Published in final edited form as:

J Struct Funct Genomics. 2004 ; 5(1-2): 111–118. doi:10.1023/B:JSFG.0000029206.07778.fc.

Automation of protein purification for structural genomics

Youngchang Kim, Irina Dementieva, Min Zhou, Ruiying Wu, Lour Lezondra, Pearl Quartey, Grazyna Joachimiak, Olga Korolev, Hui Li, and Andrzej Joachimiak*

Biosciences Division and Structural Biology Center, Argonne National Laboratory, 9700 S. Cass Ave., Bldg 202, Argonne, IL 60439, USA

Abstract

A critical issue in structural genomics, and in structural biology in general, is the availability of high-quality samples. The additional challenge in structural genomics is the need to produce high numbers of proteins with low sequence similarities and poorly characterized or unknown properties.

‘Structural-biology-grade’ proteins must be generated in a quantity and quality suitable for structure determination experiments using X-ray crystallography or nuclear magnetic resonance (NMR). The choice of protein purification and handling procedures plays a critical role in obtaining high-quality protein samples. The purification procedure must yield a homogeneous protein and must be highly reproducible in order to supply milligram quantities of protein and/or its derivative containing marker atom(s). At the Midwest Center for Structural Genomics we have developed protocols for high-throughput protein purification. These protocols have been implemented on AKTA EXPLORER 3D and AKTA FPLC 3D workstations capable of performing multidimensional chromatography. The automated chromatography has been successfully applied to many soluble proteins of microbial origin. Various MCSG purification strategies, their implementation, and their success rates are discussed in this paper.

Keywords

affinity chromatography; automation; protein purification; structural genomics

Introduction

In the past two and a half years, the NIH funded Protein Structure Initiative (PSI) pilot projects have been developing a protein structure determination pipeline [1–7] capable of producing large numbers of protein samples for structural biology applications. One of the main objectives of the PSI pilot projects is to develop technologies for production of proteins in milligram quantities reliably, reproducibly, quickly, and at low cost.

For crystallography applications the resulting protein samples must be compatible with the crystallization process. The protein in the sample must be folded and soluble, as well as chemically, conformationally, and functionally homogeneous. The sample must be free of critical contaminants that may degrade, denature, destabilize, or modify protein or interfere with crystallization or structure determination. Protein purity of >95% is typically required. Protein samples must be stable during crystallization trials, suitable for incorporation of heavy atoms to aid structure determination, and functionally relevant. The quantities of proteins in the samples must allow achieving protein concentrations in the range of 5–25 mg/ml, testing

200–500 crystallization conditions, growing X-ray-quality single crystals, establishing cryoconditions, and producing rational heavy atom derivatives for structure determination. These criteria put certain restrictions on the methods and procedures of sample preparation.

The MCSG (www.mcsg.anl.gov) has developed standard operating procedures for protein purification that make protein samples suitable for automated structure determination using synchrotron-based X-ray crystallography. These standard operating procedures are based on the following principles:

- All proteins are expressed as a fusion with a uniform, cleavable affinity tag and protected against proteolysis with several protease inhibitors.
- Proteins are purified using affinity chromatography followed by buffer-exchange chromatography, to promote protein solubility and efficient tag removal.
- The affinity tag is cleaved off by a specific tagged protease.
- The protein is further purified using affinity chromatography followed by buffer-exchange chromatography compatible with protein concentration and crystallization methods.

In the MCSG approach (Figure 1), protein samples can be obtained that are free of contaminants, including the majority of background proteins, tagged protease, affinity tags, and other low-molecular-weight contaminants, as well as uncleaved target proteins.

The MCSG standard purification procedures have been implemented on the automated robotic chromatographic platforms AKTA EXPLORER 3D and AKTA FPLC 3D (Amersham Biosciences) and successfully applied to more than 250 soluble proteins of microbial origin. MCSG is in the process of expanding the approach to well-expressed but insoluble proteins [8] and membrane proteins [9].

