

## Construction of block-shuffled libraries of DNA for evolutionary protein engineering: Y-ligation-based block shuffling

Koichiro Kitamura<sup>1</sup>, Yasunori Kinoshita<sup>1</sup>,  
Shinsuke Narasaki<sup>1</sup>, Naoto Nemoto<sup>2</sup>, Yuzuru Husimi<sup>1</sup> and  
Koichi Nishigaki<sup>1,3</sup>

<sup>1</sup>Department of Functional Materials Science, Saitama University,  
255 Shimo-Okubo, Saitama 338-8570 and <sup>2</sup>GenCom Co., 11 Minami-Oya,  
Machida 194-8511, Japan

<sup>3</sup>To whom correspondence should be addressed.

E-mail: koichi@fms.saitama-u.ac.jp

K.Kitamura and Y.Kinoshita contributed equally to this work

**Evolutionary protein engineering is now proceeding to a new stage in which novel technologies, besides the conventional point mutations, to generate a library of proteins, are required. In this context, a novel method for shuffling and rearranging DNA blocks (leading to protein libraries) is reported. A cycle of processes for producing combinatorial diversity was devised and designated Y-ligation-based block shuffling (YLBS). Methodological refinement was made by applying it to the shuffling of module-sized and amino acid-sized blocks. Running three cycles of YLBS with module-sized GFP blocks resulted in a high diversity of an eight-block shuffled library. Partial shuffling of the central four blocks of GFP was performed to obtain in-effect shuffled protein, resulting in an intact arrangement. Shuffling of amino acid monomer-sized blocks by YLBS was also performed and a diversity of more than 10<sup>10</sup> shuffled molecules was attained. The deletion problems encountered during these experiments were shown to be solved by additional measures which tame type IIS restriction enzymes. The frequency of appearance of each block was skewed but was within a permissible range. Therefore, YLBS is the first general method for generating a huge diversity of shuffled proteins, recombining domains, exons and modules with ease.**

**Keywords:** block shuffling/diversity/  
evolutionary protein engineering/GFP/library

### Introduction

A decade has passed since the first success in selection-based molecular engineering has been reported (Ellington and Szostak, 1990; Robertson and Joyce, 1990; Scott and Smith, 1990; Tuerk and Gold, 1990). Since then, RNA engineering has rapidly advanced to establish the routine technology for panning aptamers (Conrad *et al.*, 1994; Lorsch and Szostak, 1994) and ribozymes (Dai *et al.*, 1995; Wilson and Szostak, 1995), lending support to the RNA world hypothesis (Gilbert, 1986; Joyce and Orgel, 1999). On the other hand, peptides and proteins of biological functions have become able to be selected from libraries of mutated molecules by way of phage display (Parmley and Smith, 1989; Mikawa *et al.*, 1996), ribosome display (Hanes and Pluckthun, 1997), mRNA display (Cho *et al.*, 2000; or *in vitro* virus, Nemoto *et al.*, 1997) and others (Mattheakis *et al.*, 1994; Buchholz *et al.*, 1998; Kieke *et al.*, 1999). In comparison with the stage that RNA engineer-

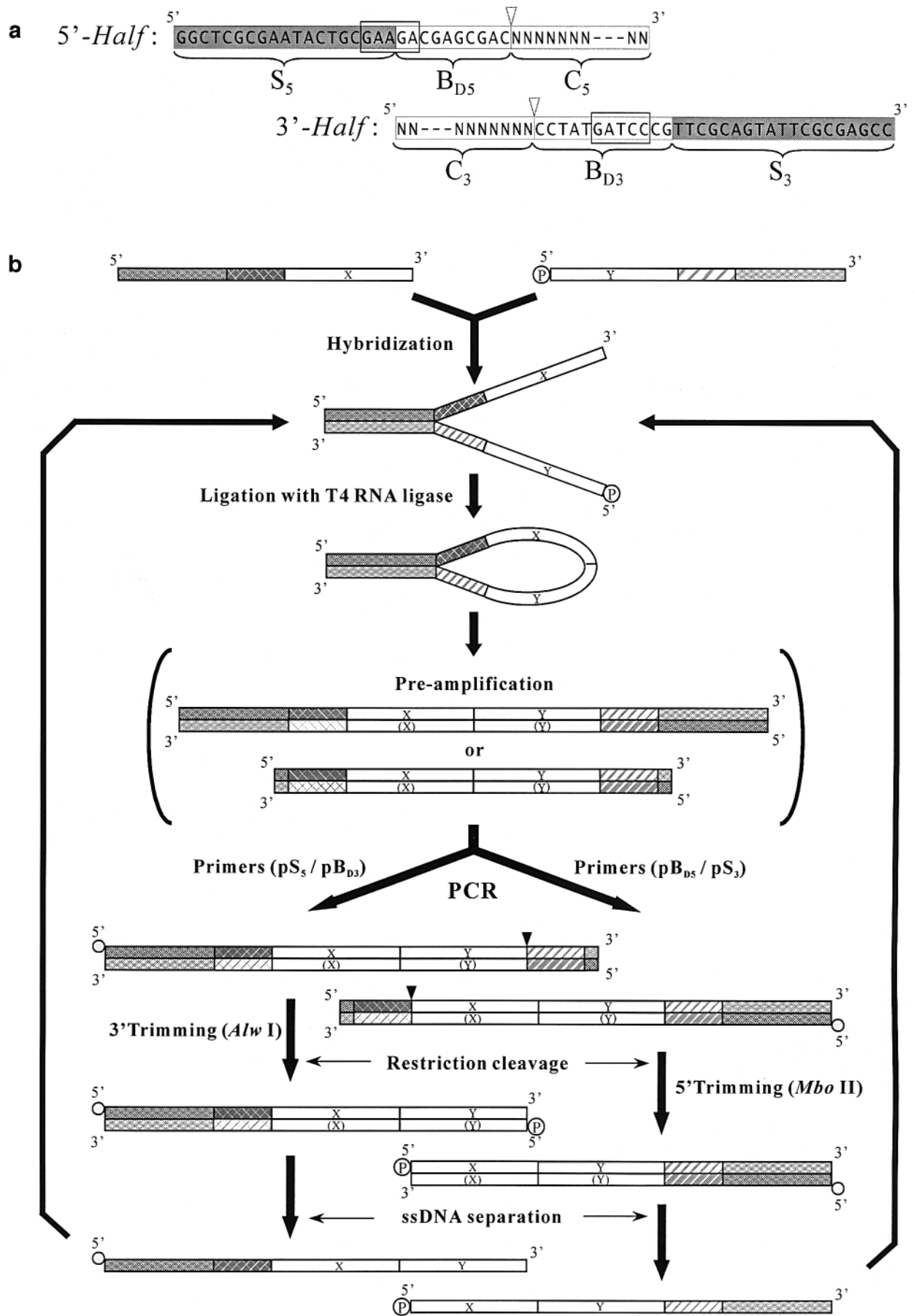
ing has attained, that of protein engineering has a long way to go, partly because the former can be easily evolved as the same molecule can work for both function and information, thus making the evolution of it much simpler and rapid (Wright and Joyce, 1997). The higher complexity of proteins is also a reason for this delayed progress. However, it is now clear that the activities performed by RNA molecules are limited and usually fall far short of that of the equivalent protein as a trade-off for simplicity (Illangasekare *et al.*, 1995). Therefore, it is now time to develop further the technology for directed evolution of proteins, and presented here is a step in this direction.

Recent findings in molecular biology, especially in the field of genome science, have been establishing that proteins have evolved through recombinations such as domain shuffling (Doolittle, 1995), exon shuffling (Kolkman *et al.*, 2001) and module shuffling (Roy *et al.*, 1999). In addition to these sophisticated shuffling mechanisms, recombination itself seems to have contributed to the molecular evolution in the form of general homologous/non-homologous mechanisms and transposable element-mediated mechanisms (Kornberg and Baker, 1992; Fedoroff, 1999). It is impressive that, with a relatively small number of genes (at most 40 000 genes) contained in the whole human genome, humans can generate a highly complex and sophisticated molecular system as a result of alternative splicing (Kondrashov and Koonin, 2001; Li *et al.*, 2001). Consequently, these facts seem to support the idea that block-shuffling mechanisms can mine functional proteins effectively.

Protein engineering can be performed based on the two fundamental mutation technologies: substitution and recombination. The former has been well developed, including site-directed mutagenesis and chemical synthesis methods (Botstein and Shortle, 1985; Sambrook and Russell, 2001a).

