STABLE: protein-DNA fusion system for screening of combinatorial protein libraries in vitro

Nobuhide Doi, Hiroshi Yanagawa*

Mitsubishi Kasei Institute of Life Sciences, 11 Minamiooya, Machida, Tokyo 194-8511, Japan

Received 24 June 1999; received in revised form 30 July 1999

Abstract We have developed a new method that permits the complete in vitro construction and selection of peptide or protein libraries. This method relies on an in vitro transcription/translation reaction compartmentalized in water in oil emulsions. In each emulsion compartment, streptavidin (STA)-fused polypeptides are synthesized and attached to the encoding DNA via its biotin label. The resulting protein-DNA fusion molecules recovered from the emulsion can be subjected to affinity selection based on the properties of the peptide portion, whose sequence can be determined from that of its DNA-tag. This method, named 'STABLE' (STA-biotin linkage in emulsions), should be useful for rapid in vitro evolution of proteins and for ligand-based selection of cDNA libraries.

Key words: In vitro transcription/translation; Water in oil emulsion; Streptavidin-biotin complex; Directed protein evolution; In vitro selection; cDNA library

1. Introduction

In vitro selection of biological macromolecules is a useful tool for the study of molecular interactions and the search for therapeutics. During the last decade, applications of in vitro selection have been extended from RNA [1–3] and DNA [4] to peptides and proteins (reviewed in [5,6]). In contrast to RNA and DNA, the in vitro selection of proteins requires the establishment of a physical linkage between each protein and its encoding nucleic acid. This linkage has been achieved by several methods, such as a phage display system [7], 'peptides on plasmids' [8] and cell-surface display systems [9]. However, these methods include a step of transformation of living cells and therefore, the library sizes and the sequence varieties in a library are limited. So far, random sequence protein libraries have been constructed by many researchers (see [10] and references cited therein), but screening for proteins with entirely new functions has not yet been successful because of the limited diversity and complexity of the libraries.

Recently, in vitro selection systems based on cell-free translation have been developed [5]. An in vitro translated polypeptide can be attached to its encoding mRNA through ribosome-mediated complexes (ribosome display) [2,11–15] or through a puromycin derivative ('in vitro virus' [16] or 'RNA-peptide fusion' [17]). These approaches can potentially overcome the limitation regarding the diversity and complexity of libraries. In these methods, however, affinity selections must be performed under RNAse-free conditions. Further, these methods require mRNA without a stop codon and thus cannot be easily applied to the screening of heterologous natural protein libraries such as cDNA libraries.

Here, we describe a new method in which a DNA is attached to the protein which it encodes through a stable linkage in a compartmentalized in vitro transcription/translation system (see Fig. 1). It is known that cell-free translation reactions work effectively in reversed micelles (water in oil emulsions) [18]. Such emulsions can be used as cell-like compartments that keep the DNAs and the proteins they encode together [19]. When an in vitro synthesized polypeptide is directly attached to its encoding DNA in each compartment, the resulting protein-DNA fusion molecules can be recovered from the emulsion and be subjected to affinity selection. We chose the streptavidin (STA)-biotin complex as a connector for the protein-DNA fusion. As a model experiment, a random peptide library was fused to STA by inserting degenerate oligonucleotides at the end of the STA gene in a biotinylated DNA fragment. The linkage between the biotinylated DNA (genotype) and STA-fused peptides (phenotype) allowed for a specific DNA to be enriched from a pool of random sequence DNAs based on the properties of its encoded peptide. We have named this method STA-biotin linkage in emulsions ('STABLE'), because (i) the STA-biotin conjugate is so stable, virtually equivalent to a covalent bond and (ii) DNA is more stable than RNA as a genetic material.

