the amounts of dimer templates, making it possible to design different types of combinatorial libraries, such as random shuffling libraries and artificial alternative splicing libraries. The block shuffling and random point mutations can be introduced simultaneously by performing RM-PCR under error-prone conditions. Most of the block sequences in the combinatorial libraries created by RM-PCR encode a long open reading frame and are suitable for protein selection experiments. Thus, RM-PCR promises to be a powerful approach in creating novel proteins, especially simulating early stages of protein evolution on a rapid timescale using an efficient and easily adaptable experimental protocol.

[8] Gene Library Synthesis by Structure-Based Combinatorial Protein Engineering

By Paul E. O'Maille, Ming-Daw Tsai, Bryan T. Greenhagen, Joseph Chappell, and Joseph P. Noel

Introduction

Structure-based combinatorial protein engineering (SCOPE) is a process for the synthesis of gene libraries that lay the genetic foundation for the exploration of the relationship between structure and function in the encoded proteins.[1] The comparative analysis both structurally and functionally of protein primary, secondary, and tertiary structure generates numerous hypotheses with which to probe the relationship between molecular structure and the ensuing functional readout. SCOPE provides a tool for constructing the gene libraries that encode rationally engineered protein variants that provide the raw material for addressing these hypothesis using both structural and functional analyses. Mechanistic hypotheses generated from structures derived from both experiment (crystallographic and nuclear magnetic resonance) and homology modeling is used to design oligonucleotides that code for crossovers between genes encoding structurally related proteins. A series of polymerase chain reactions (PCR), culminating in the selective amplification of crossover products, incorporates spatial information encoded in the oligonucleotide into a full-length gene and the resultant hybrid protein. Iteration of the process enables the synthesis of all possible combinations of desired crossovers, producing a hierarchical collection of chimeras in analogy to a Mendelian population.

[1] P. E. O'Maille, M. Bakhtina, and M. Tsai, *J. Mol. Biol.* **321,** 677 (2002).

0076-6879/04 $35.00

The principles of the process are generally applicable and the methodology is easily adapted to a range of experimental objectives. At its inception, SCOPE was developed to provide a means of generating multiple crossover gene libraries from distantly related proteins, constituting a homology-independent *in vitro* recombination approach.[1] This article presents the adaptation of SCOPE to the facile combinatorial synthesis of mutant gene libraries. The newly incorporated refinements to the originally designed SCOPE approach illustrate the underlying principles of the experimental process that make it a robust technique for the parallel exploration of protein sequence space in three dimensions. This tertiary information is embodied within the mechanistic and evolutionary underpinnings of protein structure and function, both of which are fundamental aspects of biochemical adaptive change in organisms. In addition, this information can be exploited by SCOPE for a myriad of applications in biotechnology.

Principles

The construction of gene libraries by SCOPE involves a series of PCRs. Other recombination techniques use multiple primers or random fragments in a single step, thus carrying out multiple reactions in parallel.[2,3] Separation of gene synthesis into discrete steps is an essential feature of SCOPE. This simple but critical property of SCOPE enables one to control recombination through pairing gene fragments and genes that give rise to designed and anticipated combinations of crossovers. As a consequence, libraries are constructed as a series of less complex mixtures, which reduce numerical complexity and the cost and extent of sampling required during screening, including gene sequencing and functional assays. Crossover locations and the frequency of genetically encoded crossovers are established by experimental design and are not dictated or constrained by homology between genes or the linear distance between multiple mutations.

An overview of the process illustrates the basic steps encompassing SCOPE-based recombination (Fig. 1). In step I, standard PCR amplification, using an internal and external primer pair and the appropriate template DNA, produces chimeric gene fragments. Internal primers are designed on the basis of one or more encoded three-dimensional structures viewed with reference to the variable sequence space of protein homologues and code for crossovers in the protein-coding region of genes. External primers correspond to the 5′ and 3′ termini of a given gene, as

[2] W. P. Stemmer, *Nature* **370,** 389 (1994).
[3] F. J. Perlak, *Nucleic Acids Res.* **18,** 7457 (1990).

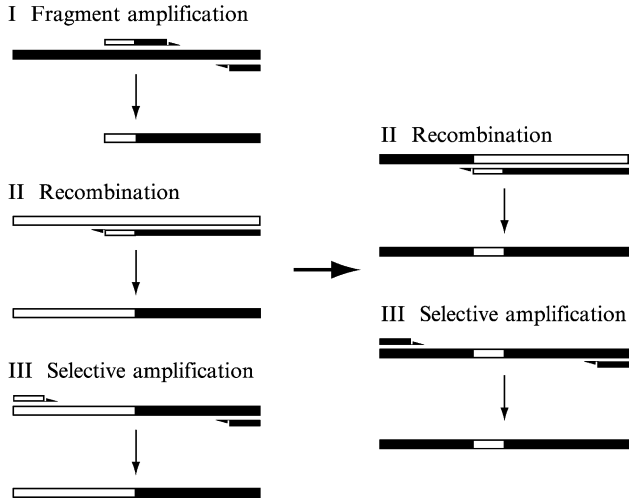


FIG. 1. Overview of the SCOPE library synthesis process.

with any typical primer pair used in a standard PCR amplification. The template consists of a plasmid or PCR product that contains the gene of interest. In step II, *in vitro* recombination occurs between a gene fragment(s) and a new template; in other words, gene fragments serve as a new set of primers, which anneal and are extended to produce single-stranded full-length chimeras. In step III, a new external primer set directs the selective amplification of recombination products by virtue of the unique genetic identity encoded at their termini. Repetition of steps II and III using various pairs of gene fragments from step I and crossover products from step III allows the production of genetically diverse, multiple crossover libraries in high yield.

