Biochemistry and Biophysics Reports 8 (2016) 284–289

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

First attempts to crystallize a non-homogeneous sample of thioredoxin from *Litopenaeus vannamei*: What to do when you have diffraction data of a protein that is not the target?

Adam A. Campos-Acevedo[a,*], Adelaida Díaz-Vilchis[a], Rogerio R. Sotelo-Mundo[b], Enrique Rudiño-Piñera[a,*]

[a] *Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología (IBT), Universidad Nacional Autónoma de México (UNAM), Avenida Universidad 2001, Colonia Chamilpa, PO Box 62210, Cuernavaca, Morelos, Mexico*

[b] *Laboratorio de Estructura Biomolecular, Centro de Investigación en Alimentación y Desarrollo, A. C. (CIAD), Carretera a Ejido La Victoria Km 0.6, PO Box 1735, Hermosillo, Sonora 83304, Mexico*

ARTICLE INFO

Keywords:
Crystal structure
Inorganic pyrophosphatase
Program *BALBES*
Protein crystallization
Protein purity

ABSTRACT

The importance of sample homogeneity and purity in protein crystallization is essential to obtain high-quality diffracting crystals. Here, in an attempt to determine the crystal structure of thioredoxin 1 from whiteleg shrimp *Litopenaeus vannamei* (*Lv*Trx), we inadvertently crystallized the hexameric inorganic pyrophosphatase of *Escherichia coli* (E-PPase) from a non-homogeneous sample product during the initial over-expression steps and partial purification of *Lv*Trx. The structure determination and identification of the crystallized protein were derived from several clues: the failures in the Molecular Replacement (MR) trials using *Lv*Trx coordinates as a search model, the unit cell parameters and space group determination, and essentially by the use of the program *BALBES*. After using the previously deposited E-PPase structure (PDB entry 1mjw) as a search model and the correct space group assignation, the MR showed an E-PPase complexed with SO_4^{-2} with small changes in the sulfate ion binding region when it compares to previously deposited E-PPases in the PDB. This work stresses the importance of protein purity to avoid the risk of crystallizing a contaminant protein or how pure need to be a protein sample in order to increase the possibility to obtain crystals, but also serves as a reminder that crystallization is by itself a purification process and how the program *BALBES* can be useful in the crystal structure determination of previously deposited structures in the PDB.

1. Introduction

To determine the three-dimensional coordinates of proteins by crystallography, it is necessary first to purify and then generate crystals of suitable size and quality for X-ray diffraction experiments. Although significant work has been performed to develop protein crystallization methodologies, protein crystallization remains as a bottleneck for structural determination [1–4]. Purity is the first variable that is essential to accomplish to obtain protein crystals [5–7] since contaminants within a protein batch may alter the crystal packing of a growing crystal [8–10]. Studies on the effect of macromolecular impurities on protein solubility and crystallizability are limited. However, Skouri et al. in 1995 [6] measured the effect of 2% (w/v) ovalbumin on lysozyme solubility over a concentration of 3–8% (w/v) NaCl. They also conducted similar experiments exploring the effect on lysozyme solubility using 1% (w/v) ovalbumin, 1% (w/v) conalbumin, and 1% (w/v) bovine serum albumin. However, no significant effects were observed in the crystalline packing.

Escherichia coli is the most used bacterial expression systems due to the in-depth knowledge of this microorganism and for the high amounts of heterologous proteins that can be produced. However, despite its many advantages, particular conditions such as incubation time or other stresses during the culture stage may favor the expression of *E. coli* native proteins more than the heterologous protein expression. Such is the case of E-PPase expressed in conditions where the energetic source is compromised to keep the culture alive [11]. In this work, we describe the expression, purification, crystallization, structural determination and coordinates analysis of the E-PPase complexed with SO_4^{-2} as a result of crystallization experiments using a non-homogeneous partially purified *Lv*Trx sample.

* Corresponding authors.
E-mail addresses: adam@ibt.unam.mx (A.A. Campos-Acevedo), rudino@ibt.unam.mx (E. Rudiño-Piñera).

http://dx.doi.org/10.1016/j.bbrep.2016.09.011
Received 25 August 2016; Received in revised form 22 September 2016; Accepted 23 September 2016
Available online 15 October 2016

2. Materials and methods

2.1. Expression and purification

The *E. coli* PPase was similarly purified by previously published studies of whiteleg shrimp thioredoxin 1 from *Litopenaeus vannamei* (*Lv*Trx) [11–13]. Cells of the host strain *E. coli* BL21 (DE3) transformed with plasmid pET11a/*Lv*Trx were grown on Luria-Bertani (LB) agar plates [1% (w/w) tryptone, 0.5% (*w/v*) yeast extract, 1% (*w/v*) NaCl] containing 200 μg ml^{-1} ampicillin at 310 K. A single colony was picked and grown for plasmid isolation. A colony was inoculated into 50 ml LB broth containing 200 μg ml^{-1} ampicillin and incubated for 12 h at 310 K; part of the culture (25 ml) was used to inoculate l L of LB medium with 200 μg ml^{-1} ampicillin. The culture was grown to an absorbance of 0.6 at OD_{600}. IPTG (isopropyl-β-thiogalactoside) was added to the broth to a final concentration of 0.4 mM and grown for an additional 12 h. The cells were then harvested by centrifugation (24,000*g*, 30 min, 277 K) using a Beckman JA-14 rotor centrifuge (Beckman Coulter, CA, USA) and washed in 0.9% NaCl (*w/v*). The cells were resuspended in cold lysis buffer (100 mM Tris–HCl, pH 8.0) containing Complete EDTA-free® protease-inhibitor cocktail (Roche Molecular Biochemical, USA) and sonicated with three pulses per 1 min and one rest interval of 5 min per pulse on an ice bath. Cell debris was removed by centrifugation (24,000*g*, 20 min, 277 K), and the supernatant containing the soluble target protein was collected for purification.

