The invention provides methods for evolving a polynucleotide toward acquisition of a desired property. Such methods entail incubating a population of parental polynucleotide variants under conditions to generate annealed polynucleotides comprising heteroduplexes. The heteroduplexes are then exposed to a cellular DNA repair system to convert the heteroduplexes to parental polynucleotide variants or recombined polynucleotide variants. The resulting polynucleotides are then screened or selected for the desired property.


Statutory Declaration of Mae Li Gan in Australian Opposition against application 703264.

Statutory Declaration of Ngarie Pepit-Young in Australian Opposition against application 703264.


* cited by examiner
FIG. 3
RESTRICTION (A, B, X AND Y)

LIGATION

TRANSFORMATION

TRANSFORMANTS

FIG. 4
**FIG. 6A**

```
Bam III
R. lupini H13-3 flaA

pRL20
(4.2 kb)

pUC18
```

**FIG. 6B**

```
R. meliloti flaA

pRM40
(4.3 kb)

Eco RI

pUC18
```
FIG. 7C

ATGACGAGCATTTCACCAACAACTCCGCAATGGCCGCGCTTTCCCGAGTGCGCTGATCTCTTCCAGCATGGAAAGACACCGGAGCCGATCTCCCTCGG
GCCTCGGTCTGCGTTCCGCGCTGCAAAAGCCCGCTACTGCGATTCGGACAATGCGTCTCCGAGACAAAGGACCCTTTTGGGCGGCAGAGCCCT
CGGCTCGGCGCCCGCAAGTTGATACCCTGTACCTGGTATGGAAT CGGATCGGAAATCCTGTTAAGGAAATCAAGCCAGCGTCTGAGTGCGCAGGAA
GACGGCGTCGAACAGGCAAGATCCAGAAGAATACCTCAGCTCAAGACCAGCGTACGAGCAAGTGGCCCGACGCGCTTCCCTTTGCGACTGGGC
TGACGGGACC...TCAGCGGGCGCGCGCTCAACAGAGGCGCGCTCGCTCGTGACGAGGCGCGTTGCTCGTACGACAGTTGACACATCGATTAC
GCTGGAAATGCTTTCAAGGTCTCTGAGTGATACCCCGGAACGCTCGGATACCCCGGACTCTGGACAGCTACAAGCTCTCGACAGCGGACGTCAGG
TGACGGG.....TGACCGGCAACGCGCGGACACGCAATGGCTGGCTCTATCTCCTTGGA.....GTCCCTACTCCAGGCCGCTGGGAGTT
CCAGGGCAATCTGCTTCTCAGGGCGCTAACAGATCTACGTCAAGGCGCCGAGCTGGTTAACGCTGACGCTGGGCCTCCATACCGCGATCA
ACACCGATGCCTGGCGAAGTTGGCCCGCGCTTACAACCGCGCGTGAAGCTCGTACTGACGCTGCGTCCGCGGATGACCCATCTCGTGAGAAACCAACAGG
GGCCGGTGCAAGTTAAACCCTCAACCAGTGCGTCTGGACATGGTACGCTGCAAGCTCGAGAATGCGGACTACCGCAAAAGG
CTCTACACAGCGCGAAGCTGGCGGCGCTGAACCTCGGCTCATTCTCGGGAGACACGCGATTCGTCGTTAACGCTGCAAGTCTCGGAGATCG
GTCGGGCGTGCGCGCTACGCGCGCTGGCCATGAACAGGAATCTCGTGCCAGAGCAGCTCAGCGCAGTGGCCCTCAAGGGCTCGCCGATCGCCAGGCGG
ATCGCCAACACTGGAGACTGCGAGGACAGACGGTCTCGCTTCCGCTAA
FIG. 7D
FIG. 11
FIG. 12
FIG. 13A
FIG. 13A
(CONTINUED)
FIG. 13A
(CONTINUED)
FIG. 13A
(continued)
FIG. 13A
(CONTINUED)
CLONED pBE3-1 AND pBE3-2

AMPLIFICATION AND DEMETHYLATION

RESTRICTION WITH ENZYME Bam HI

DEMETHYLATED TARGET MOLECULES

HETERODUPLEX POOL (SHOWN ONLY INSERT WITH PARTS OF VECTOR)

HETERODUPLEX FORMATION

TRANSFORMATION AND REPAIR

RECOMBINANTS (SHOWN ONLY THE INSERTS)

SCREENING

RECOMBINANTS (SHOWN ONLY THE INSERTS)

FIG. 14
METHOD FOR CREATING POLYNUCLEOTIDE AND POLYPEPTIDE SEQUENCES

CROSS-REFERENCES TO RELATED APPLICATIONS

This application derives priority from U.S. Ser. No. 60/067,908, filed Dec. 8, 1997, which is incorporated by reference in its entirety for all purposes.

STATEMENT OF GOVERNMENT INTEREST

The invention described herein was made in the performance of work under a NASA contract, and is subject to the provisions of Public Law 96-517 (35 USC §202) in which the contractor has elected to retain title.

COPYRIGHT NOTICE

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

TECHNICAL FIELD

The invention resides in the technical field of genetics, and more specifically, forced molecular evolution of polynucleotides to acquire desired properties.