Careful oligonucleotide design is central to the SCOPE recombination process. A discussion of the properties of the synthetic oligonucleotides, in relationship to specific applications, illustrates how SCOPE can be adapted to construct multiple crossover libraries from distantly related proteins or combinatorial mutant libraries from mechanistically related proteins.

Internal Primers

Shuffling exons or "equivalent" structural elements between homologues require *chimeric* oligonucleotides. These are composed of approximately equal halves of two distinct genes and code for a crossover region. An example of their use is illustrated in the process overview (Fig. 1).

Without *a priori* knowledge of the optimal point of fusion in regions of low identity or the compatibility of ''equivalent'' structural elements of low sequence identity, linkage variability can be introduced. This entails designing a set of chimeric oligonucleotides (for a given crossover region), which code for a series of insertions and/or deletions around a fixed point to explore alternative crossovers. The corresponding collection of gene fragments can be used together in recombination reactions (step II). Variable connections between ''equivalent'' structural elements proved to be an essential aspect of design in producing functional hybrids from distantly related DNA polymerases.[1]

Combinatorial mutagenesis by SCOPE requires *mutagenic* oligonucleotides that code for either a specific or a random set of mutations, insertions, or deletions directed at a given site in the gene of interest. Alternatively, *bridging* oligonucleotides, which code for stretches of native sequence between mutations, can be used to mediate recombination between wild-type and/or mutant genes. Mutagenic and bridging oligonucleotides are employed in PCR in a similar manner as chimeric oligonucleotides, although refinements to the process were required for their efficient incorporation into the desired chimeric sequence as described later.

External Primers

Polymerase chain reaction amplification of mutant or chimeric genes (step III, Fig. 1) is the final step of the SCOPE cycle. Like any conventional amplification, a primer set that flanks the target gene of interest is required during this final amplification step. Additionally, the inclusion of restriction or recombination sites into the final primer set for the efficient cloning of the resultant collection of genes is often desirable. However, a fundamental aspect of SCOPE is that the ''proper'' primer set be used for the selective amplification of a particular crossover product from a recombination reaction, which may contain a mixture of products. In the chimeragenesis of distantly related proteins, the termini of each gene are unique and can be exploited in this way for selective amplification.

SCOPE, as applied to the combinatorial synthesis of mutant libraries, where the termini of wild-type genes and crossover products are indistinguishable, required the design of alternative external primers and a bookkeeping system for their successful implementation and hierarchical organization and storage. Primary amplification primers (PAPs) code for DNA sequences flanking the gene (like any generic external primer), but contain an additional and unique 5′ sequence tag. Their use in gene fragment synthesis (step I) links a unique sequence to a particular mutation. Following recombination, secondary amplification primers (SAPs), which

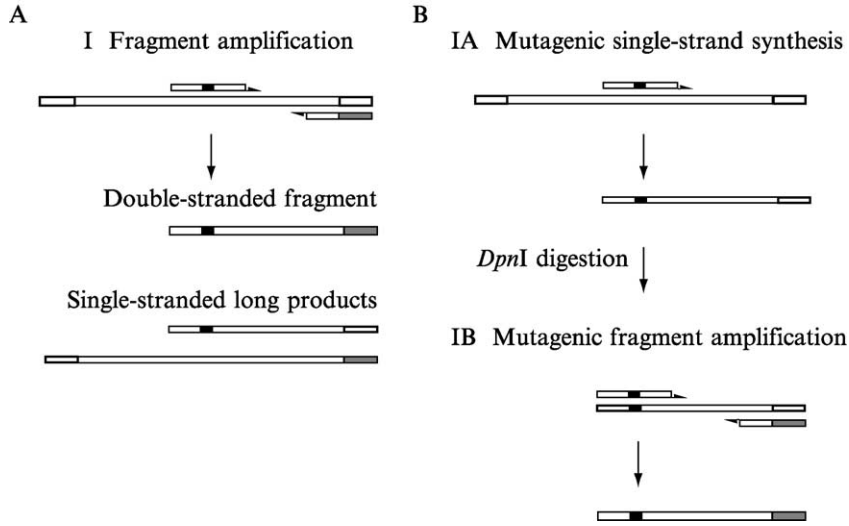
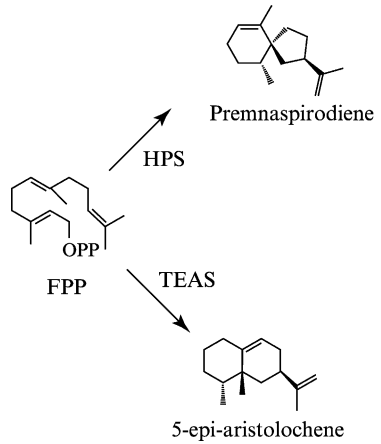


