Venuka Durani Goyal,^{a*} Pooja Yadav,^b Ashwani Kumar,^b Biplab Ghosh^b and Ravindra D. Makde^{b*}

^aChemistry Department, IIT Bombay, Powai, Mumbai 400 076, India, and ^bHigh Pressure and Synchrotron Radiation Physics Division, Bhabha Atomic Research Centre, Trombay, Mumbai 400 085, India

Correspondence e-mail:
venuka@chem.iitb.ac.in,
ravimakde@rrcat.gov.in

Received 23 July 2014

Accepted 16 September 2014

Crystallization and preliminary X-ray crystallographic analysis of an artificial molten-globular-like triosephosphate isomerase protein of mixed phylogenetic origin

A bioinformatics-based protein-engineering approach called consensus design led to the construction of a chimeric triosephosphate isomerase (TIM) protein called ccTIM (curated consensus TIM) which is as active as *Saccharomyces cerevisiae* TIM despite sharing only 58% sequence identity with it. The amino-acid sequence of this novel protein is as identical to native sequences from eukaryotes as to those from prokaryotes and shares some biophysical traits with a molten globular protein. Solving its crystal structure would help in understanding the physical implications of its bioinformatics-based sequence. In this report, the ccTIM protein was successfully crystallized using the microbatch-under-oil method and a full X-ray diffraction data set was collected to 2.2 Å resolution using a synchrotron-radiation source. The crystals belonged to space group C222₁, with unit-cell parameters $a = 107.97$, $b = 187.21$, $c = 288.22$ Å. Matthews coefficient calculations indicated the presence of six dimers in the asymmetric unit, with an approximate solvent content of 46.2%.

1. Introduction

Engineering proteins for higher thermal stability is important for facilitating their applications in industry and medicine. Because of the resource-intensive nature of library-based high-throughput methods, it is desirable to employ rational and semi-rational approaches for this purpose. With readily available sequence information, bioinformatics tools (Lockless & Ranganathan, 1999; Süel *et al.*, 2003; Magliery & Regan, 2004; Socolich *et al.*, 2005; Russ *et al.*, 2005) for elucidating various aspects of protein structure, function and activity are gaining popularity. It has been reported previously that making a consensus mutation, *i.e.* mutating a residue in a protein sequence to that observed most frequently at that position in the multiple sequence alignment of proteins of its family, could stabilize a protein (Steipe *et al.*, 1994). This concept of consensus has been used to design peptide motifs such as tetratricopeptide repeats (Main *et al.*, 2003) and ankyrin repeats (Binz *et al.*, 2003; Mosavi *et al.*, 2002) and even large proteins such as fungal phytases (Lehmann *et al.*, 2000).

The significance of the database and the methods used for consensus design have been investigated by Magliery and coworkers (Sullivan *et al.*, 2011, 2012) for the glycolytic enzyme triosephosphate isomerase (TIM). One of the variants referred to as ccTIM (curated consensus TIM) (Sullivan *et al.*, 2011), which had a consensus mutation at every position (consensus protein), was reported to be almost as active as *Saccharomyces cerevisiae* TIM despite sharing only 58% sequence identity with it. The amino-acid sequence of this protein is almost as identical to prokaryotic TIM proteins as to eukaryotic ones. For instance, it shares 61% sequence identity with the human variant (Kinoshita *et al.*, 2005), 59% with that from the thermophilic bacterium *Thermotoga maritima* (Maes *et al.*, 1999) and 58% with that from *Escherichia coli*. On the other hand, the human and *T. maritima* sequences are only 42% identical. Further, even though the closest wild-type homologue of ccTIM is from the insect *Tenebrio molitor* (64% sequence identity; Knobloch *et al.*, 2010), the database used for its design had significantly greater contribution from bacterial sequences (Sullivan *et al.*, 2011). Hence, ccTIM is neither prokaryotic nor eukaryotic.

Table 1

Macromolecule information.

