Crystallization and preliminary X-ray analysis of RabX3, a tandem GTPase from *Entamoeba histolytica*

Vijay Kumar Srivastava,‡ Mintu Chandra‡ and Sunando Datta*

Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal 462 023, India

‡ These authors contributed equally to this paper.

Correspondence e-mail: sunando@iiserb.ac.in

Received 1 March 2014
Accepted 16 May 2014

Ras superfamily GTPases regulate signalling pathways that control multiple biological processes by modulating the GTP/GDP cycle. Various Rab GTPases, which are the key regulators of vesicular trafficking pathways, play a vital role in the survival and virulence of the enteric parasite *Entamoeba histolytica*. The Rab GTPases act as binary molecular switches that utilize the conformational changes associated with the GTP/GDP cycle to elicit responses from target proteins and thereby regulate a broad spectrum of cellular processes including cell proliferation, cytoskeletal assembly, nuclear transport and intracellular membrane trafficking in eukaryotes. *Entamoeba histolytica* RabX3 (*EhRabX3*) is a unique GTPase in the amoebic genome, the only member in the eukaryotic Ras superfamily that harbours tandem G-domains and shares only 8–16% sequence identity with other GTPases. Recent studies suggested that *EhRabX3* binds to a single guanine nucleotide through its N-terminal G-domain (NTD), while the C-terminal G-domain (CTD) plays a potential role in binding of the nucleotide to the NTD. Thus, understanding the intermolecular regulation between the two GTPase domains is expected to reveal valuable information on the overall action of *EhRabX3*. To provide structural insights into the inclusive action of this unique GTPase, *EhRabX3* was crystallized by successive micro-seeding using the vapour-diffusion method. A complete data set was collected to 3.3 Å resolution using a single native *EhRabX3* crystal at 100 K on BM14 at the ESRF, Grenoble, France. The crystal belonged to monoclinic space group *C2*, with unit-cell parameters $a = 198.6$, $b = 119.3$, $c = 89.2$ Å, $\beta = 103.1^\circ$. Preliminary analysis of the data using the *Matthews Probability Calculator* suggested the presence of four to six molecules in the asymmetric unit.

1. Introduction

The GTP hydrolysis cycle provides a switch mechanism in controlling various cellular processes such as cell proliferation, cytoskeleton organization and intracellular trafficking through Ras-like small GTPases. Members of the Ras superfamily of regulatory proteins, which includes the Ras, Rho, Rab, Arf and Ran subfamilies, are able to bind to and hydrolyze GTP, thereby providing their switching capability. They alternate between GTP-bound and GDP-bound conformations in the cell. The GTP-bound conformation is biologically active and turns on cellular function through interaction with and activation of a specific effector(s). GTP hydrolysis turns off the GTPase switch by converting it to the inactive GDP-bound conformation that has low affinity for the effector(s). The ability to bind and hydrolyze GTP is determined by primary-sequence motifs that have been highly conserved evolutionarily and whose importance has been confirmed through mutational analysis (Dhar *et al.*, 1982; Valencia *et al.*, 1991; Lowry & Williamsen, 1993). For Ras proteins, the most extensively studied GTPases thus far, the identification in tumours of activating mutations that impair GTPase activity revealed a few highly conserved amino acids that are critical for normal hydrolytic function (Bos, 1989; McGrath *et al.*, 1984; Sweet *et al.*, 1984). Conservation of these amino acids is not limited to the Ras family proteins but extends to the large variety of guanine nucleotide-binding proteins (G proteins) with diverse biological functions. Such observations have led to the identification of a highly conserved catalytic site that is responsible for GTP hydrolysis.

© 2014 International Union of Crystallography
All rights reserved

The enteric pathogenic parasite *Entamoeba histolytica* is one of the causative agents of health hazards in tropical countries. It causes amoebic dysentery and liver abscesses in humans. About 50 million cases of amoebiasis and 40 000–100 000 deaths are reported annually (World Health Organization, 2012). The frequency of occurrence of this disease is higher in developing countries where sanitation is poor. Adequate sanitation and good personal hygiene may avert this infection. Various Rab GTPases, which are the key regulators of vesicular trafficking pathways, play a vital role in the survival and virulence of this enteric parasite. Over 100 Rab GTPases have been predicted to exist in the amoebic genome, three quarters of which are unique to the genus. This unique group has been designated as RabX. Some of the RabX proteins lack critical sequence motifs and hence their biochemical properties might deviate from those of their classical homologues (Kumagai *et al.*, 2004; Saito-Nakano *et al.*, 2001, 2005). *Entamoeba histolytica* RabX3 (*EhRabX3*; GenBank Accession No. AB197075) is one such unique GTPase which was identified

