Biochemical and Biophysical Research Communications 444 (2014) 480–484

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

# Combinatory use of cell-free protein expression, limited proteolysis and mass spectrometry for the high-throughput protein domain identification

Xinchun Shen [a,*], Siyuan Chen [a], Heng Ge [b]

[a] College of Food Science and Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China
[b] The Second Affiliated Hospital, School of Medicine, Xian Jiaotong University, Xian 710004, China

ARTICLE INFO

*Article history:*
Received 23 December 2013
Available online 25 January 2014

*Keywords:*
High-throughput
Cell-free protein expression
Beads-bound protein
Limited proteolysis
Mass spectrometry
Domain

ABSTRACT

The structural domains of proteins have often been identified through the use of limited proteolysis. In structural genomics studies, it is necessary to carry this out in a high-throughput manner. Here, we constructed a novel high-throughput system, which consists of cell-free protein expression and one-step affinity purification, followed by limited proteolysis using a unique new method, referred to "on beads method". All these steps were carried out on 96-well plate formats and completed in two days, even by manual handling. The merits of the new method versus the conventional one are as follows: (1) experimental times are reduced, (2) the sample preparation for limited proteolysis experiments is simplified, and (3) both protein purification and limited digestion can be performed "*in situ*" on the same sample plate. This preparation method is therefore suitable for highly automated, proteolytic analyses coupled to mass spectrometry techniques at a micro-scale protein expression level. The resulting protease-resistant fragments were analyzed by MALDI-TOF-MS and protein domains of 34 mouse cDNA products were identified with this system.



## 1. Introduction

Structural genomics aims at determining a set of protein structures that will represent all the structure units present in the biosphere. These structures can be used as the basis for the homology modeling of the majority of all remaining proteins. Structural analysis provides clues to biochemical functions of unknown proteins [1–3]. Most of large proteins are thought to be a multiple complex of functional domains. For large proteins, structural insight can be obtained by breaking them into smaller segments of amino acid sequences that can fold into native structures, even when isolated from the rest of the protein [4]. The structural domains of a protein have often been identified through the use of limited proteolysis [5–8], since the connections between protein domains are exposed and thus are better protease substrates. Therefore, limited proteolysis has been widely used to isolate domains, not only to identity function [9], but also to create smaller units for structural analysis by X-ray crystallography or, more increasingly, by NMR spectrometry [10]. Moreover, it can also provide effective information for structural and functional analyses of proteins.

On the other hand, high-throughput screening is the current focus in life science research. In the post-genomics arena, the concurrent analysis of hundreds to thousands of proteins is unavoidable [11]. Therefore, in order to minimize time costs, it is necessary to stream-line all contributing steps, including protein expression, purification, and limited proteolysis, if they are to be coupled with mass spectrometry to produce an efficient experimental means of identifying experimental domains of proteins.

Our previous group has been developing and improving the cell-free protein synthesis system [12–16]. The advantages of the cell-free system versus the conventional *in vivo* expression methods are as follows: (1) the proteins encoded by the genes of interest can be directly produced from linear DNA fragments, without cloning them into expression vectors, and (2) all reactions, including PCR and cell-free protein synthesis, can be carried out on 96-well plates and completed in one day. Thus, about one hundred proteins can be produced simultaneously on a multiwell plate. Affinity-tagged proteins have been purified semi-automatically, and it is possible to perform up to 96-purifications simultaneously using the batch system. When proteins are purified using traditional methods, they cannot be directly used in limited proteolysis and MALDI-TOF mass spectrometry experiments because of their high salt and imidazole content. Therefore exchange into a suitable buffer is necessary following protein purification; however this is a

* Corresponding author. Fax: +86 25 85870269.
*E-mail address:* shenxinchun@hotmail.com (X. Shen).

http://dx.doi.org/10.1016/j.bbrc.2014.01.067
0006-291X/

time consuming and tedious process and is a limiting factor in high throughput systems.