Source organism	None (engineered sequence)
DNA source	Gene reassembly from synthetic oligonucleotides (Sullivan <i>et al.</i> , 2011)
Expression vector	pHLIC (Durani <i>et al.</i> , 2012)
Expression host	<i>E. coli</i> BL21(DE3) pLysS
Complete amino acid sequence of the construct	MGHHHHHGGENLYFQGGSGARTPLVAGNWKMGTLAEAKEL-VEALAAKLDDVEVVVAPPFTYLDQVRELLKGSIAVGAQ-NCYKEDSGAFTGEISPAKLDLGSYVILGHSERRQYFGE-TDELVAKKVAAALAGLTPILCVGETLEEREAGKTEEVVA-RQLKAVLAGLSDEWSNVVIAEPVVAIGTGKTATPEQAQEVHAFIRKWLAELSAEVAEKVRILYGGSVKPAEAAEAQPDIDGALVGGASLKAEDFLAIINSRN

Interestingly, it shares approximately 85% sequence identity with another consensus protein, cTIM (consensus TIM; Sullivan *et al.*, 2011), characterized as a molten globular protein with very little activity. The thermal denaturation profile of ccTIM was reported to be almost identical to that of cTIM. It was further reported that, unlike the native protein, ccTIM unfolds reversibly and retains a large fraction of its secondary structure even at 368 K. Solving the crystal structure of this novel protein would help in providing a rationale for how evolutionarily diverse traits from prokaryotic and eukaryotic sequences were combined in this sequence to lead to a

structure that functions as well as a wild-type enzyme despite having similarities to a molten globular protein. Moreover, this protein belongs to the ubiquitous and functionally diverse TIM-barrel fold (Wierenga, 2001), which is well represented in important pathways such as the glycolytic pathway, amino-acid metabolism and nucleotide metabolism. Structural study of ccTIM in light of the statistical analyses could facilitate future studies for engineering TIM-barrel proteins in general. As a first step, several crystallization conditions were screened and the protein proved to be highly crystallizable. However, most of the crystals either did not diffract or lead to a diffraction pattern that suggested extreme mosaicity. In this report, we describe a successful attempt at crystallizing the ccTIM protein and collection of a full data set to 2.2 Å resolution.

2. Materials and methods

2.1. Expression and purification

A pET-based plasmid containing the gene encoding the ccTIM protein (Sullivan *et al.*, 2011) with a cleavable 6×His tag (Table 1) was received from Professor Magliery (The Ohio State University, USA). Bacterial cells transformed with this plasmid were grown at 310 K to an OD₆₀₀ of ~0.8 and then induced for 4 h at 303 K by addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The cells were harvested by centrifugation at 8000g at ambient temperature. Typically, a pellet obtained from 2 l cell culture was suspended in 50 ml lysis buffer (20 mM Tris-HCl pH 8, 300 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol) containing 1 mM phenylmethanesulfonylfluoride (PMSF) and the cells were lysed by sonication. The desired protein (6×His-ccTIM) was purified from the soluble fraction of the cell lysate by a batch method using Ni-Sepharose-based immobilized metal-ion affinity chromatography (IMAC). The cell lysate was mixed with 2 ml Ni-Sepharose beads (GE Healthcare) and incubated for 30 min at 277 K for protein binding. Subsequently, the beads were washed with 20 ml wash buffer (20 mM Tris-HCl pH 8, 300 mM NaCl, 20 mM imidazole) and the bound protein was eluted with 10 ml elution buffer (20 mM Tris-HCl pH 8, 300 mM NaCl, 500 mM imidazole). The 6×His tag was cleaved by incubating the eluted fractions with *Tobacco etch virus* (TEV) protease (1:50 molar ratio) while dialyzing against lysis buffer at 293 K for 16 h. The cleaved ccTIM protein was separated from the 6×His tag and noncleaved 6×His-ccTIM protein using Ni-Sepharose (GE Healthcare) chromatography. This protein of molecular weight 26.5 kDa was concentrated to ~17 mg ml⁻¹ using an ultrafiltration device with a 10 kDa cutoff membrane. Glycerol was added to the protein solution to 20%(v/v); it was subsequently flash-frozen in liquid nitrogen and aliquots were stored at 203 K until further use.

