PCR-linked in vitro expression: a novel system for high-throughput construction and screening of protein libraries

Suang Rungpragayphan, Hideo Nakano*, Tsuneo Yamane

Laboratory of Molecular Biotechnology, Graduate School of Biological and Agricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Received 24 January 2003; accepted 25 February 2003

First published online 14 March 2003

Edited by Julio Celis

Abstract **A novel entirely in vitro strategy for generation and screening of a combinatorial protein library in an array format has been developed and is experimentally demonstrated. The strategy exploits virtues of PCR and in vitro coupled transcription/translation. Our new approach provides high-throughput construction and screening of the in vitro protein library, and compatibility with various selection methods.**
© 2003 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: Single-molecule polymerase chain reaction; In vitro coupled transcription/translation; Combinatorial protein library; High-throughput screening

1. Introduction

Several cell-based peptide display technologies for directed evolution of protein have been developed and utilized practically in the past decade [1,2]. However, those display technologies have inherent limitations, particularly, on applications with cytotoxic proteins and the achievable library size. A few promising in vitro display technologies which could overcome those limitations have been proposed and practically demonstrated [3–5]. Inevitably, due to the fact that all prominent protein display technologies are based on affinity selection, they are inapplicable to peptides or proteins without known and suitable binding. In addition, maintaining the physical genotype–phenotype linkage throughout the processes requires laborious works on design and optimization.

With great merits of PCR [6–9] and cell-free protein synthesis [10,11] we have recently developed a combination of single-molecule DNA polymerase chain reaction amplification (SM-PCR) and in vitro coupled transcription/translation, which has lately been termed SIMPLEX (Single – Molecule – PCR – Linked in vitro Expression), for high-throughput construction of a protein library that is compatible with various screening methods [12,13]. Briefly, a gene pool is prepared to use as templates for separated single-molecule DNA amplification. Every DNA template in the pool is equipped with all components necessary for the in vitro coupled transcription/translation (Fig. 2A). After the SM-PCR, each PCR products are directly transcribed and translated in vitro yielding a protein library. Uniformities of amounts of SM-PCR products and proteins produced by the SIMPLEX which are vital for quantitative comparison were experimentally confirmed [14]. Its practicability and applicability were successfully demonstrated [14,15]. Even though the size of a protein library generated by the SIMPLEX would be ultimately unlimited, it is currently not practical at low cost. A miniaturized PCR system that could empower the SIMPLEX is not presently available. Therefore, the achievable library size is inevitably constrained by the capacity of existing thermocyclers and requirement of costly DNA polymerase. In this work, we describe a novel high-throughput SIMPLEX-based strategy for construction and screening of combinatorial protein libraries in an array format with expanded capacity and increased cost-effectiveness and its practicability is demonstrated.

2. Materials and methods

2.1. Preparation of DNA templates with random mutation

Two single chain variable fragment (scFv) encoding genes with random mutation (NNS NNS), i.e. anti-HSA-scFv with mutation at G99-A100 of V_H (Fig. 2B) and anti-HSA-scFv with mutation at Q90-W91 of V_L (Fig. 2C), were prepared by combination of conventional PCR and overlapping PCR as previously reported [14]. The original anti-HSA-scFv was from the pRSET-scFv. The DNA templates with random mutation were purified and their concentrations were determined by measuring absorbance at 260 nm.

2.2. Multiplex-PCR and SM-PCR

The DNA templates were typically diluted with 0.1% (w/v) Blue Dextran 2000 (Amersham-Pharmacia) in TE buffer to 1, 5 or 10 molecules per well (1.28×10^{-9} ng of the DNA template was estimated as one molecule). The amplification was carried out in a total volume of 10 μl with 0.5 U of the *Pfu* Turbo™ DNA polymerase (Stratagene), 0.125 mM of each dNTP, and 0.5 μM of a primer, which annealed to both ends of the template, in the cloned *Pfu* buffer. The reaction mixtures were pre-heated at 94°C for 3 min, followed by 70 cycles of: 5 s at 96°C; 5 s at 55°C; 1 min and 20 s at 72°C. All PCRs were carried out by the GeneAmp® PCR system 9700 (PE Applied Biosystems).

2.3. Examination of equality of amplification

Five-molecule PCR products were cloned into the pT7Blue-3 blunt vector (Novagen) and then electroporated into *Escherichia coli* JM109. White colonies were randomly selected and subjected to colony PCR using primers annealing to the vector (M13-RV-N and M13-M1CC). Colony PCR products were treated with the PCR product pre-sequencing kit (Amersham Pharmacia Biotech, Inc.) and sequenced using the Thermo Sequenase™ II dye terminator cycle sequencing premix kit (Amersham Pharmacia Biotech, Inc.). Clones possessing each distinct sequence were counted.