In contrast to the well developed substitution technology, the other important technology, recombination, is yet to be exploited. In general, recombination can be fulfilled by two types of technology: ligation by enzymes (Nishigaki *et al.*, 1995; Sambrook and Russell, 2001b) and homology-based PCR just as that used in DNA shuffling (Stemmer, 1994). Both are already used routinely in combining a few DNA fragments (Wakasugi *et al.*, 1997; Tsuji *et al.*, 1999; Kikuchi *et al.*, 2000) or in random-assembling of DNA fragments as in the 'microgene' method (Shiba *et al.*, 1997) and others (Shao *et al.*, 1998; Christians *et al.*, 1999; Riechmann and Winter, 2000). These technologies usually require short stretches of DNA sequences as working sequences (such as recognition sequences of restriction enzymes or homologous sequences for generating priming structures for PCR), thus imposing a constraint on these sequences except in the case of a flush-end ligation (which is governed by chance and gives a very small yield).

In this work, we report a novel technology for block shuffling of DNA, and consequently, block shuffling of protein, based



on Y-ligation [i.e. ligation of blocks with a stem and two branches (Nishigaki *et al.*, 1999)], developing a new field for evolutionary protein engineering. This technology is demonstrated to be applicable to shuffling of blocks of various sizes (amino acid monomer to polypeptide size). The significance of block shuffling for evolutionary protein engineering is also discussed.

## Materials and methods

### General procedures of Y-ligation-based block shuffling (YLBS)

Two types of single-stranded DNAs (5'-half and 3'-half strands, see Figure 1) were prepared as shuffling device sequences. These DNAs contain a single block sequence that is situated either at the 3'- or 5'-end. The strands are made complementary at their stem region and contain a D-branch region as depicted in Figure 1a, which works as a primer-binding site for PCR. The 5'-end of the 3'-half strand should be phosphorylated for a ligation reaction. Equal amounts of the 5'-half and the 3'-half strands (usually 10 pmol each in 10  $\mu$ l) were combined and hybridized through their stem regions. The 3'-end of the 5'-half strand and the 5'-end of the 3'-half strand were ligated using 50 U T4 RNA ligase (Takara, Kyoto, Japan) in the presence of 0.1 mM ATP. After pre-amplification of the ligation products (though it is possible to omit this step), two types of PCR were performed to obtain pre-5'-half and pre-3'-half PCR products. In advance, the primers containing the stem sequence (primers pS<sub>5</sub> and pS<sub>3</sub> in Figure 1) were biotinylated at the 5'-end in order to be used for preparing the ssDNAs for the next Y-ligation cycle. The PCR products were separately digested with the corresponding restriction enzyme (*Mbo*II for 3'-half and *Alw*I for 5'-half) and then ssDNAs were collected through avidin-biotin binding (streptavidin-coated magnetic beads were used). The biotinylated strands of the pre-5'-half and the non-biotinylated strands of the pre-3'-half (Figure 1, top) were then used as the 5'-half strand and the 3'-half strand, respectively, in the next Y-ligation cycle. Depending on the cycle number,  $n$ , the size and the diversity of ligated blocks were allowed to increase exponentially:  $2^n$  ( $= s$ ) for the size and  $d^n$  for the product diversity where  $d$  is the number of block diversity at the start (note that the latter increases in a series of 64, 4096,  $1.7 \times 10^7$ ,  $2.8 \times 10^{14}$  if  $d = 8$ ).

### Oligonucleotides

All oligonucleotides listed in Tables I and II were custom-synthesized [by either Nihon Bio Service (Asaka, Japan), Sawady Technology (Tokyo, Japan) or Amersham Bioscience (Tokyo, Japan)]. The starting materials for YLBS and non-labeled primers for PCR were of cartridge grade. Biotinylated and FITC-labeled primers used for PCR were of HPLC or PAGE grade. The starting materials used for the subtilisin library were prepared by PCR using oligonucleotides (Table II).

### Ligation of 5'-half and 3'-half strands

The 5'-half and the 3'-half DNAs in 10  $\mu$ l of ligation buffer containing 50 mM Tris-HCl (pH 8.0), 10 mM MgCl<sub>2</sub>, 10 mg/l BSA, 1 mM hexamminecobalt (III) chloride, and 25% polyethylene glycol 6000 (Tessier *et al.*, 1986) were annealed through heating at 94°C (5 min) and then at 60°C (15 min). Ligation was then carried out with 50 U T4 RNA ligase (Takara) in the presence of 0.1 mM ATP at 25°C for 16 h.

### Amplification of ligated products by PCR

Ligated products were purified by denaturing polyacrylamide gel electrophoresis (8 M urea, 8% acrylamide; 250 V for 45 min). Requisite bands were excised from the gel, washed with 1 ml of water and finally crushed in 50  $\mu$ l of water using a gel-crushing rod. One microliter of this extract was used for PCR to confirm the success of ligation and led to the preparation of 5'-half and 3'-half strands. PCR was performed using 10 pmol of primers (see Table II) in 50  $\mu$ l of a PCR buffer containing 200  $\mu$ M of each deoxyribonucleotide triphosphate (dNTP), 1 U *Taq* polymerase (Greiner, Tokyo, Japan), 50 mM Tris-HCl (pH 8.7) and 2.5 mM MgCl<sub>2</sub>. The cycle of pre-denaturation (90°C, 2 min), denaturation (90°C, 30 s), annealing (60°C, 1 min) and extension (72°C, 30 s) was repeated for 30 rounds. Additional steps of PCR and gel purification could be used to further purify the products.

### Restriction digestion

The pre-5'-half and the pre-3'-half PCR products were prepared using the ligation products recovered by ethanol precipitation. Each DNA was incubated with 10 U restriction enzyme [*Mbo*II (Takara), *Alw*I (New England Biolabs, Beverly, USA) or *Mbo*I (Takara)] in 10  $\mu$ l of a buffer {*Mbo*II buffer [10 mM Tris-HCl (pH 7.5), 10 mM MgCl<sub>2</sub> and 1 mM DTT], *Alw*I buffer [20 mM Tris-HCl (pH 7.9), 10 mM MgCl<sub>2</sub>, 1 mM DTT and 50 mM KCl] and *Mbo*I buffer [20 mM Tris-HCl (pH 8.5), 10 mM MgCl<sub>2</sub>, 1 mM DTT and 100 mM KCl]} at 37°C for 1 h. The surrounding sequences of the recognition sequences are shown in Figure 1.

### Preparation of ssDNAs from PCR products (pre-5'-half and pre-3'-half)

To each restriction digest (10  $\mu$ l) were added 40  $\mu$ l of water and 50  $\mu$ l of a streptavidin-coated magnetic bead [Dynabeads M-280 Streptavidin (Dyna, Oslo, Norway)] suspension. Beads were pre-treated with 0.1 M NaOH and equilibrated in 0.1 M Tris-HCl (pH 8.0), 0.5 M NaCl and 1% Tween-20 prior to use. The DNA/bead suspension was shaken at room temperature for 1 h. The collected beads were then washed three times with 100  $\mu$ l of 0.01% BSA and treated twice with 25  $\mu$ l of 25% (w/v) ammonium hydroxide at room temperature for 2 min to recover any non-biotinylated ssDNA (Jurinke *et al.*, 1997). After further washing with 100  $\mu$ l of 0.01% BSA, biotinylated ssDNA was recovered by treating the beads twice with 25  $\mu$ l of 25% (w/v) ammonium hydroxide at 65°C for 15 min, subsequently dried *in vacuo*. The biotinylated ssDNA recovered from the pre-5'-half

**Fig. 1.** (a) Y-ligation-based block shuffling (YLBS). Constructs of block shuffling devices, 5'-half and 3'-half. The stem regions (S<sub>5</sub> and S<sub>3</sub>) are shown in gray, while the branch regions (a branch is the remaining device sequence except the stem) are subdivided into a D-branch and C-branch. The typical base sequence for the stem and branch regions are shown, which include a restriction site (boxed). (b) The cycle of YLBS. After hybridization at the stem region, the two molecules (5'-half and 3'-half) are ligated at the tips of the branches by T4 RNA ligase. Then, using two sets of primers, pre-5'- and pre-3'-halves are generated by PCR, which are then cleaved by a restriction enzyme into double-stranded 5'- and 3'-halves. The sense strands (the upper strands in the Figure) are collected through biotin-avidin binding after alkaline denaturation, providing mature 5'- and 3'-halves (ssDNA) to be returned to the starting point. A circle and a circled P attached at the 5'-end denotes a biotin and a phosphate, respectively. The symbols, S<sub>5</sub>, S<sub>3</sub>, B<sub>D5</sub>, B<sub>D3</sub>, C<sub>5</sub> and C<sub>3</sub>, represent 5'-stem, 3'-stem, 5'-D-branch, 3'-D-branch, 5'-code sequence and 3'-code sequence, respectively. Restriction cleavage sites are indicated with arrowheads (actual cleavages are shown with filled arrowheads). If necessary, the pre-amplification step (braced) was inserted, employing the appropriate primers.