FIG. 2. Refinements to SCOPE. (A) Source of wild-type background. (B) Alternative fragment amplification strategy for the suppression of the wild-type sequence.

correspond to the 5′ unique sequence, are employed in the final amplification (step III) to select for the desired recombination products.

Refinements to the original SCOPE process were necessary for the efficient incorporation of unique sequence tags, their linkage to mutations, and the suppression of wild-type background genes. During step I amplification, single-stranded DNA or ''long'' product is produced from the extension of each primer on the plasmid template. When derived from PAPs, these single-stranded products code for the wild-type gene (Fig. 2A) and, if carried over into other steps of the process, give rise to a small but significant population of wild-type genes in the background. Separating step I into two reactions alleviates this contamination problem (Fig. 2B). In step IA, the internal primer and template are mixed and single-stranded DNA containing the mutation(s) is synthesized. The product of step IA is treated with *Dpn*I to digest the wild-type plasmid template, leaving only the nascent, single-stranded, mutagenic DNA. This restriction step eliminates the formation of long products that contribute to a wild-type background. A portion of step IA product is then used in step IB, where it serves as a template for conventional PCR amplification with an internal primer and a PAP. In the original development of SCOPE, gel purification of gene fragments (step I) was an essential step of the process. The aforementioned refinements enable the entire sequence of PCR reactions



SCHEME 1. Product specificity of two closely related terpene cyclases: henbane premnaspirodiene synthase (HPS) and tobacco 5-epi-aristolochene synthase (TEAS).

(steps I through III) to be conducted without purifying intermediates until the final amplification step.

An illustration of the SCOPE methodology as applied to the combinatorial synthesis of mutant gene libraries of the terpene cyclase enzyme known as tobacco 5-epi-aristolochene synthase (TEAS) is presented as a proof of principle example of the power and ease of this newly revised version of SCOPE. Previously, the product specificity of TEAS was converted from 5-epi-aristolochene to premnaspirodiene, the product of the closely related terpene cyclase premnaspirodiene synthase (HPS) from henbane, by nine sequential mutations.[4] Site-directed mutants were designed using the three-dimensional structure of TEAS[5] and homology modeling of HPS. The products of these cyclases are shown in Scheme 1. Because these mutations were made sequentially, the question remains, are all nine mutations required for product selectivity? What combinations, if any, of these nine mutations are sufficient for this property change? Moreover, in a broader evolutionary sense, what is the mechanistic and energetic landscape that links such specificity switches that are encoded by multiple changes at sites throughout a given protein? Constructing all combinations of nine mutations (2^n combinations, where n is the number of mutations) and analyzing their product specificity and kinetic properties will provide answers to some of these questions.

[4] B. T. Greenhagen, Dissertation thesis. Department of Agriculture at the University of Kentucky, 2003.
[5] C. M. Starks *et al.*, *Science* **277,** 1815 (1997).

Terpene cyclases are an ideal proof of the principal system for exploring the utility of SCOPE given their (1) unusual mechanism employing the conformationally directed production of reactive carbocation intermediates, (2) well-defined three dimensional structures, (3) ease of product identification and quantification using high-throughput GC-MS analysis, (4) evolutionarily diverse distribution of protein sequences and small molecule products across multiple kingdoms, and (5) biotechnological potential for the biosynthesis of unique small molecules representing a currently untapped region of natural product space.

Experimental Procedures

Materials

PCR components: 10× cloned pfu reaction buffer and pfu turbo DNA polymerase (Stratagene, La Jolla, CA), dNTPs (Invitrogen, Carlsbad, CA), and bovine serum albumin (BSA; New England Biolabs, Beverly, MA). PCR reactions are carried out using a PTC 200 Peltier thermal cycler (MJ Research, Waltham, MA). All PCR products are purified by gel extraction (Qiagen, Valencia, CA), cloned into pDONR 207 using Gateway cloning technology (Invitrogen) according to the manufacturer's recommended conditions. Plasmid DNA from gentamicin-resistant transformants is minipreped by the Salk Institute Microarray facility for sequencing at the Salk Institute DNA sequencing/quantitative PCR facility. The cDNA of TEAS is cloned into pH8GW (an in-house gateway destination vector) and this plasmid DNA is used as a template for PCR.