Aiming at a large-scale, high-throughput analysis of protein structures and functions, a unique new method (referred to "on beads method") has been developed on a 96-well filter plate. The protein is purified by beads affinity methods and can be directly used in mass spectrometry and limited proteolysis studies. The buffer exchange of purified proteins is avoided using this strategy. The merits of the new purification method versus the conventional one are: (1) experimental times are reduced, (2) the sample preparation for limited proteolysis experiments is simplified, and (3) both protein purification and limited digestion can be performed "*in situ*" on the same multifilter plate. This method is therefore suitable for highly automated, proteolytic analyses coupled to mass spectrometry techniques. In this research, we developed a novel system, which provides a simple, high-throughput strategy for identifying domains of proteins by limited proteolysis in a multifilter plate. Domain boundaries of 34 mouse cDNA products were identified by this system.

## 2. Materials and methods

### 2.1. PCR amplification

We used the two-step PCR amplification strategy for the template production to minimize the length of PCR primers specific for the gene of interest [17]. In the first step, an open reading frame (ORF) or fragment of an ORF was amplified using gene-specific primers. In the second step, a mixture containing a portion of the PCR product, T7 promoter fragment with tag-coding sequence, T7 terminator fragment, and a universal primer were subjected to overlapping PCR. A construct permitting expression of fusion protein under the control of T7 promoter was obtained. The PCR-amplified DNA fragments were used as templates for the expression.

### 2.2. Protein expression and affinity purification

Proteins were expressed from mouse cDNA at 37 °C for 1 h in 30 μl of *Escherichia coli* cell-free translation system [12] in 96-well microplates, and transferred into a multiscreen filtration plate (Durapore MAGV S 22, Millipore) with 20 μl of 50% slurry Ni–NTA agarose beads (QIAGEN). $His_6$-tagged proteins were bound by shaking (20 rpm on a rotary shaker) at 4 °C for 45 min. Following this, 120 μl of binding buffer (50 mM $NaH_2PO_4$, 600 mM NaCl, 15 mM imidazole, pH 8.0) was added to each well. All the following steps, including centrifugations, were also performed in the same 96-well filter plate. The agarose beads were washed three times by resuspending them in wash buffer A (twice, 50 mM $NaH_2PO_4$, 300 mM NaCl, 15 mM imidazole, pH 8.0) and wash buffer B (once, 50 mM $NaH_2PO_4$, 300 mM NaCl, 30 mM imidazole, pH 8.0), shaking for 5 min, and removal of liquid by centrifuging at 500*g* for 3 min at 4 °C (CR21G, Rotor type: R6S, Hitachi). The wash buffer was removed by washing three times with 200 μl of 10 mM Tris–HCl (pH 7.6). Followed by centrifuging at 500*g* for 3 min at 4 °C, 0.5 μl of the 50% slurry washed beads with 10 mM Tris–HCl (pH 7.6) were subjected to MALDI-TOF-MS analysis, and the remaining was analyzed by SDS–PAGE or prepared for limited proteolysis.

### 2.3. MALDI-TOF-MS analyses

Mass spectra were obtained by the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS, Voyager-DE™ STR, Perseptive Bionsystems). Aliquots (0.5 μl) of the 50% slurry washed beads were loaded onto a MALDI sample support, followed by addition of 0.5 μl sinapic acid matrix solution (saturated in 50% acetonitrile, and 0.1% TFA). All samples were analyzed in the linear mode using externally determined calibration constants.

### 2.4. Limited proteolysis and identifying protease-resistant fragments

The washed beads with bound protein obtained above were prepared in 20 μl of the 50% slurry containing 10 mM Tris–HCl buffer (pH 7.6). The limited digestion reaction with trypsin or endoproteases Glu-C (sequencing grade protease, Roche Molecular Biochemicals) was carried out in the same 96-well filter plate at 30 °C for 120 min. Digestion was stopped by adding 5 μl of 0.5% TFA, 20% acetonitrile buffer. Reaction mixtures containing beads were analyzed by MALDI-TOF-MS. The masses of protease-resistant fragments were obtained and identified using the proteins cleavage dialog by Soft GPMAW (Lighthouse data).