**Table I.** Oligonucleotides used for starting blocks in block shufflings

Sequence (5'→3')	Name	Comment <sup>a</sup>
<b>Definition<sup>b</sup></b>		
GGCTCGCGAA TACTGCGAAG ACGAGCGAC	<b>D<sub>5</sub></b>	29 <b>D<sub>5</sub>=S<sub>5</sub>+B<sub>D5</sub></b>
CCTATGATCC CGTTCCGAGT ATTCCGCGAG C	<b>D<sub>3</sub></b>	19 <b>D<sub>3</sub>=S<sub>3</sub>+B<sub>D3</sub></b>
GGCTCGCGAA TACTTTGAA	<b>S<sub>5</sub></b>	19 Stem(5')
TTCAAAGTAT TCGCGAGCC	<b>S<sub>3</sub></b>	19 Stem(3') ( <b>S<sub>3</sub>=S<sub>3</sub></b> )
GATCTCTTTT	<b>B<sub>D5</sub></b>	10 D-branch(5')
TAGGGGATCC CG	<b>B<sub>D3</sub></b>	12 D-branch(3')
<b>Starting 5'-halves for GFP</b>		
<b>D<sub>5</sub></b> -GGCGAAGGTG AAGGTGATGC AACATATGGC	5'Half-A	59
<b>D<sub>5</sub></b> -AAACTGACCC TTAATTTTAT TTGCACCACC	5'Half-B	59
<b>D<sub>5</sub></b> -GGCAAAGTGC CTGTTCCATG GCCAACCCCTG	5'Half-C	59
<b>D<sub>5</sub></b> -GTGACCACTC <b>TGACG</b> TATGG TGTTCACTGC	5'Half-D	59
<b>D<sub>5</sub></b> -TTTTCTAGAT ACCCAGATCA TATGAAACGG	5'Half-E	59
<b>D<sub>5</sub></b> -CATGATTTTT TCAAGAGTGC CATGCCGGAA	5'Half-F	59
<b>D<sub>5</sub></b> -GGTTATGTGC AGGAAAGAAC TATATCTTTT	5'Half-G	59
<b>D<sub>5</sub></b> -AAAGATGATG GGAAGTACAA GACACGCGCG	5'Half-H	59
<b>Starting 3'-halves for GFP</b>		
GGCGAAGGTG AAGGTGATGC AACATATGGC- <b>D<sub>3</sub></b>	3'Half-A	61
AAACTGACCC TTAATTTTAT TTGCACCACC- <b>D<sub>3</sub></b>	3'Half-B	61
GGCAAAGTGC CTGTTCCATG GCCAACCCCTG- <b>D<sub>3</sub></b>	3'Half-C	61
GTGACCACTC <b>TGACG</b> TATGG TGTTCACTGC- <b>D<sub>3</sub></b>	3'Half-D	61
TTTTCTAGAT ACCCAGATCA TATGAAACGG- <b>D<sub>3</sub></b>	3'Half-E	61
CATGATTTTT TCAAGAGTGC CATGCCGGAA- <b>D<sub>3</sub></b>	3'Half-F	61
GTTATGTGC AGGAAAGAAC TATATCTTTT- <b>D<sub>3</sub></b>	3'Half-G	61
AAAGATGATG GGAAGTACAA GACACGCGCG- <b>D<sub>3</sub></b>	3'Half-H	61
<b>Starting 5'-halves for aa-blocks<sup>c</sup></b>		
<b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b> -GGC	5'Half-Gly	32
<b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b> -ATC	5'Half-Ile	32
<b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b> -GAC	5'Half-Asp	32
<b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b> -AAG	5'Half-Lys	32
<b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b> -TCC	5'Half-Ser	32
<b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b> -TGC	5'Half-Cys	32
<b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b> -CCA	5'Half-Pro	32
<b>Starting 3'-halves for aa-blocks<sup>c</sup></b>		
GGC- <b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b>	3'Half-Gly	34
ATC- <b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b>	3'Half-Ile	34
GAC- <b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b>	3'Half-Asp	34
AAG- <b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b>	3'Half-lys	34
TCC- <b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b>	3'Half-Ser	34
TGC- <b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b>	3'Half-Cys	34
CCA- <b>S<sub>5</sub></b> - <b>B<sub>D5</sub></b>	3'Half-Pro	34

<sup>a</sup> Numbers shown denote the size of oligonucleotides. <sup>b</sup> For simplicity, the device (**D<sub>5</sub>**,**D<sub>3</sub>**), stem (**S<sub>5</sub>**,**S<sub>3</sub>**) and D-branch (**B<sub>D5</sub>**,**B<sub>D3</sub>**) sequences are defined and symbolized. <sup>c</sup> Blocks correspond to a single amino acid.

and the non-biotinylated ssDNA from the pre-3'-half were used as 5'-half strand and 3'-half strands for the next cycle, respectively.

#### Cloning and sequencing

The shuffled DNAs of the GFP gene fragments were digested with *Aat*II and *Nsp*V and then cloned into a specified plasmid

vector as described below. Shuffled DNAs for peptide and subtilisin libraries were cloned into plasmid pCR2.1 using a TA cloning kit (TA Cloning Kit Jr. or TOPO TA Cloning; Invitrogen, Carlsbad, CA, USA). Recombinant plasmids were transformed into *Escherichia coli* DH5 $\alpha$  or TOP10 (Invitrogen) by electroporation using Gene Pulser (Bio-Rad, Richmond, VA, USA) or the calcium chloride method (Sambrook and Russell,

Table II. Oligonucleotides used for PCR primers in block shufflings

Sequence (5'→3')	Name	Comment <sup>a</sup>
<b>Definition<sup>b</sup></b>		
GGCTCGCGAA TACTGCGAAG ACGAGCGAC	D <sub>5</sub>	29
GGCTCGCGAA TACTGCGAAC GGGATCATAG G	D' <sub>3</sub>	31
<b>PCR primers for GFP</b>		
GGCTCGCGAA TACTGCGAAG ACGAGCGAC	pD <sub>5</sub> (GFP/Subt)	29
GGCTCGCGAA TACTGCGAAC GGGATCATAG G	pD <sub>3</sub> (GFP/Subt)	31
CTGCGAAGAC GAG	pB <sub>D5</sub> (GFP/Subt)	13
GCGAACGGGA TC	pB <sub>D3</sub> (GFP/Subt)	12
ACTGACAGAA AATTTGTGCC CATTGACGTC ACCATC	GFP-D1	36
GAAGTCAAGT TCGAAGGTGA TACCCTTGTT	GFP-D2	30
GATGGTGACG TCAATGGGCA CAAATTTTCT GTCAGT	GFP-P1	36
AACAAGGGA TCACCTTCGA ACTTGACTTC	GFP-N1	30
<b>PCR primers for amino acid blocks</b>		
GGCTCGCGAA TACTTTGAAG ATCTCTTTT	pD <sub>5</sub> (aa)	29
GGCTCGCGAA TACTTTGAAC GGGATCCCCT A	pD <sub>3</sub> (aa)	31
TTGAAGATCT CTTT	pB <sub>D5</sub> (aa)	15
TTGAACGGGA TCCCCTA	pB <sub>D3</sub> (aa)	17
GGCTCGCGAA TACTTTGAA	pS <sub>5</sub>	19
GGCTCGCGAA TACTTTGGGG ATCTCTTTT	pD <sub>5</sub> (aa-ab)	29
TTGACGAAGA TCCCCTA	pB <sub>D3</sub> (aa-in)	17
<b>Primer pairs for subtilisin<sup>c</sup></b>		
D <sub>5</sub> -CCTGGGGAA TCCCTTATAT / GCTACTTTAC	D <sub>5</sub> -A / a'	49 / 19 [A,75]
D <sub>5</sub> -GTACTTGATA CAGGAGTGGC / CCCAATACGC CATATGA	D <sub>5</sub> -B / b'	49 / 17 [B,156]
D <sub>5</sub> -GTGGCTCCTG GAGCTGAACT / AAATCATGT TGGCAAT	D <sub>5</sub> -C / c'	49 / 17 [C,129]
D <sub>5</sub> -GGAAGTCCTT CTGGGTCTAC / AATAAGACAC CTGC	D <sub>5</sub> -D / d'	49 / 14 [D,69]
D <sub>5</sub> -ATTGGGGCGG CTGGAAACTC / TCTGATCCAT AGCTGC	D <sub>5</sub> -E / e'	49 / 16 [E,138]
D <sub>5</sub> -CTTGAGATTA TGGCCCTGG / GCCATTGACG TACC	D <sub>5</sub> -F / f'	49 / 14 [F,84]
D <sub>5</sub> -TCTCCACATG TTGCTGGGGT / TGAGCCGCAT ACTCAGC	D <sub>5</sub> -G / g'	49 / 17 [G,156]
D <sub>5</sub> -TAATCCCTAA GGATGTACTG / ATATCCCTGT ACTCATA	D <sub>5</sub> -H / h'	49 / 17 [H,72]
CCTTGGGGAA TCCCTTATAT / D' <sub>3</sub> -GCTACTTTAC	A / D' <sub>3</sub> -a'	20 / 50 [A,75]
GTACTTGATA CAGGAGTGGC / D' <sub>3</sub> -CCCAATACGC	B / D' <sub>3</sub> -b'	20 / 48 [B,156]
GTGGCTCCTG GAGCTGAACT / D' <sub>3</sub> -AAATCATGT	C / D' <sub>3</sub> -c'	20 / 48 [C,129]
GGAAGTCCTT CTGGGTCTAC / D' <sub>3</sub> -AATAAGACAC	D / D' <sub>3</sub> -d'	20 / 45 [D,69]
ATTGGGGCGG CTGGAAACTC / D' <sub>3</sub> -TCTGATCCAT	E / D' <sub>3</sub> -e'	20 / 47 [E,138]
CTTGAGATTA TGGCCCTGG / D' <sub>3</sub> -GCCATTGACG	F / D' <sub>3</sub> -f'	20 / 45 [F,84]
TCTCCACATG TTGCTGGGGT / D' <sub>3</sub> -TGAGCCGCAT	G / D' <sub>3</sub> -g'	20 / 48 [G,156]
TAATCCCTAA GGATGTACTG / D' <sub>3</sub> -ATATCCCTGT	H / D' <sub>3</sub> -h'	20 / 48 [H,72]