All PCR are carried out using a master mix of a standard set of PCR components for a 50-μl scale reaction:

5 μl of 10× cloned pfu reaction buffer to give 1×
1 μl of pfu turbo DNA polymerase (Stratagene) (2.5 U/μl) to give 0.05 U/μl
0.5 μl of BSA (10 mg/ml) to give 0.1 mg/ml
8 μl of dNTP mix (1.25 m*M*) to give 200 μM each dNTP

Primers

Oligonucleotides are from Integrated DNA Technologies (IDT) and are listed in Table I. For both mutagenic and chimeric primers, the mutation(s) or crossover point(s) is located in the center of the oligonucleotide, such that the flanking sequence is complementary to a given gene; ideally, this should be 18 to 24 nucleotides (or have a T_m greater than or equal to 50°) for effective PCR. SAPs are designed to consist of 21 nucleotides

TABLE I
OLIGONUCLEOTIDES USED FOR SCOPE COMBINATORIAL MUTAGENESIS

Primer	T_m (°)	Sequence[a]
Internal mutagenic primers[b]		
A275T+	60.2	GAATGCTACTTTTGG**A**CATTAGGAGTTTATTTTGAG
V291A+	62.8	CTCGCGTCATGCTCG**C**TAAGACCATATCAATG
V372I+	60.1	GAAAGAATGAAAGAA**A**TAGTAAGAAATTATAATGTCGAGTCAAC
T402S+	62.8	CCTAAGCAATGCACTAGCAACT**T**CCACATATTAC
Y406L+	59.9	CCACATATTACT**TG**CTCGCGACAAC
T402S/Y406L+	67.9	GCACTAGCAACT**T**CCACATATTACT**TG**CTCGCGACAAC
S436N+	60.3	GAATCCAAAAATTCTTGAAGCTA**A**TGTAATTATATGTCGAG
I438T+	59.9	CTTGAAGCTAGTGTAA**C**TATATGTCGAGTTATCG
I439L+	64.2	CTTGAAGCTAGTGTAATT**C**T**G**TGTCGAGTTATCGATGAC
S436N/I438T+	61.4	AATTCTTGAAGCTA**A**TGTAA**C**TATATGTCGAGTTATCGATG
S436N/I439L+	64.1	AATTCTTGAAGCTA**A**TGTAATT**C**T**G**TGTCGAGTTATCGATGAC
S436N/I438T/I439L+	64.9	AATTCTTGAAGCTA**A**TGTAA**C**T**C**T**G**TGTCGAGTTATCGATGAC
I438T/I439TL+	65.1	CTTGAAGCTAGTGTAA**C**T**C**T**G**TGTCGAGTTATCGATGAC
V516I+	60.3	CTCAATCTTGCTCGTATT**A**TTGAGGTTACATATATACAC
Primary amplification primers[c]		
A_b1	72.1	ATGCTTTAAGGCTCTGGGCCGACAAGTTTGTACAAAAAAGCAGGC
B_b1	71.7	GTCACGCATATGATTCGGCGGACAAGTTTGTACAAAAAAGCAGGC
C_b1	74.9	CCTGGCTTGCTCGGATAGAACACAAGTTTGTACAAAAAAGCAGGC
D_b1	73.4	CTACAGAGAATGCCGCGGTCCACAAGTTTGTACAAAAAAGCAGGC
E_b1	68.6	GGTCGTCGACCCCAGCGTGAGACAAGTTTGTACAAAAAAGCAGGC
F_b1	70.8	TCGCAATTCACGGCTTGACCCACAAGTTTGTACAAAAAAGCAGGC
1 _b2	70.6	AAATGCAGGTAGCAGAGCTGTACCACTTTGTACAAGAAAGCTGGG
2 _b2	71.6	TTAGGACGGGACTGCTGTAGCACCACTTTGTACAAGAAAGCTGGG
3 _b2	73.4	ACATTCTGACGTGAAACGCGCACCACTTTGTACAAGAAAGCTGGG
4 _b2	72.0	GGACACACTTAGCCTTCCAGGACCACTTTGTACAAGAAAGCTGGG
5 _b2	74.1	GGCCTGGAGTAGGATCTTTGCACCACTTTGTACAAGAAAGCTGGG
6 _b2	70.2	CCCGTCCCACTTCGTGACCGCACCACTTTGTACAAGAAAGCTGGG
Secondary amplification primers[d]		
A	61.3	ATGCTTTAAGGCTCTGGGCCG
B	59.7	GTCACGCATATGATTCGGCGG
C	58.7	CCTGGCTTGCTCGGATAGAAC
D	60.9	CTACAGAGAATGCCGCGGTCC
E	65.1	GGTCGTCGACCCCAGCGTGAG
F	61.7	TCGCAATTCACGGCTTGACCC
1	58.2	AAATGCAGGTAGCAGAGCTGT
2	59.7	TTAGGACGGGACTGCTGTAGC
3	59.9	ACATTCTGACGTGAAACGCGC
4	58.4	GGACACACTTAGCCTTCCAGG
5	58.6	GGCCTGGAGTAGGATCTTTGC
6	65.7	CCCGTCCCACTTCGTGACCGC

