J. Biochem. **132**, 689–696 (2002)

Guide Oligonucleotide-Dependent DNA Linkage That Facilitates Controllable Polymerization of Microgene Blocks[1]

Kiyotaka Shiba,*,†,2 Tomoko Hatada,† Yuki Takahashi,† and Tetsuo Noda†,‡

**Department of Protein Engineering and †Department of Cell Biology, Cancer Institute, Japanese Foundation for Cancer Research, Toshima, Tokyo 170-8455; and ‡Department of Molecular Genetics, University of Tohoku School of Medicine, Aoba, Sendai 980-8575*

Received July 18, 2002; accepted August 2, 2002

Faster and more efficient searches of a huge protein sequence space for the purpose of conducting experiments in protein evolution can be achieved through the development of a block shuffling-based evolution system. One of the key components of such a system is the accurate and efficient linkage of gene units. Here we introduce a new method that allows accurate and controllable linkage of microgene blocks. This method employs a thermostable DNA ligase that links two single-stranded microgene blocks when they hybridize a complementary guide oligonucleotide. At high temperature, the ligation of the microgene units is fully dependent on the guide oligonucleotide, which can exclude undesired polymer formation, including the incorporation of microgenes having illegitimate sizes and "head-to-head" and "tail-to-tail" ligation of blocks. We were also able to assemble three microgene units using two guide oligonucleotides. Using this method of controllable linkage should facilitate further development of a step-by-step system for the polymerization of gene blocks, leading to a versatile block shuffling-based protein evolution system.

Key words: combinatorial polymers, evolutionary biotechnology, *in vitro* evolution, protein engineering.

Selection from molecular diversity libraries is a promising strategy for *de novo* creation of novel, functional proteins. This *in vitro* selection approach has already yielded remarkable successes in evolving functional RNAs, DNAs and short peptides whose structures are unrelated to existing biomolecules (*1–5*). In these experiments, large pools of random sequences prepared by combinatorial polymerization of four nucleotides were used as starting libraries from which functional molecules were evolved by neo-Darwinian selection. By following a similar strategy, Keefe and Szostak recently succeeded in evolving artificial 80–amino acid ATP-binding proteins with sequences unrelated to any existing protein from a random library containing 6×10^{12} proteins (*6*). Their success confirms that a local search of a huge protein sequence space (sequences of 80 amino acids are able to form 10^{38} different proteins) is sufficient to encounter functional new proteins.

Nevertheless, the results of theoretical experiments suggest that "a hierarchical approach" would allow a much more efficient search of protein sequence space and provide a greater likelihood of generating new protein folds (*7*). This theory proposes that combinatorial linkage between preexisting gene blocks *via* nonhomologous DNA recombination, rearrangement or insertion is the rate-limiting genetic event that has infused the existing protein world with structural diversity. This theory is fully compatible with the early hypothesis of Gilbert, "the exon theory of genes" (*8, 9*), which proposes that polymerization of exons (*i.e.*, microgenes coding for structured short peptides) through their flanking introns gave birth to the first set of genes. Although there is some dispute about the antiquity of introns (*10, 11*), it is now evident from observations of the structures of existing genes that such "exon shuffling" played a pivotal role in creating new genes in metazoan phyla, especially genes related to multicellularity (*12, 13*). Thus, both from a theoretical point of view and from analytical observations of the structures of existing genes, a hierarchical evolution of proteins—*i.e.*, mature proteins evolving from assemblages of microgenes having primordial biological functions—is a more plausible scenario for protein evolution than the hypothesis that insists proteins originated from random sequences of nucleotides (*14, 15*).

With the aim of establishing an efficiently directed protein evolution system, efforts are ongoing in many laboratories to apply the principle of "exon shuffling" or "a hierarchical approach" as an underlying concept of the system. DNA shuffling among a family of related genes has already enabled a sparser sampling of protein sequence space, and accelerated directed evolution of existing proteins (*16*). This powerful technique, however, can be applied only to genes or gene blocks that share enough sequence similarity to enable recombination using PCR. To recombine gene blocks that do not share any sequence similarity, special protocols need to be devised. Examples of such artifices include: (i) the use of engineered group II introns having trans-splicing activity to shuffle RNA cassettes of exons (*17*); (ii) the use of lox-Cre recombination sites inserted into self-splicing group I introns to assemble a large repertoire of peptides

[1]This work was supported in part by an HFSP grant to K.S.

[2] To whom correspondence should be addressed. Tel: +81-3-3918-0111, Fax: +81-3-5394-3903, E-mail: kshiba@jfcr.or.jp

from two different exons (*18*); (iii) the PCR-based assemblage of gene blocks pre-made so that they contain regions that partially (*19, 20*) or completely (*21, 22*) overlap regions in other blocks; (iv) the "SCRATCHY" method (*23*), which enables creation of multiple-crossover DNA libraries in which nonhomologous genes are first recombined by "ITCHY" (incremental truncation for creating hybrid enzymes) (*24*) or "SHIPREC" (sequence homology–independent protein recombination) (*25*), and then the resultant fusions are shuffled by "DNA shuffling" (*26*); and (v) "Y-ligation," which uses T4 RNA ligase to ligate single-stranded DNA blocks that form one stem and two branches (Y-shape) (*27*). In this paper, we introduce a new protocol that enables the controllable linkage of microgene blocks using thermostable DNA ligase and one or two guide oligonucleotides. This approach should facilitate the development of a protein evolution system based on "exon shuffling" or "a hierarchical approach."

MATERIALS AND METHODS

Preparation of Microgenes—The microgenes used in these experiments and the respective oligonucleotide primer sets used to prepare them are listed in Table I; the sequences of the oligonucleotide primers are shown in Table II. Microgene blocks were prepared by PCR using the indicated primer pairs (Table I) and Taq or HiFi DNA polymerase (Roche Diagnostics GmbH, Mannheim), with 1 ng of the 0.8 kb *Bst*EII–*Bam*HI fragment of pKS21 [containing the *ileS* gene (*34*)] serving as a template. The PCR was run for 35 cycles; the annealing temperature was 60°C. Microgene products were then separated in 2% agarose gels, and DNAs were recovered using a QIAquick Gel Extraction Kit (Qiagen GmbH, Hilden). Purified microgene blocks were quantified by ethidium bromide staining (*28*).