<sup>a</sup> Sizes in nucleotide. In parenthesis, subtilisin block name and its size. <sup>b</sup> For simplicity, the device sequences (D<sub>5</sub> and D'<sub>3</sub>) are defined and symbolized. <sup>c</sup> Primers used to generate starting blocks (5'-halves and 3'-halves) are shown: forward primers(left) and reverse primers (right) .

2001c). The plasmid DNA was purified from 5 ml cultures using a plasmid extraction kit, Wizard Plus SV Minipreps DNA Purification System (Promega, Madison, WI, USA). DNA sequences were determined using a DNA sequencing kit, Thermo Sequenase fluorescently-labeled primer cycle sequencing kit with 7-deaza-dGTP (Amersham Pharmacia Biotech, Little Chalfont, WI, USA) and a DNA sequencer, DSQ2000 (Shimazu, Kyoto, Japan). In a few cases, custom-sequencing (Sawaday Technology) was adopted.

#### Vector preparation of the GFP library

Plasmid pGFPuv4-NF, which contains a non-fluorescent GFP gene coding sequence, was prepared by restriction digestion with *Nco*I (Takara) followed by treatment of pGFPuv4 (Ito *et al.*, 1999) with T4 DNA polymerase (Takara) and re-ligation with T4 DNA ligase (Takara). The expression vector was prepared from plasmid pGFPuv4-NF by PCR using GFP-D1 and GFP-D2 (10 pmol each, Table II) as primers. These primers contain *Aat*II and *Nsp*V restriction sites and are designed so as not to

modify the open reading frame of the GFP gene. PCR assays consisted of 30 cycles of denaturation (94°C, 30 s), annealing (55°C, 30 s) and extension (72°C, 2.5 min) after pre-denaturation (94°C, 2 min). The PCR products were finally treated with *Aat*II (New England Biolabs) and *Nsp*V (Takara) and purified using a QIAquick column (QIAGEN, Hilden, Germany).

#### Selection of GFP library

Block-shuffling products of four blocks (C–F) were attached with blocks A + B (at the upstream of the four blocks) and blocks G + H (downstream of them) through PCR procedures. For cloning into an expression vector (see above), the products were attached with linkers by PCR using primers, GFP-P1 and GFP-N1, which contain restriction sequences for *Aat*II and *Nsp*V, respectively, to fit with the expression vector. Then, they were treated with the restriction enzymes *Aat*II and *Nsp*V and incubated in a solution containing 4 fmol of expression vector and 4 Weiss units of T4 DNA ligase (Takara) at 16°C for 18 h. The ligation mixture was subjected to electroporation with 80  $\mu$ l of a competent cell (*E. coli* strain DH5 $\alpha$ ) suspension using an electropulser (Bio-Rad) under the conditions of 1.8 kV and 4 ms of pulse width. To these was added 2 ml of SOB medium. The samples were incubated at 37°C for 1 h, then plated in 1 ml aliquots onto an A4 size LB agar plate (210 $\times$ 280 mm) which contained 100  $\mu$ g/ml ampicillin, and incubated at 37°C for 18 h. The culture plates were analyzed using a fluoroimager, Molecular Imager FX (Bio-Rad, Hercules, CA, USA) using wavelengths of 488 nm (excitation) and 515–545 nm (emission).

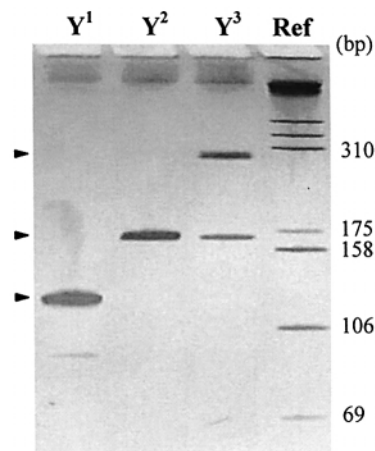
#### PAGE and temperature-gradient gel electrophoresis (TGGE)

Denaturing gel electrophoresis (8 M urea, 8% acrylamide) was performed at 60°C (at 250 V for 45min). The running buffer contained 40 mM Tris-acetate (pH 8.0), 20 mM sodium acetate and 2 mM EDTA. To evaluate the diversity of shuffled DNAs, PCR products were analyzed using TGGE. Gels consisted of 8 M urea and 6% acrylamide (acrylamide: bis-acrylamide, 19:1) in 1 $\times$  TBE buffer [90 mM Tris-borate and 2 mM EDTA (pH 8.0)]. TGGE was performed under a temperature gradient of 20–70°C using Thermo Gradient TG (TAITEC, Saitama, Japan). DNA bands were detected with silver staining.

## Results and discussion

### Shuffling of module-sized blocks (30 nucleotides)

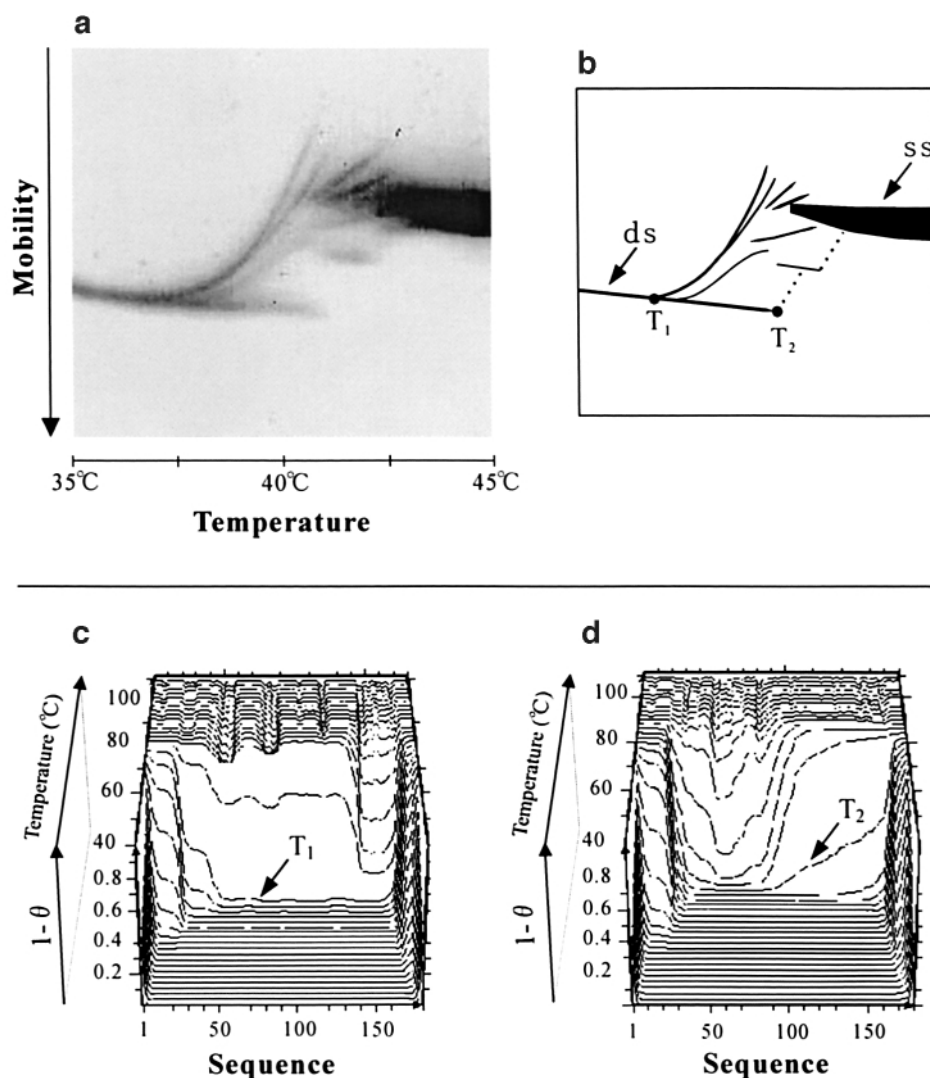
**Products of each intermediate step and their diversity.** A serial procedure of YLBS (Figure 1) was devised and performed on GFP blocks (see Table I) following the protocols described in Materials and methods. GFP was selected because of its fluorescent nature which is useful for screening; it was divided into eight blocks of 30 nucleotides (Table I). The ligation products of GFP blocks at each step ( $Y^1$ – $Y^3$ ) were obtained in a yield of approximately 5–50%, extracted from a gel and then PCR amplified (Figure 2). [Note that the amount of the seed DNA for PCR extracted from a gel, which was estimated to be more than 1 fmol ( $\sim 10^9$  molecules), exceeds the whole diversity of  $Y^3$  ( $= 8^8 = 1.7 \times 10^7$ ) in this case.] A population of shuffled DNAs could be purified to an apparently single band in gel electrophoresis (except  $Y^3$  in this case, which needed another purification step) due to the same degree of polymerization. The diversity of the products at each step can be readily monitored by using TGGE as shown in Figure 3. The complicated transition pattern indicates that the apparently single band consists of various DNA strands. From the range of the initial transition of these strands, we can make a rough estimation of the diversity