BACKGROUND

A variety of approaches, including rational design and directed evolution, have been used to optimize protein functions (1, 2). The choice of approach for a given optimization problem depends, in part, on the degree of understanding of the relationships between sequence, structure and function. Rational redesign typically requires extensive knowledge of a structure-function relationship. Directed evolution requires little or no specific knowledge about structure-function relationship; rather, the essential features is a means to evaluate the function to be optimized. Directed evolution involves the generation of libraries of mutant molecules followed by selection or screening for the desired function. Gene products which show improvement with respect to the desired property or set of properties are identified by selection or screening. The gene(s) encoding those products can be subjected to further cycles of the process in order to accumulate beneficial mutations. This evolution can involve few or many generations, depending on how far one wishes to progress and the effects of mutations typically observed in each generation. Such approaches have been used to create novel functional nucleic acids (3, 4), peptides and other small molecules (5), antibodies (3), as well as enzymes and other proteins (5, 6, 7). These procedures are fairly tolerant to inaccuracies and noise in the function evaluation (7).


A PCR-based group of recombination methods consists of DNA shuffling [5, 6], staggered extension process [89, 90] and random-priming recombination [87]. Such methods typically involve synthesis of significant amounts of DNA during assembly/recombination step and subsequent amplification of the final products and the efficiency of amplification decreases with gene size increase.

Yeast cells, which possess an active system for homologous recombination, have been used for in vivo recombination. Cells transformed with a vector and partially overlapping inserts efficiently join the inserts together in the regions of homology and restore a functional, covalently-closed plasmid [91]. This method does not require PCR amplification at any stage of recombination and therefore is free from the size considerations inherent in this method. However, the number of crossovers introduced in one recombination event is limited by the efficiency of transformation of one cell with multiple inserts. Other in vivo recombination methods entail recombination between two parental genes cloned on the same plasmid in a tandem orientation. One method relies on homologous recombination machinery of bacterial cells to produce chimeric genes [92]. A first gene in the tandem provides the N-terminal part of the target protein, and a second provides the C-terminal part. However, only one crossover can be generated by this approach. Another in vivo recombination method uses the simultaneous organization of substrates in a vector [93]. Before transformation into E. coli cells, plasmids are linearized by endonuclease digestion between the parental sequences. Recombination is performed in vivo by the enzymes responsible for double-strand break repair. The ends of linear molecules are degraded by a 5' exonuclease activity, followed by annealing of complementary single-strand 3' ends and restoration of the double-strand plasmid [94]. This method has similar advantages and disadvantages of tandem recombination on circular plasmid.

SUMMARY OF THE INVENTION

The invention provides methods for evolving a polynucleotide toward acquisition of a desired property. Such methods entail incubating a population of parental polynucleotide variants under conditions to generate annealed polynucleotides comprises heteroduplexes. The heteroduplexes are then exposed to a cellular DNA repair system to convert the heteroduplexes to parental polynucleotide variants or recombined polynucleotide variants. The resulting polynucleotides are then screened or selected for the desired property.

In some methods, the heteroduplexes are exposed to a DNA repair system in vitro. A suitable repair system can be prepared in the form of cellular extracts.

In other methods, the products of annealing including heteroduplexes are introduced into host cells. The heteroduplexes are thus exposed to the host cells' DNA repair system in vivo.

