Purification, Refolding, Crystallization and Diffraction Analysis of the Native and Selenomethionine-Substituted Rat Epididymal-Specific Lipocalin†

Tao Wang,‡ Dao-Wei Zhu,§ Yong-Shui Zou,‡ Arezki Azzi,§ Yi Shi,‖ Yu-xin Jin,‖ Rong-Guang Zhang,⊥ and Sheng-Xiang Lin*†§

Structural Biology Platform and State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Laboratory of Oncology and Molecular Endocrinology, CHUL Research Center (CHUQ) and Laval University, Canada, and Biosciences Division, Structural Biology Center, Argonne National Laboratory, Argonne, Illinois

Received September 27, 2007

ABSTRACT: We describe a straightforward crystallogenesis protocol leading to the preparation of protein crystals suitable for structure determination that involves protein expression and purification, refolding of the overexpressed protein, search of optimal crystallization conditions and diffraction data collection on native and selenomethionine substituted crystals. The protocol is exemplified with epididymal-specific lipocalin (rLcn6), a newly discovered monomeric protein of 19 kDa that may play an important role in sperm maturation. This protein was cloned from Norway rat (Rattus norvegicus) genome and expressed as insoluble inclusion bodies in Escherichia coli. After refolding of the purified protein, microcrystals were obtained after sparse matrix screening. Optimization of conditions after stepwise incremental initial conditions led to single crystals belonging to space group P2_12_1 and data set at 1.90 Å. After attempts with several models, initial phasing was not found by molecular replacement. A similar methodological scheme was used to grow quality crystals of the selenomethionine-substituted rLcn6 protein and collect diffraction data at 2.0 Å, allowing phasing and structure resolution. This protocol may be of particular help when overproduction results in denatured proteins within inclusion bodies, a situation that often occurs especially with proteins from eukaryotes, as well as with structural genomic projects.

1. Introduction

Crystallization drawbacks often originate from the poor quality of the protein preparations used for crystallization purposes. Here, we describe a protocol leading to crystals suitable for X-ray analysis that could be of general use when the purified proteins are of poor biochemical quality. It was worked out with the epididymal-specific lipocalin from Norway rat (rLcn6), a newly found lipocalin that may play a role in sperm maturation.

The lipocalin (Lcn) family consists of structurally conserved hydrophobic ligand binding proteins, which can also form macromolecular complexes and serve as cell-surface receptors. The lipocalin genes have undergone repeated duplications during evolution, thus expanding to more than 110 known genes. They have been found in all the major taxonomic groups from prokaryotes to plants, invertebrates, and vertebrates. To date, the presence of a six- or eight-stranded β-barrel is common in their tertiary structures. Because of their ability to bind and transport small hydrophobic molecules, lipocalins participate in the distribution of such substances. However, the function of lipocalin is not limited to transfer processes. Lipocalins function in a broad range of systems including taste and odor chemoreception and transportation, immune modulation, pros-

1 Part of the special issue (Vol 7, issue 11) on the 11th International Conference on the Crystalization of Biological Macromolecules, Québec, Canada, August 16–21, 2006 (preconference August 13–16, 2006).
2 Corresponding author and visiting scientist at SIBS. E-mail: structuralbiology@sibs.ac.cn.
3 Structural Biology Platform, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences.
4 CHUL Research Center (CHUQ) and Laval University.
5 State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences.
6 Argonne National Laboratory.

10.1021/cg700945a CCC: $37.00 © 2007 American Chemical Society
Published on Web 11/07/2007
expressed in *Escherichia coli* BL21-CodonPlus (DE3)-RIL, and Seleno-
methionine-substituted (Se-Met) rLcn6 was expressed in *Escherichia coli*
strain B834 (DE3). Rat Lcn6 was expressed in *E. coli* BL21-
CodonPlus (DE3)-RIL cells in 2×YTA medium incubated at 37°C with
shaking until the A600 reached 0.6 (O.D.). IPTG was added to a
final concentration of 0.1 mM and the incubation was continued for an
additional 6 h at 20°C. Se-Met rLcn6 was expressed in *E. coli* B834
(DE3) following a similar protocol except that a complete amino acid
medium with Se-Met replaced 2×YTA medium. Both native and
derivative proteins were expressed in insoluble inclusion bodies.
The cells induced by IPTG were precipitated by centrifugation at
7700 g for 10 min at 4°C. Cell sediments were resuspended on ice in
0.05 M Na2HPO4 containing 1 mM EDTA, 0.25 M NaCl, and 0.05%(v/
v)-Triton X-100 at 5 mL per gram (wet weight). Lysozyme was added
to 1 mg/mL and incubated on ice for 30 min. The cells were lysed by
ultrasonication and centrifuged at 40000 g (JA25.5, Beckman) for 30
min at 4°C. The Pellet was retained and dissolved in 5 M guanidine
hydrochloride (Gu-HCl) in buffer B (0.05 M Na2HPO4 at pH 5.5, 5%
glycerol and 0.1 mM PMSF). rLcn6 protein was refolded by gradual
dilution with 10× volumes of buffer B. The solution was concentrated
and mixed with Ni-sepharose 6 (agarose) over 2 h. The protein with
Ni-sepharose 6 mixture was filled into an empty nickel column, washed
with buffer B containing 100 mM imidazole (pH 5.5), and then eluted
with the same buffer in the presence of 400 mM imidazole. After gel
filtration chromatography, homogeneous proteins were obtained. During
the gel filtration, 5% isopropanol was added to effectively reduce the
hydrophobic interactions between the desired protein and other proteins.
The fractions were collected and concentrated to 3 mg/mL by Centricon
YM-30 (Millipore) with pH 7.0 HEPES-Na buffer containing 0.02%
β-octylglucoside.