in an *in silico* screening for small GTPases in the amoebic genome. The presence of tandem GTPase domains in *EhRabX3* intrigued us and prompted us to investigate the GTP-binding and hydrolytic properties of this novel GTPase. The other proteins that possess double GTPase domains belong to EngA, a distinctive member of the Era subfamily of GTPases, and its homologue the Der family (Robinson *et al.*, 2002). These latter groups of proteins bind to two nucleotides, with one in each domain. Recent studies have shown that, unlike its prokaryotic counterpart which binds to GTP with 1:2 stoichiometry (Lamb *et al.*, 2007), *EhRabX3* binds to only one guanine nucleotide molecule through its NTD (N-terminal G-domain). The CTD (C-terminal G-domain) plays an indispensable role in the binding of the nucleotide to the NTD, indicating a disparate protein–nucleotide interaction. In addition, it has been reported that this unique GTPase has extremely low intrinsic GTPase activity and that it could be enhanced by up to 20-fold upon the introduction of site-specific mutations (Chandra *et al.*, 2014). Here, we report the crystallization and preliminary X-ray diffraction analysis of full-length wild-type *EhRabX3* (349 amino acids). Structural insights into *EhRabX3* will provide valuable information to aid understanding of the intermolecular regulation between its two GTPase domains as well as the molecular mechanisms underlying the low catalytic activity of *EhRabX3*.

2. Materials and methods

2.1. Cloning, expression and purification of full-length *EhRabX3*

Full-length *EhRabX3* was amplified from cDNA, extracted from the reference strain *E. histolytica* HM1:IMSS cl6. A single fragment of 1050 base pairs was amplified from the PCR reaction using

Figure 1

Cloning, expression and purification of full-length *EhRabX3*. (a) PCR amplification of full-length *EhRabX3* cDNA. Specific oligonucleotides designed for the amplification of full-length *EhRabX3* cDNA were used to amplify the 1050 bp single band as indicated by the arrowhead. (b) Affinity purification profile of 6×His-tagged *EhRabX3*. The whole cell lysate and eluted fractions (100–250 mM imidazole concentration) from the affinity column were analyzed using SDS–PAGE. (c) Size-exclusion chromatogram of *EhRabX3*. Fractions from affinity-purified *EhRabX3* were pooled and subjected to a Superdex 75 prep-grade gel-filtration column (16/60). *EhRabX3* was eluted at a retention volume of 57.2 ml, which corresponds to a molecular mass of ~40 kDa. (d) Fractions (24–32) corresponding to the chromatogram peak were analyzed using SDS–PAGE.

Table 1

Recombinant protein-production information.

In the primers, the sequence given in upper-case letters denotes the insert-specific sequence and the lower-case letters denote the restriction sites. In the protein sequence, the *EhRabX3* sequence is indicated in bold and the 6×His tag is in italics.

Source organism	<i>Entamoeba histolytica</i>
Strain (DNA source)	HM1:IMSS cl6
Accession code	AB197075
Forward primer (5'–3')	ggatccATGATTGACCAGCACAATAA
Reverse primer (5'–3')	ctcgagTTACTCATATGATTGTTCAATCTTTAAACG
Cloning vector	pET-28a (+)
Expression vector	BL21(DE3)
Expression host	<i>Escherichia coli</i>
Complete amino-acid sequence of the construct produced	MGSSHHHHSSGLVPRGSHMASMTGGQMGRGSMAMIRPAH-MIRPAHSLRTRSQIIGKKEIRLLVVGSGVGKTTLCDCF-FESHQSQGEETREKHVQIDNDFIRISISDIVGKQSFYACD-NPYDGYDAILVMYDITELKSFTDLKTMWLPDIFLYCNIDT-QIIIGNKKDQIEDRIITRKEAEQFAQDRLCQFYESTKD-DSCQLLFDICISRDFLQCDIKIRMLMVGDNVQKTTFIRKF-ALQDPTGHDPMNAITTRFEMEIKIYEIIMDWGFYNKLLQ-TNPAISRTIEAILIVYDITNEESFQNIHRKYYPLINNKFS-DVAGVIVGYKTDLEAQRKITMGDALTLADWLGYKYVEMSS-KDTEHSSIIKALAHSTRINRLKIEQSYE

EhRabX3-specific primers (Fig. 1*a*). To express recombinant full-length *EhRabX3* (~40 kDa), the protein-coding region was subcloned into bacterial cloning vector pET-28a (Novagen) within *Bam*HI and *Xho*I restriction sites and expressed as a 6×His fusion protein. The recombinant protein-production information is provided in Table 1.