Ligation of Microgene Blocks—Ligation was carried out in 20 μl of a mixture containing the indicated amounts of microgene blocks and guide oligonucleotides, 20 mM Tris-HCl (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 0.1% Nonidet P-40 (NP40), 0.1 mM ATP, 1 mM dithiothreitol, and 4 units of *Pfu* DNA ligase (Stratagene, La Jolla). The enzyme was added to the reaction mixture at 65°C, and in most experiments a thermocycle of 92°C for 1 min and 65°C for 1 min was then repeated seven times. The resultant microgene polymers were detected by amplifying the contents of 1 μl of ligation mixture using the indicated primers in 50 μl of PCR mixture. In the experiment shown in Fig. 5, lanes 1–4, the incubation conditions were 95°C for 5 min followed by 65°C for 2 h instead of thermal cycling. In the experiments shown in Fig. 5, lanes 5–8, and in Fig. 6, the thermostable DNA ligase Ampligase (Epicentre Technologies, Madison) was used instead of *Pfu* DNA ligase. The ligation mixture for Ampligase contained 20 mM Tris-HCl (pH 8.3), 25 mM KCl, 10 mM $MgCl_2$, 0.01% Triton X-100, 0.5 mM nicotin-

TABLE I. **Abbreviations of microgenes and primer pairs used for preparations.**

Microgene blocks	Primer pairs[b]
ΔA[a]	KY-606/KY-543
pFΔ	KY-605/KY-607
ΔAΔ	KY-606/KY-632
pΔFΔ	KY-633/KY-607
pB	KY-601/KY-545
pC	KY-602/KY-547
pD	KY-603/KY-549
pE	KY-604/KY-551
p(B-E)b	KY-721/KY-722
ΔAΔb	KY-606/KY-723
pΔFΔb	KY-633/KY-724

[a]Δ indicates that the microgene lacks WS-2 (left Δ) or WS-1 (right Δ) sequences. [b]Nucleotide sequences are shown in Table II.

TABLE II. **Primer sequences.**

Primers	Sequences (5′ to 3′)
KY-543	CCTCCCTCAGAGCCGGGCCCGGACGGAGAAGTTTTGTCGTA
KY-545	CCTCCCTCAGAGCCGGGCCCGGTCCAGATTACCAGCGAGA
KY-547	CCTCCCTCAGAGCCGGGCCCTTTCGCCAGAATCACGGCCT
KY-549	CCTCCCTCAGAGCCGGGCCCGGTACCGGCATCCAGGGTA
KY-551	CCTCCCTCAGAGCCGGGCCCGTTCACGCCATCCAGCGTC
KY-601	pCGCCAACGCCGGCAAGGGGGGTATCGACGTTGCTTTCCAGGC[a]
KY-602	pCGCCAACGCCGGCAAGGGGGGTACCACGCCGTGGACTCTGCC
KY-603	pCGCCAACGCCGGCAAGGGGGGATCTGGTTGAAAGCGTAATGC
KY-604	pCGCCAACGCCGGCAAGGGGGGTGCCGTTCACACCGCG
KY-605	pCGCCAACGCCGGCAAGGGGGGTCTTCAAAGCGAACGACATC
KY-606	GGACGGTCACCTGCACAAAGGCGCGA
KY-607	GGCGGGATCCACTGCACGCCTTTGAT
KY-632	GACGGAGAAGTTTTGTCGTAAT
KY-633	pCGTCTTCAAAGCGAACGACATC
KY-638	CCCCCTTGCCGGCGTTGGCGCCTCCCTCAGAGCCGGGAAA
KY-640	CCCCCTTGCCGGCGTTGGCGGACGGAGAAGTTTTGTCGTCCG
KY-641	GATGTCGTTCGCTTTGAAGACGCCTCCCTCAGAGCCGGGAAA
KY-642	TTGCCGGCGTTGGCGCCTCCCTCAGAGCCGAAA
KY-643	CCGGCGTTGGCGCCTCCCTCAGAGAAA
KY-644	GCGTTGGCGCCTCCCTCAACA
KY-645	TTGGCGCCTCCCAAT
KY-703	AAGTTGAGTATTACGACAAAACTTCTCC
KY-704	CTTTGATCTCTTTCAGTGACTGCG
KY-721	pCGCCAACGCCGGCAAGGGGG
KY-722	bCCTCCCTCAGAGCCGGGCCC[b]
KY-723	bGACGGAGAAGTTTTGTCGTAAT
KY-724	bCGGGATCCACTGCACGCCTTTGAT

[a]"p" indicates a phosphorylated 5′-end. [b]"b" indicates a biotinylated 5′-end.

amide adenine dinucleotide (NAD), and 5 units of enzyme. For the experiment shown in Fig. 5, lanes 5–8, the incubation conditions were 95°C for 5 min followed by 45°C for 2 h; those for Fig. 6A were three thermocycles of 95°C for 1 min and 45°C for 10 min; and those for Fig. 6B were 95°C for 5 min followed by 45°C for 10 h. In Fig. 5, lanes 9–12, 4 units of T4 DNA ligase (Roche Diagnostics GmbH) were used in a mixture containing 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, and 5% (w/v) polyethylene glycol (MW 8,000). The incubation was at 37°C for 2 h.