## 3. Results and discussion

### 3.1. Constructing a high-throughput limited proteolysis system

We constructed a high-throughput limited proteolysis process as shown in Figs. 1A and 2B described in detail as follows. The cell-free system is more suited to automated and high throughput expression than other systems [13,14,16,18]. The *E. coli* cell-free reaction, with the 2-step PCR-amplified mouse cDNA as the templates, was carried out at 37 °C for 1 h in the 96-well plate. The expressed proteins were transferred into 96-well filter plates containing nickel chelate affinity beads, and were purified with binding and washing steps only. Sizes of expressed beads-bound protein were directly determined by SDS–PAGE and MALDI-TOF-MS. In parallel, the expressed beads-bound proteins were directly used in the limited digestion reaction. The proteolysis with trypsin (or endoproteases Glu-C) was carried out in 10 mM Tris–HCl buffer (pH 7.6) at 30 °C for 2 h in the 96-well filter plate as described in Section 2. After limited digestion, the beads-containing reaction mixtures were analyzed using MALDI-TOF-MS.

### 3.2. Effect of beads on mass spectrometry of proteins on beads

To examine the effect of recording mass spectra of proteins in the presence of affinity beads, experiments were carried out as shown in Fig. 1A. Expressed bead-linked proteins were suspended in 10 mM Tris–HCl buffer (pH 7.6) on the same 96-well filter plate, then directly analyzed using MALDI-TOF-MS. For comparison, expressed proteins were also purified using a traditional method. The expressed proteins were eluted from the beads using an elution buffer of 50 mM $NaH_2PO_4$, 300 mM NaCl, 250 mM imidazole, pH 8.0 as shown in Fig. 1A. Subsequently, the eluted proteins were exchanged into 10 mM Tris–HCl buffer (pH 7.6) on the 96-well filtration plate using an ultrafiltration (UF) membrane with a 10-kDa cut off. Identical mass spectra (A–D shown in Fig. 2) were obtained for proteins purified by both methods. We concluded that the affinity beads had little effect on the mass spectra of the proteins. Low intensity noise appeared in the low molecular region of some spectra, as may be observed in Fig. 2C and D. The problem was pronounced for some proteins and was a feature of both novel and traditional methods. One possible explanation might be the presence of impurities in some protein samples, which the one step purification had failed to remove.

### 3.3. Effect of beads on limited proteolysis and examining optimal size of filter

Additionally, we investigated the optimal amount of trypsin required for limited digestion of bead-linked proteins using the new

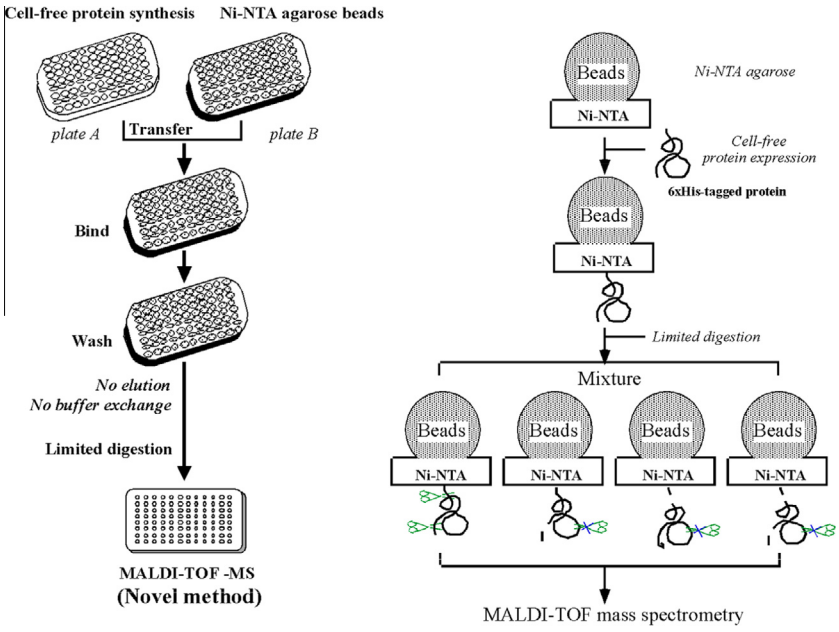


**Fig. 1.** (A) A flowchart of a novel high-throughput protein limited digestion system. Plate A: general multiscreen plate; Plate B: multiscreen filtration plate with a lower protein binding 0.22 μm pore size durapore polyvinylinylidene difluoride (PVDF) membrane filter (Millipore). (B) A model showing the identification of protease-resistant fragments by limited proteolysis. The example shows a protein containing one domain only.