**Fig. 2.** Ligation products of GFP blocks at each YLBS stage ( $Y^1$ – $Y^3$ ). Ligation products at each stage ( $Y^i$  strands for the  $i$ -th Y-ligation product) were electrophoretically purified and PCR amplified, resulting in a discrete band (indicated with an arrowhead). The minor band appearing in the lane for  $Y^3$  is a contaminant derived from the  $Y^2$  product and can be ruled out by not selecting here. Each Y-ligation product has a construct of stem (5')/D-branch (5')/C-branch (5')/C-branch (3')/D-branch (3')/stem (3') as shown in Figure 1. Thus,  $Y^1$  consists of 120mers (19/10/60/60/12/19);  $Y^2$ , 300mers (19/10/120/120/12/19), respectively (see Figure 1 and Table 1 for details).

of the shuffled DNAs as shown in Figure 3. Namely, a possible DNA sequence which has the lowest or highest melting temperature among the shuffled DNAs was predicted, mainly based on G+C content, and then subjected to a prediction program for the DNA melting profile (Steger, 1994) to obtain the theoretical transition temperatures. Though the values obtained are known to deviate from the true ones with a definite relationship (Nishigaki *et al.*, 1984, 2000; Abrams and Stanton 1992), the relative values are sufficiently reliable (Nishigaki *et al.*, 1984). The difference between the highest and the lowest melting temperatures,  $\Delta T (= T_2 - T_1)$ , for a population of shuffled DNAs, could be theoretically obtained as 4.0°C, and was comparable with the experimental result (4.8°C) obtained with TGGE (Figure 3). As theoretically expected, the transition profile of shuffled DNAs can be taken as a composite of many transitions occurring at various temperatures, providing a qualitative confirmation of diversity. Thus, TGGE is a convenient, though less accurate, way of diversity monitoring.

**Properties of the library obtained.** The sequences of the shuffled DNAs were determined after cloning as shown in Figure 4a. This analysis finally confirmed the success of block shuffling for eight blocks. In Figure 5, the frequency of appearance of each block is shown in two modes; independent appearance mode and neighboring mode. It is evident that all blocks are appearing within reasonable frequencies (1.3–29.4%) and the neighboring frequencies are consistent with those which are combinatorially expected from the frequencies of each block: the neighboring of E-E is the highest and so on. Basically, a similar tendency was observed with the other confirmation experiments (data not shown), meaning that a different ratio of blocks as starting materials may be necessary to obtain the even mole ratio of blocks in the final products. This deviation problem is an important issue from the technological viewpoint since it is related to the size of diversity. Nevertheless, we think the library thus provided can be used as it is (without changing the starting mole ratio) for most panning purposes since species of the lowest frequency, if not zero at the initial stage, can be relatively increased through selection processes to the level



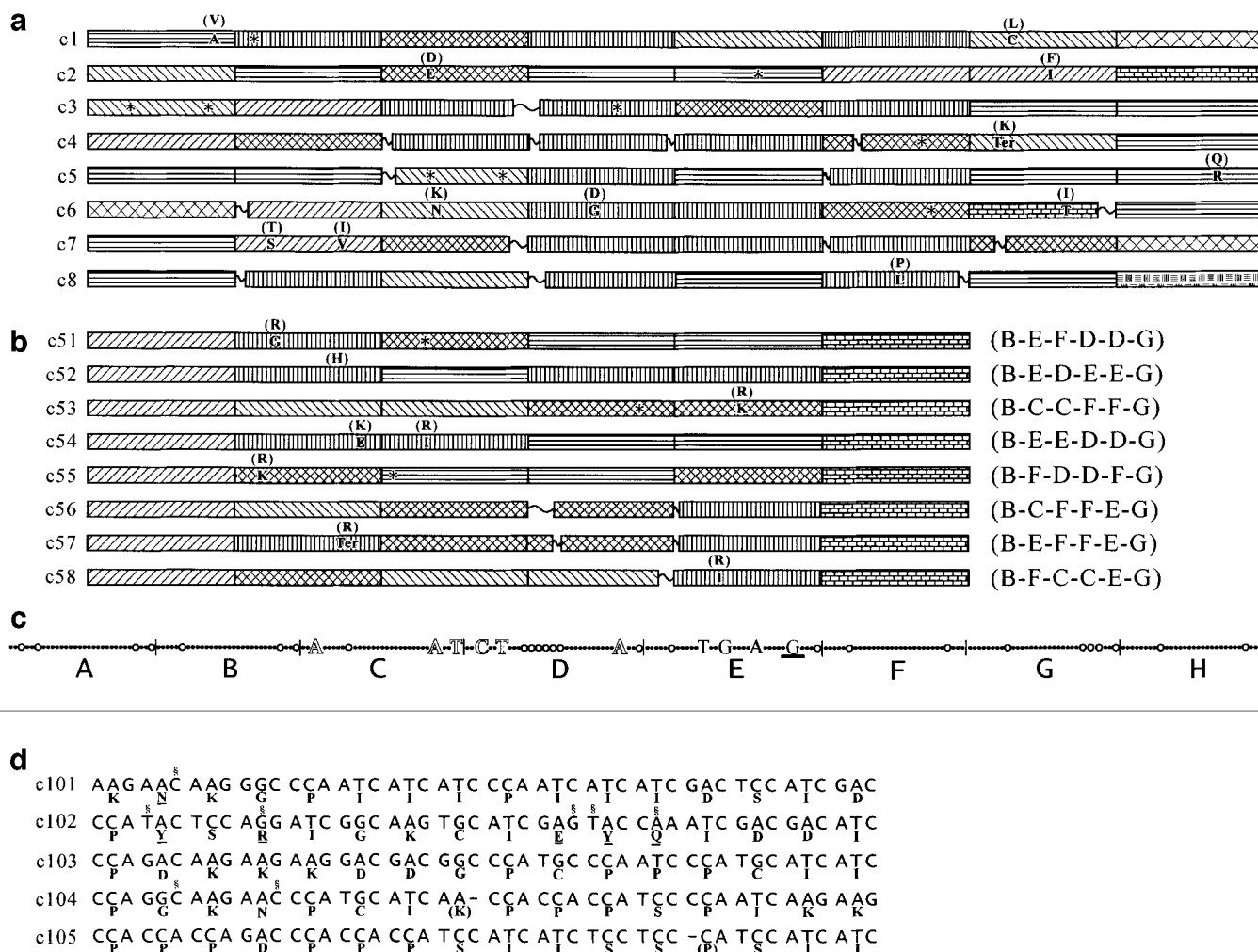
**Fig. 3.** Diversity analysis of ligation products by TGGE. (a) TGGE analysis. Ligation products, which contain a population of shuffled DNAs, were subjected to electrophoresis with a temperature gradient of 20–70°C. (b) A schematic presentation for the composite transitions observed in Figure 3a.  $T_1$  and  $T_2$  stand for the lowest and the highest initial transition temperatures from a double-stranded to a single-stranded state, respectively. (c) The melt map of the DNA (made of blocks A-E-A-F) which is estimated to be the lowest initial melting temperature. (d) The melt map for the DNA of the highest initial melting temperature (B-G-G-B). Each axis of the 3D representation denotes: position of the nucleotide along the DNA sequence (left to right), temperature (front to back), and  $1 - \theta$  where  $\theta$  is the helix probability ( $0 \leq \theta \leq 1$ ) (Steger, 1994). At  $T_1$  and  $T_2$ , melting occurs since the drastic change in helix probability can be detected in the graph. [These temperatures theoretically obtained are known to be correlated with those experimentally obtained in a fixed relation (Nishigaki *et al.*, 2000).]

of their selection (fitness) value (Husimi *et al.*, 1982; Voigt *et al.*, 2000).