The reconstructed vector was transformed into *Escherichia coli* BL21 (DE3) competent cells. The cells were grown in 2 l Luria broth medium supplemented with 30 µg ml⁻¹ kanamycin at 310 K until the OD₆₀₀ reached 0.6. Protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside to a final concentration of 100 µM. After overnight culture, the cells were harvested by centrifugation (5000g, 10 min) at 277 K and the bacterial pellet was dissolved in lysis buffer consisting of 50 mM Tris pH 8.0, 300 mM NaCl, 5 mM β-mercaptoethanol, 1% Triton X-100, 5 mM MgCl₂, 10 µg ml⁻¹ protease-inhibitor cocktail (leupeptin hydrochloride, pepstatin A and aprotinin; Sigma–Aldrich, India), 200 µM phenylmethanesulfonyl fluoride (PMSF; Sigma–Aldrich, India), 100 µM GDP (Sigma–Aldrich, India).

After sonication, cell debris was removed by centrifugation at 30 000g for 45 min at 277 K and the supernatant was loaded onto a column containing Ni-NTA metal-affinity resin (GE Healthcare). The column was thoroughly washed with lysis buffer containing 100–500 mM salt. Finally, the protein of interest was eluted using a linear gradient of imidazole (100–250 mM) in 50 mM Tris pH 8.0, 300 mM NaCl (Fig. 1*b*) at 277 K. The pooled fractions were further purified using a Superdex 75 prep-grade (16/60) gel-filtration column (GE Healthcare) equilibrated with a buffer consisting of 50 mM Tris pH 8.0, 50 mM NaCl, 5 mM MgCl₂, 5 mM dithiothreitol (DTT), 7% glycerol (Figs. 1*c* and 1*d*) at 277 K. The yield of pure protein was 5 mg per litre of *E. coli* culture. The purified protein was concentrated to 10 mg ml⁻¹ using a Centricon Ultra-10 kDa centrifugal filter (Millipore, USA) for crystallization. The concentration of protein was determined by the Bradford assay (Bradford, 1976) after each purification step.

2.2. Crystallization

Crystallization experiments were carried out using the hanging-drop vapour-diffusion method at 293 K. Initial crystallization screening was performed using commercial screens (Crystal Screen, Crystal Screen 2 and Index from Hampton Research, and Structure

Screen 1, Structure Screen 2 and ProPLEX Screen from Molecular Dimensions). 2 µl protein solution was mixed with 1 µl reservoir solution and equilibrated against 500 µl reservoir solution. The initial crystallization conditions were set up using different protein

(a)

(b)

(c)

Figure 2

Snapshot of *EhRabX3* crystals. (a) Micro-crystals obtained after optimizing the condition (ProPLEX Screen 2, Molecular Dimensions) containing 1.4 M sodium malonate pH 7.0. (b) Crystals obtained after micro-seeding. (c) Final diffraction-quality crystals obtained after repeated micro-seeding with a protein concentration of 6.5 mg ml⁻¹ (5 µl protein:1 µl reservoir) using 500 µl reservoir solution (0.7 M sodium malonate pH 6.0).

Table 2
Crystallization details of *EhRabX3*.

Method	Hanging-drop vapour diffusion
Plate type	SPL 24-well plates
Temperature (K)	293
Protein concentration (mg ml ⁻¹)	6.5
Buffer composition of protein solution	50 mM Tris pH 8.0, 50 mM NaCl, 5 mM MgCl ₂ , 5 mM dithiothreitol (DTT), 7% glycerol
Composition of reservoir solution	0.7 M sodium malonate pH 6.0
Volume and ratio of drop	5 µl protein:1 µl reservoir
Volume of reservoir (µl)	500

concentrations (5.5, 6, 6.5 and 7.0 mg ml⁻¹) and the hits were optimized by grid screening. After 21 d, bundles of micro-crystals were observed in a drop equilibrated against 1.4 M sodium malonate pH 7.0 (ProPlex Screen 2, Molecular Dimensions; Fig. 2*a*). Micro-crystals from this droplet were crushed and used to micro-seed further screening experiments using varying concentrations of sodium malonate (0.2–2.0 M) within the pH range 6.0–7.5. Optimization around these conditions and micro-seeding with the improved crystals led to the production of optimized single crystals (Figs. 2*b* and 2*c*). Crystallization details are given in Table 2.