2. Materials and methods

2.1. DNA preparation

The STA gene was amplified from the genome DNA of *Streptomyces avidinii* (obtained from the Japan Collection of Microorganisms in RIKEN) by polymerase chain reaction (PCR) with *Ex Taq* DNA polymerase (Takara Shuzo) and cloned into pUC18, resulting in pUC-STA. A mutant STA gene was amplified from pUC-STA by PCR using a primer containing a N-terminal T7·tag sequence (i.e. the initial 11 amino acids of the T7 gene *10* protein) [20] and digested with *Nhe*I and *Hin*dIII. Similarly, a PCR fragment containing a T7 RNA polymerase promoter was amplified from pET20b (Novagen) and digested with *Nde*I and *Nhe*I. These fragments were simultaneously ligated into the *Nde*I-*Hin*dIII backbone vector pUC19, resulting in pT7-STA. To construct a STA-fused random sequence library (STA-Random; Fig. 2), the STA gene was amplified from pT7-STA by PCR with a sense primer containing a T7 promoter and a reverse biotinylated primer containing a C-terminal random sequence (10 codons of the form NNY: N = A, G, C or T; Y = C or T). A mutant STA gene (STA-His; Fig. 2) was also amplified from pT7-STA by PCR with a reverse biotinylated primer containing an additional hexahistidine-encoding sequence. A *gfp* gene amplified from pND101 [21] by PCR was employed as a carrier DNA for dilution of all DNA used for the in vitro transcription/translations. All DNAs were purified using Recochip (Takara Shuzo) and/or Wizard PCR Preps (Promega).

*Corresponding author. Fax: (81) (42) 724-6317.
E-mail: hyana@libra.ls.m-kagaku.co.jp

Abbreviations: FITC, fluorescein isothiocyanate; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; STA, streptavidin

0014-5793/99/$20.00
PII: S0014-5793(99)01041-8

2.2. In vitro transcription/translation in reversed micelles

In vitro transcription/translation reactions in water in oil emulsions were performed as described by Tawfik and Griffiths [19] with minor modifications. 50 μl reaction mixtures of the *Escherichia coli* S30 extract system for linear DNA (Promega) were prepared on ice according to the manufacturer's specifications, but supplemented with 10 nM carrier DNA, 1 nM DNA library, 40 U of RNAse inhibitor (Toyobo), 100 U of T7 RNA polymerase (Promega) and 0.5% w/v sodium deoxycholate (Nacalai Tesque). This reaction mixture was added to 950 μl of the oil phase at 4°C whilst stirring as described [19] and the mixture was incubated for 1 h at 25°C.

2.3. Enrichment procedure

To recover the protein-DNA fusions, the emulsions were broken as previously described [19], but with quenching buffer (10 mM Tris-HCl, pH 8.0, 1 M NaCl, 10 mM 2-mercaptoethanol, 10 mM imidazole, 1 μM D-biotin). The protein-DNA fusions were captured on Ni-NTA agarose resin (Qiagen). The resins were washed five times with buffer (0.1 M NaH_2PO_4, 10 mM Tris-HCl, pH 8.0, 2.25 M GuHCl, 2 M NaCl, 20 mM 2-mercaptoethanol, 1% Tween-20, 20 mM imidazole). Aliquots of the washed resins were dissolved in 2 M 2-mercaptoethanol and incubated at 95°C for 3 min to free the biotinylated DNA from STA [22]. The DNA was amplified by PCR, cloned into pUC18 vector and sequenced using a CEQ2000 DNA analysis system (Beckman Coulter).

3. Results

3.1. Library design

Biotinylated linear DNA fragments for expressing STA-fused peptides under the control of a T7 promoter were constructed (Fig. 2). We cloned the STA gene from *S. avidinii* by PCR based on its reported nucleotide sequence [23]. At the N-terminus of STA, a T7·tag sequence (MASMTGGQQMG) that is known to enhance the protein solubility [20] was ligated. A random decapeptide library was constructed by fusing a degenerate oligonucleotide at the 3′-end of the STA gene (STA-Random in Fig. 2). We also constructed a STA mutant gene with a C-terminal hexahistidine sequence (STA-His in Fig. 2). The protein expression of this DNA was confirmed in a non-emulsified *E. coli* S30 extract reaction system (Fig. 3).

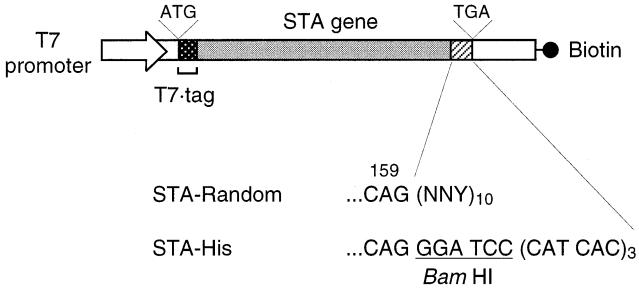

Fig. 2. Schematic diagram of STA genes used in this study.

