

Retrospective analyses of the bottleneck in purification of eukaryotic proteins from *Escherichia coli* as affected by molecular weight, cysteine content and isoelectric point

Won Bae Jeon*

Laboratory of Biochemistry and Biodefense Program, Daegu Gyeongbuk Institute of Science and Technology, Daegu 704-230, Korea

Experimental bioinformatics data obtained from an *E. coli* cell-based eukaryotic protein purification experiment were analyzed in order to identify any bottleneck as well as the factors affecting the target purification. All targets were expressed as His-tagged maltose-binding protein (MBP) fusion constructs and were initially purified by immobilized metal affinity chromatography (IMAC). The targets were subsequently separated from the His-tagged MBP through TEV protease cleavage followed by a second IMAC isolation. Of the 743 total purification trials, 342 yielded more than 3 mg of target proteins for structural studies. The major reason for failure of target purification was poor TEV proteolysis. The overall success rate for target purification decreased linearly as cysteine content or isoelectric point (pI) of the target increased. This pattern of pI versus overall success rate strongly suggests that pI should be incorporated into target scoring criteria with a threshold value. [BMB reports 2010; 43(5): 319-324]

INTRODUCTION

The purification of many proteins in sufficient quantity and purity is a major challenge in this new era of high-throughput structural genomics and functional proteomics (1). *E. coli* is the most widely used expression host for high-throughput protein production from prokaryotic and eukaryotic genomes (2). The advantages of an *E. coli* host include low cost, ease of culture, high expression (3) and availability of numerous vectors and strains. In many cases, however, eukaryotic proteins generated in *E. coli* cells frequently accumulate as inclusion bodies, which are insoluble aggregates of unfolded or misfolded proteins (4). A simple but efficient approach for promoting accumulation of recombinant proteins in soluble form is to ex-

press foreign proteins as a chimera with a solubility-enhancing fusion partner (5).

The aims of structural genomics are to expand the knowledge of fold-function space and ultimately predict three-dimensional structures of proteins from their gene sequences (6). Potential candidates for structural study are selected by extensive bioinformatic analysis of annotated open reading frames from a variety of gene collections in which scoring is mainly based on predicted characteristics; for example, dissimilarity to known structures, percentage of low complexity and intrinsic protein disorder (7). Biochemical properties of a target such as molecular weight (MW) and the number of cysteines are also considered in deciding the overall selection tier of the targets (8). However, the inclusion of such parameters in the target selection and scoring process has yet to be validated by experimental bioinformatic analyses. In addition, although pI is one of the most important biochemical properties of proteins, the relationship between target pI and purification success rate has not yet been addressed.

In the present study, an *E. coli* cell-based protein purification system that provides sufficient amounts of pure proteins for structural analyses is evaluated. All protein targets were from Arabidopsis, human or mouse genomes and were expressed as His-tagged MBP fusion constructs in order to exploit high-throughput IMAC capture (9) as well as to increase the solubility of targets in *E. coli* (10). Overall success rate is the percentage of target proteins out of the total number subjected to purification of which more than 3 mg were isolated with a purity greater than 90% followed by concentration to higher than 10 mg/ml without precipitation. Step success rate is the percentage of successfully obtained product out of the total number of trials performed for an individual step. In our results, a bottleneck in target purification was identified, and the effects of MW, cysteine content and pI on the overall purification success rate of target proteins are shown. Finally, possible ways for improving the overall efficiency of target purification are discussed.

*Corresponding author. Tel: 82-53-430-8630; Fax: 82-53-431-8605; E-mail: wbjeon@dgist.ac.kr

Received 8 February 2010, Accepted 3 March 2010

Keywords: Cysteine content, Experimental bioinformatics, High-throughput purification, Isoelectric point, Structural genomics

RESULTS AND DISCUSSION

Statistics and bottleneck in target purification

E. coli cells were grown in 2 liter culture scale, and target proteins were expressed as His-tagged MBP fusion constructs. Out of 783 total cell cultures performed using Terrific Broth media, the average wet cell mass harvested was 15.7 g. The expression and solubility of each fusion protein were scored according to a three-tiered scoring protocol. It was found that 558 (82%) out of 684 samples showed a high or medium level of expression while 512 (86%) out of 597 were found in the soluble fraction. Out of 743 total purification trials performed, 565 (76%) yielded an average of 162 mg of His-tagged MBP fusion proteins while 342 (46%) trials provided more than 3 mg of target proteins. General features of each purification step were described in detail elsewhere (11), and statistics of the overall target purification as well as each individual step are summarized in Table 1.