In several methods, the introduction of annealed products into host cells selects for heteroduplexes relative to transformed cells comprising homoduplexes. Such can be achieved, for example, by providing a first polynucleotide variant as a component of a first vector, and a second polynucleotide variant is provided as a component of a second vector. The first and second vectors are converted to linearized forms in which the first and second polynucleotide variants occur at opposite ends. In the incubating step, single-stranded forms of the first linearized vector reanneal with each other to form linear first vector, single-stranded forms of the second linearized vector reanneal with each other to form linear second vector, and single-stranded
variants comprises at least 5 polynucleotides having at least
by treatment with DNA ligase.
In some methods, the polynucleotide variants encode variants of a polypeptide. In
some methods, the population of variant polynucleotide variants are at least 10 kb in length. In
some methods, the polynucleotides are provided in double stranded form and are converted to single stranded form before the annealing step. Optionally, such conversion is by conducting asymmetric amplification of the first and second double stranded polynucleotide variants to amplify a first strand of the first polynucleotide variant, and a second strand of the second polynucleotide variant. The first and second strands anneal in the incubating step to form a heteroduplex.
In some methods, a population of polynucleotides comprising first and second polynucleotides is provided in double stranded form, and the method further comprises incorporating the first and second polynucleotides as components of first and second vectors, whereby the first and second polynucleotides occupy opposite ends of the first and second vectors. In the incubating step single-stranded forms of the first linearized vector reanneal with each other to form linear first vector, single-stranded forms of the second linearized vector reanneal with each other to form linear second vector, and single-stranded linearized forms of the first and second vectors anneal with each to form a circular heteroduplex bearing a nick in each strand. In the introducing step selects for transformed cells comprises the circular heteroduplexes relative to the linear first and second vector.
In some methods, the first and second polynucleotides are obtained from chromosomal DNA. In some methods, the polynucleotide variants encode variants of a polypeptide. In some methods, the population of polynucleotide variants comprises at least 20 variants. In some methods, the population of polynucleotide variants are at least 10 kb in length.
In some methods, the polynucleotide variants comprises natural variants. In other methods, the polynucleotide variants comprise variants generated by mutagenic PCR or cassette mutagenesis. In some methods, the host cells into which heteroduplexes are introduced are bacterial cells. In some methods, the population of polynucleotide variants comprises at least 20 variants. In some methods, the population of polynucleotide variants are at least 10 kb in length.
In some methods, the polynucleotide variants comprises natural variants. In other methods, the polynucleotide variants comprise variants generated by mutagenic PCR or cassette mutagenesis. In some methods, the host cells into which heteroduplexes are introduced are bacterial cells. In some methods, the population of variant polynucleotide variants comprises at least 5 polynucleotides having at least 90% sequence identity with one another.
Some methods further comprise a step of at least partially demethylating variant polynucleotides. Demethylation can be achieved by PCR amplification or by passing variants through methylation-deficient host cells.
Some methods include a further step of sealing one or more nicks in heteroduplex molecules before exposing the heteroduplexes to a DNA repair system. Nicks can be sealed by treatment with DNA ligase.
Some methods further comprise a step of isolating a screened recombinant polynucleotide ariant. In some methods, the polynucleotide variant is screened to produce a recombinant protein or a secondary metabolite whose production is catalyzed thereby.
In some methods, the recombinant protein or secondary metabolite is formulated with a carrier to form a pharmaceutical composition.
In some methods, the polynucleotide variants encode enzymes selected from the group consisting of proteases, lipases, amylases, cutinases, cellulases, amylases, oxidases, peroxidases and phytases. In other methods, the polynucleotide variants encode a polypeptide selected from the group consisting of insulin, ACTH, glucose, somatostatin, somatotropin, thymosin, parathyroid hormone, pigmented hormones, somatomedin, erythropoietin, luteinizing hormone, chorionic gonadotropin, hyperalgal releasing factors, antidiuretic hormones, thyroid stimulating hormone, relaxin, interferon, thrombopoietic (TPO), and prolactin.
In some methods, each polynucleotide in the population of variant polynucleotides encodes a plurality of enzymes forming a metabolic pathway.

BRIEF DESCRIPTION OF THE DRAWINGS
FIG. 1 illustrates the process of heteroduplex formation using polymerase chain reaction (PCR) with one set of primers for each different sequence to amplify the target sequence and vector.
FIG. 2 illustrates the process of heteroduplex formation using restriction enzymes to linearize the target sequences and vector.
FIG. 3 illustrates a process of heteroduplex formation using asymmetric or single primer polymerase chain reaction (PCR) with one set of primers for each different sequence to amplify the target sequence and vector.
FIG. 4 illustrates heteroduplex recombination using unique restriction enzymes (X and Y) to remove the homoduplexes.
FIG. 5 shows the amino acid sequences of the FlaA from R. lupini (SEQ ID NO: 1) and R. mellii (SEQ ID NO: 2).
FIG. 6 shows the locations of the unique restriction sites utilized to linearize pRL20 and pRM40.
FIGS. 7A, B, C and D show the DNA sequences of four mosaic flaA genes created by in vitro heteroduplex formation followed by in vivo repair (a) is SEQ ID NO:3, (b) is SEQ ID NO:4; (c) is SEQ ID NO:5 and (d) is SEQ ID NO:6.
FIG. 8 illustrates how the heteroduplex repair process created mosaic flaA genes containing sequence information from both parent genes.
FIG. 9 shows the physical maps of Actinoplanes utuhensis ECB deacylase mutants with enhanced specific activity ((a) is pM7-2 for Mutant 7-2, and (b) is pM16 for Mutant 16).
FIG. 10 illustrates the process used for Example 2 to recombine mutations in Mutant 7-2 and Mutant 16 to yield ECB deacylase recombinant with more enhanced specific activity.
FIG. 11 Specific activities of wild-type ECB deacylase and improved mutants Mutant 7-2, Mutant 16 and recombinant Mutant 15.
FIG. 12. Positions of DNA base changes and amino acid substitutions in recombinant ECB deacylase Mutant 15 with respect to parental sequences of Mutant 7-2 and Mutant 16.
FIGS. 13A, B, C, D and E show the DNA sequences of A. utuhensis ECB deacylase gene mutant M-15 genes created by in vitro heteroduplex formation followed by in vivo repair (SEQ ID NO:7).
FIG. 14 illustrates the process used for Example 3 to recombine mutations in RC1 and RC2 to yield thermostable subtilisin E.
FIG. 15 illustrates the sequences of RC1 and RC2 and the ten clones picked randomly from the transformants of the reaction products of duplex formation as described in Example 3. The x’s correspond to base positions that differ between RC1 and RC2. The mutation at 995 corresponds to amino acid substitution at 181, while that at 1107 corre-
sponds to an amino acid substitution at 218 in the subtilisin protein sequence.

FIG. 16 shows the results of screening 400 clones from the library created by heteroduplex formation and repair for initial activity (A) and residual activity (A). The ratio A/A was used to estimate the enzymes' thermostability. Data from active variants are sorted and plotted in descending order. Approximately 12.9% of the clones exhibit a phenotype corresponding to the double mutant containing both the N181D and the N218S mutations.