Materials and methods

Cloning and expression

The proteins were cloned in pMCSG7 vector [10,13] and expressed in *Escherichia coli* BL21 (DE3)-Gold (Stratagene) harboring a plasmid encoding three rare tRNAs [2,5,11,12]. The pMCSG7 vector creates a construct with cleavable His₆-tag fused into the N-terminus of the target protein. The pMCSG7 construct bearing a TEV protease cleavage site adds three artificial residues (SerAsnAla) on the N-terminus of the target protein. Target proteins were expressed at the scale of 2 L of low-density culture or 250 mL of high-density culture [14].

Basic protein purification protocol

Isolated cell pellets were resuspended in five volumes of the lysis buffer containing 50 mM HEPES 8.0, 500 mM NaCl, 10 mM imidazole, 10 mM β -ME and 5% glycerol (buffer A) and inhibitors of proteases (Sigma, P8849) and incubated for 30 min on ice with lysozyme (Sigma) at 1 mg/ml followed by the sonication (6×30 s, on ice). All samples were clarified by centrifugation at $30,000 \times g$ (RC5C-Plus centrifuge, Sorval) for 20 min followed by filtration through 0.4 μ m and 0.22 μ m in-line filters (Gelman).

The standard purification protocol includes the following chromatographic steps:

- IMAC-I (immobilized metal affinity chromatography) using a 5-ml HiTrap Chelating HP column (Amersham Biosciences) charged with Ni⁺² following factory-recommended procedures.

- b. Buffer-exchange chromatography on a HiPrep 26/10 desalting column (Amersham Biosciences).
- c. His₆-tag cleavage using the recombinant TEV protease expressed from the vector pRK508 (a gift from Dr D. Waugh, NCI) and purified using a procedure described earlier [15]. The protease is added at an approximate ratio of 1 mg protease per 50 mg of target protein and incubated at 4 °C for 16–24 h.
- d. IMAC-II using a 1-ml HiTrap Chelating column (Amersham Biosciences) charged with Ni⁺² following factory-recommended procedures.
- e. Buffer-exchange chromatography on a customized desalting column, the Sephadex G-25 Fine 26/20 XK (Amersham Biosciences).

Steps (a) and (b) were performed on the AKTA EXPLORER 3D system and steps (d) and (e) on AKTA FPLC 3D (see Results and discussion).

IMAC-I and buffer exchange steps

All chromatography experiments were performed at 4 °C. Crude extracts of six proteins (typically 15–50 mL) were applied by the sample pump (flow rate 1 mL/min) sequentially onto six 5-mL HiTrap chelating HP columns charged with Ni⁺². The columns were washed with 10 column volumes (CV) of buffer A, followed by 15 CV of buffer A containing 20 mM imidazole (flow rate 5 mL/min). Each protein was first eluted to a 10-mL loop with buffer A containing 250 mM imidazole (flow rate 2 mL/min), then applied to a HiPrep 26/10 desalting column pre-equilibrated with buffer A. Just prior to injecting protein onto the desalting column, 2 mL of 5 mM EDTA in buffer A was injected onto the desalting column to create a slow-moving EDTA zone on the desalting column and sequester any Ni⁺² ions released from the chelating column. The buffer exchange step was run at a flow rate of 8 mL/min.

The desalting column was washed and re-equilibrated prior to the next purification cycle. The tubing and loop were washed between chromatography steps to avoid cross-contamination. The final peak fractions and all solutions that could contain target protein were collected.

Throughout the purification process, several parameters, including UV absorbance, pressure, flowrate, pH, and ionic strength, were monitored and logged (Figure 2). All fractions were analyzed and documented and all data stored in a single results file.

The purification processes in this experiment took 12–15 h for six proteins, depending on the initial sample volumes. The chelating columns were recycled four to five times using an automated procedure by metal stripping with 50 mM EDTA and charging with 100 mM NiSO₄.