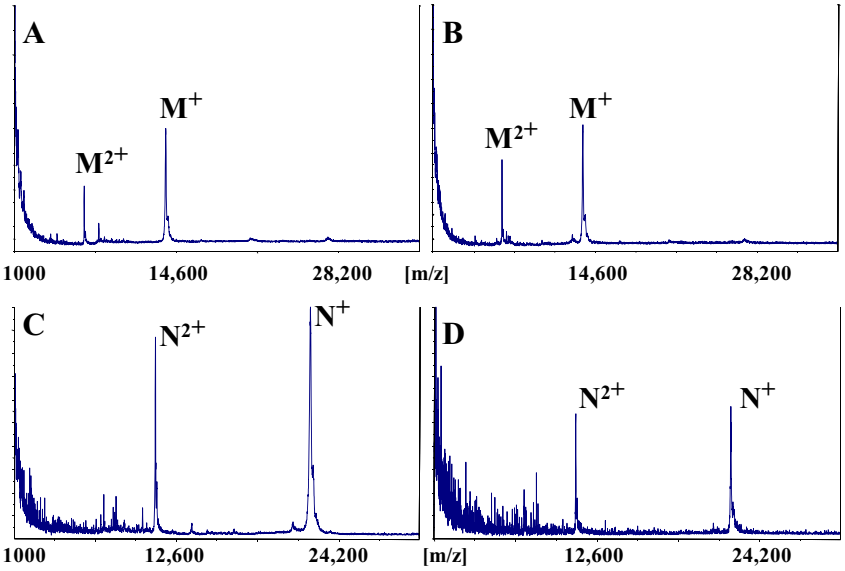


**Fig. 2.** Mass spectrometry of purified protein on beads. $M^+$ (or $N^+$) and $M^{2+}$ (or $N^{2+}$), singly and doubly charged molecular ions of expression products, respectively. Left (A and C): using new method as described in Section 2. Right (B and D): using traditional method (ultrafiltration membrane, UF membrance) with buffer exchange. The buffer exchange of eluted proteins was carried out three times using 200 μl of 10 mM Tris–HCl buffer (pH 7.6) in the 96-well filtration plate with a 10-kDa cut off UF membrane by centrifuging at 3000*g* for 40 min at 4 °C. Masses of $His_6$-tag fusion 8F4 (M) and $His_6$-tag fusion Ras (N) were determined as 13678.98 and 22101.3 by MALDI-TOF-MS, respectively. $His_6$-tag fusion 8F4 (8F4, plate ID, mouse cDNA) is a 126 amino acid protein with a MW of 13676.5; $His_6$-tag fusion Ras (human c-Ha-Ras protein, Kigawa et al. [12]) is a 197 amino acid protein with a MW of 22103.3.

method as shown in Fig. 1A. Fig. 1B summarizes the limited proteolysis procedure adopted in the new method. Following the addition of trypsin to various final concentrations, the reaction mixtures, in 20 μl of 10 mM Tris–HCl buffer (pH 7.6), were incubated at 30 °C for 2 h. Completed reaction mixtures were analyzed using MALDI-TOF-MS. An example is shown in Fig. 3 for the 136 amino acids, $His_6$-tag fusion 2B7 protein (refer to A. 2B7: plate ID, mouse cDNA). $His_6$-tag fusion 2B7 was degraded slowly with increasing concentration of trypsin. Peaks A and B appeared at *m*/*z* 14549.5 and 12537.9, respectively, in the mass spectrum. Peak A represents full length His-tag fusion 2B7 of MW 14547.4. Fragment B of MW 12540.38 and sequence of 19–136 amino acids is a proteolytic fragment resulting from A. Here, we identified the protease-resistant fragment (fragment B) by Soft GPMAW. At the optimal protease: 2B7 ratio, despite the relatively even distribution of theoretical cleavage sites, the trypsin cleaves 2B7 only one site shown in Fig. 3E. This indicated that 2B7 synthesized in cell-free system was one folded domain. The optimal trypsin concentration for the formation of fragment B was 8 ng/μl for 2B7, when the reaction mixture was incubated at 30 °C for 2 h. Overall, our results