DEFINITIONS

Screening is, in general, a two-step process in which one first physically separates the cells and then determines which cells do and do not possess a desired property. Selection is a form of screening in which identification and physical separation are achieved simultaneously by expression of a selection marker, which, in some genetic circumstances, allows cells expressing the marker to survive while other cells die (or vice versa). Exemplary screening members include luciferase, β-galactosidase and green fluorescent protein. Selection markers include drug and toxin resistance genes. Although spontaneous selection can and does occur in the course of natural evolution, in the present methods selection is performed by man.

An exogenous DNA segment is one foreign (or heterologous) to the cell or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides.

The term gene is used broadly to refer to any segment of DNA associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. Genes also include non-expressed DNA segments that, for example, form recognition sequences for other proteins.

The term “wild-type” means that the nucleic acid fragment does not comprise any mutations. A “wild-type” protein means that the protein will be active at a level of activity found in nature and typically will comprise the amino acid sequence found in nature. In an aspect, the term “wild type” or “parental sequence” can indicate a starting or reference sequence prior to a manipulation of the invention.

“Substantially pure” means an object species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual macromolecular species in the composition), and preferably a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present in the composition. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species.

Percentage sequence identity is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison. Optimal alignment of sequences for aligning a comparison window can be conducted by computerized implementations of algorithms GAP, BESTFIT, EUSTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.

The term naturally-occurring is used to describe an object that can be found in nature as distinct from being artificially produced by man. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. Generally, the term naturally-occurring refers to an object as present in a non-pathological (undiseased) individual, such as would be typical for the species.

A nucleic acid is operably linked when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it increases the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by several kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous.

A specific binding affinity between, for example, a ligand and a receptor, means a binding affinity of at least 1 x 10^-16 M^-1.

The term “cognate” as used herein refers to a gene sequence that is evolutionarily and functionally related between species. For example but not limitation, in the human genome, the human CD4 gene is the cognate gene to the mouse CD4 gene, since the sequences and structures of these two genes indicate that they are highly homologous and both genes encode a protein which functions in signaling T cell activation through MHC class II-restricted antigen recognition.

The term “heteroduplex” refers to hybrid DNA generated by base pairing between complementary single strands derived from the different parental duplex molecules, whereas the term “homoduplex” refers to double-stranded DNA generated by base pairing between complementary single strands derived from the same parental duplex molecules.

The term “nick” in duplex DNA refers to the absence of a phosphodiester bond between two adjacent nucleotides on one strand. The term “gap” in duplex DNA refers to an absence of one or more nucleotides in one strand of the duplex. The term “loop” in duplex DNA refers to one or more unpaired nucleotides in one strand.

A mutant or variant sequence is a sequence showing substantial variation from a wild type or reference sequence that differs from the wild type or reference sequence at one or more positions.

DETAILED DESCRIPTION

I. General

The invention provides methods of evolving a polynucleotide toward acquisition of a desired property. The substrates for the method and a population of at least two polynucleotide variant sequences that contain regions of similarity with each other but, which also have point(s) or regions of divergence. The substrates are annealed in vitro at the
regions of similarity. Annealing can regenerate initial substrates or can form heteroduplexes, in which component strands originate from different parents. The products of annealing are exposed to enzymes of a DNA repair, and optionally a replication system, that repairs unmatched pairings. Exposure can be in vivo as when annealed products are transformed into host cells and exposed to the host's DNA repair system. Alternatively, exposure can be in vitro, as when annealed products are exposed to cellular extracts containing functional DNA repair systems. Exposure of heteroduplexes to a DNA repair system results in DNA repair at bulges in the heteroduplexes due to DNA mismatching. The repair process differs from homologous recombination in promoting nonreciprocal exchange of diversity between strands. The DNA repair process is typically effected on both component strands of a heteroduplex molecule and at any particular mismatch is typically random as to which strand is repaired. The resulting population can thus contain recombinant polynucleotides encompassing an essentially random reassortment of points of divergence between parental strands. The population of recombinant polynucleotides is then screened for acquisition of a desired property. The property can be a property of the polynucleotide sequence or thereof, such as mRNA or a protein.

II. Substrates For Shuffling

The substrates for shuffling are variants of a reference polynucleotide that show some region(s) of similarity with the reference and other region(s) or point(s) of divergence. Regions of similarity should be sufficient to support annealing of polynucleotides such that stable heteroduplexes can be formed. Variants forms often show substantial sequence identity with each other (e.g., at least 50%, 75%, 90% or 99%). There should be at least sufficient diversity between substrates that recombination can generate more diverse products than there are starting materials. Thus, there must be at least two substrates differing in at least two positions. The degree of diversity depends on the length of the substrate being recombined and the extent of the functional change to be evolved. Diversity at between 0.1–25% of positions is typical. Recombination of mutations from very closely related genes or even whole sections of sequences from more distantly related genes or sets of genes can enhance the rate of evolution and the acquisition of desirable new properties. Recombination to create chimeric or mosaic genes can be useful in order to combine desirable features of two or more parents into a single gene or set of genes, or to create novel functional features not found in the parents. The number of different substrates to be combined can vary widely in size from two to 10, 100, 1000, to more than 10^2, 10^3, or 10^4 members.