Preparation of Single-Stranded Microgene Blocks—To exclude the DNA strand corresponding to the nonreading frame of the microgenes from the ligation reaction mixture, single-stranded microgene blocks corresponding to the reading frame were prepared as follows. First, microgene blocks were reamplified by PCR using each of the microgenes as a template with primers KY-721 and KY-722. The KY-721 oligonucleotide is complementary to the WS-2 region (see "RESULTS") and has a phosphorylated 5′-end, while the KY-722 oligonucleotide is complementary to the WS-1 region and has a biotinylated 5′-end. Consequently, after reamplification of the microgene blocks with KY-721 and KY-722, the 5′-ends of the reading strands were phosphorylated, while the 5′-ends of non-reading strands were biotinylated (designated as pNb).

The KY-606/KY-723 and KY-633/KY-724 primer pairs were used to prepare ΔAΔb and pΔFΔb, respectively. Aliquots (250 μl) of re-amplified PCR reaction solution containing approximately 100 pmol of microgene block were mixed with 250 μl of buffer containing 5 mg/ml streptavidin-coated superparamagnetic polystyrene beads (Dynabeads M-280, Dynal, Oslo) in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 2.0 M NaCl, and then incubated for 15 min at room temperature. During this incubation, biotinylated double-stranded microgene blocks were bound to the streptavidin moiety of the bead. Using a magnetic particle concentrator (MPC-E, Dynal), the beads were washed with 250 μl of buffer containing 10 mM Tris-HCl (pH 7.5), 1 mM EDTA and 2.0 M NaCl followed by 250 μl of water. The water was completely removed from the tube, and the beads were incubated in 40 μl of freshly prepared 0.1 N NaOH for 10 min at room temperature. Double-stranded DNAs were denatured to single-stranded DNAs in the alkaline solution, and the reading strand without the biotin moiety was liberated from the beads. The supernatant enriched with the single-stranded reading frame was then separated from the beads and neutralized by adding 12 μl of 0.333 N HCl.

RESULTS

Microgene Blocks—The microgene blocks used in these experiments were designed from *ileS*, the gene for *Escherichia coli* isoleucyl-tRNA synthetase (*29*). Six microgenes, MG-A to MG-F, encoded polypeptides ranging from 30 to 69 amino acids in length, one of which constituted the CP1 (connective peptide 1) region of isoleucyl-tRNA synthetase (Fig. 1A). The CP1 region connects the first and second halves of the N-terminal nucleotide binding fold (Rossmann fold) of the enzyme (*30*) and is folded as an independent globular domain (*31, 32*) (Fig. 1B). This region is responsible for editing the activity of the enzyme (*31–33*) and is important for avoiding the incorporation of valine instead of isoleucine during the translation process. A mutant isoleucyl-tRNA synthetase lacking the CP1 region showed aminoacylation activity *in vitro* (*30*), but did not support cell growth of an *ileS* null strain (K. S. unpublished result).

We previously showed that isoleucyl-tRNA synthetase can be cleaved into two fragments at various points throughout the enzyme structure without eliminating its ability to support growth *in vivo* (*34–36*). The reconstitution of activity was achieved through noncovalent assembly of enzyme fragments within the cell. Because this productive assembly is sufficiently flexible to accommodate the fusion of short segments of unrelated sequences at fragment junctions, we reasoned that cleavage points that did not eliminate the enzyme activity in the reconstitution experiment would be best suited for defining microgene boundaries. Therefore, all boundaries between microgene blocks correspond to cleavage points in CP1 where reconstitution occurred.

Welting Sequences—Each microgene block contained a 20-nt "welting sequence" at its 5′- and 3′-termini (WS-1 and WS-2, respectively). This sequence facilitated discriminative ligation of the microgene blocks and allowed the recreation of a 40-nt "WS-1 + 2" sequence when ligated in WS-1–WS-2 order (Fig. 2A). These welting sequences were designed so that: (i) the WS-1 + 2 sequence would encode a glycine-rich, flexible tridecapeptide that was in frame with respect to the microgene products; (ii) WS-1 and WS-2 contained the *Nae*I and *Apa*I recognition sites, respectively; (iii) linkage between WS-1 and WS-2 would recreate the *Nar*I recognition site at its junction; and (iv) illegitimate "head-to-head" or "tail-to-tail" linkage of WS-1 or WS-2 would recreate the *Stu*I and *Bss*HII recognition sites, respectively, at their junctions. These restriction enzyme recognition sites within the WS sequences enabled analysis and manipulation of the polymerization products.

Guide Oligonucleotide-Dependent Ligation Reaction—Microgenes with welting sequences were ligated under conditions similar to those used in the LDR, LAR, or LCR (ligation "detection," "amplification," or "chain" reaction) method (*37–39*), which is used to detect single-base substitutions on a chromosome using two adjacent oligonucleotides complementary to a target and a thermostable ligase. A key advantage of this approach relates to the fact that thermostable ligase will preferentially ligate two oligonucleotides that are perfectly complementary to their target sequence and that hybridize at a position directly adjacent to each other; thus discriminating against mismatches at primer junctions.

Adopting the principle of LDR, we designed a novel linkage strategy in which microgene blocks with complete WS-1 and WS-2 sequences at their respective 3′ and 5′ ends can be selectively ligated by thermostable ligase, whereas those having shorter or longer ends will be eliminated. The scheme of the system is illustrated in Fig. 2B. First, microgene blocks with WS-1 and WS-2 attached are prepared by PCR. Only the strands corresponding to the reading frame are phosphorylated at their 5′ ends (see "MATERIALS AND METHODS"). Double-stranded microgene blocks are denatured into single-stranded DNAs at high temperature, after which an oligonucleotide complementary to the "WS-1 + 2" sequence (Guide Oligo) hybridizes to the WS-1 sequence of the 3′ portion of block 1 and the WS-2 sequence of the 5′ portion of block 2. Because the two-microgene blocks are