Close inspection of the sequencing results reveals that there are deletion/insertion and point mutations involved. Table III shows the statistic scores for the phenomena involved in this experiment [together with the result of four-block shuffling performed using a subset (C, D, E and F) of the same eight blocks]. The deletion rate is remarkably high (49.4%) in this eight-block shuffling. Therefore, this point was intensively investigated and led to an affirmative conclusion discussed later (see Solution for deletion problems). Several technical alterations such as reducing PCR events and making the reaction time a minimum were employed in the whole procedure of four-block shuffling (Figure 4), drastically reducing the per-event deletion rate (49.4–10.6%) (Table III). Although this score cannot be said to be sufficient as it is from the engineering viewpoint, the poorer score (5.7%) for the population of deletion-free has already reached the level ( $=3.4 \times 10^9$ ) sufficient for the whole

diversity ( $=1.7 \times 10^7$ ) assuming pmole ( $=6 \times 10^{11}$ ) or more to be dealt. Of course, this is the simplest estimation. Deletion and insertion alterations often lead to a frame-shift and contribute to an increase in the diversity of the library just as in primordial B cells. The low level of point mutation caused by PCR has a negligible effect on the property of the block-shuffled library since it walks only a step or so in the sequence space while any pair out of the shuffled DNAs are far more separated.

*Expression of the shuffled GFP proteins.* To check the effectiveness of block shuffling, four blocks (blocks C, D, E and F) cut out of the GFP gene were shuffled and integrated into the remaining moiety of the GFP gene and then processed to expression in bacterial cells of *E. coli*. Out of  $10^6$  colonies (10 A4 plates) examined, only a single clone turned out to be fluorescent. The DNA sequence of this fluorescent protein was analyzed, confirming the genuine block-shuffled product (Figure 4c). This result is supported by the fact that the probability of finding the same number of point mutations at the specific sites where base



**Fig. 4.** Results for YLBS. (a) GFP-derived eight blocks, (b) GFP-derived four blocks, (c) a block-shuffled fluorescent product, and (d) aa-blocks. Silent point mutations are shown with an asterisk and a one letter amino acid abbreviation is used to show the mutations that caused amino acid substitution (the original amino acid is shown in parenthesis) in (a) and (b). *Ter* stands for a termination code assigned assuming each block is independently decoded. (Throughout the decoding the frame-shift effect was not considered.) Different boxes designate different blocks, A–H. Deletions are shown with curved link lines. In the case of four-block shuffling (b), the blocks, C, D, E and F, were selected for shuffling. The resultant DNAs of four linked blocks were integrated to offer the construct  $ABX_1X_2X_3X_4GH$  where  $X_1$ – $X_4$  are one of C, D, E or F. Only B and G blocks at both sides are shown here (alignments are shown beside the block representation). (c) Those 30 nucleotides which are intentionally substituted with the other nucleotide as signatures of blocks, are shown in circles and do not change the amino acid coding. Point mutations which occurred during the shuffling experiment are shown in capitals and are silent except one underlined conversion from AAA (Lys) to AGA (Arg) at the 238th amino acid residue of GFP. The hemmed figures represent the nucleotides mutated from the signature ones. (d) Five C-branch sequences out of the  $Y^4$  library. Point mutation and deletion are shown by a chapter mark (§) and a hyphen, respectively. Amino acids are shown as a one letter abbreviation.

substitutions were intentionally introduced as a signature (30 nucleotides in total), was too small ( $1.5 \times 10^{-68}$ ) for it to occur by chance [to obtain this probability, the rate of point substitution was assumed to be  $1.5 \times 10^{-4}$ /replication/nucleotide (Table III)]. The fluorescent product was examined to have half intensity and the same spectrum of GFPuv4 fluorescence (Ito *et al.*, 1999). This must be the result caused by a single amino acid substitution (Lys to Arg) at the 238th amino acid residue of the GFP protein. The ratio of fluorescent products was unexpectedly low (1 against  $10^6$ ) since the product of the original block arrangement could be expected at a frequency of 2000 against  $10^6$  colonies, considering that the number of diversity was  $256 (=4^4)$  and 43.8% of the product was deletion-free (Table III). This may be interpreted as being a result of the uneven frequency of appearance of each block, which makes a full set of blocks less probable, or the susceptibility of GFP fluorescence to point mutations of five to 10 amino acid residues per shuffled product on average (Tsien and Prasher, 1998). Furthermore, it may reflect

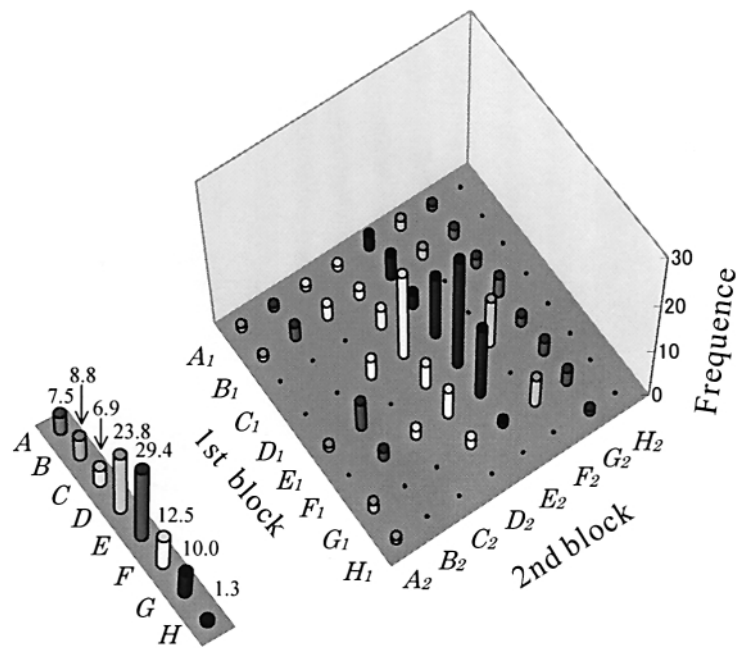
that it is extremely difficult, if not impossible, to obtain active block-substituted (or block-shuffled) proteins from GFP (Dopf and Horiagon, 1996). Whatever the case may be, this kind of research, now easily accessible owing to YLBS technology, provides valuable information on the convertibility of segments within a protein without loss of function.

#### Shuffling of amino acid monomer-sized blocks and others

There are intriguing points in dealing with a single amino acid (or a minimum-sized block of peptide) as a block since it can provide us with another way to synthesize, in theory, an arbitrary sequence of protein with a favorable occurrence of amino acids and without interference of stop codons. Therefore, it was examined whether we can shuffle trinucleotides (we call this 'aa-block' here) by YLBS technology.

*Shuffled products of aa-blocks.* Seven species of trinucleotides corresponding to Gly, Ile, Asp, Lys, Ser, Cys and Pro were chosen as starting blocks (Table I) by considering the codon





**Fig. 5.** Frequency of appearance of each GFP-derived block. The relative frequency of each single block appearance (normalized to 100% in total) is depicted as the height of a bar (left). The bars on the square mat sheet represent the frequency of the consecutive appearance of blocks as  $X_1X_2$  (from 5' to 3') where  $X_1$  and  $X_2$  represent any of the eight blocks. No appearance is shown with a dot.

**Table III.** Statistics on mutations observed in the shuffling of GFP-derived blocks

	Point mutation			Deletion <sup>a</sup>	
	Mutation rate (nucleotide/replication)	Transition	Transversion	Percent deletion (/event)	Deletion-free products
Eight blocks	$1.6 \times 10^{-4b}$	65.7% (90/137) <sup>c</sup>	34.3% (47/137)	49.4% (121/245)	5.7% (2/35)
Four blocks	$1.5 \times 10^{-4b}$ (63/87)	72.4% (24/87)	27.6% (17/160)	10.6% (14/32)	43.8%

<sup>a</sup>The insertion was found only twice in eight-block shuffling.