2.3. X-ray diffraction data collection and processing

Optimized crystals were briefly soaked in cryoprotectant comprising 2 M sodium malonate pH 6.0, scooped up in a CryoLoop (Hampton Research) and then flash-cooled in a stream of cold N₂ (100 K). Initially, data were collected at the home source to 6 Å resolution using a MAR345 detector mounted on a Bruker RU300 Cu Kα rotating-anode X-ray generator. Subsequently, high-resolution diffraction data to 3.3 Å resolution were collected from a single crystal with dimensions of around 150 µm on the BM14 synchrotron

Table 3
Crystal parameters and data-collection statistics.

Values in parentheses are for the last shell.	
Beamline	BM14, ESRF
Wavelength (Å)	0.95
Detector	MAR225 CCD
Temperature (K)	100
Space group	C2
Unit-cell parameters (Å, °)	$a = 198.6$, $b = 119.3$, $c = 89.2$, $\beta = 103.1$
Matthews coefficient (Å ³ Da ⁻¹)	2.24–3.36
Solvent content (%)	45–63
Resolution range (Å)	50.6–3.3 (3.43–3.30)
Total No. of reflections	257452 (37347)
No. of unique reflections	34790 (5116)
Multiplicity	7.4 (7.3)
$\langle I/\sigma(I) \rangle$	12.4 (2.4)
Completeness (%)	99.0 (98.7)
$R_{\text{merge}}^{\dagger}$	0.043 (0.79)

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i th observation of reflection hkl and $\langle I(hkl) \rangle$ is the average intensity of the i observations.

X-ray beamline at the European Synchrotron Radiation Facility, Grenoble, France, using a MAR225 CCD detector. The intensity data were indexed, integrated and scaled using *DENZO* and *SCALEPACK* in the *HKL-2000* suite (Otwinowski & Minor, 1997). The crystal parameters, data-collection parameters and processing statistics are summarized in Table 3.

3. Results and discussion

In the initial crystallization screening, 400 commercial screening conditions were tested. Precipitate formation and crystal growth were observed in a few of these conditions over three weeks, and these conditions were further optimized to obtain microcrystals. These micro-crystals were further micro-seeded repeatedly with a protein

Figure 3
Diffraction pattern of *EhRabX3*. X-ray diffraction image from an *EhRabX3* crystal. The circles indicate resolutions of 50.0, 20.0, 10.0, 6.0 and 3.3 Å. The image corresponds to 1° oscillation with 7 s exposure time and a crystal-to-detector distance of 336 mm at a wavelength of 0.95 Å. The diffraction images were collected on a MAR225 CCD detector. The resolution at the edge of the diffraction pattern is 3.25 Å.

Figure 4
Self-rotation function calculated using *GLRF* (Tong & Rossmann, 1990) using the resolution range of 10–3.5 Å with radius of integration 25 Å. Shown here is the self-rotation function at $\kappa = 180^\circ$. The plot shows two peaks 90° apart, both normal to c^* , suggesting 222 point-group symmetry with one twofold noncrystallographic symmetry axis parallel to the space-group symmetry axis.

Figure 5

Sequence alignment of *EhRabX3* with human Rab4a (186 amino acids; PDB entry 2bmd). Conserved residues are highlighted in red boxes. The sequence alignment was produced using *ClustalW* (Thompson *et al.*, 1994) and the figure was generated using *ESPrpt* (Gouet *et al.*, 1999).

concentration of 6.5 mg ml^{-1} ($5 \mu\text{l}$ protein solution: $1 \mu\text{l}$ reservoir) using $500 \mu\text{l}$ reservoir solution (0.7 M sodium malonate pH 6.0), which yielded diffraction-quality crystals within 7–10 d. A complete native data set was collected and processed to $3.3 \text{ \AA}$ resolution (Fig. 3) using a single native *EhRabX3* crystal on BM14. The crystal belonged to space group *C2*, with unit-cell parameters $a = 198.6$, $b = 119.3$, $c = 89.2 \text{ \AA}$, $\beta = 103.1^\circ$. Preliminary analysis of the calculated Matthews coefficient (Matthews, 1968) suggested the presence of six molecules in the asymmetric unit ($V_M = 2.24 \text{ \AA}^3 \text{ Da}^{-1}$, corresponding to a solvent content of 45%); however, the existence of four ($V_M 3.36 \text{ \AA}^3 \text{ Da}^{-1}$, corresponding to a solvent content of 63%) or five ($V_M 2.69 \text{ \AA}^3 \text{ Da}^{-1}$, corresponding to a solvent content of 54%) molecules in the asymmetric unit, cannot be ruled out.