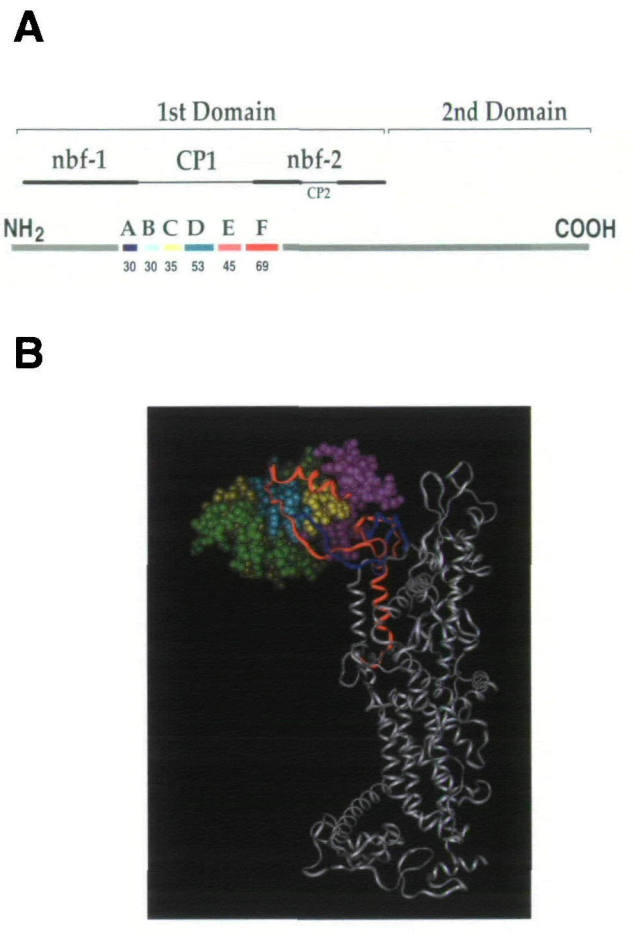

Fig. 1. **The microgene blocks used in this study were designed from *ileS*, the gene for *E. coli* isoleucyl-tRNA synthetase (*29*).** (A) The gene codes for the 938-amino acid-long monomeric enzyme composed of an N-terminal catalytic domain (1st Domain) and a C-terminal anticodon-recognition domain (2nd Domain). The numbers represent the length of the peptides produced from the respective microgenes (without welting sequences). (B) A space-filling model showing the regions corresponding to MG-B to MG-E using the same color code as in (A). The catalytic domain is divided into two halves (nbf-1 and nbf-2, where nbf is the nucleotide binding fold) by the CP1 (connective peptide 1) (*30*), which is folded as an independent domain within the tertiary structure of the enzyme (*32*). The nbf-2 has an additional short insert, CP2. Microgene blocks, MG-A to MG-F, were defined from the CP1 region so that their boundaries correspond to the points at which the enzyme can be cleaved into two fragments without losing its *in vivo* activity (*34–36*).

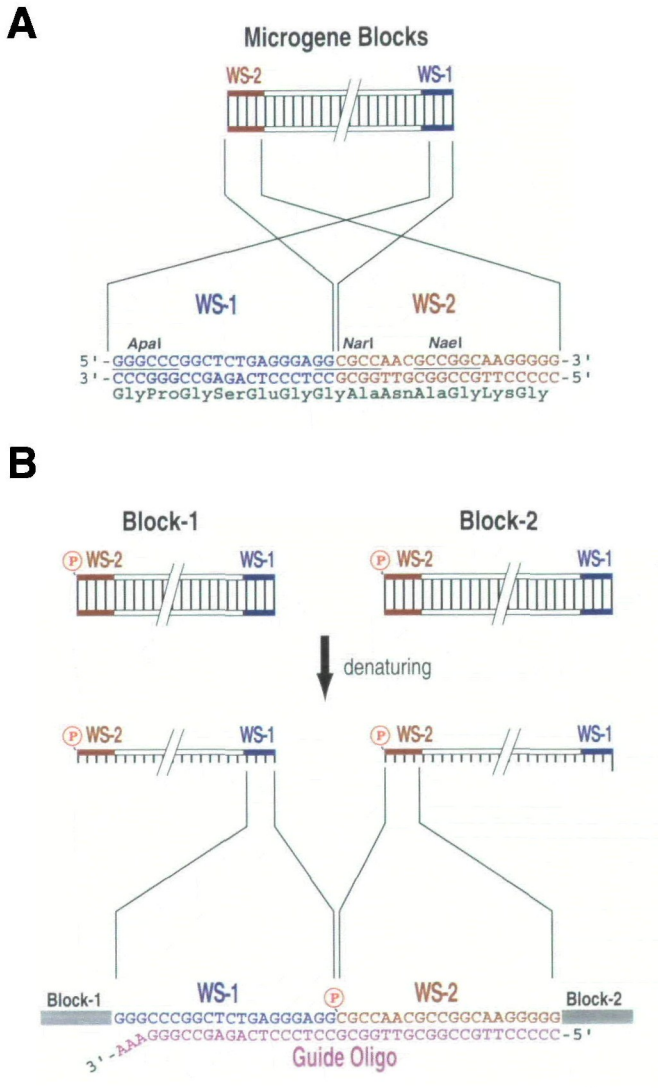

Fig. 2. **Schematic diagram showing the protocol used to ligate microgenes with the assistance of a Guide Oligo.** Welting sequences-1 (WS-1) and -2 (WS-2) were added to each microgene block (A). The translational product, which is in-frame with respect to the isoleucyl-tRNA synthetase peptide, is shown. The restriction enzyme recognition sites in WS-1 and WS-2 are shown. Upon hybridization with a Guide Oligo, WS-1 and WS-2 are aligned so that they are directly adjacent to each other. (B) When the 5′-end of WS-2 is phosphorylated, the two sequences will be covalently joined by thermostable ligase at high temperature. The Guide Oligo has three mismatched bases at the 3′ end that will be needed for "chain reaction" ligation (*39*), although that reaction was not explored in the study.

aligned so that they are precisely adjacent to each other, and the 5′-end of WS-2 is phosphorylated, these two blocks can be covalently joined by thermostable ligase at high temperature. If there is a mismatch, a gap or overhang at a block junction, the blocks are not linked by the enzyme, thereby avoiding the accumulation of by-products stemming from illegitimate PCR products. Likewise, "head-to-head" and "tail-to-tail" ligation of blocks is eliminated.