2.2. Crystallization and Data Collection. rLcn6 crystals were
obtained by vapor-diffusion in hanging drops at 298 K. The initial
crystallization screen was performed with Sparse-matrix screening and
Screen II from Hampton Research, USA. The crystallization was
initiated by mixing 1 μL of protein with 0.5 μL of well solution. Many
small crystals gradually appeared within 1 week. After the crystallization
conditions were optimized, crystals of reasonable size were chosen and
shipped to APS (Advanced Photon Source, Chicago). The best native
data set of the crystals was collected at 1.90 Å in a 19-BM Beam-line,
and the best data from the Se-Met-rLcn6 crystal was collected at 2.0
Å at 100 K. The cryo-protectant solution was the same as the reservoir
solution containing 20% glycerol (v/v).

3. Results and Discussion
A serious hurdle for structure determination in structure
genomics is the acquisition of pure, soluble, and concentrated
protein samples. Many proteins fail to adopt their three-
dimensional structure and end up as inclusion bodies when
expressed in the more conventional bacterial expression sys-
tems. In this study, a successful protocol for rLcn6 expression,
purification, and refolding was established. rLcn6 Se-Met
derivative was also obtained by the same protocol.

3.1. Protein Preparation and Refolding. rLcn6 was cloned
into the pET24 expression vector and verified by sequencing.
A 6× His-tag in pET24 allowed efficient purification of the
expressed fusion protein by metal affinity chromatography. The
plasmid rLcn6-pET24a was transfected into *Escherichia coli*
BL21-CodonPlus (DE3)-RIL. The protein was expressed in
insoluble inclusion bodies in BL21 (DE3) cells. Many attempts,
involve decreasing the concentration of IPTG, expression at
lower temperatures, and the use of various *Escherichia coli*
cell lines, etc., were made to express the protein in the soluble
fraction. The fusion protein was highly expressed after induction
at 37°C with 0.1 mM IPTG. SDS-PAGE of cell lysates
demonstrated a major protein band of the expected 20 kDa size
(Figure 1).

The recombinant rLcn6–pET24a protein was nevertheless
expressed as inclusion bodies in both BL21-CodonPlus (DE3) -
RIL and B834. The induced cells were collected by centrifuga-
tion and homogenized by ultrasonication. Inclusion bodies were completely dissolved in buffer B containing 5 M Gu-HCl that contained more than 80% protein in rLcn6. The protein solution was slowly diluted with 10× volume buffer to decrease the concentration of denaturant.6,7 This is more convenient in terms of handling, time, and cost than other methods,6,7 and is useful for large-scale purifications. The solutions were then concentrated and centrifuged to separate the precipitate for the subsequent chromatography. The latter was mixed with Ni-NTA material and incubated at room temperature for 2 h. The mixture was loaded onto the column, which was then washed with buffer containing decreasing concentrations of Gu-HCl from 1 to 0 M. The refolded protein bound onto the column and was then washed with a buffer containing 100 mM imidazole and the targeted protein was eluted at 400 mM imidazole. During this step, on-column refolding was carried out.9 The refolding and purification thus occurred on the same column.10,11 After an additional gel filtration chromatography, homogeneous protein was obtained.

The N-terminal amino acids were sequenced (Institutes for Biological Sciences, SIBS, CAS) and were determined to identify with the rLcn 6 sequence. The homogeneous protein was collected and concentrated to about 2 mg/ml in the absence of detergent due to its hydrophobic nature (Figure 2). From this figure, many amino acids could be identified in the hydrophobic zone. Several detergents were screened and β-O.G was found to increase the protein solubility to about 3 mg/mL. When we carried out the crystal screening, the protein was mixed with well solution in a 2:1 ratio to facilitate its supersaturation in the crystallization drop.

Initially, microcrystals of rLcn6 were obtained by vapor-diffusion. They were identified as being rLcn6 crystals by dissolving the washed crystals and performing SDS-PAGE analysis. The molecular mass correlated with that of the purified rLcn6. The initial crystallization condition consisted of 0.1 M HEPES-Na pH 7.5, 20% (w/v) PEG 4000, and 10% isopropanol. The condition was optimized by varying the PEG concentration from 6 to 24% (w/v) by 2% (w/v) increments, increasing the pH from 5.6 to 8.5 by 0.1 units, and varying isopropanol concentration from 5 to 15% by 1% increments. This screen suggested that the best crystals were obtained under the following conditions: pH 7.7 HEPES-Na, 5% PEG4000, 10% isopropanol, and 24 °C (Figure 3).