The initial small population of the specific nucleic acid sequences having mutations may be created by a number of different methods. Mutations may be created by error-prone PCR. Error-prone PCR uses low-fidelity polymerization conditions to introduce a low level of point mutations randomly over a long sequence. Alternatively, mutations can be introduced into the template polynucleotide by oligonucleotide-directed mutagenesis. In oligonucleotide-directed mutagenesis, a short sequence of the polynucleotide is removed from the polynucleotide using restriction enzyme digestion and is replaced with a synthetic polynucleotide in which various bases have been altered from the original sequence. The polynucleotide sequence can also be altered by chemical mutagenesis. Chemical mutagens include, for example, sodium bisulfite, nitrous acid, hydroxylamine, hydrazine or formic acid. Other agents which are analogues of nucleotide precursors include nitrosoanidine, 5-bromouracil, 2-aminopurine, or acridine. Generally, these agents are added to the PCR reaction in place of the nucleotide precursor thereby mutating the sequence. Intercalating agents such as proflavine, acriflavine, quinacrine and the like can also be used. Random mutagenesis of the polynucleotide sequence can also be achieved by irradiation with X-rays or ultraviolet light. Generally, plasmid DNA or DNA fragments so mutagenized are introduced into E. coli and propagated as a pool or library of mutant plasmids.

Alternatively the small mixed population of specific nucleic acids can be found in nature in the form of different alleles of the same gene or the same gene from different related species (i.e., cognate genes). Alternatively, substrates can be related but nonallelic genes, such as the immunoglobulin genes. Diversity can also be the result of previous recombination or shuffling. Diversity can also result from resynthesizing genes encoding natural proteins with alternative codon usage.

The starting substrates encode variant forms of sequences to be evolved. In some methods, the substrates encode variant forms of a protein for which evolution of a new or modified property is desired. In other methods, the substrates can encode variant forms of a plurality of genes constituting a multigene pathway. In such methods, variation can occur in one or any number of the component genes. In other methods, substrates can contain variants segments to be evolved as DNA or RNA binding sequences. In methods, in which starting substrates containing coding sequences, any essential regulatory sequences, such as a promoter and polyadenylation sequence, required for expression may also be present as a component of the substrate. Alternatively, such regulatory sequences can be provided as components of vectors used for cloning the substrates.

The starting substrates can vary in length from about 50, 250, 1000, 10,000, 100,000, 10^2 or more bases. The starting substrates can be provided in double- or single-stranded form. The starting substrates can be DNA or RNA and analogs thereof. If DNA, the starting substrates can be genomic or cDNA. If the substrates are RNA, the substrates are typically reverse-transcribed to cDNA before heteroduplex formation. Substrates can be provided as cloned fragments, chemically synthesized fragments or PCR amplification products. Substrates can derive from chromosomal, plasmid or viral sources. In some methods, substrates are provided in concatemeric form.

III. Procedures for Generating Heteroduplexes

Heteroduplexes are generated from double stranded DNA substrates, by denaturing the DNA substrates and incubating under annealing conditions. Hybridization conditions for heteroduplex formation are sequence-dependent and are different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, hybridization conditions are selected to be about 25°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium.

Exemplary conditions for denaturation and renaturation of double stranded substrates are as follows. Equimolar concentrations (1.0–5.0 mM) of the substrates are mixed in
plexes relative to reformed parental homoduplexes. Homoduplexes of any other substrate, but strands of the same vector with the substrate and the P1P2 primer pair initiates are then respectively digested with different restriction points on two vectors bearing substrates provided that the molecules cut with three or more enzymes, cutting at three products generated by this amplification are double-stranded different restriction enzymes that cut at different places.

The process is the same for single stranded DNA substrates, amplification product might have most of the vector on the primers in each primer pair prime amplification in opposite vectors. In the general case, such amplification generates forms differing in the division of vector sequences flanking steps. In general, selection is achieved by designing substrates such that heteroduplexes are formed in open-circles, whereas homoduplexes are formed as linear molecules. A subsequent transformation step results in substantial enrichment (e.g., 100-fold) for the circular heteroduplexes.

FIG. 1 shows a method in which two substrate sequences in plasmid vectors are PCR-amplified using two different sets of primers (P1, P2 and P3, P4). Typically, first and second substrates are inserted into separate copies of the same vector. The two different pairs of primers initiate amplification at different points on the two vectors. FIG. 1 shows an arrangement in which the P1/P2 primer pairs initiate amplification at one of the boundaries of the vector with the substrate and the P1/P2 primer pair initiates amplification at the other boundary in a second vector. The two primers in each primer pair prime amplification in opposite directions around a circular plasmid. The amplification products generated by this amplification are double-stranded linearized vector molecules in which the first and second substrates occur at opposite ends of the vector. The amplification products are mixed, denatured and annealed. Mixing and denaturation can be performed in either order. Reannealing generates two linear homoduplexes, and an open circular heteroduplex containing one nick in each strand, at the initiation point of PCR amplification. Introduction of the amplification products into host cells selects for the heteroduplexes similar to that shown in FIG. 1, except that the nicks occur within the vector component rather than at the interface between plasmid and substrate. Initiation of amplification outside the substrate component of a vector has the advantage that it is not necessary to design primers specific for the substrate borne by the vector.