To confirm the appropriateness of the designed strategy, we first tested the effect of varying the amount of Guide Oligo present during high-temperature ligation catalyzed by thermostable ligase. In the experiment shown in Fig. 3A, 2 fmol of two microgenes were ligated at 65°C using thermostable *Pfu* ligase in the presence of 0 to 200 fmol of Guide Oligo. For the sake of simplicity, one microgene block had only WS-1 at its 3′ end (ΔA) and the other had only WS-2 at its phosphorylated 5′ end (pFΔ). The Guide Oligo aligned the two blocks in ΔA-pFΔ order, and the ligated dimer was detected as a 355 bp DNA after PCR using primers specific to each block. The 355 bp DNA was not observed when Guide Oligo was omitted from the reaction. But as the concentration of Guide Oligo was increased from 0.2 fmol to 20 fmol, efficient dimerization indicative of ligation of the single-stranded DNAs occurred. The dependency of ligation on Guide Oligo was not observed when there was

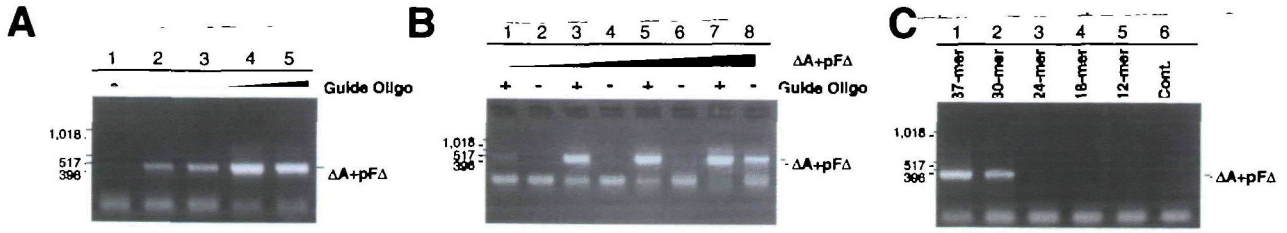

Fig. 3. **Two microgene blocks, ΔA and pFΔ, were ligated under various conditions.** (A) The effect of the amount of Guide Oligo (KY-638) on ligation in a reaction mixture containing 2 fmol of each microgene was investigated under the following conditions: lane 1, 0 fmol of Guide Oligo; lane 2, 0.2 fmol of Guide Oligo; lane 3, 2 fmol; lane 4, 20 fmol; lane 5, 200 fmol. (B) Effect of varying the amount of microgene block: lanes 1 and 2, 0.2 fmol of each microgene block; lanes 3 and 4, 2 fmol of each microgene block; lanes 5 and 6, 20 fmol; and lanes 7 and 8, 200 fmol. Lanes 1, 3, 5, and 7 contained 400 fmol of Guide Oligo KY-638 (+). (C) Effect of the length of the Guide Oligo. Two fmol of ΔA and pFΔ were ligated in the presence of 20 fmol of KY-638 (lane 1), KY-642 (lane 2), KY-643 (lane 3), KY-644 (lane 4), or KY-645 (lane 5). In lane 6, Guide Oligo was not added. Each Guide Oligo has a sequence that bridges WS-2 and WS-1. The lengths of the complementary regions are shown. In all experiments, the ligated products were detected as a 355 bp PCR product amplified using KY-606 and KY-607 as primers (indicated by "ΔA + pFΔ" on the right of the gels).

a high concentration of microgene blocks, however. In the experiment shown in Fig. 3B, 200 fmol each of the two microgenes were added to 20 μl of reaction mixture. At that concentration, the two microgenes were ligated even without Guide Oligo (lane 8). Therefore, to ensure Guide Oligo-dependent linkage, the concentration of the microgene in the ligation reaction must be kept below the threshold where Oligo-independent linkage begins to occur.

The effect of the length of the Guide Oligo was investigated in the experiment shown in Fig. 3C. The standard Guide Oligo was a 37-nt sequence complementary to the WS-1 + 2 sequence (Fig. 2B). When ligations were carried out using shorter derivatives with 30-, 24-, 18-, or 12-nt sequences complementary to the junction region of WS-1 and WS-2, the results showed that the 30-nt Guide Oligo supported ligation of microgene blocks under our conditions, while shorter oligonucleotides did not.

Three Block Ligation Using Two Guide Oligos—The experiments shown in Fig. 3 indicate that ligation of microgene blocks is dependent on Guide Oligo under the appropriate conditions—*i.e.*, a low concentration of microgene blocks and an excess of Guide Oligo. Using the same conditions, we attempted to ligate three microgene blocks using two different Guide Oligos. The blocks used were ΔAΔ, pN (N = B – E) and pΔFΔ, where ΔAΔ and pΔFΔ had welting sequences at both ends. In addition, two new Guide Oligos were used in this experiment: KY-640, which is complementary to the sequence of the 3′ region of ΔAΔ followed by WS-2, and KY-641, which is complementary to the sequence of WS-1 followed by the 5′ region of pΔFΔ. KY-640 would ensure linkage between ΔAΔ and pN (N = B – E), while KY-641 would ensure linkage between pN (N = B – E) and pΔFΔ. When 20 fmol each of the three microgene blocks (ΔAΔ, pΔFΔ and each of pB – E) and 400 fmol each of KY-640 and KY-641 were incubated with thermostable ligase, the expected microgene polymers were detected as PCR products with 453 bp (ΔAΔ-B-ΔFΔ), 459 bp (ΔAΔ-C-ΔFΔ), 516 bp (ΔAΔ-D-ΔFΔ) or 492 bp (ΔAΔ-E-ΔFΔ) (Fig. 4A). In the absence of Guide Oligos, no polymers were produced (data not shown).