*Corresponding author. Fax: (81)-52-7894145.
E-mail address: hnakano@agr.nagoya-u.ac.jp (H. Nakano).

Abbreviations: SM-PCR, single-molecule polymerase chain reaction; SIMPLEX, single-molecule-PCR-linked in vitro expression; anti-HSA-scFv, anti-human serum albumin single chain variable fragment

0014-5793/03/$22.00 © 2003 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.
doi:10.1016/S0014-5793(03)00251-5

2.4. Anti-HSA-scFv(QW) library construction and screening

A pool of anti-HSA-scFv with random mutation at Q90-W91 of V_L was prepared as described above. Ten-molecule PCR was performed in 96-well plates. The in vitro coupled transcription/translation of all PCR products were carried out in 40 μl reactions with components as previously reported [14], also in 96-well plates. The reactions were incubated at 30°C for 90 min, subsequently put on ice to stop the reaction. Control reactions were performed under the identical condition without the DNA template.

The scFv library was screened by means of ELISA. Briefly, 96-well plates (Beckton Dickinson and Company) were coated with the human serum albumin (HSA) (Nakarai, Japan), and blocked by 3% bovine serum albumin (Sigma). Each cell-free reaction mixture in the protein library was diluted and added into the pre-coated plate. Penta-His horseradish peroxidase (HRP) conjugate (Qiagen) was used as a secondary antibody interacting with the histidine-tag of the anti-HSA-scFv. Finally, orthophenylenediamine (Wako Pure Chemical Industries, Japan) solution, a substrate for the enzyme HRP, was added to each well. The enzymatic reaction was terminated by 2 M H_2SO_4 and ELISA signals were determined by measuring the absorbance at 492 nm by the SPECTRA MAX250 (Molecular Devices). PCR products of positive reactions were collected, purified and subsequently used as templates for the next round of construction and screening, where the SM-PCR is employed instead of the 10-molecule PCR. Sequences of positive clones were analyzed.

3. Result

3.1. SIMPLEX-based library with expanded capacity

In order to increase the achievable size of the SIMPLEX-based library, a PCR of multiple molecules of templates (e.g. 10 molecules), denoted multiplex-PCR, is implemented for an initial construction of the library (Fig. 1). Compared with a library generated by the SM-PCR, a library that is primarily generated by 10-molecule PCR should have its size enlarged 10 times, providing that each molecule of templates is amplified equally. The primary PCR library is directly transformed to a primary protein library by means of in vitro coupled transcription/translation. Thereafter, the protein library is screened. PCR products of positive reactions from the primary PCR library are collected and subjected to the next round of construction and screening, where the SM-PCR is employed in order to obtain every single positive clone.

3.2. Equality of amplification of each template in a multiplex-PCR

According to our strategy, equal amplification of each DNA template in the multiplex-PCR is crucial for library-size expansion. In principle, amplification by PCR occurs in a non-linear manner. When the starting number of each template in a PCR reaction is imbalanced, the quantitative relationship among each kind of DNA in the PCR product is generally not the same as their relation before the amplification. However, in our multiplex-PCR where total number of DNA template is relatively very small and the starting number of every DNA template is equal (i.e. one molecule each), and with the use of a single primer that can effectively prevent by-product formation [9,13] the amplification of every DNA template is expected to be equivalent. In order to investigate the equality of amplification of every DNA template, considering the number of sequencing works, we chose five-molecule PCR as a representative of the multiplex-PCR to be studied on an assumption that the PCR of a greater number of templates will behave correspondingly. Five-molecule PCRs of anti-