[a] Bold and underlined characters indicate the sites of designed mutations. Shading is used to indicate part of the attB1 and attB2 recombination sequences; complete attB sites are generated by amplification of the target gene(s) from the destination vector pH8GW.
[b] Mutageneic primers are named according to the amino acid substitutions they code for.

and have a T_m greater than or equal to 55°. PAPs contain 24 bases (in addition to their unique sequence), which correspond to partial Gateway attB sites; the remaining attB sequence becomes incorporated into PCR products by amplification from pH8GW. T_m values are calculated based on nearest-neighbor thermodynamic parameters.[6]

Gel Electrophoresis

Analysis of PCR fragments and separation of products for gel purification are performed using 2% (w/v) agarose gels in 1× TAE buffer containing 0.1 μg/ml ethidium bromide. Concentrations of PCR products (steps IB and III) are estimated by comparison to a standard of known concentration, such as the low DNA mass ladder (Invitrogen) using densitometry software such as ImageJ (http://rsb.info.nih.gov/ij/).

Method

Prior to library construction, all primers are tested to ensure that they result in unique amplification products of the expected size. Like any standard PCR amplification, optimization of cycling parameters for specific template and primer sets may be necessary.

Step IA: Mutagenic/Chimeric Single-Stranded DNA Synthesis

Procedure. Reactions are mixed on ice using the following:
14.5 μl of PCR master mix (as defined earlier)
1 μl internal primer (5 μ*M* stock) to give 0.1 μ*M*
1 μl plasmid DNA template (10 n*M* stock) to give ~200 p*M*
33.5 μl filter-sterilized H_2O added to give a 50-μl reaction volume

The master mix is added last and the resultant reaction is mixed by pipetting. Cycling program: 96° for 5 min, followed by 50 cycles of 96° for 30 s, 55° for 30 s, and 72° for 1 min/kb of product followed by incubation at 4° at the completion of cycling.

Comments. The amount of single-stranded product formed is limited by the amount of template DNA and the number of cycles performed.

[6] H. T. Allawi and J. Santa Lucia, Jr., *Biochemistry* **36,** 10581 (1997).

[c] Primary amplifications primers are named according to their unique sequence tag (A through F for forward and 1 through 6 for reverse as listed in Fig. 4) and gateway recombination sequence (b1 for attB1 and b2 for attB2).

[d] Unique sequence tags are labeled according to their corresponding primary amplification primer (A through F for forward and 1 through 6 for reverse as listed in Fig. 4).

Estimated yields for the aforementioned reaction (using 50 cycles and ~200 p*M* plasmid) are about 10 fmol of the final single-stranded product; this amount is well in excess of what is required for subsequent amplification reactions. A 0.1 μM concentration of internal primer ($>10^3$ molar excess of plasmid template) is sufficient; higher primer concentrations may promote the formation of alternative products in subsequent amplification steps (step IB).

Dpn*I Digestion of Plasmid DNA*

Procedure. The addition of 1 μl of *Dpn*I (20 U/μl, New England Biolabs) with mixing is followed by incubation at 37° for 1 h for digestion of the original DNA template and 20 min at 80° for heat inactivation.

Step IB: Mutagenic/Chimeric Double-Stranded DNA Fragment Amplification

Procedure. Reactions are mixed on ice using the following:

14.5 μl of PCR master mix (as defined earlier)
2 μl internal primer (5 μM stock) to give 0.2 μM
1 μl primary amplification primer (5 μM stock) to give 0.1 μM
1 μl of step IA reaction as template to give ~1–10 p*M* single-stranded DNA
31.5 μl filter-sterilized H_2O added to give a 50-μl reaction volume

The master mix is added last with pipetting to mix reactions. Cycling program: 96° for 5 min, followed by 40 cycles of 96° for 30 s, 55° for 30 s, and 72° for 1 min/kb of product followed by incubation at 4° at the completion of cycling. Amplification products are verified by agarose gel electrophoresis.

Comments. Internal primers should be in excess of external primers; keeping the concentration of external primers below saturation and increasing the number of cycles ensure their depletion. This is sufficient to suppress wild-type background arising from ''long'' products generated during subsequent amplification steps from carryover of the external primer. The step IA product can be diluted up to 10,000-fold and still provide enough template for robust amplification.

Step II: Recombination

Single Mutants/Crossovers

Procedure. Reactions are mixed on ice using the following:

5.8 μl of PCR master mix (as defined earlier)
1 μl of step IB reaction to give ~10 n*M* (or 1–5 ng/μl) gene fragment

1 μl plasmid DNA template (10 n*M* stock) to give ~200 p*M* final (1 ng/μl for a 7-kb plasmid)
12.2 μl filter-sterilized H_2O added to give a 20-μl reaction volume

The master mix is added last with pipetting to mix reactions. Cycling program: 96° for 5 min, followed by 15 cycles of 96° for 30 s (+2″/cycle), 55° for 30 s, and 72° for 1 min/kb of product followed by incubation at 4° at the completion of cycling.