IMAC-II and buffer-exchange steps

Proteins purified with IMAC-I and buffer exchange were treated with the His₇-tagged TEV protease to remove the His₆-tag for 16–24 h at 4 °C following the basic protocol (see above). Cleavage was monitored by SDS-PAGE and Coomassie Brilliant Blue R (Amersham Biosciences) staining. After the cleavage, the reaction mixture containing target protein (cleaved and some uncleaved), His₇-tagged-TEV protease and His₆-tag was applied to a 1-ml chelating affinity column and the column was washed with 3 CV of buffer A. All chromatographic steps were performed at 22 °C. The column flow-through and wash fraction was first collected onto a 20-mL loop and then applied to a customized desalting column Sephadex G-25 fine XK 26/20 equilibrated with storage buffer containing 20 mM Tris/HCl 7.5, 500 mM NaCl, and 2 mM DTT. Protein was eluted with storage buffer, protein peaks were collected in 2-mL fractions (Figure 3) and analyzed by the SDS-PAGE stained with Coomassie

Brilliant Blue R (Figure 4). Purification of six proteins takes about 9 h. The 1-mL chelating columns were recycled four to five times using the automated procedure described above.

Protein characterization

We have used several methods to characterize protein samples. Table 1 indicates the method (s) used for the various aspects of protein characterization.

Protein concentration and storage

All proteins were concentrated with Centricon Plus Centrifugal Filter Units (Millipore), using molecular weight cutoff as recommended by the manufacturer. All proteins were flash frozen in ~ 50 μ L aliquots in liquid nitrogen temperature in the storage buffer and stored in an LS6000 liquid nitrogen storage system (Taylor-Wharton) for an extended period of time.

Results and discussion

Proteins encoded by microbial genomes represent a highly diverse population of amino acid sequences. As a result, these proteins are highly dissimilar in their properties, making design of standard purification protocols a rather challenging undertaking. To address the protein diversity issue, a common affinity marker can be attached to all proteins that allows selective purification of tagged protein from crude extracts using single-step affinity chromatography. The affinity tag should be unique, accessible, and preferably small, have high capacity to bind to a matrix and excellent conditional affinity (ON/OFF binding), and should be low cost. The His₆-tag and its variants appear to meet all of these criteria and are highly effective for protein purification [16]. However, the His₆-tag-based approach still has a few drawbacks such as, not all proteins can be labeled with His₆-tag on their N- or C-terminus, the tag may interfere with protein folding or oligomerization, and the tag may be inaccessible or lead to protein aggregation [17].

Protein expression system for automated purification

As a precondition to establishing generic standard operating procedures for automated protein expression, we focused on selecting:

- Proper expression construct, with high level of target protein expression and solubility.
- Effective protease for tag removal.
- A buffer system that would promote the solubility of most proteins.
- Chromatography media and hardware.

We tested several different affinity tags, proteases, and buffer conditions with multiple target proteins for their efficiency and adaptability to the structural genomics pipeline. Five constructs containing His₆-tag and His₆-S-tag and different proteolytic sites were used for target protein expression:

- pET15b (His₆-tag – thrombin site) [18]
- pET30LIC (His₆ – thrombin site–S-tag –factor Xa site) [19,20]
- pMCSG3 (His₆-tag – factor Xa site) [unpublished]
- pProEX (His₆-tag – TEV protease cleavage site: ENLYFQ ↓ G) [21]
- pMCSG7 (His₆-tag – TEV protease cleavage site: ENLYFQ ↓ S) [10,13]

We found the pMCSG7 vector (a derivative of pET vector) to be most compatible with our standard operating procedures [10].

Three proteases (human thrombin, factor Xa from bovine plasma, and recombinant TEV protease) were tested for efficiency of tag removal using a standard protocol. Parameters evaluated were: efficiency of tag cleavage, level of nonspecific cleavage, optimum temperature, and fraction of successfully processed proteins. Our results show that TEV protease is most suited for MCSG targets (Table 2). TEV protease offers several advantages:

- It is highly specific, recognizing a seven-aminoacid sequence.
- It shows virtually no nonspecific proteolysis of target proteins.
- It is active under a wide range of conditions, including low temperature (4 °C), broad range of pH, and high ionic strength [22].

The TEV protease expressed from the vector pRK508 carries noncleavable His₆-tag and can be removed from protein samples by IMAC. Moreover, TEV protease was highly effective at removing His₆-tags for more than 96% of tested MCSG target proteins. TEV protease failed completely in only a few cases (Table 2).