We found that a higher concentration of microgene blocks was needed for three-block ligation than was needed for two-block ligation. For example, under the conditions at which two-block ligation proceeded well, *i.e.*, 2 fmol of microgene blocks and 20- to 200-fold excess of Guide Oligo, ligation of ΔAΔ-E-ΔFΔ was not observed (Fig. 4B, lanes 1 and 2). At least 20 fmol of each microgene block was required for three-block polymerization (Fig. 4B, lane 3).

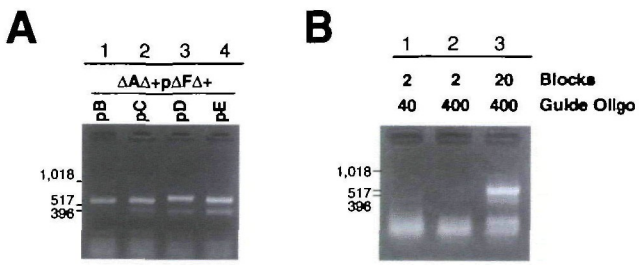

Fig. 4. **Three microgene blocks were ligated using two Guide Oligos.** (A) 20 fmol of ΔAΔ and pΔFΔ were ligated with 20 fmol of pB (lane 1), pC (lane 2), pD (lane 3), or pE (lane 4) in the presence of 400 fmol of Guide Oligos KY-640 and KY-641. KY-640 bridged between ΔAΔ and pN, while KY-641 bridged between pN and pΔFΔ. (B) Microgene blocks ΔAΔ, pE, and pΔFΔ were ligated using KY-640 and KY-641. The amounts of microgene and Guide Oligo used for each ligation were 2 fmol of microgene and 40 fmol of Guide Oligo (lane 1); 2 fmol of microgene and 400 fmol of Guide Oligo (lane 2); or 20 fmol of microgene and 400 fmol of Guide Oligo (lane 3). In all experiments, the ligated products were detected by PCR using KY-606 and KY-607 as primers.

Polymerization Using Ampligase and T4 DNA Ligase—Several thermostable DNA ligases with differing enzymatic properties are commercially available. Therefore, we used single-strand-enriched microgene blocks (see "MATERIALS AND METHODS"), to test the ability of an ATP-dependent *Pfu* ligase derived from *Pyrococcus furiosus*, a hyperthermophilic marine archaebacterium, and NAD-dependent Ampligase derived from a thermophilic bacterium, the precise source of which is unknown, to catalyze the polymerization of ΔAΔ, pN (N = B – E) and pΔFΔ guided by KY-640 and KY-641. Both enzymes efficiently ligated the three fragments (Fig. 5, lanes 1–8). T4 DNA ligase, in contrast, produced illegitimate products that likely resulted from Guide Oligo-independent ligation (Fig. 5, lanes 9–12).

Multiple Block Ligation—Finally, we attempted to carry out polymerization of four or more microgene blocks. In one experiment the ligation mixture contained microgene blocks ΔAΔ, pFΔ and pE and Guide Oligos KY-638 and KY-

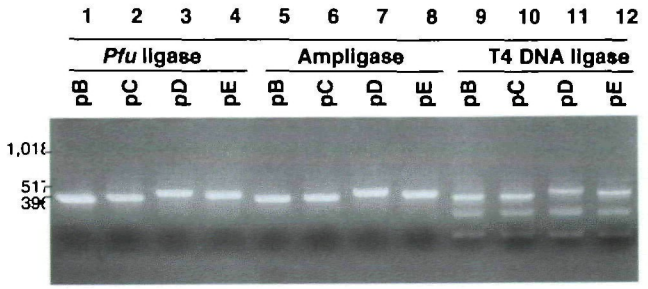

Fig. 5. **Polymerization using Ampligase and T4 DNA ligase.** Three microgene blocks, ΔAΔ, pN, and pΔFΔ, were ligated using *Pfu* ligase (lanes 1–4), Ampligase (lanes 5–8), or T4 DNA ligase (lanes 9–12). Ten fmol of microgene and 40 nmol of Guide Oligo were used. The ligated products were detected by PCR using KY-703 and KY-704 as primers. Single-stranded-enriched microgene blocks were used in these experiments (see "MATERIALS AND METHODS").

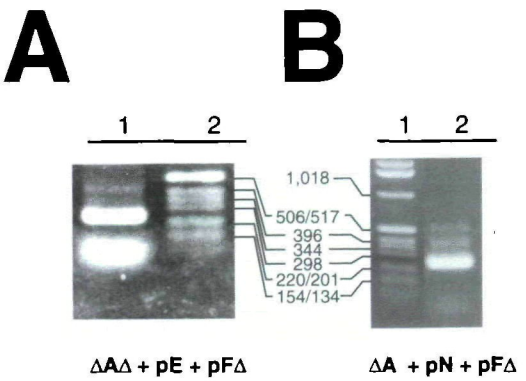

Fig. 6. **Microgenes were ligated under conditions that allow multiple block polymerization.** (A) Ten fmol each of microgene blocks ΔAΔ, pE, and pFΔ were ligated in the presence of 10 fmol of Guide Oligos KY-638 and KY-640. KY-638 mediated the linkages between pE–pE as well as pE–pFΔ. (B) Twenty fmol each of ΔA, pB, pC, pD, pE, and pFΔ were ligated using Ampligase in the presence of 100 fmol of KY-638. The ligated products were detected by PCR using KY-703 and KY-704 as primers.