Multiple Mutants/Crossovers

Procedure. Same as just described, except the gel-purified full-length mutant/chimeric gene (step III product) at ~1.0 ng/μl (~1 n*M* final concentration) is substituted for plasmid DNA.

Multiplex Recombination

Procedure. A mixture of gene fragments (step IB products) corresponding to a collection of mutations or alternative crossovers is pooled, and 1 μl (to give ~10 n*M*) is used with either a plasmid or a full-length mutant/chimeric gene (step III product) as the template in a recombination reaction.

Comments. The amount of full-length, single-stranded recombination product produced in step II is limited by the amount of gene fragment from step IB added to the reaction mixture. Gene fragments should ideally be 1- to 10-fold molar excess of the plasmid or mutant gene that it is recombining with. This is particularly important in the case of *single mutants/crossovers*, where only one terminus can be exploited in the following step for selective amplification. The plasmid concentration should be kept to a minimum; about 10 p*M* is the lowest concentration that can be used to give the amplifiable recombination product in step III.

Step III: Selective Amplification of Recombination Products

Amplification of Single Mutants/Crossovers

Procedure. Reactions are mixed on ice using the following:
14.5 μl of PCR master mix (as defined earlier)
2 μl secondary amplification primer (5 μ*M* stock) to give 0.2 μ*M*
2 μl primary amplification primer (5 μ*M* stock) to give 0.2 μ*M*
1 μl of step II reaction as template to give ~100–200 p*M* single-stranded DNA
30.5 μl filter-sterilized H_2O added to give a 50-μl reaction volume

The master mix is added last with pipetting to mix reactions. Cycling program: 96° for 5 min, followed by 30 cycles of 96° for 30 s, 55° for 30 s,

and 72° for 1 min/kb of product followed by an additional 10 min at 72° and incubation at 4° at the completion of cycling. Amplification products are verified by agarose gel electrophoresis.

Amplification of Multiple Mutants/Crossovers

Procedure. Same as *amplification of single mutants/crossovers*, but only secondary amplification primers are used.

Comments. The final step in a cycle of SCOPE is a standard PCR amplification of full-length mutant/chimeric genes with unique sequence tags at both 5′ and 3′ ends. In synthesis of the first generation of mutants, only one SAP can be used for selective amplification. This corresponds to the unique sequence of the PAP used in step IB. A PAP is directed at the opposite terminus, where it incorporates unique sequence at this terminus. Because this primer is directed to the flanking sequence of the gene, it can efficiently prime any carryover long product (single-stranded wild-type DNA) from step IB or any plasmid from step II. Therefore, it is important to eliminate the long product from step IB and minimize the amount of plasmid in step II, as the single-stranded product generated at this step has the potential to carry over into subsequent rounds of synthesis. In the step III amplification of multiple mutants/crossovers, SAP combinations are chosen to allow selective amplification of desired recombination products.

Product Isolation and Cloning

Full-length mutant genes from step III are gel purified using the Qiagen gel extraction kit according to the manufacturer's recommended procedures. Gel-purified attB PCR products are cloned into pDONR207 via the gateway BP reaction according to the manufacturer's recommendations.

Controls

Dpn*I Treatment.* Digestion is omitted and the step IB reaction is performed using the undigested step IA product as the template. Because plasmid DNA is carried over into the step IB reaction, PAPs can be extended to produce wild-type single-stranded DNA as described previously (Fig. 2A). As a result, wild-type genes can be amplified efficiently using a 1-μl portion of step IB as the template and a PAP and SAP primer pair. If the step IB reaction is performed using a 10-fold molar excess of mutagenic primer, then the amount of amplifiable wild-type gene decreases markedly. In fact, the combination of increasing the number of cycles in step IA to 100 (resulting in 2-fold more template) and using a 10-fold excess of internal mutagenic primer in step IB enables the suppression of

wild-type background and a mutagenesis efficiency of 80%, as apparent from terpene cyclase libraries produced in this manner.

Step IB Products. The selectivity of amplification (or the suppression of wild-type sequences) is evaluated using the step IB reaction as the template. If *Dpn*I digestion is complete, no amplifiable wild-type product is observed. In the case where restriction digestion is omitted, wild-type product is observed.

Results and Discussion

SCOPE was applied to create a library representing all possible combinations of nine point mutations in the terpene cyclase, TEAS. The location of mutations in the amino acid and nucleotide sequences of TEAS are indicated in Fig. 3. The nine positions were recombined as six units (shown in boxes). Some mutations were clustered, requiring a plurality of internal primers. For example, amino acid positions 436, 438, and 439 required a collection of seven internal primers to code for all permutations: three single, three double, and a triple mutant. A system was developed to introduce unique sequence tags; PAPs link mutation (or a collection thereof) to a unique sequence during gene fragment amplification and SAPs enable selective amplification of the desired combinations of mutations. An illustration of their use and the nomenclature system is described in Fig. 4.