Platform for automated multidimensional chromatography

MCSG collaborated with Amersham Biosciences to adapt the AKTA Explorer 100A for multidimensional automated chromatography needs. The complete AKTA EXPLORER 3D system consists of the AKTA Explorer 100A, UNICORN software version 4.0 or higher, sample pump P-950, fraction collector Frac-950, a 3D kit that allows attachment of multiple columns and loops, multi-channel UV/Vis spectrometer, two sample loops, and an air sensor.

The AKTA EXPLORER 3D system executes a series of commands to perform automated purification and sample collection. The sample pump includes an in-line air sensor that allows unattended direct loading of crude protein extracts. The system is capable of purifying up to seven protein samples. Samples are loaded serially on (1) up to seven singlestep IMAC columns, (2) up to six IMAC columns, each followed by a buffer-exchange chromatography step, or (3) up to five IMAC columns, followed by buffer-exchange chromatography and another chromatographic step. In between chromatographic steps, the samples are stored in sample loops (10 mL and 20 mL). Automated peak detection allows collection of target protein peaks and other relevant fractions into appropriate loops or in the fraction collector. AKTA FPLC is similarly outfitted for multidimensional chromatography (AKTA FPLC 3D).

Using UNICORN software, we have developed several methods for automated protein purification, as well as for automated charging of chelating columns that utilize the chemistry of metal stripping followed by recharging of the matrix with Ni⁺². Potential problems with leaching of Ni⁺² during purification have been addressed (see Materials and methods).

Large-scale evaluation of purification protocol

The strategy for automated protein purification is outlined in Figure 1. Using steps 1 and 2, a large number of His₆-tagged proteins (24–30 per week) can be produced on a scale of 20–200 mg and 85–90% purity for well-expressed, soluble proteins (Figures 2 and 4). These proteins are suitable for initial crystallization screening. However, the level of impurities and the presence of the His₆-tag may affect protein stability, solubility, and aggregation, thereby affecting the protein's ability to crystallize and reducing the quality of crystals (as discussed earlier). The use for crystallization screening of such samples is recommended only when the protein yield is very low. Typically, the initial screening of crystallization conditions leads to crystals in about 25% of such protein samples.

Including additional purification steps – tag cleavage by TEV protease and IMAC-II followed by buffer exchange – resulted in much higher quality protein samples, typically 95–98% pure.

In some cases, persistent contaminants must be removed by additional chromatography (ion exchange or/and gel filtration) to improve purity and crystal quality. Figure 5 shows the distribution of protein yield for 253 proteins purified using the process described in this paper and summarized in Figure 1.

Conclusions

We and others have shown that the automation of protein chromatographic steps is feasible using commercially available products [23]. The MCSG automated protein purification process has been tested using manual approaches to evaluate various purification steps' reliability, robustness, cost, and labor savings. The process has since been ported to the robotic workstation, and the resulting purification data have been deposited using manual and automated entry into the MCSG Protein Purification Database and integrated with the central MCSG repository for public access (www.mcsg.anl.gov).

Among the many chromatographic workstations currently available, the AKTA EXPLORER 3D workstation from Amersham Biosciences could best accommodate our protocols with respect to multiple column steps, extract volumes, protein yields, flow properties, buffer compatibility, cold-box operations, and data management. The AKTA EXPLORER 3D system provides up to eight column slots, one of which is used for a bypass, and two loops, where the intermediate protein-containing solutions are held between the two chromatographic steps.

The system offers several advantages. Its multitasking capabilities allow for simultaneous applications and pump washes. All chromatographic steps are run under optimal conditions, purifications are highly reproducible (Figures 2 and 3), and protein exposure to air is limited. The software offers high flexibility; for example, purification can be run manually or in automated mode by programs (scripts).

Several programs were scripted starting from templates provided with the purification workstation. More than 200 proteins have been purified using this automated system and over 30% produced crystals for the MCSG structural genomics program.