3.2. STA-biotin linkage in emulsions

Water in oil emulsions were prepared according to the protocol of Tawfik and Griffiths [19]. The size of the emulsion compartments was in the range of 1–5 μm in diameter and the number of compartments was approximately 10^{10}/ml of emul-

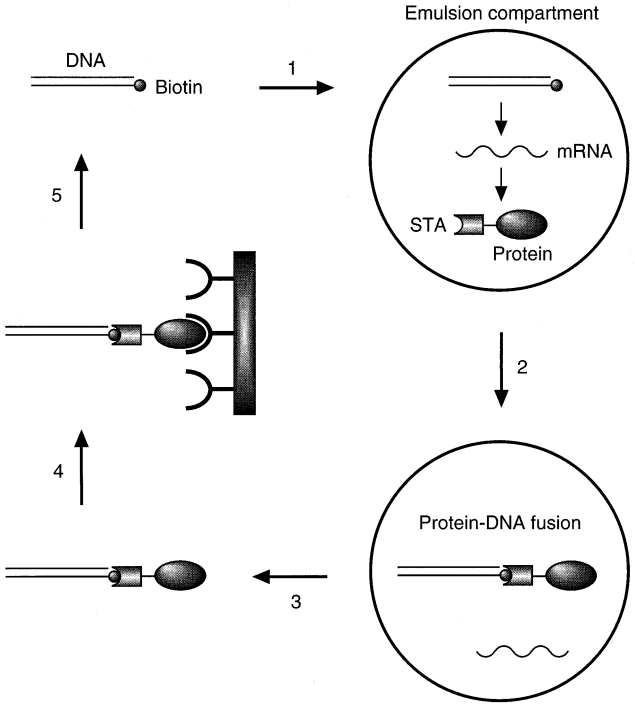

Fig. 1. Schematic representation of the STABLE selection system. (1) A biotinylated DNA library encoding STA-fused random sequence proteins is introduced into an in vitro transcription/translation system. (2) A single DNA molecule compartmentalized in a water in oil emulsion is transcribed and translated in vitro. In each emulsion compartment, the translated protein binds to its encoding DNA through STA-biotin linkage. (3) The protein-DNA fusions are recovered from the emulsion. (4) A desired protein-DNA fusion is affinity-selected by binding of the random sequence protein portion to a target molecule. (5) The tagging DNA is amplified by PCR and used for the next cycle of enrichment.

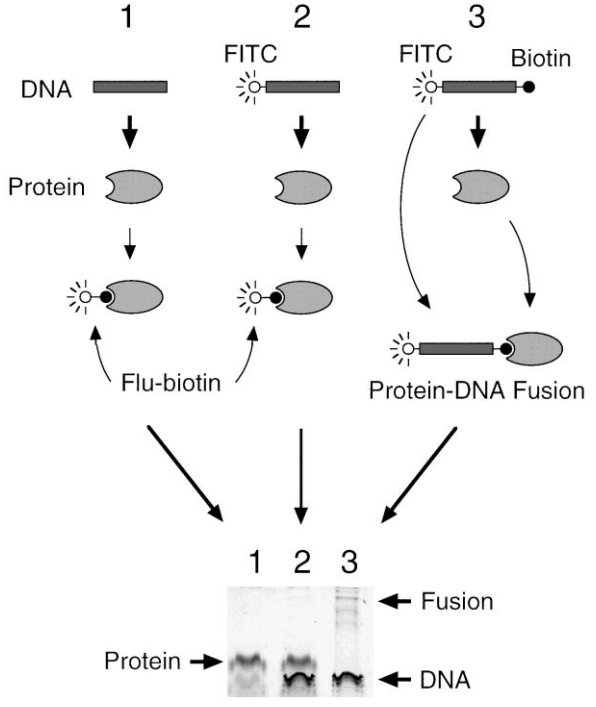

Fig. 3. Expression of STA gene and formation of a protein-DNA fusion. The STA-His DNA (Fig. 2) without a label (lane 1), with a 5′-end FITC label (lane 2) or with both 5′-FITC and 3′-biotin labels (lane 3) was transcribed and translated in vitro. The products were analyzed by 10% sodium dodecyl sulfate-PAGE. The STA-His proteins bound to fluorescein-biotin conjugate (Flu-biotin; lanes 1 and 2), the FITC-labelled DNA (lanes 2 and 3) and the protein-DNA fusion (lane 3) were detected with a FluorImager 595 (Molecular Dynamics).