As shown in Table 1, the most common reason for failure of target purification was incomplete TEV protease cleavage. In general, removal of His-tagged MBP from the target domains was completed within 3-5 h. However, about 40% of fusion proteins were cleaved incompletely (less than 70%) even after incubation for more than 12 h with TEV protease. Analytical gel filtration used to separate uncleaved fusion proteins from the target and His-tagged MBP found that they were eluted at void volume with no further cleavage observed even after addition of new TEV protease (data not shown). SDS-PAGE performed to investigate the mobilities and molecular sizes of these uncleaved fusion proteins under denaturing conditions showed that they migrated to a position corresponding to the calculated MW of their individual monomeric forms. The soluble aggregates eluted by gel filtration were capable of binding an amylose affinity column although uncleaved by TEV protease. When TEV proteolysis was further attempted in the presence of 0.1% sodium dodecyl sulphate (SDS), removal of His-tagged MBP was detected. This SDS-induced susceptibility to TEV protease is due to exposure of the TEV protease cleavage site, which indicates there is no damage or mutation during

expression. All together, these results suggest that uncleaved fusion proteins form soluble aggregates in which the His-tagged MBP is exposed outside the shell and the target domains are hidden inside the core. Although more than 90% of the targets were cleaved, approximately 5% were precipitated immediately upon removal of the His-tagged MBP. This outcome is judged as a failed purification. It is unclear at present why such proteins were precipitated since the target scores of their sequences predicted high solubility. It may be that the precipitated targets do not have the proper native conformation or that they require missing cofactors or biologically relevant partners for solubility.

The correlation between percentage of TEV proteolysis and success rate of the 2nd IMAC step was further analyzed and summarized in Table 2. When the percentage of His-MBP fusion protein cleavage was more than 90%, the percentage of targets purified after the 2nd IMAC was about 73%. In contrast, when the percentage of TEV proteolysis was less than 90%, the percentage of target purification was drastically dropped to an average of 16%. TEV proteolysis is the penultimate step just before final isolation of the target. A great deal of effort has been dedicated to the expression and purification of fusion proteins. However, such efforts become futile if cleavage of fusion proteins is not efficient or if targets precipitate upon removal of the solubility partner. Therefore, these results emphasize the requirement for early screening of the small-scale culture in order to identify abortive targets that are poorly cleaved or insoluble following TEV proteolysis. Therefore, any abortive targets would be denied entry into the purification pipeline until the triage methods can rescue them.

Molecular weight vs. purification

The targets subjected to purification ranged from 7 to 60 kDa in molecular size, with an average MW of 27 kDa. The overall success rates of target purification are summarized according to MW; MW (total number of targets, percentage of success), 10 < MW (19, 16), 10 < MW ≤ 20 (226, 41), 20 < MW ≤ 30 (228, 37), 30 < MW ≤ 40 (162, 35), 40 < MW ≤ 50 (70, 38),

Table 1. Statistics for purification steps

Step	No.	Overall success rate (%) [*]	Step success rate (%) [†]
Purification tried	743	100	100
1st IMAC purified	565	76	76
TEV cleaved more than 90%	381	51	67
2nd IMAC purified	342	46	87

^{*}Numbers of products obtained/total number of targets subjected to purification × 100. [†]Numbers of products obtained at each step/total number of targets brought through from the previous step × 100.

Table 2. Correlation between TEV proteolysis and success rate at 2nd IMAC step

TEV proteolysis (%)	No. (%) [*]	No. of target purified after 2nd IMAC	Success rate at 2nd IMAC step (%)
% ≤ 30	110 (16.2)	9	8.2
30 < % ≤ 40	47 (6.9)	1	2.1
40 < % ≤ 50	43 (6.3)	8	18.6
50 < % ≤ 60	31 (4.6)	6	19.4
60 < % ≤ 70	33 (4.8)	3	9.1
70 < % ≤ 80	36 (5.3)	13	36.1
90 ≤ %	381 (55.9)	278	73.0

^{*}Percentage as a function of the total number of targets subjected to TEV proteolysis.

and $50 < MW \leq 85$ (38, 23). The data shows that the purification protocols applied in this report work best for targets with MWs in the range from 10 to 50 kDa. Therefore, new methods are needed to increase the production of small (< 10 kDa) and large (> 50 kDa) proteins. The target selection process in structural genomics generally discriminates against proteins with MWs greater than 60 kDa, which is supported by our purification success data.