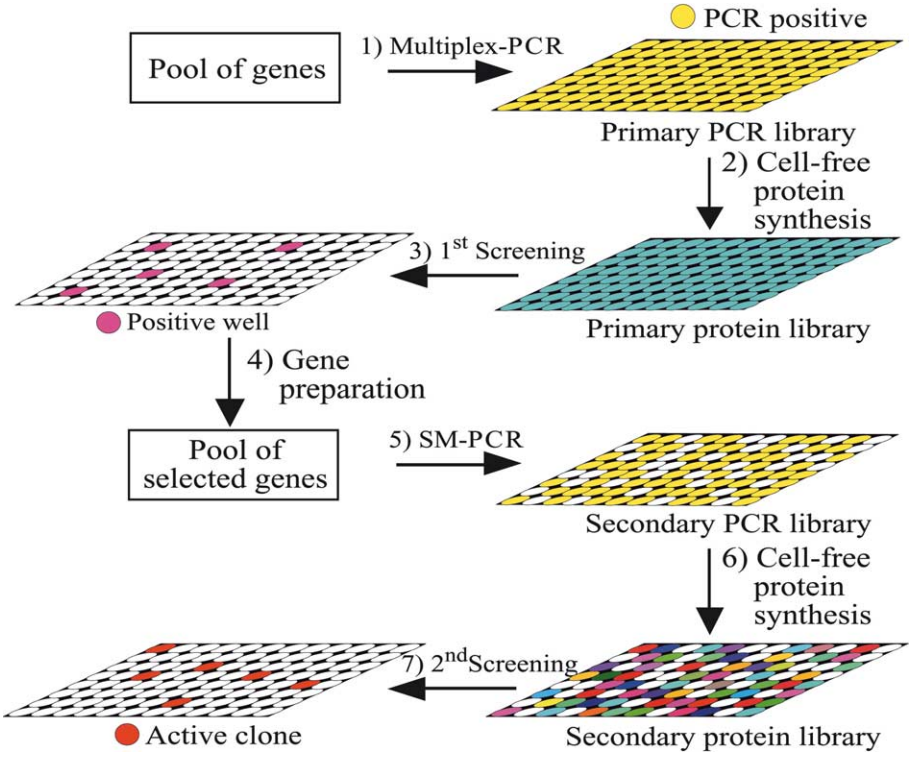

Fig. 1. Schematic representation of SIMPLEX-based protein library construction and screening. 1: DNA templates in the gene pool are diluted to a specified number of molecules per well and amplified by the multiplex-PCR yielding the primary PCR library. 2: The PCR library is converted to the primary protein library by means of cell-free protein synthesis. 3: The protein library is screened for clones with desired properties. 4: DNAs encoding for positive clones are collected, and prepared for the second round of screening where the SM-PCR is performed instead of the multiplex-PCR in order to isolate single genes (5–7).

HSA-scFv genes with random mutation in its light chain at G99 and A100 (NNS NNS) were performed. Each of five PCR products was cloned and transformed. Colonies were picked randomly and their sequences at the randomized sites were analyzed. Results are shown in Table 1. Kinds of observed sequences represent numbers of templates in each five-molecule PCR reaction and number of clones possessing each sequence represents its molecular ratio in the PCR product. From each five-molecule PCR product, an approximately equivalent number of each sequence was observed demonstrating an equal amplification of each template in the reaction. Divergence of all sequences observed indicates diversity of the PCR library. When PCRs from five molecules on average were performed, more or less numbers of templates were predictably observed in some reactions owing to the stochastic dilution.

3.3. Library of anti-HSA-scFv with random mutation

Since there are multiple proteins in one well in the primary protein library, a detected signal in the first screening process is an accumulative value of those molecules. Hence, it is not suitable for a comparative screening for an improvement in some properties of proteins. On the other hand, the strategy is extremely useful for an on–off screening or a screening for new targets. In one of our preliminary experiments, we noticed that binding affinity of the anti-HSA-scFv decreased dramatically when Q90 and W91 of the light chain were randomly mutagenized (data not shown). We thought that one or both of these amino acid residues might be conserved and therefore a library of the anti-HSA-scFv with random mutation at Q90 and W91 was a good example for an on–off screening.

In order to demonstrate a widen application of the strategy, we constructed and screened the library [Lib. anti-HSA-scFv(QW)] aiming at obtaining a clone that has the same sequence with the wild type. With NNS NNS mutation, there were 961 variants in the library. The library was initially generated by 10-molecule PCRs in two 96-well plates, and subsequently transformed to a protein library in the same plate format. After the first screening by ELISA, nine PCR products (average of 90 variants) were selected, mixed and purified. SM-PCRs of selected PCR products were carried out in three 96-well plates. The PCR library was converted to a protein library and screened in the same manner with the first round. Sequences of positive variants were analyzed and the results are shown in Table 2. The clone X1A, which shows the highest ELISA signal, has the sequence at the mutation site identical to the wild type as we have anticipated.