640. Because KY-638 will bridge both pE-pFΔ and pE-pE, polymers with the structure ΔAΔ-$(E)_{0-n}$–FΔ were expected. PCR of the ligated products with primers KY-703 and KY-704, which hybridized, respectively, to ΔAΔ and pFΔ, revealed that by far the main product was ΔAΔ-$(E)_0$–FΔ (241 bp), although it was accompanied by a trace amount of ΔAΔ-$(E)_1$–FΔ (418 bp) (Fig. 6A). A similar experiment was carried out using pN (N = B – E) along with ΔA and pFΔ, and, although a ΔA-$(N)_2$–FΔ product from ΔA-$(D)_1$–FΔ with a size exceeding 462 bp was observed, the major product was ΔA-FΔ. No polymers longer than four blocks were obtained (Fig. 6B).

DISCUSSION

We have introduced a novel method that allows accurate and controllable linkage of microgene blocks *in vitro*. The method employs a thermostable DNA ligase that links two single-stranded DNA molecules after they hybridize a complementary partner such that they are positioned directly adjacent to one another. Because the ligation reaction proceeds efficiently at high temperature (but very slowly at low temperature), stringent conditions for hybridization of the three DNA molecules (the two adjacent sequences and their complementary sequence) can be selected. This avoids nonspecific ligation derived from undesired hybrid formation under relaxed conditions. Previously, these properties were successfully applied to the diagnostic method termed LDR, LAR, or LCR (*37–39*). With this method, two primers are designed so that the primer junction corresponds to a target site on a chromosome. If a mismatch exists between the primer junction and its target, it will be discriminated against by the thermostable DNA ligase, and the primers will not be ligated. This enables detection of a single nucleotide change within a genome.

Accurate linkage between DNA molecules is particularly important for *in vitro* assembly of microgene blocks. Enzymatically or chemically prepared DNA blocks inevitably contain illegitimate products with longer or shorter ends (*40*) that will change a reading fame of a block and often result in pre-termination of translation due to stop codons scattered in the second and third reading frames. Therefore, to avoid the incorporation of illegitimate microgene units during block shuffling experiments, we employed a thermostable DNA ligase in the protocol diagrammed in Fig. 2. With this protocol, complexes comprising a Guide Oligo and the 5′- and 3′-regions of two microgene blocks are formed such that the blocks are aligned directly adjacent to each other (Fig. 2B). The fact that Guide Oligos shorter than 24-nt do not support ligation is consistent with the idea that stable hybrid formation between a Guide Oligo and microgene blocks is a prerequisite. In addition, because a thermostable DNA ligase will not ligate two DNA molecules that have a gap or an overhang at their junction, illegitimate microgene blocks are excluded from the resultant polymers.

We first evaluated this protocol in a simple experiment in which two microgene blocks were ligated with the assistance of a Guide Oligo (Fig. 3). As expected, the ligation depended on the presence of Guide Oligo (Fig. 3A), although a relatively low concentration of microgene was required, as Guide Oligo-independent ligation did occur at higher concentrations (Fig. 3B). Guide Oligo-dependent ligation was then extended to three blocks, a reaction mediated by two different Guide Oligos (Fig. 4). In this case, we noticed the presence of a minor product whose size corresponded to ΔAΔ-ΔFΔ, suggesting the conditions allowed Guide Oligo-independent linkage of blocks. This by-product was completely eliminated, however, by using single-strand-enriched microgene blocks and by avoiding thermal cycling during the incubation step (Fig. 5). Accurate linkage among three microgene blocks was confirmed by sequencing the microgene polymers (data not shown). Moreover, ligation at 37°C using T4 DNA ligase resulted in higher rates of by-products (Fig. 5, lanes 9–12), confirming that the use of thermostable DNA ligase is critical for this system. Finally, we attempted multiple block polymerization (Fig. 6), but although multiple polymers were detected as minor products, the majority of the product was comprised of the smallest polymer. Additional artifices—solid-phase-assisted assembly of gene blocks, perhaps (*41*)—will need to be incorporated into the method before multiple polymerization can be achieved.

In the experiments described in this paper, all microgene

blocks had welting sequences added to their ends, which allowed polymerization of the blocks using a single Guide Oligo. The drawback of this strategy is that an additional tridecapeptide is inserted at every microgene junction. To minimize the effect of this peptide on the folding of whole proteins, the welting sequences were designed so that they encode a glycine-rich flexible peptide in-frame with respect to the microgenes. Definition of the microgene blocks is another critical component of a "hierarchical approach" to protein evolution. We defined the units by referring to previous enzyme cleavage experiments (*34–36*). Boundaries between microgenes corresponded to the sites at which the enzyme could be cleaved into two complementary fragments without eliminating its *in vivo* activity. This means that the boundaries of these microgenes were sufficiently flexible to incorporate the extra peptides derived from the welting sequences.

A key technology that allows accurate and efficient linkage of microgene units is a prerequisite for developing a versatile protein evolution system based on "exon shuffling" or a "hierarchical approach." The Guide Oligo-dependent ligation method described here provides features necessary for that key technology. The next phase will be the development of a step-by-step polymerization system by incorporating solid-phase-assisted assembly (*41*).

We would like to thank Ms. Tamiko Minamisawa for her technical assistance.