Cysteine content vs purification

The number of cysteines per target showed unimodal distribution with an average of 3.5. For targets with a cysteine count greater than 3, *i.e.*, having at least 1 disulfide bond, the overall rate of purification decreased as the number of cysteines was increased (Fig. 1A). However, when the number of cysteines relative to the total number of amino acid residues in a target was considered, the relationship between cysteine content and overall success rate was more clearly observed. The distribution of cysteine content was observed to be unimodal with an average value of 1.6%. For targets with cysteine content greater than 0.5%, the overall success rate of target purification decreased linearly ($r=0.987$) as the cysteine content increased (Fig. 1B). This relationship clearly demonstrates that cysteine content should be used instead of cysteine count in scoring cysteine-rich targets for structural genomics.

Cysteine disulfide bonds play important roles in maintaining the stability and activity of proteins as well as refolding. Hence, *E. coli* has evolved to have cytoplasmic redox control mechanisms as well as periplasmic protein disulfide isomerase (PDI) to ensure correct formation of covalent bonds (12). The observation that the rate of target purification decreases as cysteine content increases is similar to other reports on the production of murine Wint-1 protein (13), human tissue plasminogen (hTPA) (14) and horseradish peroxidase (HRP) (15). These eukaryotic proteins possessing multiple disulfide-bonds accumulate as insoluble aggregates upon overexpression in *E. coli* cells. However, soluble and biologically active Wint-1 can be prepared when expressed as a fusion with an OmpA signal sequence and *c-myc* tag in the cytoplasm followed by trans-

location into the periplasm where an oxidizing environment and PDI are provided (13). Likewise, efficient folding of hTPA was achieved by cytoplasmic expression in *trxB gor* mutant *E. coli* cells, which are defective in maintaining the reduced states of thioredoxins and glutathione (14). Similarly, coexpression of the folding accessory proteins DsbA, DsbB, DsbC and DsbD in the cytoplasm of *E. coli* promotes the production of correctly folded soluble HPR (15). Therefore, specific coexpression techniques along with utilization of redox mutant *E. coli* strains as expression hosts might increase the production of biologically active stable proteins with native disulfide bonds and help rescue cysteine-rich targets that were not successfully purified by current methods.

Isoelectric point vs purification

Since pI has long been a standard measure for distinguishing between proteins, the effect of pI on target purification must be analyzed. A histogram of pI values calculated from the amino acid sequences of targets shows bimodal distribution with the highest modes centered at pI values of 5.5 and 9.5 for acidic and basic targets, respectively (Fig. 2A). The overall success rate of target purification decreased linearly ($r=0.983$) as the pI values of targets increased (Fig. 2B). For basic targets with pI values higher than 9, the overall success rate was about 23% on average, which was only half of that for acidic targets with pI values lower than 5.0.

The most striking observation from our study is that the overall purification of basic targets was much less successful compared to that of acidic targets. The distribution of pI values for proteins from *E. coli*, plants and human revealed major clusters with major peaks at pH 5.5 and pH 9.5 and minor peaks at pH 7.8 and pH 12 (16). Although it is not yet clear whether this multimodal pI distribution reflects the subcellular localization of proteins or rather is the result of discontinuous amino acid ionization values (17), systemic analysis of the pI values of eukaryotic proteins provides important evidence that cytosolic proteins are responsible for the acidic modality and integral membrane proteins for the basic modality (18). In addition, the minor modality centered at pH 12 was presumably

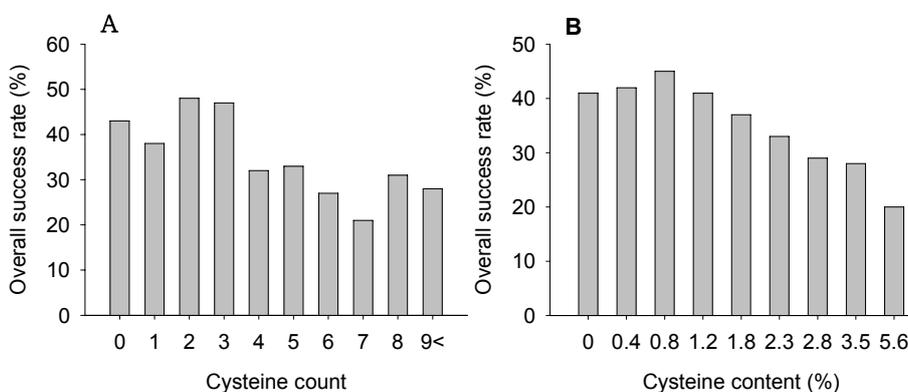


Fig. 1. Overall success rates for each cysteine count group (A) and individual cysteine content group (B).