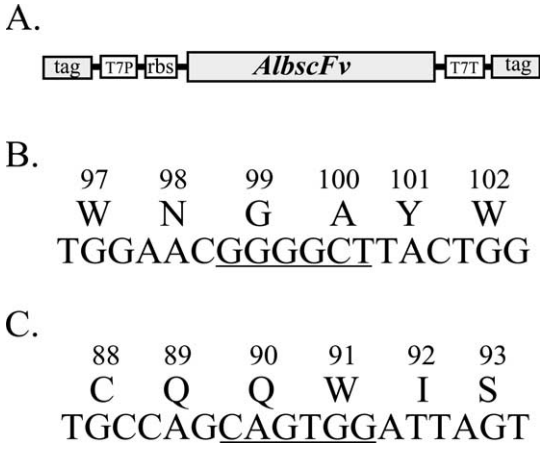

Fig. 2. Design of the DNA templates for experiments. A: Schematic representation of the DNA template ready for amplification by PCR and in vitro coupled transcription/translation. It contains a DNA tag (tag) serving as primer annealing sites at both ends, T7 promoter (T7P), ribosome binding site (RBS), anti-HSA-scFv encoding gene (*AlbscFv*), and T7 terminator (T7T). B: The sequence at the mutation site (underline) in the V_H of the DNA templates that were used in the investigation of equality of amplification in the multiplex-PCR. C: The sequence at the mutation site (underline) in the V_L of the library anti-HSA-scFv(QW).

Table 1
Equality of amplification of each anti-HSA-scFv mutant in five-molecule PCR

PCR product number	Nucleotide sequence	Number of clones
1	TCC GCC	7
	CGC TCA	7
	ATC ACA	6
2	TAC GAC	5
	GTC GCT	4
	GGC GAC	6
	ATC AGC	4
	CCG TTC	5
3	TGC ACG	5
	TAC ACG	4
	TGG CAC	2
	TGC GCG	5
	CCC ACC	5
4	TTC GTC	4
	CCG GTC	5
	TGC ACC	4
	TCC TCA	4
	CAG AAG	4
	ATC ACG	4
5	CCC GCC	4
	ACC TTC	4
	CAG GCC	3
	TAC CAC	3
	AGC CTG	3
	ATG ACC	3

The observed sequences at the randomized site and the numbers of clones that possessed the same sequence are shown. Sequence of the wild type is GGG GCT.

4. Discussion

We have developed a new strategy for high-throughput construction and screening of combinatorial protein libraries based on the SIMPLEX. The uniformities of amounts of both PCR products and proteins produced by the SIMPLEX, which are critical for comparative screening, were previously confirmed [14]. In this work, we have expanded the achievable size of the SIMPLEX-based library by employing the multiplex-PCR. Equality of amplification of every template in the multiplex-PCR that is vital for library size expansion was validated. Regarding the NNS NNS randomization where S represents C or G, there were unexpected As or Ts at the sixth nucleotide of a few observed sequences as the results show in Table 1. We examined flanking sequences of the mutation site and did not observe any spontaneous mutation. Spontaneous mutation rate of the SM-PCR, which is performed in a nearly identical condition with the multiplex-PCR, was reported to

Table 2
Nucleotide and amino acid sequences at the mutation site and ELISA signals (absorbance at 492 nm) of the positive clones from the Lib. anti-HSA-scFv(QW) after the secondary screening

Clone name	Nucleotide sequence	Amino acid sequence	ELISA signal
Ctrl WT	CAG TGG	Q W	0.510
X1A	**CAG TGG**	**Q W**	**0.529**
Z6A	TAC TGG	Y W	0.479
Z7G	TAC TGG	Y W	0.472
X2C	CTC ATG	L M	0.375
Z1B	TCG GTC	S V	0.303
X3D	AGC ATC	S I	0.281
X3F	CTG ACC	L T	0.235
X4A	TGC GTC	C V	0.200
Z1C	ATC TCC	I S	0.129
X3H	TGC AGC	C S	0.123
Z1G	ACC GTG	T V	0.107

be very low [13]. Thus, the disagreement is likely due to poor quality of primers that were used to introduce the random sequences.

By a typical dilution, although number of templates in some PCR reactions may be more or less, the average number of templates per reaction is five. Accordingly, an achievable size of the library generated by multiplex-PCR will be presumably enlarged X times, as the number of templates in the initial multiplex-PCR is X. Considering the efficiency of the SM-PCR, although the template concentration is precise and amplification of one molecule is 100% efficient, only 63% of all reactions will be successfully amplified according to the Poisson distribution, and the other 36% is wasted [14]. As the initial construction, the use of multiplex-PCR where nearly 100% of reaction is effective will substantially increase not only the library size but also the cost–effectiveness.