<sup>b</sup>Eight blocks (= 137/8400/100), four blocks (= 87/5760/100).

<sup>c</sup>The corresponding score experimentally obtained / the total number.

usage of wheat germ (Ikemura, 1985). These blocks were attached with the shuffling device sequence of 5'-half or 3'-half (see Figure 1), which has a minor change in their sequence to adapt to the restriction enzyme employed (Table I). Using the essentially identical procedures used in YLBS of GFP blocks (Figure 1), aa-blocks were ligated to dimers ( $Y^1$ ), tetramers ( $Y^2$ ), octamers ( $Y^3$ ) and hexadecamers ( $Y^4$ ). The diversities of these libraries ( $Y^1$ – $Y^4$ ) could be checked in the same way as shown in Figure 3. The final products ( $Y^4$ ) were cloned and sequenced (68 clones) as partly shown in Figure 4d. Ligation of 16 blocks after the fourth step was confirmed to be approximately 8% of the final products (thus,  $4.8 \times 10^{10}$  diversity), showing the capability of aa-block shuffling. However, there also happened to be a high frequency of deletion as in the previous experiments with GFP blocks. The same situation continued for four similar independent experiments (the statistics of these are shown in Table IV). The occurrence of each aa-block was within a statistically permissible range of fluctuation. Therefore, we concluded that aa-blocks could be shuffled by YLBS technology. This technology is more powerful in generating a diversity of DNAs encoding polypeptides than the conventional method that synthesizes nucleotide by nucleotide since YLBS can syn-

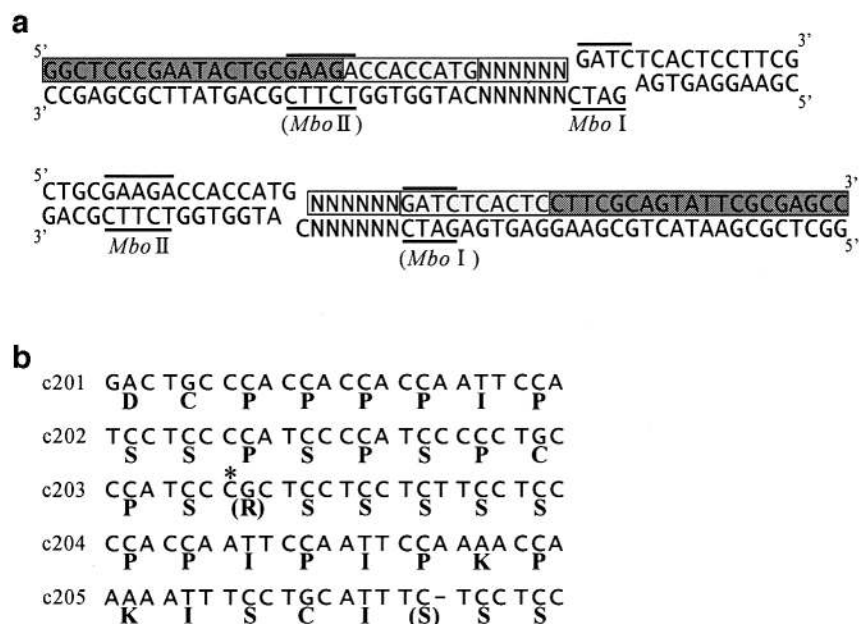
**Table IV.** Statistics on mutations observed in the shuffling of 16 blocks, each of which corresponds to an amino acid monomer

Point mutation			Deletion	
Mutation rate (nt/replication)	Transition	Transversion	Percent deletion (/event)	Deletion-free products
$1.9 \times 10^{-4a}$	35.5% (27/76)	64.5% (49/76)	10.3% (105/1020)	7.4% (5/68)

<sup>a</sup>Sixteen blocks (= 76/3264/120), eight blocks (= 7/624/100).

thesize DNAs by the unit of trinucleotides or hexanucleotides, and so on.

*Shuffling of longer blocks.* Hitherto, shufflings of the smallest and the medium-sized blocks were successfully performed. Thus, we additionally examined the possibility with a larger and uneven size of blocks. For this purpose, the subtilisin gene was divided into eight blocks of sizes 69–156 (shown in Table II) and subjected to block shuffling using basically identical shuffling device sequences. As a result, the shuffling of this size, which finally attained the ligated length of approximately 900 nucleo-



**Fig. 6.** Deletion problem solved. (a) The revised constructs for the stem sequences are shown for the restriction cleavage stage: 5'-half (top) and 3'-half (bottom). Both the 5'- and 3'-halves proper are boxed and slightly detached from their precursor DNAs in order to show the cleavage mode: *Mbo*I and *Mbo*II were adopted for generating the 5'-half and the 3'-half, respectively. Restriction enzymes not used here are shown braced. (b) Shuffled products ( $Y^3$ ) obtained after the construct improvement. Point mutation and deletion are shown by an asterisk and a hyphen, respectively. Substituted amino acids are shown with parenthesis. Only five clones with a deletion-containing result are shown here [23 out of 26 (=88%) were deletion-free].

tides (C-G-C-A-A-B-E; Table II), also proved possible (data not shown), though the yield of products at each step was low in this preliminary experiment probably due to the lack of optimization in operations and the uneven sizes of the blocks. Taking these into consideration, we can safely say that an average-sized protein of approximately 300 amino acids can be shuffled by this technology, although the recovery of products becomes smaller as the size of the branch in the Y-ligation construct grows larger (Nishigaki *et al.*, 1999). (We are now hopefully challenging the additional technology that can provide us with a sufficient amount of ligation products even in the case of long branches so that YLBS can be easily applied to the shuffling of the higher molecular weight proteins.) It is technically preferable to use blocks of an even size since it enables us to collect the products in a single band in gel electrophoresis.

#### Solution for deletion problems

The stubborn problem of deletion encountered in the shuffling of both module-sized blocks and aa-blocks has finally been solved by our independent experiments. The main causes for deletion phenomena were determined to be: (i) impurity of starting blocks and (ii) anomalous excision activity of type IIS restriction enzymes. From the close inspection of deletion products (partly shown in Figure 4), in which deletion was strongly associated with the second strand scission by type IIS restriction enzyme *Mbo*II, a hypothesis was presented that an anomalous cleavage occurs in the scission of the second strand, which follows after the cleavage of one strand, probably due to the instability of the nick-containing structure of the substrate DNA. Based on this hypothesis and assuming the strand to be cleaved secondly, we made up a new construct of a Y-ligation device sequence which does not depend on the abnormal second strand scission (Figure 6a). Using this construct, we could build up YLBS libraries of <10% deletion (as partly shown in Figure 6b), leading to the final solution for the deletion phenomena.

*Related problems to be tackled in the future.* For the purpose of diversity, the technology with a final deletion rate of <10% must be sufficiently usable since each species contained in a library can be produced in multi-copies (say, 100 copies or more) so as not to fail to express. However, if the deletion rate is zero, then one can keep the possibility, in theory, to generate a diversity of molecules as many as the number of the whole molecules recruited for the experiment. In this sense, it is technologically significant to reduce the deletion rate as much as possible, especially in such rare cases where all the members of a set must be assorted.

On the other hand, the deviation problem is rather serious since it means that some poorly-distributed block species cannot make a substantial contribution to the formation of a shuffled library. In addition, the deviation may reflect the essential property of the molecular device employed, i.e. T4 RNA ligase, on the interaction with various substrate DNA sequences (e.g. the trinucleotide of CCC may be more favored as a substrate in the ligation reaction than that of GGG). This seems to be the case and, fortunately, is not so extreme, based on our experimental data (partly shown in this work). This leads us to take two types of measures against these problems: (i) try not to use nucleotide sequences which provide poor substrates for T4 RNA ligase and (ii) try to raise the yield of Y-ligation as much as possible since the yield of 100% stands for no deviation as all the molecules have been involved in the reaction. The former is more realistic and the latter is a desirable aim. Therefore, our big technological challenge can be said to be to increase the yield of Y-ligation.

#### Block-shuffled libraries essential for protein engineering

Block shuffling can generate a well defined, well distributed library of proteins, which is definitely important for protein engineering as stated below. As a result of the studies on evolutionary molecular engineering (Eigen and Gardiner, 1984; Voigt *et al.*, 2000; and other works), it is now evident that once

a protein of a function is discovered (however weak its activity may be), then it can be evolved to provide a better function by hill-climbing through point mutations. In particular, it is known to be effective to accumulate advantageous point mutations by sexual PCR (Stemmer, 1994) or the like. Unfortunately, there is no general rule established that enables us to find a *de novo* functional protein used for evolution. Here, the well defined and well distributed libraries generated by block shuffling can work as convenient initial materials to be examined. This is because the well defined library enables us to design the next library, of which members are the most different in sequence from those of the former library, allowing us to evade testing the same or close sequences which have already been tested before. Therefore, point mutation is a way to walk in a small stride whereas block shuffling is a way to walk in a wide stride with controlled walking. (Even point mutations can make a wide stride by virtue of their high frequency, but the resultant products are unpredictable and uncontrollable in general.)