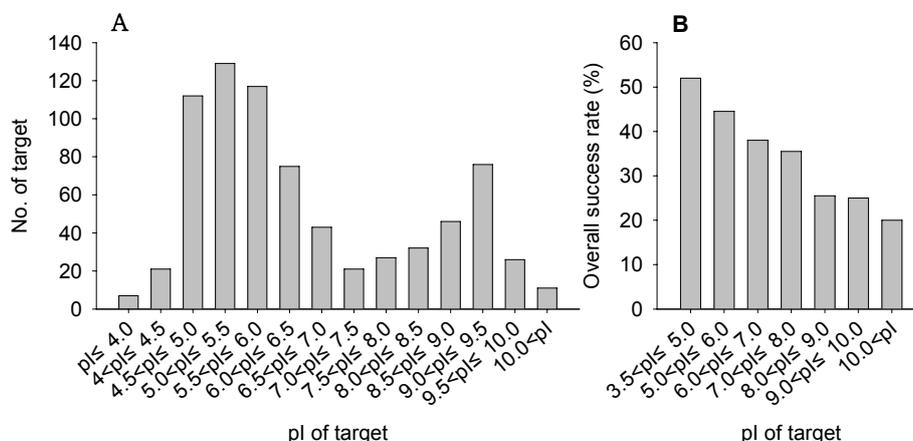


Fig. 2. Histogram of the calculated pI values of targets (A) and overall success rate for each pI group (B). The pI values represented are pI values of target proteins. The calculated pI of His-MBP tag is 5.5.

due to nucleic acid-associated proteins as well as other predicted proteins of unknown function (16). At present, there is no clear explanation for the lower rate of purification of basic targets, especially those with pI values higher than 9. However, based on the fact that membrane proteins are of a basic modality, it is possible that a certain proportion of the basic targets that were subjected to purification include membrane proteins. One can also speculate that basic targets associated with nucleic acids require acidic partners to maintain their stability under the non-native conditions of IMAC capture and TEV proteolysis.

In conclusion, an *E. coli* cell-based system produces a broad range of eukaryotic targets with high yield and purity sufficient for structural analysis (19, 20). However, maximizing the purification of target proteins is still a challenge. Filtering a significant fraction of highly basic proteins from the priority target list as well as optimizing the protocols of tag cleavage are both expected to significantly increase the production efficiency of the *E. coli* expression system.

MATERIALS AND METHODS

Materials

E. coli Rosetta or B834 (DE3)/pLacI/RARE strains were obtained from Novagen (Madison, WI, USA). AKTA purifier with Unicorn 4.12 software, HiTrap chelating HP, HiPrep 26/10 desalting and MBP Trap HP columns were purchased from GE Healthcare (Piscataway, NJ, USA), The 0.8 μ m pore size filters, Centriprep or Centricon, were purchased from Millipore (Billerica, MA, USA). The 2 liter polyethylene terephthalate (PET) beverage bottles were from Ball Corporation (Chicago, IL, USA).

Target selection

Arabidopsis, human, and mouse proteins were scored and prioritized based on a variety of quantitative and predicted characteristics identified by software packages. Characteristics (and the relevant algorithm or tools) considered in categorizing and

scoring targets were homologous structure (BLAST) (21), fragment score (BLAST), external structural genomics target, structure class, transmembrane segments (SCOP classification (22), TMHMM (23), HMMTOP (24), signal peptide (SignalP (25) and TargetP (26)), cysteine count, Pfam score (27), low complexity (seg) (28), new fold prediction, gene chip results (29), disorder prediction (PONDR) (30), solubility prediction and black list. Based on the results of the bioinformatic analyses, each open reading frame (ORF) was allocated a 14 digit target score, and these scores are then used to divide the ORFs into priority tiers. ORFs in the first, second or third tiers were considered prime targets. More than 69,886 unique targets from the *Arabidopsis thaliana*, human and mouse genomes have been evaluated with the target scoring protocol. The target selection criteria and target list are given on the CESG website (www.uwstructuralgenomics.org).

Cloning, expression, purification and laboratory information manage system

The detailed protocols for cloning (31), expression (32, 33) and purification (11) were described in separate papers. Below is a brief outline of the protocols implemented in this study.