Although FIG. 1 is exemplified for two substrates, the above scheme can be extended to any number of substrates. For example, an initial population of vector bearing substrates can be divided into two pools. One pool is PCR-amplified from one set of primers, and the other pool from another. The amplification products are denatured and annealed as before. Heteroduplexes can form containing one strand from any substrate in the first pool and one strand from any substrate in the second pool. Alternatively, three or more substrates cloned into multiple copies of a vector can be subjected to amplification with amplification in each vector starting at a different point. For each substrate, this process generates amplification products varying in how flanking vector DNA is divided on the two sides of the substrate. For example, one amplification product might have most of the vector on one side of the substrate, another amplification product might have most of the vector on the other side of the substrate, and a further amplification product might have an equal division of vector sequence flanking the substrate. In the subsequent annealing step, a strand of substrate can form a circular heteroduplex with a strand of any other substrate, but strands of the same substrate can only reanneal with each other to form a linear homoduplex. In a still further variation, multiple substrates can be performed by performing multiple iterations of the scheme in FIG. 1. After the first iteration, recombinant polynucleotides in a vector, undergo heteroduplex formation with a third substrate incorporated into a further copy of the vector. The vector bearing the recombinant polynucleotides and the vector bearing the third substrate are separately PCR amplified from different primer pairs. The amplification products are then denatured and annealed. The process can be repeated further times to allow recombination with further substrates.

An alternative scheme for heteroduplex formation is shown in FIG. 2. Here, first and second substrates are incorporated into separate copies of a vector. The two copies are then respectively digested with different restriction enzymes. FIG. 2 shows an arrangement in which the restriction enzymes cut at opposite boundaries between substrates and vector, but all that is necessary is to use two different restriction enzymes that cut at different places. Digestion generates linearized first and second vector bearing first and second substrates, the first and second substrates occupying different positions relative to the remaining vector sequences. Denaturation and reannealing generates open circular heteroduplexes and linear homoduplexes. The scheme can be extended to recombination between more than two substrates using analogous strategies to those described with respect to FIG. 1. In one variation, two pools of substrates are formed, and each is separately cloned into vector. The two pools are then cut with different enzymes, and reannealing proceeds as for two substrates. In another variation, three or more substrates can be cloned into three or more copies of vector, and the three or more result molecules cut with three or more enzymes, cutting at three or more sites. This generates three different linearized vector forms differing in the division of vector sequences flanking the substrate moiety in the vectors. Alternatively, any number of substrates can be recombined pairwise in an iterative fashion with products of one round of recombination annealing with a fresh substrate in each round.

In a further variation, heteroduplexes can be formed from substrates molecules in vector-free form, and the heteroduplexes subsequently cloned into vectors. Such can be achieved by asymmetric amplification of first and second substrates as shown in FIG. 3. Asymmetric or single primer PCR amplifies only one strand of a duplex. By appropriate selection of primers, opposite strands can be amplified from two different substrates. On reannealing amplification products, heteroduplexes are formed from opposite strands of the two substrates. Because only one strand is amplified from each substrate, reannealing does not reform homoduplexes (other than for small quantities of unamplified substrate). The process can be extended to allow recombination of any number of substrates using analogous strategies to those described with respect to FIG. 1. For example,
substrates can be divided into two pools, and each pool subject to the same asymmetric amplification, such that amplification products of one pool can only anneal with amplification products of the other pool, and not with each other. Alternatively, shuffling can proceed pairwise in an iterative manner, in which recombinants formed from heteroduplexes of first and second substrates, are subsequently subjected to heteroduplex formation with a third substrate. Point mutations can also be introduced at a desired level during PCR amplification.

FIG. 4 shows another approach of selecting for heteroduplexes relative to homoduplexes. First and second substrates are isolated by PCR amplification from separate vectors. The substrates are denatured and allowed to anneal forming both heteroduplexes and reconstructed homoduplexes. The products of annealing are digested with restriction enzymes X and Y. This site is in the first substrate but not the second substrate, and vice versa for Y. Enzyme X cuts reconstructed homoduplex from the first substrate and enzyme Y cuts reconstructed homoduplex from the second substrate. Neither enzyme cuts heteroduplexes. Heteroduplexes can effectively be separated from restriction fragments of homoduplexes by further cleavage with enzymes A and B having sites proximate to the ends of both the first and second substrates, and ligation of the products into vector having cohesive ends compatible with ends resulting from digestion with A and B. Only heteroduplexes cut with A and B can ligate with the vector. Alternatively, heteroduplexes can be separated from restriction fragments of homoduplexes by size selection on gels. The above process can be generalized to N substrates by cleaving the mixture of heteroduplexes and homoduplexes with N enzymes, each one of which cuts a different substrate and no other substrate. Heteroduplexes can be formed by directional cloning. Two substrates for heteroduplex formation can be obtained by PCR amplification of chromosomal DNA and joined to opposite ends of a linear vector. Directional cloning can be achieved by digesting the vector with two different enzymes, and digesting or adapting first and second substrates to be respectively compatible with cohesive ends of only of the two enzymes used to cut the vector. The first and second substrates can thus be ligated at opposite ends of a linearized vector fragment. This scheme can be extended to any number of substrates by using principles analogous to those described for FIG. 1. For example, substrates can be divided into two pools before ligation to the vector. Alternatively, recombinant products formed by heteroduplex formation of first and second substrates, can subsequently undergo heteroduplex formation with a third substrate.