2.2. Crystallization

For crystallization, the required amount of ccTIM protein was obtained from the freezer at 203 K and subjected to gel-filtration chromatography using Superdex 200 10/300 GL (GE Healthcare) with 2×M50 buffer (20 mM MES pH 6.5, 100 mM NaCl). Gel filtration was performed to remove impurities and heterogeneity and also for buffer equilibration. The eluted fractions containing ccTIM protein were diluted with an equal volume of water and 1 mM DTT was added. The protein was subsequently concentrated to 16 mg ml⁻¹ and used to set up crystallization trials employing commercially available screens (Crystal Screen, Crystal Screen 2 and Index from Hampton Research and JCSG-*plus* from Molecular Dimensions). Crystallization trials were carried out using the microbatch-under-oil method (Chayen *et al.*, 1992) in 96-well U-bottom plates at 294 K.

Figure 1

(a) Crystals of ccTIM obtained using the buffer condition 0.2 M ammonium acetate, 0.1 M Tris-HCl pH 8.5, 25% PEG 3350 at 294 K. (b) Crystals of ccTIM grown at 277 K.

Table 2

Crystallization.

Method	Microbatch-under-oil
Plate type	96-well
Temperature (K)	294
Protein concentration (mg ml ⁻¹)	14–16
Buffer composition of protein solution	10 mM MES pH 6.5, 50 mM NaCl
Volume and ratio of drop	2 µl (1:1)

Typically, 1 µl each of crystallization solution and protein solution were mixed and overlaid with 50 µl Al's oil [a 1:1(v:v) mixture of silicone and mineral oil]. Crystal hits were observed from seven different crystallization conditions in 7 d. Most of these conditions contained polyethylene glycol (PEG) as precipitant, with PEG 3350 yielding crystals at pH values ranging from 5.5 to 8.5. Acetate anion was observed in two conditions, one containing PEG 300 and the other containing PEG 3350 as precipitant. Three crystallization conditions containing PEG 3350 and one with PEG 300 were further optimized to produce crystals that were of suitable size for diffraction studies. Several crystals were tested for X-ray diffraction quality; however, all of them showed poor diffraction patterns such as low resolution, high mosaicity or multiple lattices. To improve the diffraction quality of these crystals, various crystal-growth conditions were investigated. Finally, a crystal that showed good-quality diffraction was obtained from a plate that was set up using protein solution of concentration 14 mg ml⁻¹ and incubated for ~4 months at 294 K using a buffer composition of 0.2 M ammonium acetate, 0.1 M Tris-HCl pH 8.5, 25% PEG 3350 (Fig. 1*a*; Table 2).

2.3. Data collection and processing

Crystals were cryoprotected using crystallization solution supplemented with 10–15% glycerol and flash-cooled in liquid nitrogen or directly mounted in a cryostream at 100 K. Single-crystal X-ray diffraction data (Fig. 2) were collected on the protein crystallography beamline (PX-BL21) of the Indus-2 Synchrotron, India. The beam-

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	PX-BL21, Indus-2
Wavelength (Å)	0.979470
Temperature (K)	100
Detector	MAR225 CCD
Crystal-to-detector distance (mm)	250
Rotation range per image (°)	1
Exposure time per image (s)	40
Total rotation range (°)	1–133
Space group	<i>C</i> 22 ₁
Unit-cell parameters (Å, °)	<i>a</i> = 107.97, <i>b</i> = 187.21, <i>c</i> = 288.22, <i>α</i> = <i>β</i> = <i>γ</i> = 90
Resolution range (Å)	48.04–2.20 (2.24–2.20)
Total No. of reflections	762580 (23404)
No. of unique reflections	146513 (6822)
Completeness (%)	99.4 (95.2)
Multiplicity	5.2 (3.4)
<i>I</i> / <i>σ</i> (<i>I</i>)	7.6 (2.3)
<i>R</i> _{meas} [†] (%)	17.9 (56.0)
<i>R</i> _{meas} for inner resolution shell (48.04–12.05 Å) (%)	6.8
<i>CC</i> _{1/2} [‡]	0.99 (0.75)
Overall <i>B</i> factor from Wilson plot (Å ²)	19.2
Mosaicity (°)	0.17