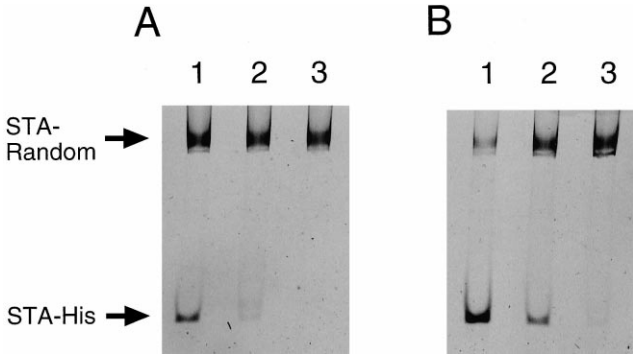

Fig. 4. Enrichment of a hexahistidine sequence (STA-His) from a pool of random sequences (STA-Random). (A) Before and (B) after in vitro selection. The input ratios of STA-His to STA-Random were 1:10 (lane 1), 1:100 (lane 2) and 1:1000 (lane 3).

sified reaction [19]. Since 10^{10} DNA molecules (in 50 μl of S30 reaction mixture supplemented with 1 nM DNA library) were added to 1 ml of emulsified reaction mixture, each compartment was expected to contain on average one DNA molecule. Thus, an in vitro synthesized STA-fused peptide can be linked with the corresponding genetic information (the biotinylated DNA).

As STA forms a tetramer [24], four copies of peptides can be displayed on a STA-biotinylated DNA fusion particle. Since the efficiency of the protein-DNA fusion formation was estimated as 1% of the total DNA (Fig. 3), a large amount of free biotinylated DNA would still remain in the emulsified reactions. To prevent crosslinking of such DNA with the three empty sites of a protein-DNA fusion, a quenching buffer containing free biotin was added before emulsions were broken.

3.3. In vitro selection of random peptide library

We have tested the STABLE selection system using protein-DNA fusions in a model experiment. To estimate the efficiency of enrichment in a selective step, protein-DNA fusions were constructed from a series of dilutions of the STA-His DNA into the STA-Random library in emulsified reactions and were recovered from the emulsions. Aliquots of each mixture of protein-DNA fusions pooled before and after affinity selection with Ni-agarose resin were amplified by PCR with a reverse fluorescein isothiocyanate (FITC)-labelled primer and digested with *Bam*HI, which cleaves only the STA-His DNA (see Fig. 2). The DNA was subjected to 10% polyacrylamide gel electrophoresis (PAGE) and analyzed with a Molecular Imager FX (Bio-Rad). As shown in Fig. 4, STA-His was enriched 10-fold relative to STA-Random after a single round of affinity selection. The STA-His sequence was confirmed by cloning and sequencing of the enriched DNA.

4. Discussion

We have developed a new method that permits the complete in vitro construction and selection of peptide or protein libraries by using a stable linkage of the peptide to its encoding DNA (Fig. 1). This method relies on a cell-free protein synthesis system in reversed micelles. Although Tawfik and Griffiths developed the compartmentalized in vitro system for screening of enzyme catalysts [19], we have applied it for in vitro protein selection based on physical association. In this study, we used STA as a fusion partner of the random peptide library, but many other proteins (e.g. DNA-binding proteins [8]) could also be used as adapters for the protein-DNA linkage.

The advantages of our STABLE selection system compared to other in vitro systems (ribosome display [11–15] and in vitro virus or RNA-peptide fusion [16,17]) are as follows. First, the STABLE method is technically simpler because it does not require reverse transcription steps and a spacer ligation step (required for RNA-peptide fusion). The simpler procedure allows for faster in vitro protein evolution. Further, the genetic information is carried on DNA but not on RNA, which would be less robust with respect to the selection process.

One of the most useful applications of the STABLE method would be ligand-based selection of cDNA libraries. So far, protein-protein interactions have been screened by using yeast two-hybrid systems [25] or phage display systems [26,27]. However, some cDNAs are harmful or lethal when expressed in yeast or *E. coli* [28]. This limitation can be overcome by using totally in vitro systems such as ribosome display and RNA-peptide fusion [5]. In these methods, however, full-length cDNAs cannot be used because they contain stop codons. In contrast, there is no technical difficulty in displaying cDNA libraries on protein-DNA fusions using the in vitro STABLE system. For expression of cDNAs, coupled wheat germ extract or rabbit reticulocyte lysate systems would be more suitable than the *E. coli* S30 extract system used here. Experiments along this line are in progress in our laboratory and the results will be published elsewhere.