IV. Vectors and Transformation

In general, substrates are incorporated into vectors either before or after the heteroduplex formation step. A variety of cloning vectors typically used in genetic engineering are suitable. The vectors containing the DNA segments of interest can be transferred into the host cell by standard methods, depending on the type of cellular host. For example, calcium chloride transformation is commonly utilized for prokaryotic cells, whereas calcium phosphate treatment, Lipofection, or electroporation may be used for other cellular hosts. Other methods used to transform mammalian cells include the use of Polybrene, protoplast fusion, liposomes, electroporation, and microinjection, and biolistics (see, generally, Sambrook et al., supra). Viral vectors can also be packaged in vitro and introduced by infection. The choice of vector depends on the host cells. In general, a suitable vector has an origin of replication recognized in the desired host cell, a selection marker capable of being expressed in the intended host cells and/or regulatory sequences to support expression of genes within substrates being shuffled.

V. Types of Host Cells

In general any type of cells supporting DNA repair and replication of heteroduplexes introduced into the cells can be used. Cells of particular interest are the standard cell types commonly used in genetic engineering, such as bacteria, particularly, *E. coli* (16, 17). Suitable *E. coli* strains include *E. coli* mutS, mutL, dam-, and/or recA*, *E. coli* XL-1-Gold ([Tet/ΔmerA]183 ΔmerCB-λisdsmr-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 recA1 lacHPro [FproAB lacZAM15 Tn10 (Tet') Amy Cam']*, *E. coli* S1301 mutS [Genotype: lacZ53, mutS201::Tn5, thyA36, rha-5, metH1, deoC, In(rrnD-rrnE)], (20, 24, 28–42). Preferred *E. coli* strains are *E. coli* SCS 110 [Genotype: rpsL (Sir), thr, leu, endA, thi-1, lacY, galk, galt, ara, tna, tss, dam, dcm, supE44, Δlac-proAB], [E. coli ES1301 mutS [Genotype: lacZ53, mutS201::Tn5, thyA36, rha-5, metH1, deoC, In(rrnD-rrnE)]] (20, 24, 28–42).

Eukaryotic organisms are also able to carry out mismatch repair (43–48). Mismatch repair systems in both prokaryotes and eukaryotes are thought to play an important role in the maintenance of genetic fidelity during DNA replication, Some of the genes that play important roles in mismatch repair in prokaryotes, particularly mutS and mutL, have homologs in eukaryotes, in the outcome of genetic recombinations, and in genome stability. Wild-type or mutant S. cerevisiae has been shown to carry our mismatch repair of heteroduplexes (49–56), as have COS-1 monkey cells (57). Preferred strains of yeasts are *Pichia* and *Saccharomyces*. Mammalian cells have been shown to have the capacity to repair G-T to G-C base pairs by a short-patch mechanism (38, 58–63). Mammalian cells (e.g., mouse, hamster, primate, human), both cell lines and primary cultures can also be used. Such cells include stem cells, including embryonic stem cells, zygotes, fibroblasts, lymphocytes, Chinese hamster ovary (CHO), mouse fibroblasts (NIH3T3), kidney, liver, muscle, and skin cells. Other eukaryotic cells of interest include plant cells, such as maize, rice, wheat, cotton, soybean, sugarcane, tobacco, and arabisopsis; fish, algae, fungi (aspergillus, podospora, neurospora), insect (e.g., baculo lepidopiera) (see, Winnacker, “From Genes to Clones,” VCH Publishers, New York, (1987), which is incorporated herein by reference).

In vivo repair occurs in a wide variety of prokaryotic and eukaryotic cells. Use of mammalian cells is advantageous in certain application in which substrates encode polypeptides that are expressed only in mammalian cells or which are intended for use in mammalian cells. However, bacterial and yeast cells are advantageous for screening large libraries due to the higher transformation frequencies attainable in these strains.

V. IN VITRO DNA Repair Systems

As an alternative to introducing annealed products into host cells, annealed products can be exposed a DNA repair
system in vitro. The DNA repair system can be obtained as extracts from repair-competent E. coli, yeast or any other cells (64-67). Repair-competent cells are lysed in appropriate buffer and supplemented with nucleotides. DNA is incubated in this cell extract and transformed into competent cells for replication.

VI. Screening and Selection

After introduction of annealed products into host cells, the host cells are typically cultured to allow repair and replication to occur and optionally, for genes encoded by polynucleotides to be expressed. The recombinant polynucleotides can be subject to further rounds of recombination using the heteroduplex procedures described above, or other shuffling methods described below. However, whether after one cycle of recombination or several, recombinant polynucleotides are subjected to screening or selection for a desired property. In some instances, screening or selection is performed in the same host cells that are used for DNA repair. In other instances, recombinant polynucleotides, their expression products or secondary metabolites produced by the expression products are isolated from such cells and screened in vitro. In other instances, recombinant polynucleotides are isolated from the host cells in which recombination occurs and are screened in selected other host cells. For example, in some methods, it is advantageous to allow DNA repair to occur in a bacterial host strain, but to screen an expression product of recombinant polynucleotides in eucaryotic cells. The recombinant polynucleotides surviving screening or selection are sometimes useful products in themselves. In other instances, such recombinant polynucleotides are subjected to further recombination with each other or other substrates. Such recombination can be effected by the heteroduplex methods described above or any other shuffling methods. Further round(s) of recombination are followed by further rounds of screening or selection on an iterative basis. Optionally, the stringency of selection can be increased at each round.