[†] Diederichs & Karplus (1997). [‡] Karplus & Diederichs (2012).

line is located on a bending-magnet port (1.502 T) of the 2.5 GeV synchrotron and is equipped with a MARdtb with automated sample changer, a CCD detector (Rayonix 225) and a liquid-nitrogen cryostream (Oxford Cryostream). Crystals were centred by a semi-automatic method using the in-line camera of the MARdtb and the MARdtb GUI. In this method, live crystal images were dragged to the centre of the cross-wire of the display at different 90° *φ* intervals. Data were collected at a wavelength of 0.979470 Å with 1° oscillation and 40 s exposure for each image. The crystals diffracted to 2.2 Å resolution. Data were indexed and integrated using *XDS* (Kabsch, 2010) and were subsequently scaled using *AIMLESS* from the *CCP4* suite (Winn *et al.*, 2011). Detailed data-collection statistics are shown in Table 3.

Figure 2

An X-ray diffraction image (labelled in Å) of the ccTIM crystal obtained at the Indus-2 synchrotron (RRCAT), India.

Figure 3

Sequence alignment of ccTIM with its closest natural homologue (PDB entry 2i9e). Secondary-structure elements as per the 2i9e structure are indicated below the sequences. Helical regions are depicted as rectangular boxes, sheet regions as arrows and loop regions as lines. Regions where the two sequences are different are highlighted in bold.

3. Results and discussion

Crystals of the ccTIM protein were observed from seven different crystallization conditions with different pH values, salts and precipitants. Morphologically, the crystals grown at 294 K appeared to be perfect (Fig. 1a); however, most of them either did not diffract or produced diffraction patterns that were suggestive of multiple lattices. Many of the ccTIM crystals grew as hexagonal plates or as hexagonal prisms with weaker crystal contacts in the direction of the hexagonal axes than in the hexagonal planes. As a result, the crystals developed lattice defects that were not quite visible in the morphology of the crystals grown at 294 K. However, at 277 K the growth defects in the crystals were quite prominent and were indicative of problems in stacking of the hexagonal plates. It is possible that at lower temperatures the protein molecules lacked the entropy required to re-adjust once defective contacts had been made in the crystal lattice, leading to the propagation of defects and their reflection in the morphology of the crystals (Fig. 1b). Finally, data were collected successfully from a crystal of ccTIM observed after incubation at 294 K for 4 months. The higher intrinsic order of this particular crystal could have been achieved by very slow growth. A full data set was collected to 2.2 Å resolution (Fig. 2) using a synchrotron-radiation source (Table 3). The crystals belonged to space group *C222*₁, with unit-cell parameters *a* = 107.97, *b* = 187.21, *c* = 288.22 Å. The Matthews coefficient (*V*_M; Matthews, 1968) calculated from the unit-cell parameters was 2.29 Å³ Da⁻¹ for six dimers in the asymmetric unit, corresponding to an approximate solvent content of 46.2%. The dimeric nature of this protein has been demonstrated previously by size-exclusion chromatography and analytical ultracentrifugation experiments (Sullivan *et al.*, 2011).