Acknowledgements: We thank Drs H. Kinashi and N. Nemoto for technical advice on *Streptomyces* DNA purification and the in vitro experiments, respectively.

References

[1] Robertson, D.L. and Joyce, G.F. (1990) Nature 344, 467–468.
[2] Tuerk, C. and Gold, L. (1990) Science 249, 505–510.
[3] Bartel, D.P. and Szostak, J.W. (1993) Science 261, 1411–1418.
[4] Ellington, A.D. and Szostak, J.W. (1992) Nature 355, 850–852.
[5] Jermutus, L., Ryabova, L.A. and Plückthun, A. (1998) Curr. Opin. Biotechnol. 9, 534–548.
[6] Doi, N. and Yanagawa, H. (1998) Cell. Mol. Life Sci. 54, 394–404.
[7] Smith, G.P. (1985) Science 228, 1315–1317.
[8] Cull, M.G., Miller, J.F. and Schatz, P.J. (1992) Proc. Natl. Acad. Sci. USA 89, 1865–1869.
[9] Georgiou, G., Stathopoulos, C., Daugherty, P.S., Nayak, A.R., Iverson, B.L. and Curtiss III, R. (1997) Nat. Biotechnol. 15, 29–34.
[10] Doi, N., Yomo, T., Itaya, M. and Yanagawa, H. (1998) FEBS Lett. 427, 51–54.
[11] Mattheakis, L.C., Bhatt, R.R. and Dower, W.J. (1994) Proc. Natl. Acad. Sci. USA 91, 9022–9026.
[12] Gersuk, G.M., Corey, M.J., Corey, E., Stray, J.E., Kawasaki, G.H. and Vessella, R.L. (1997) Biochem. Biophys. Res. Commun. 232, 578–582.
[13] Hanes, J. and Plückthun, A. (1997) Proc. Natl. Acad. Sci. USA 94, 4937–4942.
[14] He, M. and Taussig, M.J. (1997) Nucleic Acids Res. 25, 5132–5134.
[15] Hanes, J., Jermutus, L., Schaffitzel, C. and Plückthun, A. (1999) FEBS Lett. 450, 105–110.
[16] Nemoto, N., Miyamoto-Sato, E., Husimi, Y. and Yanagawa, H. (1997) FEBS Lett. 414, 405–408.

[17] Roberts, R.W. and Szostak, J.W. (1997) Proc. Natl. Acad. Sci. USA 94, 12297–12302.
[18] Nametkin, S.N., Kolosov, M.I., Ovodov, S.Y., Alexandrov, A.N., Levashov, A.V., Alakhov, V.Y. and Kabanov, A.V. (1992) FEBS Lett. 309, 330–332.
[19] Tawfik, D.S. and Griffiths, A.D. (1998) Nat. Biotechnol. 16, 652–656.
[20] Gallizia, A., de Lalla, C., Nardone, E., Santambrogio, P., Brandazza, A., Sidoli, A. and Arosio, P. (1998) Protein Expr. Purif. 14, 192–196.
[21] Doi, N. and Yanagawa, H. (1999) FEBS Lett. 453, 305–307.
[22] Jenne, A. and Famulok, M. (1999) BioTechniques 26, 249–254.
[23] Argarana, C.E., Kuntz, I.D., Birken, S., Axel, R. and Cantor, C.R. (1986) Nucleic Acids Res. 14, 1871–1882.
[24] Bayer, E.A., Ben-Hur, H. and Wilchek, M. (1990) Methods Enzymol. 184, 80–89.
[25] Fields, S. and Song, O. (1989) Nature 340, 245–246.
[26] Crameri, R., Jaussi, R., Menz, G. and Blaser, K. (1994) Eur. J. Biochem. 226, 53–58.
[27] Jespers, L.S., Messens, J.H., De Keyser, A., Eeckhout, D., Van den Brande, I., Gansemans, Y.G., Lauwereys, M.J., Vlasuk, G.P. and Stanssens, P.E. (1995) Bio/Technology 13, 378–382.
[28] Allen, J.B., Walberg, M.W., Edwards, M.C. and Elledge, S.J. (1995) Trends Biochem. Sci. 20, 511–516.