collins-Racie, L. A. & McCoy, J. M. (2000). *Methods Enzymol.* **326**, 322–340.
- Leslie, A. G. W. (1999). *Acta Cryst.* **D55**, 1696–1702.
- Lim, D., Gregorio, G. G., Bingman, C., Martinez-Hackert, E., Hendrickson, W. A. & Goff, S. P. (2006). *J. Virol.* **80**, 8379–8389.
- Lima, W. F., Nichols, J. G., Wu, H., Prakash, T. P., Migawa, M. T., Wyrzykiewicz, T. K., Bhat, B. & Crooke, S. T. (2004). *J. Biol. Chem.* **279**, 36317–36326.
- Lindahl, G. & Lindahl, T. (1984). *Mol. Gen. Genet.* **196**, 283–289.
- Liu, Y., Manna, A., Li, R., Martin, W. E., Murphy, R. C., Cheung, A. L. & Zhang, G. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 6877–6882.
- Nowotny, M., Gaidamakov, S. A., Crouch, R. J. & Yang, W. (2005). *Cell*, **121**, 1005–1016.
- Nowotny, M., Gaidamakov, S. A., Ghirlando, R., Cerritelli, S. M., Crouch, R. J. & Yang, W. (2007). *Mol. Cell*, **28**, 264–276.
- Ogawa, T., Pickett, G. G., Kogoma, T. & Kornberg, A. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 1040–1044.
- Richarme, G. & Caldas, T. D. (1997). *J. Biol. Chem.* **272**, 15607–15612.
- Rigaud, J.-L., Levy, D., Mosser, G. & Lambert, O. (1998). *Eur. Biophys. J.* **27**, 305–319.
- Rivas, F. V., Tolia, N. H., Song, J. J., Aragon, J. P., Liu, J., Hannon, G. J. & Joshua-Tor, L. (2005). *Nature Struct. Mol. Biol.* **12**, 340–349.
- Sachdev, D. & Chirgwin, J. M. (2000). *Methods Enzymol.* **326**, 312–321.
- Smith, D. B. (2000). *Methods Enzymol.* **326**, 254–270.
- Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. & Kobe, B. (2003). *Protein Sci.* **12**, 1313–1322.
- Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
- Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992). *J. Cryst. Growth*, **122**, 273–285.
- Wu, H., Lima, W. F. & Crooke, S. T. (2001). *J. Biol. Chem.* **276**, 23547–23553.
- Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. (1990). *Science*, **249**, 1398–1405.
- You, D. J., Chon, H., Koga, Y., Takano, K. & Kanaya, S. (2007). *Biochemistry*, **46**, 11494–11503.