An *R*_{meas} (Diederichs & Karplus, 1997) of about 17.9% was observed for the overall resolution range of 48.04–2.20 Å and of 6% for the lowest resolution shell (48.04–12.05 Å). The higher *R*_{meas} values for these data are probably due to the poor diffraction quality of the crystal. The closest sequence homologue of this engineered protein in the Protein Data Bank is the TIM protein from the insect *T. molitor* (PDB entry 2i9e; Knobloch *et al.*, 2010) with a sequence identity of ~64%. A model based on the 2i9e structure was used for molecular replacement, which was carried out by *Phaser* (McCoy, 2007) and led to a solution with final log-likelihood gain (LLG) of 12 039 for six dimers in an asymmetric unit. Further structure refinement using *PHENIX* (Adams *et al.*, 2010) resulted in a model with *R*_{work} and *R*_{free} values of 25 and 27%, respectively. In general, the differences between ccTIM and 2i9e are distributed throughout the sequences (Fig. 3) and it is difficult to predict how the structure would be affected by them. Further, there are five insertions and deletions at positions 34, 61, 159, 199 and 244 of the alignment. It would be interesting to know how the structure is modified in these locations to accommodate the changes. Analysis of the ccTIM structure and comparison with various eukaryotic and prokaryotic TIM structures will unveil how phylogenetically diverse features have been combined in this protein. It would also provide a structural basis for the similarity of some of its biophysical properties to those of a molten globular protein.

The authors thank Drs S. M. Sharma (BARC), P. D. Gupta (RRCAT) and G. S. Lodha (RRCAT) and Professor Ruchi Anand (IIT Bombay) for their encouragement and support for this project. We also thank Professor T. J. Magliery from The Ohio State

University for providing the ccTIM DNA construct. We acknowledge BARC and RRCAT for providing infrastructure and IIT Bombay and the DST Fast Track Scheme (grant No. SB/FT/CS-139/2012) for financial support of VDG.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003). *J. Mol. Biol.* **332**, 489–503.
- Chayen, N. E., Shaw Stewart, P. D. & Blow, D. M. (1992). *J. Cryst. Growth*, **122**, 176–180.
- Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
- Durani, V., Sullivan, B. J. & Magliery, T. J. (2012). *Protein Expr. Purif.* **85**, 9–17.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Kinoshita, T., Maruki, R., Warizaya, M., Nakajima, H. & Nishimura, S. (2005). *Acta Cryst.* **F61**, 346–349.
- Knobeloch, D., Schmidt, A., Scheerer, P., Krauss, N., Wessner, H., Scholz, C., Küttner, G., von Rintelen, T., Wessel, A. & Höhne, W. (2010). *Insect Mol. Biol.* **19**, 35–48.
- Lehmann, M., Kostrewa, D., Wyss, M., Brugger, R., D'Arcy, A., Pasamontes, L. & van Loon, A. P. (2000). *Protein Eng.* **13**, 49–57.
- Lockless, S. W. & Ranganathan, R. (1999). *Science*, **286**, 295–299.
- Maes, D., Zeelen, J. P., Thanki, N., Beaucamp, N., Alvarez, M., Dao Thi, M. H., Backmann, J., Martial, J. A., Wyns, L., Jaenicke, R. & Wierenga, R. K. (1999). *Proteins*, **37**, 441–453.
- Magliery, T. J. & Regan, L. (2004). *J. Mol. Biol.* **343**, 731–745.
- Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. (2003). *Structure*, **11**, 497–508.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J. (2007). *Acta Cryst.* **D63**, 32–41.
- Mosavi, L. K., Minor, D. L. Jr & Peng, Z.-Y. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 16029–16034.
- Russ, W. P., Lowery, D. M., Mishra, P., Yaffe, M. B. & Ranganathan, R. (2005). *Nature (London)*, **437**, 579–583.
- Socolich, M., Lockless, S. W., Russ, W. P., Lee, H., Gardner, K. H. & Ranganathan, R. (2005). *Nature (London)*, **437**, 512–518.
- Steipe, B., Schiller, B., Plückthun, A. & Steinbacher, S. (1994). *J. Mol. Biol.* **240**, 188–192.
- Stiel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. (2003). *Nature Struct. Biol.* **10**, 59–69.
- Sullivan, B. J., Durani, V. & Magliery, T. J. (2011). *J. Mol. Biol.* **413**, 195–208.
- Sullivan, B. J., Nguyen, T., Durani, V., Mathur, D., Rojas, S., Thomas, M., Syu, T. & Magliery, T. J. (2012). *J. Mol. Biol.* **420**, 384–399.
- Wierenga, R. K. (2001). *FEBS Lett.* **492**, 193–198.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.