Acknowledgments

We would like to thank Linda Henry and Jennifer Gerdin from Amersham Biosciences for setting up and debugging the AKTA EXPLORER 3D system and helping with programming; Luke Maj, Allison Mo, Mike Straza, Dave Popiel, Thomas Rivera, Elena Vinokour, and Kelly Peterson for contributing to the development of the initial protein purification procedures; Lindy Keller for help in preparation of this manuscript; and Mark I. Donnelly for useful comments. This work was supported by National Institutes of Health Grant GM62414 and by the U.S. Department of Energy, Office of Biological and Environmental Research, under contract W-31-109-Eng-38.

Abbreviations

MCSG, Midwest Center for Structural Genomics; IMAC, immobilized metal affinity chromatography; TEV, tobacco etch virus; β -ME, B-mercaptoethanol; DTT, dithiothreitol; EDTA, ethylenedi-aminetetraacetate; SDS-PAGE, polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate.

References

1. Burley SK, Bonanno JB. *Methods Biochem. Anal* 2003;44:591–612. [PubMed: 12647406]

2. Yee A, Pardee K, Christendat D, Savchenko A, Edwards AM, Arrowsmith CH. *Acc. Chem. Res* 2003;36:183–189. [PubMed: 12641475]
3. Chance MR, Bresnick AR, Burley SK, Jiang JS, Lima CD, Sali A, Almo SC, Bonanno JB, Buglino JA, Boulton S, Chen H, Eswar N, He G, Huang R, Ilyin V, McMahan L, Pieper U, Ray S, Vidal M, Wang LK. *Protein Sci* 2002;11:723–738. [PubMed: 11910018]
4. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, McPhillips TM, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Elsliger MA, Wooley J, Taylor SS, Hodgson KO, Wilson IA, Schultz PG, Stevens RC. *Proc. Natl. Acad. Sci. USA* 2002;99:11664–11669. [PubMed: 12193646]
5. Christendat D, Yee A, Dharamsi A, Kluger Y, Gerstein M, Arrowsmith CH, Edwards AM. *Prog. Biophys. Mol. Biol* 2000;73:339–345. [PubMed: 11063779]
6. Pedelacq JD, Piltch E, Liong EC, Berendzen J, Kim CY, Rho BS, Park MS, Terwilliger TC, Waldo GS. *Nat. Biotechnol* 2002;20:927–932. [PubMed: 12205510]
7. Heinemann U, Frevert J, Hofmann K, Illing G, Maurer C, Oschkinat H, Saenger W. *Prog. Biophys. Mol. Biol* 2000;73:347–362. [PubMed: 11063780]
8. Kim Y, Dementieva I, Joachimiak G, Lezondra L, Li H, Laury N, Quartey P, Wu R, Zhou M, Joachimiak A. Protein Structure Initiative Workshop on Protein Production and Crystallization, NIH. April 9–11;2003
9. Laible PD, Scott HN, Hofman SJ, Hanson DK. *J. Struct. Funct. Genomics*. 2003submitted
10. Stols L, Gu M, Dieckman L, Raffin R, Collart FR, Donnelly MI. *Protein Expr. Purif* 2002;25:8–15. [PubMed: 12071693]
11. Studier FW. *J. Mol. Biol* 1991;219:37–44. [PubMed: 2023259]
12. Pan SH, Malcolm BA. *Biotechniques* 2000;29:1234–1238. [PubMed: 11126126]
13. Zhang R, Grembecka J, Vinokour E, Collart F, Dementieva I, Minor W, Joachimiak A. *J. Struct. Biol* 2002;139:161–170. [PubMed: 12457846]
14. Millard, C. Sanville; Stols, L.; Quartey, P.; Kim, Y.; Dementieva, I.; Donnelly, MI. *Protein Expr. Purif* 2003;29:311–320. [PubMed: 12767824]
15. Kapust RB, Waugh DS. *Protein Expr. Purif* 2000;19:312–318. [PubMed: 10873547]
16. Bornhorst JA, Falke JJ. *Methods Enzymol* 2000;326:245–254. [PubMed: 11036646]
17. Hammarstrom M, Hellgren N, van Den Berg S, Berglund H, Hard T. *Protein Sci* 2002;11:313–321. [PubMed: 11790841]
18. Osipiuk J, Górnicki P, Maj L, Dementieva I, Laskowski R, Joachimiak A. *Acta Crystallogr. D, Biol. Crystallogr* 2001;57:1747–1751. [PubMed: 11679764]
19. Zhang R, Dementieva I, Duke N, Collart F, Quate-Randall E, Alkire R, Maltsev N, Korolev O, Joachimiak A. *Proteins* 2002;48:423–426. [PubMed: 12112707]
20. Korolev S, Koroleva O, Petterson K, Collart F, Dementieva I, Joachimiak A. *Acta Crystallogr* 2002;D58:2116–2121.
21. Wu, R-y.; Zhang, R-g.; Dementieva, I.; Maltsev, N.; Laskowski, R.; Gornicki, P.; Joachimiak, A. *J. Biol. Chem* 2003;278:20240–20244. [PubMed: 12649270]
22. Dougherty WG, Parks TD. *Virology* 1989;172:145–155. [PubMed: 2672562]
23. Page R, Grzechnik SK, Canaves JM, Spraggon G, Kreusch A, Kuhn P, Stevens RC, Lesley SA. *Acta Crystallogr. D, Biol. Crystallogr* 2003;59(Pt 6):1028–1037. [PubMed: 12777766]