Selenomethionine-substituted (Se-Met) rLcn6 derivative was expressed in E. coli strain B834 (DE3). The cells were cultured in the complete amino acid medium containing selenomethionine. Similarly to the native protein, it was also expressed in inclusion bodies. The cells were observed to grow more slowly using this particular medium. To obtain sufficient protein (from the inclusion bodies), the expression time was prolonged to 14 from 6 h. The following SDS-PAGE of cell lysates showed a major protein band consistent with the native rLcn6 protein (Figure 1).

The Se-Met rLcn6 was purified and refolded following the same protocol. It displayed biochemical characteristics similar to those of the native protein. Its low solubility could be improved by the presence of 0.02% β-O.G. The homogeneous Se-Met protein was used for crystal screening and the crystals appeared under conditions similar to those of the native protein. After optimization, suitable crystals were obtained. In contrast to the native rLcn6, it was more difficult to reproduce Se-Met rLcn6 crystals. Typically, the Se-Met rLcn6 protein lost its crystallization ability after one week. When the protein was loaded onto an SDS-polyacrylamide gel, no degraded bands were identified. Fortunately, the Se-Met crystals appeared after three days, so there was an opportunity for optimization.

### 3.2. Data Collection and Processing.

Both the native rLcn6 crystals and Se-Met crystals were packed in a capillary and sent to APS (Advanced Photon Source, Chicago). The diffraction data of the native crystal were processed and scaled with Denzo/HKL,12 shown in Table 1. The crystals were indexed to space group $P_{2_1}2_12_1$ with unit-cell parameters $a = 48.59$ Å, $b = 51.71$ Å, $c = 58.68$ Å. From the Matthews coefficient, the solvent fraction was estimated to be 33.31%, showing that there is only one molecule in the unit cell, whereas the crystal packing parameter $V_m$ is 1.84 Å³/Da. Molecular replacement was attempted by Molrep,13 Amore,14 Beast,15 and Phaser16,17 using several starting models such as Boar salivary lipocalin,18 Major urinary protein,11 alpha 2u-globulin,19 Major Horse Allergen Equ c1,20 and other related proteins. However, the true peak

| Table 1. Data Process Statistics of rLcn6 and Se-Met rLcn6 |
|-----------------|-----------------|
|                | rLcn6           | Se-Met          |
| $a$ (Å)        | 48.59           | 47.32           |
| $b$ (Å)        | 51.71           | 50.20           |
| $c$ (Å)        | 58.69           | 59.27           |
| $a = b = c$ (deg) | 90              | 90              |
| space group    | $P_{2_1}2_12_1$ | $P_{2_1}2_12_1$|
| average redundancy | 8.7 (4.8)      | 15.3 (8.1)      |
| $R_{merge}$    | 0.067(0.346)    | 0.110(0.604)    |
| $I / (o)$      | 28.8 (2.3)      | 24.9 (1.5)      |
| no. of independent reflns | 10267 (629) | 9123 (674) |
| completeness (%) | 84 (52.7)      | 91.6 (69.3)    |
| resolution (Å) | 1.90 (1.97–1.90) | 2.0 (2.07–2.0) |

* For high resolution shell.
was not found because of their low identity. This urged us to express the Se-Met protein. The purification and refolding procedures were similar to the native protein protocol. The Se-Met rLcn6 crystals were obtained under the same crystallization conditions as the native protein and the structure was determined by single anomalous diffraction (SAD). The diffraction data of Se-Met rLcn6 were collected and processed (Table 1). Se-Met crystal was also indexed to space group with P2_12_12_1 and its unit-cell dimensions were a = 47.32 Å, b = 50.20 Å, and c = 59.27 Å. There is little difference between their unit-cell parameters. The Matthews coefficient was calculated as 1.86 Å^3/Da with an estimated solvent content assuming one molecule in the asymmetric unit. The structure determination is being carried out.

4. Conclusion

Here we report how crystals of Se-Met rLcn6 were obtained after protein denaturation, purification, and refolding using the same protocol as the native protein. The crystals were obtained under identical conditions. The diffraction data also show that the native and Set-Met Lcn6 crystal forms have the same space group with very similar unit cells and resolution. This protocol may be of particular help when overproduction results in protein accumulation in inclusion bodies, a situation that often occurs especially with proteins from eukaryotes, or a situation often unencountered in structural genomic projects.

Acknowledgment. We thank Prof. Yong-lian Zhang for her interest in this work. The study is supported by IBCB and SIBS for Structural Biology Platform, and a grant from the National Natural Foundation of China (30670448).

References


CG700945A