REFERENCES

1. Ellington, A.D. and Szostak, J.W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822
2. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment—RNA ligands to bacteriophage-T4 DNA polymerase. *Science* **249**, 505–510
3. Scott, J.K. and Smith, G.P. (1990) Searching for peptide ligands with an epitope library. *Science* **249**, 386–390
4. Cwirla, S.E., Peters, E.A., Barrett, R.W., and Dower, W.J. (1990) Peptides on phage—a vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. USA* **87**, 6378–6382
5. Devlin, J.J., Panganiban, L.C., and Devlin, P.E. (1990) Random peptide libraries: a source of specific protein binding molecules. *Science* **249**, 404–406
6. Keefe, A.D. and Szostak, J.W. (2001) Functional proteins from a random-sequence library. *Nature* **410**, 715–718
7. Bogarad, L.D. and Deem, M.W. (1999) A hierarchical approach to protein molecular evolution. *Proc. Natl. Acad. Sci. USA* **96**, 2591–2595
8. Gilbert, W. (1978) Why genes in pieces? *Nature* **271**, 501
9. Gilbert, W. (1987) The exon theory of genes. *Cold Spring Harbor Symp. Quant. Biol.* **52**, 901–905
10. Logsdon, J.M., Tyshenko, M.G., Dixon, C., Jafari, J.D., Walker, V.K., and Palmer, J.D. (1995) Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. *Proc. Natl. Acad. Sci. USA* **92**, 8507–8511
11. de Souza, S.J., Long, M., Klein, R.J., Roy, S., Lin, S., and Gilbert, W. (1998) Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient protein. *Proc. Natl. Acad. Sci. USA* **95**, 5094–5099
12. Patthy, L. (1999) Genome evolution and the evolution of exon-shuffling—a review. *Gene* **238**, 103–114
13. Long, M. (2001) Evolution of novel genes. *Curr. Opin. Genet. Dev.* **11**, 673–680
14. Ptitsyn, O.B. and Volkenstein, M.V. (1986) Protein structure and neutral theory of evolution. *J. Biomol. Struct. Dyn.* **4**, 137–156
15. White, S.H. and Jacobs, R.E. (1993) The evolution of proteins from random amino acid sequences. I. Evidence from the lengthwise distribution of amino acids in modern protein sequences. *J. Mol. Evol.* **36**, 79–95
16. Crameri, A., Raillard, S.A., Bermudez, E., and Stemmer, W.P.C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288–291
17. Mikheeva, S. and Jarrell, K.A. (1996) Use of engineered ribozymes to catalyze chimeric gene assembly. *Proc. Natl. Acad. Sci. USA* **93**, 7486–7490
18. Fisch, I., Kontermann, R.E., Finnern, R., Hartley, O., Solergonzalez, A.S., Griffiths, A.D., and Winter, G. (1996) A strategy of exon shuffling for making large peptide repertoires displayed on filamentous bacteriophage. *Proc. Natl. Acad. Sci. USA* **93**, 7761–7766
19. Crameri, A., Cwirla, S., and Stemmer, W.P.C. (1996) Construction and evolution of antibody-phage libraries by DNA shuffling. *Nature Med.* **2**, 100–102
20. Kolkman, J.A. and Stemmer, W.P. (2001) Directed evolution of proteins by exon shuffling. *Nature Biotech.* **19**, 423–428
21. Tsuji, T., Yoshida, K., Satoh, A., Kohno, T., Kobayashi, K., and Yanagawa, H. (1999) Foldability of barnase mutants obtained by permutation of modules or secondary structure units. *J. Mol. Biol.* **286**, 1581–1596
22. Tsuji, T., Onimaru, M., and Yanagawa, H. (2001) Random multi-recombinant PCR for the construction of combinatorial protein libraries. *Nucleic Acids Res.* **29**, E97
23. Lutz, S., Ostermeier, M., Moore, G.L., Maranas, C.D., and Benkovic, S.J. (2001) Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl. Acad. Sci. USA* **98**, 11248–11253
24. Ostermeier, M., Shim, J.H., and Benkovic, S.J. (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. *Nature Biotech.* **17**, 1205–1209
25. Sieber, V., Martinez, C.A., and Arnold, F.H. (2001) Libraries of hybrid proteins from distantly related sequences. *Nature Biotech.* **19**, 456–460
26. Stemmer, W.P.C. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389–391
27. Nishigaki, K., Taguchi, K., Kinoshita, Y., Aita, T., and Husimi, Y. (1998) Y-ligation: an efficient method for ligating single-stranded DNAs and RNAs with T4 RNA ligase. *Mol. Divers.* **4**, 187–190
28. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
29. Webster, T.A., Tsai, H., Kula, M., Mackie, G.A., and Schimmel, P. (1984) Specific sequence homology and three-dimentional structure of an aminoacyl transfer RNA synthetase. *Science* **226**, 1315–1317
30. Starzyk, R.M., Webster, T.A., and Schimmel, P. (1987) Evidence for dispensable sequences inserted into a nucleotide fold. *Science* **237**, 1614–1618
31. Nureki, O., Vassylyev, D.G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T.L., Schimmel, P., and Yokoyama, S. (1998) Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science* **280**, 578–582
32. Silvian, L.F., Wang, J., and Steitz, T.A. (1999) Insights into editing from an ile-tRNA synthetase structure with tRNAile and mupirocin. *Science* **285**, 1074–1077
33. Lin, L., Hale, S.P., and Schimmel, P. (1996) Aminoacylation error correction. *Nature* **384**, 33–34
34. Shiba, K. and Schimmel, P. (1992) Functional assembly of a randomly cleaved protein. *Proc. Natl. Acad. Sci. USA* **89**, 1880–1884
35. Shiba, K. and Schimmel, P. (1992) Tripartite functional assembly of a large class I aminoacyl tRNA synthetase. *J. Biol. Chem.* **267**, 22703–22706
36. Shiba, K. (1995) Dissection of an enzyme into two fragments at intron-exon boundaries in *Tracing Biological Evolution in Pro-*

tein and Gene Structures (Go, M. and Schimmel, P., eds.) pp. 11–21, Elsevier Science Publishers, Amsterdam
37. Landegren, U., Kaiser, R., Sanders, J., and Hood, L. (1988) A ligase-mediated gene detection technique. *Science* **241**, 1077–1080
38. Wu, D.Y. and Wallace, R.B. (1989) The ligation amplification reaction (LAR)—amplification of specific DNA sequences using sequential rounds of template-dependent ligation. *Genomics* **4**, 560–569
39. Barany, F. (1991) The ligase chain reaction in a PCR world. *PCR Methods Appl.* **1**, 5–16
40. Clark, J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* **16**, 9677–9686
41. Nord, K., Nilsson, J., Nilsson, B., Uhlén, M., and Nygren, P.A. (1995) A combinatorial library of an alpha-helical bacterial receptor domain. *Protein Eng.* **8**, 601–608