An attribute of SCOPE synthesis is the fractionation of complex mixtures into many simpler ones. This has the benefit of reducing the numerical complexity and hence the screening requirements necessary to verify and identify the collection of desired changes. This effect arises from sampling probability as described by the following mathematical expression:

$$p(n) = 1 - \sum_{i=1}^{n-1} (-1)^{i+1} \frac{n!}{i!(n-i)!} \left[\frac{(n-i)}{n}\right]^k \tag{1}$$

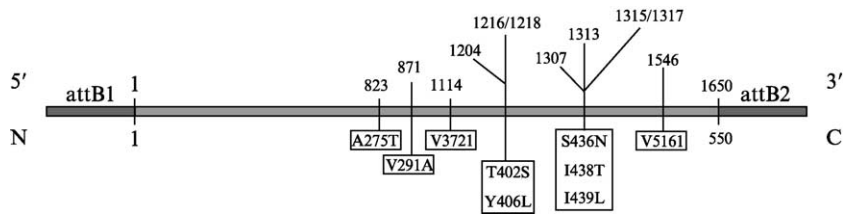


FIG. 3. Location of mutations in TEAS recombined by SCOPE. The nucleotide positions and corresponding amino acid changes are shown above and below, respectively

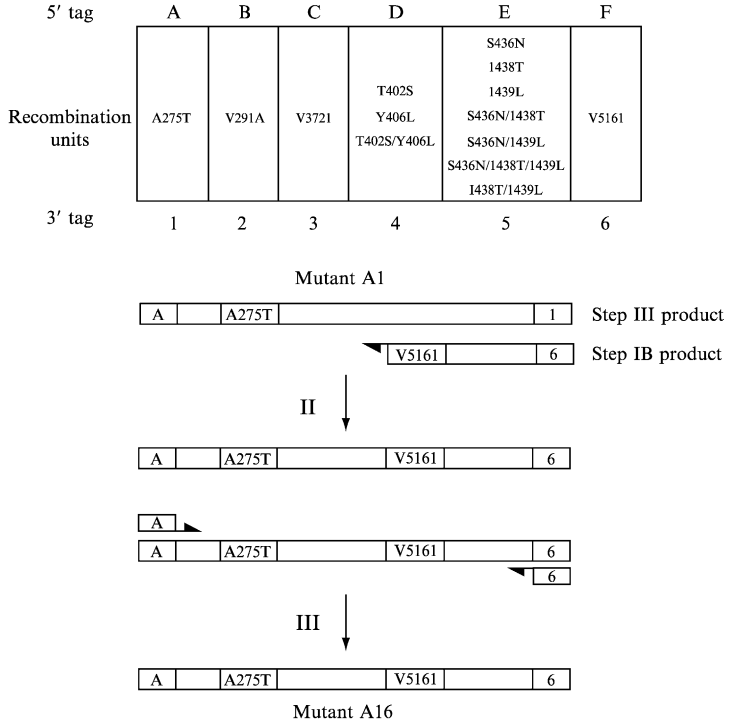


FIG. 4. Recombination units and tagging system. The recombined positions and their associated unique sequence tags give rise to a naming system to describe the recombination product created by SCOPE.

where k is the sample size, n is the number of unique members, and p is the probability that a sample of size k contains at least one representative of each unique member. As complexity increases, the amount of oversampling required to achieve the same probability of screening the library increases, where oversampling refers to sample size *(k)* in multiples of library complexity *(n)*. This can be shown graphically in Fig. 5.

Each iteration of the process ends with a conventional PCR amplification step, and after multiple iterations, additional mutations accumulate. The overall frequency of nondesired additional mutations in the population analyzed is 5.5%. No strong bias for the type of error or its location within the gene was observed. The nondesired mutation rate after the first round was 2.67%, which matches previous measures of pfu error frequency.[7] However, the random mutation rate increases as a function of

[7] J. Cline, J. C. Braman, and H. H. Hogrefe, *Nucleic Acids Res.* **24,** 3546 (1996).

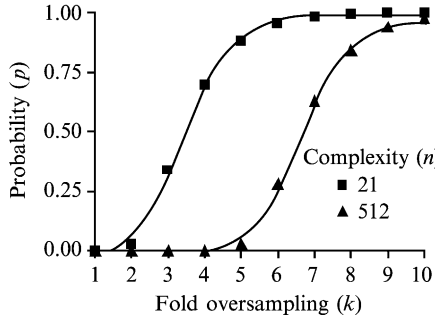


FIG. 5. Sampling probability as a function of oversampling. The probability that a sample contains one copy of each unique clone for a given complexity *(n)* is calculated using Eq. (1). Probability is calculated for a range of sample sizes *(k)* that are in multiples of a fixed library complexity *(n)* and the results are fit to a sigmoidal curve.

iterations of SCOPE and, after four iterations, reached 8.9%. Using a higher fidelity polymerase can minimize random mutation rates. Alternatively, products from step III amplification reactions can be cloned and the SCOPE cycle started anew (from step IA). Bridging oligonucleotides may be useful in this case to recombine various mutations. Also, gene fragments (from step IB) can be made to include multiple mutations from previous cycles.