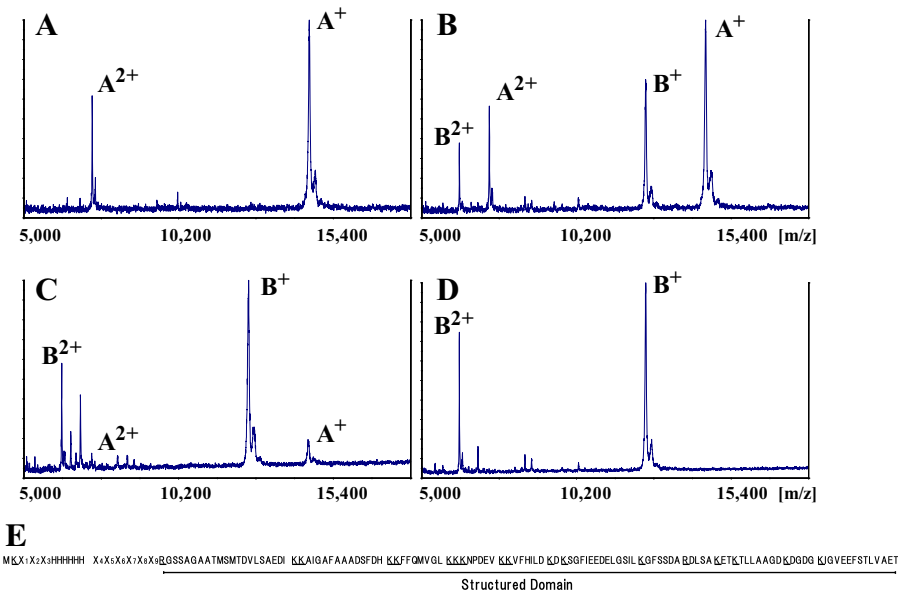


**Fig. 3.** Mass spectrometry analysis of the tryptic digestion of $His_6$-tag fusion 2B7 "on beads". After the addition of trypsin to various final concentrations, the reaction mixtures in 20 μl of 10 mM Tris–HCl buffer (pH 7.6), were incubated for limited proteolysis at 30 °C for 2 h. The reaction mixtures were analyzed using MALDI-TOF-MS. (A) control, no trypsin addition; (B) 0.5 ng/μl trypsin; (C) 6 ng/μl trypsin; (D) 8 ng/μl trypsin. $His_6$-tag fusion 2B7 (2B7, plate ID, mouse cDNA): 136 amino acid. $A^+$ (or $B^+$) and $A^{2+}$ (or $B^{2+}$), singly and doubly charged molecular ions of the protein (or fragment), respectively. (E) Linear representation of $His_6$-tag fusion 2B7 primary sequence: theoretical trypsin cleavage sites are *underlined*. Where *X* is an amino acid.

suggested that the optimal concentration of trypsin for the limited proteolysis of most proteins was in the range 2–8 ng/μl. In contrast, under identical reaction conditions, the optimal concentration of trypsin required to digest proteins prepared via a traditional method was in the range 0.25–1 ng/μl, and the same protease-resistant fragment was obtained by both methods (data not shown). We hypothesize that the presence of affinity-beads sterically hinders proteolytic cleavage of proteins. Moreover, we also investigated the limited digestion of 2B7 by endoproteases Glu-C instead of trypsin. The similar results were obtained, indicating that other enzymes could also be used in the process (data not shown). In these studies several types of 96-well filtration plate assembly, a hydrophilic durapore membrane with a 0.22 μm pore size (rather than 0.65 or 0.45 μm) was used to avoid possible passage of the reaction mixture through the filter during limited proteolysis (data not shown).

### 3.4. Summary of data obtained using the new method

We summarized the experimental results obtained by the new method from Supplemental Table 1 (Fig. 4). Proteins expressed in the cell-free system were purified by nickel chelate affinity beads, and their sizes were determined with SDS–PAGE and MALDI-TOF-MS. The limited digestion reaction with trypsin (and/or endoproteases Glu-C) was carried out at 30 °C for 120 min. For 59 (49%) of the total 120 clones, expressed proteins were detectable using SDS–PAGE after one-step of affinity purification. The molecular masses of $His_6$-tagged proteins, affinity-purified using Ni–NTA agarose beads, were also determined by MALDI-TOF-MS. As a result, Masses of 43 (73%) of the total 59 clones expressed proteins could be obtained by MALDI-TOF-MS; and their masses corresponded with the calculated values from amino acid sequence data.