The supernatant was fractionated by two consecutive precipitation steps at 50% and 85% ammonium sulfate saturation. The precipitate was resuspended in buffer (10 mM Tris–HCl, pH 7.5) and heated at 343 K for 20 min. Cell debris was then removed by centrifugation (24,000*g*, 20 min, 277 K). The supernatant was dialyzed two times at 277 K in ten times its volume in buffer (10 mM Tris–HCl, pH 7.5) and the supernatant was loaded onto a 15 ml ion exchange column (Q-Sepharose™ GE Healthcare, Sweden) pre-equilibrated and washed with three column volumes of buffer (10 mM Tris–HCl, pH 7.5). The sample was eluted with a pulse of 300 mM NaCl in the same buffer at a flow rate of 1 ml min^{-1}. Column fractions were collected and analyzed by SDS-PAGE, and fractions with the enzyme were pooled and dialyzed against buffer 10 mM Tris–HCl, pH 7.5, and concentrated. Protein concentration was determined with the Bradford dye reagent (Bio-Rad Laboratories, USA) [14].

2.2. Protein crystallization

Crystallization trials were performed using Crystal Screen and Crystal Screen II kits from Hampton Research (Aliso Viejo, CA, USA) by the hanging drop vapor-diffusion method at 291 K. The drops were prepared manually in 24-well crystallization plates by mixing the enzyme (2 μl) at 30 mg ml^{-1} with the reservoir solution (2 μl) containing 0.1 M sodium acetate, pH 4.6 and 2.0 M ammonium sulfate. DTT was added to a final concentration of 5 mM directly in the crystallization drops. Suitable crystals for diffraction were obtained in one month. Crystals for data collection were then flash-cooled by immersion in liquid nitrogen using 30% (v/v) glycerol into the mother solution as a cryoprotectant.

2.3. Data collection and structure determination

Diffraction data were collected on beamline X6A of the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory (BNL), USA, using an ADSC Quantum 270 detector. The X-ray diffraction data were collected from a single crystal at 100 K (incident wavelength, λ=0.975 Å). The diffraction images data were integrated using XDS [15] and scaled with SCALA from the CCP4 suite (Collaborative Computational Project, Number 4) [16]. Molecular replacement was done with a cross-rotational search followed by a translational search using the coordinates of E-PPase was performed using the program PHASER [17] to obtain initial models and phases (LLG=8383, RFZ=9.9, TFZ=11.9, being a trimer found in the asymmetric unit). The models were improved based on manual inspection of the $2F_o–F_c$ map after rigid-body refinement and geometric constraint performed in REFMAC [18]. All further refinement was done using the PHENIX suite [19]. The final model was completed and refined using the programs PHENIX and COOT to a final R_{work}/R_{free} of 19.3/23.5% [20]. Data collection and refinement statistics are summarized in Table 1.

Table 1
Summary of crystallographic data collection and refinement. Values in parentheses are for the highest resolution shell.

Parameters	*E-PPase PDB (4UM4)*
Data collection statistics	
X-ray source	*BNL NSLS Beamline X6A*
Wavelength (Å)	*0.975*
Space group	*C 1 2 1*
Unit-cell dimensions	
a, b, c (Å)	*120.0, 108.9,81.0*
α, β, γ angles (°)	*120.0, 108.9, 81.0*
Resolution range (Å)	*90.0, 97.0, 90.0*
No. de reflections	*19.10–2.65*
No. of unique reflections	*61,030*
Completeness (%)	*27,508*
R_{sym} *(%)*[1]	*92.0 (94.0)*
R_{meas} *(%)*[2]	*6.0 (43)*
I/σ(I)	*7.9 (56)*
Multiplicity	*10.3 (2.3)*
Asymmetric unit content	*2.2 (2.2)*
	Trimer
Refinement statistics	
R_{work}/R_{free} *(%)*	*19.3/23.5*
B-value (Å²)	
Protein	*40.8*
Ion/Ligand	*83*
Water	*38*
All atoms	*50.8*
Wilson plot B-value (Å²)	*45.8*
RMSD from ideal stereochemistry	
Bond lengths (Å)	*0.017*
Bond angles (°)	*1.88*
Coordinate error (Maximum-Likelihood Base) (Å)	*0.12*
Ramachandran plot (%)	
Most favored regions	*92.3*
Additional allowed regions	*4.8*
Disallowed regions	*2.9*

[1] $R_{sym}=\Sigma_i \Sigma_{hkl} |I_i(hkl)–(I(hkl)| \Sigma_{hkl} \Sigma_i I_i (hkl)$, where $I_i(hkl)$and $(I(hkl))$ represent the diffraction-intensity values of the individual measurements and the corresponding mean values. The summation is over all unique measurements.

[2] R_{meas} is a redundancy-independent version of R_{sym}, $R_{meas} = \sum_h \sqrt{n_h/n_h–1} \sum^{nh}{}_i | \hat{I}_h–I_{h,i}| / \sum_h \sum^{nh}{}_i I_{h,i}$, where $\hat{I}_h=1/n_h \sum^{nh}{}_i Ih, i$.

3. Results and discussion

3.1. Co-expression between a heterologous enzyme and E-PPase host enzyme

During the initial attempts to establish a protocol for expressing and purifying recombinant enzyme *Lv*Trx [13], high levels of the overexpressed *Lv*Trx were not found among the presence of *E. coli* proteins as a result of the long incubation period. After the ion exchange chromatography step had been performed, a denaturing electrophoresis (15% SDS-PAGE) was applied, revealing different protein populations and proportions (Figs. 1a, 1b).

During the purification process and despite several attempts to