The most significant benefit provided by this technology must be in the readiness to integrate a meaningful block, irrespective of its size [from amino acid monomer to domain (approximately 100 amino acids) or more], into proteins at any site at any frequency. Now we can examine what is the effect if we dope a particular peptide sequence at various locations of a protein or what will happen if a null block (or nothing block) is inserted (in other words, introduction of block deletion) at various sites of a protein. Meaningful blocks can be either exons, domains or modules. Obviously, YLBS is the first general technology that has enabled us to shuffle exons and domains at will, with ease, and in a huge scale of diversity, opening the gate wide for evolutionary protein engineering. This is especially important when modularity is becoming a more important key concept in evolution (Go, 1981; Dover, 2000).

## Acknowledgements

This study was supported in part by JSPS Grant 'Research for the Future' 96100101. Some parts of this paper are derived from the PhD thesis of Y.K.

## References

- Abrams, E.S. and Stanton, V.P., Jr. (1992) *Methods Enzymol.*, **212**, 71–104.  
 Botstein, D. and Shortle, D. (1985) *Science*, **229**, 1193–1201.  
 Buchholz, C.J., Peng, K.W., Morling, F.J., Zhang, J., Cosset, F.L. and Russell, S.J. (1998) *Nat. Biotechnol.*, **16**, 951–954.  
 Cho, G., Keefe, A.D., Liu, R., Wilson, D.S. and Szostak, J.W. (2000) *J. Mol. Biol.*, **297**, 309–319.  
 Christians, F.C., Scapozza, L., Cramer, A., Folkers, G. and Stemmer, W.P. (1999) *Nat. Biotechnol.*, **17**, 259–264.  
 Conrad, R., Keranen, L.M., Ellington, A.D. and Newton, A.C. (1994) *J. Biol. Chem.*, **51**, 32051–32054.  
 Dai, X., De Mesmaeker, A. and Joyce, G.F. (1995) *Science*, **267**, 237–240.  
 Doolittle, R.F. (1995) *Annu. Rev. Biochem.*, **64**, 287–314.  
 Dopf, J. and Horiagon, T.M. (1996) *Gene*, **173**, 39–44.  
 Dover, G. (2000) *Bioessays*, **22**, 1153–1159.  
 Eigen, M. and Gardiner, W. (1984) *Pure Appl. Chem.*, **56**, 967–978.  
 Ellington, A.D. and Szostak, J.W. (1990) *Nature*, **346**, 818–822.  
 Fedoroff, N.V. (1999) *Ann. NY Acad. Sci.*, **18**, 251–264.  
 Gilbert, W. (1986) *Nature*, **319**, 618.  
 Go, M. (1981) *Nature*, **291**, 90–92.  
 Hanes, J. and Pluckthun, A. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 4937–4942.  
 Husimi, Y., Nishigaki, K., Kinoshita, Y. and Tanaka, T. (1982) *Rev. Sci. Instrum.*, **53**, 517–522.  
 Ikemura, T. (1985) *Mol. Biol. Evol.*, **2**, 13–34.  
 Illangasekare, M., Sanchez, G., Nickles, T. and Yarus, M. (1995) *Science*, **267**, 643–647.  
 Ito, Y., Suzuki, M. and Husimi, Y. (1999) *Biochem. Biophys. Res. Commun.*, **264**, 556–560.

- Joyce, G.F. and Orgel, L.E., (1999) In Gesteland, R.F. and Atkins, J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, New York, pp.1–26.  
 Jurinke, C., van den Boom, D., Collazo, V., Luchow, A., Jacob, A. and Koster, H. (1997) *Anal. Chem.*, **69**, 904–910.  
 Kieke, M.C., Shusta, E.V., Boder, E.T., Teyton, L., Wittrup, K.D. and Kranz, D.M. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 5651–5656.  
 Kikuchi, M., Ohnishi, K. and Harayama, S. (2000) *Gene*, **243**, 133–137.  
 Kolkman, J.A. and Stemmer, W.P. (2001) *Nat. Biotechnol.*, **19**, 423–428.  
 Kondrashov, F.A. and Koonin, E.V. (2001) *Hum. Mol. Genet.*, **10**, 2661–2669.  
 Kornberg, A. and Baker, T.A. (1992) In Kornberg, A. and Baker, T.A. (eds), *DNA Replication. Repair, Recombination, Transformation, Restriction and Modification*. W.H. Freeman and Company, New York, pp. 771–832.  
 Li, W.H., Gu, Z., Wang, H. and Nekrutenko, A. (2001) *Nature*, **409**, 847–849.  
 Lorsch, J.R. and Szostak, J.W. (1994) *Biochemistry*, **33**, 973–982.  
 Mattheakis, L.C., Bhatt, R.R. and Dower, W.J. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 9022–9026.  
 Mikawa, Y.G., Maruyama, I.N. and Brenner, S. (1996) *J. Mol. Biol.*, **262**, 21–30.  
 Nemoto, N., Miyamoto-Sato, E., Husimi, Y. and Yanagawa, H. (1997) *FEBS Lett.*, **414**, 405–408.  
 Nishigaki, K., Husimi, Y., Masuda, M., Kaneko, K. and Tanaka, T. (1984) *J. Biochem.*, **95**, 627–635.  
 Nishigaki, K., Kinoshita, Y. and Kyono, H. (1995) *Chem. Lett.*, **1995**, 131–132.  
 Nishigaki, K., Taguchi, K., Kinoshita, Y., Aita, T. and Husimi, Y. (1999) *Mol. Div.*, **5**, 1–4.  
 Nishigaki, K., Saito, A., Takashi, H. and Naimuddin, M. (2000) *Nucleic Acids Res.*, **28**, 1879–1884.  
 Parmley, S.F. and Smith, G.P. (1989) *Adv. Exp. Med. Biol.*, **251**, 215–218.  
 Riechmann, L. and Winter, G. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 10068–10073.  
 Robertson, D.L. and Joyce, G.F. (1990) *Nature*, **344**, 467–468.  
 Roy, S.W., Nosaka, M., Souza, S.J. and Gilbert, W. (1999) *Gene*, **238**, 85–91.  
 Sambrook, J. and Russell, D.W. (2001a) In Irwin, N. and Janssen, K.A. (eds), *Molecular Cloning*. Cold Spring Harbor Laboratory Press, New York, pp. 13.40–13.80.  
 Sambrook, J. and Russell, D.W. (2001b) In Irwin, N. and Janssen, K.A. (eds), *Molecular Cloning*. Cold Spring Harbor Laboratory Press, New York, pp. 4.82–4.85.  
 Sambrook, J. and Russell, D.W. (2001c) In Irwin, N. and Janssen, K.A. (eds), *Molecular Cloning*. Cold Spring Harbor Laboratory Press, New York, pp. 1.116–1.122.  
 Scott, J.K. and Smith, G.P. (1990) *Science*, **249**, 386–390.  
 Shao, X., Zhao, H., Giver, L. and Arnold, F.H. (1998) *Nucleic Acids Res.*, **26**, 681–683.  
 Shiba, K., Takahashi, Y. and Noda, T. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 3805–3810.  
 Steger, G. (1994) *Nucleic Acids Res.*, **22**, 2760–2768.  
 Stemmer, W.P. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 10747–10751.  
 Tessier, D.C., Brousseau, R. and Vernet, T. (1986) *Anal. Biochem.*, **158**, 171–178.  
 Tsien, R. and Prasher, D. (1998) In Chalfie, M. and Kain, S. (eds), *Green Fluorescent Protein: Properties, Applications and Protocols*. Wiley-Liss, New York, pp. 97–116.  
 Tsuji, T., Yoshida, K., Satoh, A., Kohno, T., Kobayashi, K. and Yanagawa, H. (1999) *J. Mol. Biol.*, **268**, 1581–1596.  
 Tuerk, C. and Gold, L. (1990) *Science*, **249**, 505–510.  
 Voigt, C.A., Kauffman, S. and Wang, Z.G. (2000) *Adv. Protein Chem.*, **55**, 79–160.  
 Wakasugi, K., Ishimori, K. and Morishima, I. (1997) *Biophys. Chem.*, **68**, 265–273.  
 Wilson, C. and Szostak, J.W. (1995) *Nature*, **374**, 777–782.  
 Wright, M.C. and Joyce, G.F. (1997) *Science*, **276**, 614–617.

Received March 18, 2002; revised June 17, 2002; accepted July 2, 2002