As shown in Fig. 4, we identified domains of 34 out of a total of 120 expressed proteins by limited proteolysis using our new method. Kuroda et al. [19] have developed an automated domain prediction tool based on sequence similarities. This enables genome-wide searches and has allowed a database of predicted protein domains to be established. Comparing our experimental results with this protein domain databank, two-fifths of the proteins we studied correlated well, while one-third showed partial correlation. Limited proteolysis is a classical strategy to isolate discretely folded domains of proteins taking advantage that the exposed regions between folded domains are the most sensitive protease cleavage sites. Limited proteolysis occurs at "hinges and fringes" such as exposed surface loops, domain linking segments [20], and disordered regions [7]. For a proportion of the proteins investigated, the exposed surface loops in domains were possibly cleaved by trypsin (data not shown), where local unfolding of the α-helix existed. However, in some cases the linker between domains cannot be broken, for example, in large multidomain proteins where the tryptic cleavage sites are buried. Limited proteolysis may therefore be used to monitor surface regions of

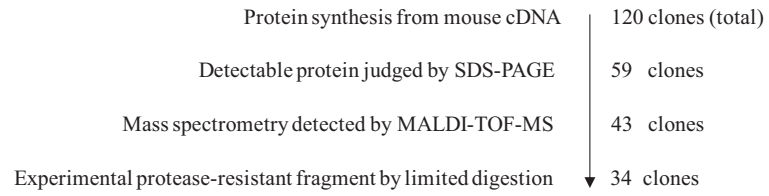


**Fig. 4.** Summary of the data obtained using the new method from Supplemental Table 1. An assessment of the quantity and A judgment as to whether a protein or not was purified based on the presence of a band by SDS–PAGE. The sizes of proteins or protease-resistant fragments were analyzed with MALDI-TOF-MS. All protease-resistant fragments were identified by Soft GPMAW.

proteins, ligand-induced conformational changes, protein domains, disordered regions and protein unfolding/refolding [7,21]. In contrast, a computing tool designed to automatically define domain structures based on sequence similarities is limited and is unable to produce a comprehensive list of domains. We believe that the method of identifying protein domains using limited proteolysis together with mass spectrometry is an effective tool, and will enhance our understanding of protein structural determinants.

### 3.5. Features of the new method

Using the "on beads method", the buffer exchange of purified proteins into a buffer providing suitable proteolytic conditions was avoided, and both purification of proteins and their limited digestion could be performed in the same multifilter plate. This new approach took less time, a typical buffer exchange process (for example, by ultrafiltration) requiring at least 180 min (Fig. 1A). The "on beads method" requires only one general multiscreen plate for protein expression and one multiscreen filtration plate. So it is not only simpler, but also possible to identify domain on a less than microgram level of protein expression. This simplified strategy is advantageous for an automated high-throughput expression, purification, and limited proteolysis system compared to traditional methods.

Altogether, our results suggest the method would be well suited for automation and high-throughput limited proteolysis process. Combining this process with automated MALDI-TOF-MS allows us to identify protein domains more simply and rapidly than previous approaches. Currently, the high-throughput protein expression and limited digestion system for protein domain identification is playing an important role in the protein domain project.

## Conflict of interest

The authors have declared that no conflict of interest exists.

## Acknowledgments

We thank members of Protein Research Group, RIKEN Systems and Structural Biology Center, Yokohama (JAPAN) for their discussion. This work was supported by the National Natural Science Foundation of China (No. 31271983), in part by the RIKEN Structural Genomics/Proteomics Initiative (RSGI) and 2013 Jiangsu University Outstanding Science and Technology Innovation Team.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2014.01.067.