Generally, the efficiency of amplification by each DNA polymerase depends on sequences of templates and primers. In our system, every template has the same primer annealing site. For a library that the difference among templates is relatively small such as a mutant library or a naive scFv library, amplification of every template in the multiplex-PCR should be balance. However, for a general gene library such as cDNA in which the divergence of sequence may be large, the amplification will be bias, affecting the correlation between the number of templates in the multiplex-PCR and the achievable library size. In spite of this, the use of multiplex-PCR is still helpful for the capacity expansion.

In the SM-PCR, amounts of PCR products are uniform because every reaction reaches a plateau stage even though some might be amplified from multiple molecules [14,16]. As number of genes in the multiplex-PCR increases, amount of each gene in the PCR product decreases, consequently reducing amount of each protein produced in the subsequent step. Therefore, sensitivity of the screening method and the number of molecules in the initial multiplex-PCR must be carefully optimized before construction of the library.

In conclusion, our system has a number of appealing characteristics. It is performed entirely in vitro and the protein generation is highly parallel. The genotype–phenotype linkage is through the position of wells on plates, which is easy to maintain. DNA is more stable than RNA which is used in other in vitro systems. As numerous copies of a gene ($\sim$50 ng/μl of a PCR product) are used as the template for the in vitro expression, a large amount of each protein is produced which enhances the sensitivity of detection. The library can be screened by not only affinity selection but also enzymatic activity. Significantly, it is speedy and completely suits automation. With a precise liquid handling system and an array technology [17–20], this strategy could be a powerful tool for high-throughput construction and screening of various protein libraries.

Acknowledgements: This work was financially supported by a Grant-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science (no. 12450332).

References

[1] Rodi, D.J. and Makowski, L. (1999) Curr. Opin. Biotechnol. 10, 87–93.
[2] Wittrup, K.D. (2001) Curr. Opin. Biotechnol. 12, 395–399.
[3] Amstutz, P., Forrer, P., Zahnd, C. and Plückthun, A. (2001) Curr. Opin. Biotechnol. 12, 400–405.
[4] Roberts, R.W. (1999) Curr. Opin. Chem. Biol. 3, 268–273.
[5] Arnold, F.H. and Volkov, A.A. (1999) Curr. Opin. Chem. Biol. 3, 54–59.
[6] Vogelstein, B. and Kinzler, W. (1999) Proc. Natl. Acad. Sci. USA 96, 9236–9241.
[7] Lagally, E.T., Medintz, I. and Mathies, R.A. (2001) Anal. Chem. 73, 565–570.
[8] Jena, P.K., Liu, A.H., Smith, D.S. and Wysocki, L.J. (1996) J. Immunol. Methods 190, 199–213.
[9] Brownie, J., Shawcross, S., Theakar, J., Whitcombe, D., Ferrie, R., newton, C. and Little, S. (1997) Nucleic Acids Res. 25, 3235–3241.
[10] Jermutus, L., Ryabova, L.A. and Plückthun, A. (1998) Curr. Opin. Biotechnol. 9, 534–548.
[11] Tabuchi, M., Hino, M., Shinohara, Y. and Baba, Y. (2002) Proteomics 2, 430–435.
[12] Ohuchi, S., Nakano, H. and Yamane, T. (1998) Nucleic Acids Res. 26, 4339–4346.
[13] Nakano, H., Kobayashi, K., Ohuchi, S., Sekiguchi, S. and Yamane, T. (2000) J. Biosci. Bioeng. 90, 456–458.
[14] Rungpragayphan, S., Kawarasaki, Y., Imaeda, T., Kohda, K., Nakano, H. and Yamane, T. (2002) J. Mol. Biol. 318, 395–405.
[15] Koga, U., Kobayashi, K., Yang, J., Nakano, H. and Yamane, T. (2002) J. Biosci. Bioeng. 94, 84–86.
[16] Innis, M.A. and Gelfand, D.H. (1990) in: PCR Protocols: A Guide to Methods and Applications (Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J., Eds.), pp. 10, Academic Press, California.
[17] Schober, A., Günther, R., Schwienhorst, A., Döring, M. and Lindermann, B.F. (1993) BioTechniques 15, 324–329.
[18] Nagai, H., Murakami, Y., Morita, Y., Yokoyama, K. and Tamiya, E. (2001) Anal. Chem. 73, 1043–1047.
[19] Khandurina, J. and Guttman, A. (2002) Curr. Opin. Chem. Biol. 6, 359–366.
[20] Battersby, B.J. and Trau, M. (2002) Trends Biotechnol. 20, 167–173.