In the example used here, 512 mutants were made from a series of simpler mixtures, the most complex of which contained 21 unique members. To achieve a 50% probability ($p = 0.5$) of identifying by screening every unique member in a mixture of that complexity requires 3.38-fold oversampling. To achieve the same probability of identifying all unique members of a library by screening a mixture of 512 unique possibilities requires 6.6-fold oversampling. Given the exponential relationship between sample size and library complexity, this difference equates to a reduction in numerical complexity of a factor of 25 for the entire library.

Library Analysis

Over 600 colonies from discrete mixtures, representing about half of the complexity of the TEAS library (241 unique members), were picked and their sequences determined. A summary of the results is listed in Table II. Of the clones sequenced, only 24 wild-type genes (3.5%) were found. This library was synthesized prior to addition of the *Dpn*I restriction step (as described earlier), and while the efficiency of the first round of mutagenesis was ~80%, the overall efficiency of the entire process reached 96.5%

TABLE II
SEQUENCE ANALYSIS RESULTS

Library statistics	
Clones sequenced	692
Wild-type genes	24
Percentage of mutants	96.5%
Complexity screened	241
Unique clones identified	193
Fold oversampled	2.8
Complexity covered	80.1%
Total library complexity	512
Percentage of verified mutants	37.70%
Additional mutations	
Silent	9
Frameshift	16
Point mutants	13
Total	38
Mutation rate	5.49%

(Table II). Mutations become incorporated into the wild-type sequence during recombination reactions in subsequent iterations of the process. As a result, wild-type sequences vanish in multiple crossover populations.

Aside from the low-level appearance of wild-type sequence and random mutations likely arising from PCR errors, the actual distribution of mutations in a given mixture was as designed experimentally. Some recombination reactions produce a single product having several designed mutations, such as A1236. In reactions containing multiple mutations, the reaction distribution appears random.

Concluding Remarks

Adaptation of SCOPE to combinatorial mutant library design and construction demonstrates the broader utility of these library construction principles. While various techniques have been developed for either homology-independent recombination or combinatorial mutagenesis, none can efficiently do both. SCOPE provides an effective means for the creation of both global or local sequence space as demonstrated by the synthesis of DNA libraries representing the genetically encoded information spanned by distant homologues[1] or closely related members of a gene family.

Acknowledgments

We are grateful to the National Institutes of Health for the grants that supported this work (GM43268 to M.D.T. and GM54029 to J.C. and J.P.N.). P.E.O. is an NIH Postdoctoral Research Fellow (GM069056-01). Additionally, we thank Marina Bakhtina and Brandon Lamarch for valuble consultations during the early phases of this work.

[9] New Enzymes from Combinatorial Library Modules

By WERNER BESENMATTER, PETER KAST, and DONALD HILVERT

Introduction

Current strategies for the construction of combinatorial gene libraries for directed evolution experiments generally make use of cassette mutagenesis[1,2] to insert library modules[3–6] into plasmids. We have applied this technique in a variety of formats to investigate chorismate mutase, a key enzyme in the biosynthesis of aromatic amino acids.[7] Active variants are directly selected from gene libraries transformed into a chorismate mutase-deficient *Escherichia coli* strain (Fig. 1).[8] Because catalytic activity is an extremely sensitive probe for protein integrity, a wealth of information on structural and functional aspects of this enzyme can be derived from sequence patterns in selected variants.[9]

The extent of randomization of the gene library cassettes depends on the questions asked. For instance, to investigate the roles of individual active site residues, one or two codons were randomized at a time.[8,10] When loops connecting secondary structural elements were (re-)designed, we opted for formats mutagenizing three to seven codons

[1] J. A. Wells, M. Vasser, and D. B. Powers, *Gene* **34,** 315 (1985).
[2] J. F. Reidhaar-Olson and R. T. Sauer, *Science* **241,** 53 (1988).
[3] S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, and M. H. Hecht, *Science* **262,** 1680 (1993).
[4] G. Cho, A. D. Keefe, R. Liu, D. S. Wilson, and J. W. Szostak, *J. Mol. Biol.* **297,** 309 (2000).
[5] S. V. Taylor, K. U. Walter, P. Kast, and D. Hilvert, *Proc. Natl. Acad. Sci. USA* **98,** 10596 (2001).
[6] T. Matsuura, A. Ernst, and A. Plückthun, *Protein Sci.* **11,** 2631 (2002).
[7] E. Haslam, "Shikimic Acid: Metabolism and Metabolites." Wiley, Chichester, UK, 1993.
[8] P. Kast, M. Asif-Ullah, N. Jiang, and D. Hilvert, *Proc. Natl. Acad. Sci. USA* **93,** 5043 (1996).
[9] S. V. Taylor, P. Kast, and D. Hilvert, *Angew. Chem. Int. Ed.* **40,** 3310 (2001).
[10] P. Kast, J. D. Hartgerink, M. Asif-Ullah, and D. Hilvert, *J. Am. Chem. Soc.* **118,** 3069 (1996).

0076-6879/04 $35.00