The target genes were cloned into a pDONR221 plasmid and then transferred into a pQE80-derived expression vector, pVP13-GW, which allows expression of the target as an N-terminal fusion in combination with (His)₆-MBP. The TEV protease site is present in the linker region between the MBP and target. Isolated targets contain a non-native N-terminal serine residue in place of the normal N-terminal methionine. Constructs were tested for expression and solubility by gel electrophoresis of 0.4 ml cultures prior to large-scale expression. Either *E. coli* Rosetta or B834 (DE3)/pLacI/RARE strains were grown at 25°C for 22-24 h in a 2 liter PET bottle containing 500 ml of Terrific Broth, auto-induction or minimal medium for unlabeled or unlabeled proteins. After harvesting, cells were disrupted by programmed sonication on ice followed by centrifugation at 5,000 g for 20 min. Supernatant was used for

purification of fusion proteins. Prior to loading onto HiTrap chelating HP columns, samples were filtered through a 0.8 µm pore size filter. Automated purification was performed on an AKTA purifier controlled by Unicorn 4.12 software. The purity and concentrations of the targets were determined by SDS-PAGE and absorbances were measured at 280 nm. Purified targets were desalted using a HiPrep 26/10 desalting column against buffers specific to the pI of the target and were concentrated to 10-15 mg/ml using Centrprep or Centricon. Concentrated targets were then drop-frozen as 30 µl droplets for storage until use. The SESAME program was used for data harvesting, storage and management (34).

Acknowledgements

This work was supported by the National Institutes of Health, Protein Structure Initiative Grants P50 GM-64598 and U54 GM074901.

REFERENCES

- Lesley, S. A. (2001) High-Throughput proteomics: protein expression and purification in the postgenomic world. *Protein Expr. Purif.* **122**, 159-164.
- Mus-Veteau, I. (2002) Heterologous expression and purification systems for structural proteomics of mammalian membrane proteins. *Comp. Funct. Genomics* **3**, 511-517.
- Gathmann, S., Rupprecht, E. and Schneider, D. (2006) High level expression of a protein precursor for functional studies. *BMB Reports* **39**, 717-721.
- Chrnyk, B. A., Evans, J., Lillquist, J., Young, P. and Wetzel, R. (1993) Inclusion body formation and protein stability in sequence variants of interleukin-1 beta. *J. Biol. Chem.* **268**, 18053-18061.
- Kapust, R. B. and Waugh, D. S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668-1674
- Norvell, J. C. and Machalek, A. Z. (2000) Structural genomics programs at the US National Institute of General Medical Sciences. *Nature Struct. Biol.* **7**, 31.
- Rodrigues, A. P. C., Grant, B. J. and Hubbard, R. E. (2006) sgTarget: a target selection resource for structural genomics. *Nucleic Acids Res.* **34**, W225-W230.
- Wrobel R. L., Bingman C. A., Jeon W. B., Song J., Vinarov D. A., Frederick R. O., Aceti D. J., Sreenath H. K., Zolnai Z., Vojtik F. C., Bitto E., Fox B. G., Phillips Jr., G. N. and Markley J. L. (2006) Structural Proteomics; in *Plant Proteomics*, Finnie, C. (ed.), pp. 99-128, Blackwell Publishing, Oxford, USA.
- Porath, J., Carlsson, J., Olsson, I. and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598-599.
- di Guan, C., Li, P., Riggs, P. D. and Inouye, H. (1988) Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **67**, 21-30.
- Jeon, W. B., Aceti, D. J., Bingman, C. A., Vojtik, F. C., Olson, A. C., Ellefson, J. M., McCombs, J. E., Sreenath, H. K., Blommel, P. G., Seder, K. D., Burns, B. T., Geetha, H. V., Harms, A. C., Sabat, G., Sussman, M. R., Fox, B. G. and Phillips Jr., G. N. (2005) High-throughput purification and quality assurance of *Arabidopsis thaliana* proteins for eukaryotic structural genomics. *J. Struct. Funct. Genomics* **6**, 143-147.
- Kadokura, H., Katzen, F. and Beckwith, J. Protein disulfide bond formation in prokaryotes. *Annu. Rev. Biochem.* **72**, 111-135.
- Fahnert, B., Veijola, J., Roel, G., Karkkainen, M. K., Railo, A., Destree, O., Neubauer, P. and Vainio, S. (2004) Murine Wnt-1 with an internal c-myc tag recombinantly produced in *Escherichia coli* can induce intracellular signaling of the canonical Wnt pathway in eukaryotic cells. *J. Biol. Chem.* **279**, 47520-47527.
- Bessette, P. H., Aslund, F., Beckwith, J. and Georgiou, G. (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13703-13708.
- Kondo, A., Kohda, J., Endo, Y., Shiromizu, T., Kurokawa, Y., Nishihara, K., Yanagi, H., Yura, T. and Fukuda, H. (2000) Improvement of productivity of active horseradish peroxidase in *Escherichia coli* by coexpression of Dsb proteins. *J. Biosci. Bioeng.* **90**, 600-606.
- Wu, S., Wan, P., Li, J., Li, D., Zhu, Y. and He, F. (2006) Multi-modality of pI distribution in whole proteome. *Proteomics* **6**, 449-455.
- Weiller, G. F., Caraux, G. and Sylvester, N. (2004) The modal distribution of protein isoelectric points reflects amino acid properties rather than sequence evolution. *Proteomics* **4**, 943-949.
- Schwartz, R., Ting, C. S. and King, J. (2001) Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. *Genome Res.* **11**, 703-709.
- Jeon, W. B., Allard, S. T. M., Bingman, C. A., Bitto, E., Han, B. W., Wesenberg, G. E. and Phillips Jr., G. N. (2006) X-ray crystal structures of the conserved hypothetical proteins from *Arabidopsis thaliana* gene loci At5g11950 and At2g37210. *Proteins* **65**, 1051-1054.
- Phillips, G. N., Jr, Fox, B. G., Markley, J. L., Volkman, B. F., Bae, E., Bitto, E., Bingman, C. A., Frederick, R. O., McCoy, J. G., Lytle, B. L., Pierce, B. S., Song, J. and Twigger, S. N. (2007) Structures of proteins of biomedical interest from the Center for Eukaryotic Structural Genomics. *J. Struct. Funct. Genomics* **8**, 73-84.
- Altschul, S. F., Gish, W., Miller, W. and Myers, E. W. (1990) Lipman DJ: Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
- Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536-540.
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567-580.
- Tusnady, G. E. and Simon, I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849-850.

25. Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1-6.
26. Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005-1016.
27. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C. and Eddy, S. R. (2004) The Pfam protein families database. *Nucleic Acids Res.* **32**, D138-D141.
28. Wootton, J. C. and Federhen, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases. *Comput. Chem.* **17**, 149-163.
29. Stolc, V., Samanta, M. P., Tongprasit, W., Sethi, H., Liang, S., Nelson, D. C., Hegeman, A., Nelson, C., Rancour, D., Bednarek, S., Ulrich, E. L., Zhao, Q., Wrobel, R. L., Newman, C. S., Fox, B. G., Phillips Jr., G. N., Markley, J. L. and Sussman, M. R. (2005) Identification of transcribed sequences in *Arabidopsis thaliana* by using high-resolution genome tiling arrays. *Proc. Natl. Acad. Sci.* **102**, 4453-4458.
30. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C. and Brown, C. J. (2000) Intrinsic protein disorder in complete genomes. *Genome Inform.* **11**, 161-171.
31. Thao, S., Zhao, Q., Kimball, T., Steffan, E., Blommel, P. G., Ritters, M., Newman, C. S., Fox, B. G. and Wrobel, R. L. (2004) Results from high-throughput DNA cloning of *Arabidopsis thaliana* target genes using site-specific recombination. *J. Struct. Funct. Genomics* **5**, 267-276.
32. Sreenath, H. K., Bingman, C. A., Buchan, B. W., Seder, K. D., Burns, B. T., Geetha, H. V., Jeon, W. B., Vojtik, F. C., Aceti, D. J., Frederick, R. O., Phillips Jr., G. N. and Fox, B. G. (2005) Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Protein Expr. Purif.* **40**, 256-267.
33. Tyler, R. C., Sreenath, H. K., Aceti, D. J., Bingman, C. A., Singh, S., Markley, J. L. and Fox, B. G. (2005) Auto-induction medium for the production of [U-15N] and [U-13C, U-15N]-labeled proteins for NMR screening and structure determination. *Protein Expr. Purif.* **40**, 268-278.
34. Zolnai, Z., Lee, P. T., Li, J., Chapman, M. R., Newman, C. S., Phillips Jr., G. N., Rayment, I., Ulrich, E. L., Volkman, B. F. and Markley, J. L. (2003) Project management system for structural and functional proteomics: Sesame. *J. Struct. Funct. Genomics* **4**, 11-23.