## References

[1] K.Y. Hwang, J.H. Chung, S.H. Kim, Y.S. Han, Y. Cho, Structure-based identification of a novel NTPase from Methanococcus jannaschii, Nat. Struct. Biol. 6 (1999) 691–696.
[2] T.I. Zarembinski, L.W. Hung, H.J. Mueller-Dieckmann, K.K. Kim, H. Yokota, et al., Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics, Proc. Natl. Acad. Sci. USA 95 (1998) 15189–15193.
[3] F. Ataei, S. Hosseinkhani, K. Khajeh, Limited proteolysis of luciferase as a reporter in nanosystem biology: a comparative study, Photochem. Photobiol. 85 (2009) 1162–1167.
[4] I.D. Campbell, A.K. Downing, Building protein structure and function from modular units, Trends Biotechnol. 12 (1994) 168–172.
[5] J.A. Barwell, A. Bochkarev, R.A. Pfuetzner, H. Tong, D.S. Yang, et al., Overexpression, purification, and crystallization of the DNA binding and dimerization domains of the Epstein-Barr virus nuclear antigen 1, J. Biol. Chem. 270 (1995) 20556–20559.
[6] S.L. Cohen, A.R. Ferre-D'Amare, S.K. Burley, B.T. Chait, Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry, Protein Sci. 4 (1995) 1088–1099.
[7] A. Fontana, P.P. de Laureto, B. Spolaore, E. Frare, Identifying disordered regions in proteins by limited proteolysis, Methods Mol. Biol. 896 (2012) 297–318.
[8] R.A. Pfuetzner, A. Bochkarev, L. Frappier, A.M. Edwards, Replication protein A. Characterization and crystallization of the DNA binding domain, J. Biol. Chem. 272 (1997) 430–434.
[9] V. Redeker, S. Lachkar, S. Siavoshian, E. Charbaut, J. Rossier, et al., Probing the native structure of stathmin and its interaction domains with tubulin. Combined use of limited proteolysis, size exclusion chromatography, and mass spectrometry, J. Biol. Chem. 275 (2000) 6841–6849.
[10] R. Fukunaga, S. Fukai, R. Ishitani, O. Nureki, S. Yokoyama, Crystal structures of the CP1 domain from Thermus thermophilus isoleucyl-tRNA synthetase and its complex with L-valine, J. Biol. Chem. 279 (2004) 8396–8402.
[11] Y.P. Shih, W.M. Kung, J.C. Chen, C.H. Yeh, A.H. Wang, et al., High-throughput screening of soluble recombinant proteins, Protein Sci. 11 (2002) 1714–1719.
[12] T. Kigawa, Y. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, et al., Cell-free production and stable-isotope labeling of milligram quantities of proteins, FEBS Lett. 442 (1999) 15–19.
[13] S. Yokoyama, H. Hirota, T. Kigawa, T. Yabuki, M. Shirouzu, et al., Structural genomics projects in Japan, Nat. Struct. Biol. 7 (Suppl.) (2000) 943–945.
[14] S. Yokoyama, Y. Matsuo, H. Hirota, T. Kigawa, M. Shirouzu, et al., Structural genomics projects in Japan, Prog. Biophys. Mol. Biol. 73 (2000) 363–376.
[15] T. Kigawa, T. Yabuki, N. Matsuda, T. Matsuda, R. Nakajima, et al., Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression, J. Struct. Funct. Genomics 5 (2004) 63–68.
[16] M. Aoki, T. Matsuda, Y. Tomo, Y. Miyata, M. Inoue, et al., Automated system for high-throughput protein production using the dialysis cell-free method, Protein Expr. Purif. 68 (2009) 128–136.
[17] T. Yabuki, Y. Motoda, K. Hanada, E. Nunokawa, M. Saito, et al., A robust two-step PCR method of template DNA production for high-throughput cell-free protein synthesis, J. Struct. Funct. Genomics 8 (2007) 173–191.
[18] T. Sawasaki, T. Ogasawara, R. Morishita, Y. Endo, A cell-free protein synthesis system for high-throughput proteomics, Proc. Natl. Acad. Sci. USA 99 (2002) 14652–14657.
[19] Y. Kuroda, K. Tani, Y. Matsuo, S. Yokoyama, Automated search of natively folded protein fragments for high-throughput structure determination in structural genomics, Protein Sci. 9 (2000) 2313–2321.
[20] D.P. Bloxham, L.H. Ericsson, K. Titani, K.A. Walsh, H. Neurath, Limited proteolysis of pig heart citrate synthase by subtilisin, chymotrypsin, and trypsin, Biochemistry 19 (1980) 3979–3985.
[21] S.J. Hubbard, The structural aspects of limited proteolysis of native proteins, Biochim. Biophys. Acta 1382 (1998) 191–206.