To find noncrystallographic symmetry within the asymmetric unit we used *GLRF* (Tong & Rossmann, 1990) to calculate the self-rotation function. The rotation function was calculated at $\kappa = 60, 90, 120$ and 180° . Significant peaks were observed only at $\kappa = 180^\circ$. The plot clearly indicates two peaks that are 90° apart and both normal to c^* (Fig. 4), which suggests the presence of an oligomer in the asymmetric unit with 222 point-group symmetry, assuming the presence of a twofold noncrystallographic symmetry axis parallel to the space-group symmetry axis. Four molecules per asymmetric unit are also consistent with the relatively high solvent content, resulting in limited diffraction by the *EhRabX3* crystals.

Structure determination via the molecular-replacement method employing the *CCP4* suite (*Phaser* and *MOLREP*) (Winn *et al.*, 2011) using human Rab4a (186 amino acids; PDB entry 2bmd; Huber & Scheidig, 2005) as search model did not provide an unambiguous solution, probably due to the low sequence identity (16%) between these two proteins (Fig. 5).

Currently, a selenomethionine derivative of *EhRabX3* is being prepared for experimental phasing. The determination of the three-dimensional structure of *EhRabX3* will be useful to aid understanding of the molecular basis of intermolecular interaction between the two GTPase domains in this novel protein. The structure would also potentially improve the understanding of nucleotide exchange and GTP hydrolysis by *EhRabX3* and its implications in vesicular trafficking in *E. histolytica*.

We thank Dr Tomoyoshi Nozaki and Dr Yumiko Saito-Nakano for providing the *EhRabX3* construct. We also thank the DBT-Well-

come Trust for access to beamline BM14 at the ESRF, Grenoble, France. This work was supported by the Japan Society for the Promotion of Science (JSPS) and the Japan-India Science Cooperative Program (JSPS-DST) under the Joint Research Projects (13039221-000423), Grants-in-Aid for Scientific Research (24590514) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and the Department of Science and Technology (SERB/F/1648/2012-13), India.

References

- Bos, J. L. (1989). *Cancer Res.* **49**, 4682–4689.
- Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
- Chandra, M., Mukherjee, M., Srivastava, V. K., Saito-Nakano, Y., Nozaki, T. & Datta, S. (2014). *Biochemistry*, **7**, 1191–1205.
- Dhar, R., Ellis, R. W., Shih, T. Y., Oroszlan, S., Shapiro, B., Maizel, J., Lowy, D. & Scolnick, E. (1982). *Science*, **217**, 934–936.
- Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
- Huber, S. K. & Scheidig, A. J. (2005). *FEBS Lett.* **579**, 2821–2829.
- Kumagai, M., Makioka, A., Takeuchi, T. & Nozaki, T. (2004). *J. Biol. Chem.* **279**, 2316–2323.
- Lamb, H. K., Thompson, P., Elliott, C., Charles, I. G., Richards, J., Lockyer, M., Watkins, N., Nichols, C., Stammers, D. K., Bagshaw, C. R., Cooper, A. & Hawkins, A. R. (2007). *Protein Sci.* **16**, 2391–2402.
- Lowry, D. R. & Williamsen, B. M. (1993). *Annu. Rev. Biochem.* **62**, 4637–4648.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McGrath, J. P., Capon, D. J., Goeddel, D. V. & Levinson, A. D. (1984). *Nature (London)*, **310**, 644–649.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Robinson, V. L., Hwang, J., Fox, E., Inouye, M. & Stock, A. M. (2002). *Structure*, **10**, 1649–1658.
- Saito-Nakano, Y., Loftus, B. J., Hall, N. & Nozaki, T. (2005). *Exp. Parasitol.* **110**, 244–252.
- Saito-Nakano, Y., Nakazawa, M., Shigeta, Y., Takeuchi, T. & Nozaki, T. (2001). *Mol. Biochem. Parasitol.* **116**, 219–222.
- Sweet, R. W., Yokoyama, S., Kamata, T., Feramisco, J. R., Rosenberg, M. & Gross, M. (1984). *Nature (London)*, **311**, 273–275.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
- Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* **A46**, 783–792.
- Valencia, A., Chardin, P., Wittinghofer, A. & Sander, C. (1991). *Biochemistry*, **30**, 4637–4648.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- World Health Organization (2012). *Vaccine*, **30**, 4717–4718.