Figure 1.
Strategy for automation of protein purification steps for proteins expressed in *E. coli*.

Figure 2.

Example of chromatograms (as part of a results file) of IMAC-I and buffer-exchange steps using AKTA EXPLORER 3D for a six-protein (APC35594, APC35601, APC35609, APC35617, APC35624, APC35625) run. A: The chromatogram showing the progress of sample loading and column wash of six proteins with buffer A. (B, C) The chromatograms showing the first two proteins, (a) wash with buffer A containing 20 mM imidazole, (b) elution of His₆-tagged target proteins with buffer A containing 250 mM imidazole, (c) His₆-tagged target proteins after buffer exchange. Target protein names are indicated as APC numbers. In each chromatogram, UV absorbance at 280 nm is plotted versus milliliters of buffer solution flow. (d) Progress of the step gradient is indicated by the curve of %B, in green.

Figure 3.

Example of chromatogram (as part of a results file of a four-protein run) of IMAC-II and buffer exchange using AKTA FPLC 3D. Shown here is one protein (APC36103). (a) Sample loading and column wash with buffer A. (b) Elution of cleaved His₆-tags, His₇-tagged TEV protease, and uncleaved target protein with buffer A containing 250 mM imidazole. (c) Cleaved target protein after buffer exchange. In the chromatogram, UV absorbance at 280 nm is plotted *versus* milliliters of buffer solution flow.

Figure 4. SDS-PAGE of 30-kDa target protein (APC234), purified by the process described in Figure 1: *lane 1* – crude extract; *lanes 2 and 3* – IMAC-I flow through; *lane 4* – IMAC-I elution; *lane 5* – after TEV protease cleavage and IMAC-II; *lane 6* – low-molecular-weight markers (Amersham Biosciences), which run with apparent molecular weights of 97, 66, 45, 30, 20.1, and 14.4 kDa.

Figure 5. Distribution of protein production levels using the automated chromatography process. Total number of proteins was 253. The numeral on top of each column corresponds to the number of proteins purified in the amount indicated below the column (in milligrams).

Table 1

Protein characterization methods in the MCSG protocol.

Protein parameter	Method of characterization
Purity	SDS-PAGE stained with Coomassie Brilliant Blue and lab-on-the-chip 2100 Bioanalyzer (Agilent)
Concentration	Coomassie Plus Protein Assay (Pierce, Catalog No. 23236) and UV spectrometry
Poly-dispersity	Dynamic light scattering (DynaPro, Protein Solutions)
Estimated molecular weight in solution	Size exclusion chromatography
Suspected chemical heterogeneity and bound ligands	Mass spectrometry (MALDI-TOF Biflex III, Bruker)
Bound ligands	UV/Vis spectrometry

Table 2

Efficiency of His-tag cleavage by TEV protease.

% of cleavage	99–80%	70–50%	0%
Number of proteins ^a	200	31	8

^aProteins (total 239) were incubated with 1:50 ratio of protease to target protein at 4 °C for 16–24 h.