The nature of screening or selection depends on the desired property sought to be acquired. Desirable properties of enzymes include high catalytic activity, capacity to confer resistance to drugs, high stability, the ability to accept a wider (or narrower) range of substrates, or the ability to function in nonnatural environments such as organic solvents. Other desirable properties of enzymes include capacity to bind to a selected target, secretion capacity, capacity to generate an immune response to a given target, lack of immunogenicity and toxicity to pathogenic microorganisms. Desirable properties of DNA or RNA polynucleotides sequences include capacity to specifically bind a given protein target, and capacity to regulate expression of operably linked coding sequences. Some of the above properties, such as drug resistance, can be selected by plating cells on the drug. Other properties, such as the influence of a regulatory sequence on expression, can be screened by detecting appearance of the expression product of a reporter gene linked to the regulatory sequence. Other properties, such as capacity of an expressed protein to be secreted, can be screened by FACS™ using a labelled antibody to the protein. Other properties, such as immunogenicity or lack thereof, can be screened by isolating protein from individual cells or pools of cells, and analyzing the protein in vitro or in a laboratory animal.

VII. Variations

1. Demethylation

Most cell types methylate DNA in some manner, with the pattern of methylation differing between cells types. Sites of methylation include 5-methylcytosine (m5C), N4-methylcytosine (m4C) and N6-methyladenine (m6A), 5-hydroxymethylcytosine (hm5C) and 5-hydroxymethyluracil (hm5U). In E. coli, methylation is effected by Dam and Dcm enzymes. The methylase specified by the dam gene methylates the N6-position of the adenine residue in the sequence GATC, and the methylase specified by the dem gene methylates the C5-position of the internal cytosine residue in the sequence CCWGG. DNA from plants and mammal is often subject to CG methylation meaning that CG or CNG sequences are methylated. Possible effects of methylated on cellular repair are discussed by references 18–20.

In some methods, DNA substrates for heteroduplex formation are at least partially demethylated on one or both strands, preferably the latter. Demethylation of substrate DNA promotes efficient and random repair of the heteroduplexes. In heteroduplexes formed with one strand dam-methylated and one strand unmethylated, repair is biased to the unmethylated strand, with the methylated strand serving as the template for correction. If neither strand is methylated, mismatch repair occurs, but shows insignificant strand preference (23, 24).

Demethylation can be performed in a variety of ways. In some methods, substrate DNA is demethylated by PCR-amplification. In some instances, DNA demethylation is accomplished in one of the PCR steps in the heteroduplex formation procedures described above. In other methods, an additional PCR step is performed to effect demethylation. In other methods, demethylation is effected by passing substrate DNA through methylation deficient host cells (e.g. an E. coli dam dcm” strain). In other methods, substrate DNA is demethylated in vitro using a demethylating enzymes. Demethylated DNA is used for heteroduplex formation using the same procedures described above. Heteroduplexes are subsequently introduced into DNA-repair-proficient but restriction-enzyme-defective cells to prevent degradation of the unmethylated heteroduplex.

2. Sealing Nicks

Several of the methods for heteroduplex formation described above result in circular heteroduplexes bearing nicks in each strand. These nicks can be sealed before introducing heteroduplexes into host cells. Sealing can be effected by treatment with DNA ligase under standard ligation conditions. Ligation forms a phosphodiester bond to link two adjacent bases separated by a nick in one strand of double helix of DNA. Sealing of nicks increases the frequency of recombination after introduction of heteroduplexes into host cells.

3. Error Prone PCR Attendant To Amplification

Several of the formats described above include a PCR amplification step. Optionally, such a step can be performed under mutagenic conditions to induce additional diversity between substrates.

VIII. Other Shuffling Methods

The methods of heteroduplex formation described above can be used in conjunction with other shuffling methods. For example, one can perform one cycle of heteroduplex shuffling, screening or selection, followed by a cycle of shuffling by another method, followed by a further cycle of screening or selection. Other shuffling formats are described by WO 95/22575; US. Patent No. 5,605,793; U.S. Patent No. 5,811,238; WO 96/19256; Stemmer, Science, 270, 1510 (1995); Stemmer et al., Gene, 164, 49–53 (1995); Stemmer, Bio/Technology, 13, 549–553 (1995); Stemmer, Proc. Natl. Acad. Sci. USA 91, 10747–10751 (1994); Stemmer, Nature...
Pounds have spatial electronic properties which are comparable as a lead protein. In particular, the non-peptic components of amino acids or replace amino acids with chemical structures. The analogs should have a stabilized electronic configuration and molecular conformation that allows key functional groups to be presented in substantially the same way as a lead protein. In particular, the non-peptic compounds have spatial electronic properties which are comparable to the polypeptide binding region, but will typically be much smaller molecules than the polypeptides, frequently having a molecular weight below about 2 CHD and preferably below about 1 CHD. Identification of such non-peptic compounds can be performed through several standard methods such as self-consistent field (CSF) analysis, configuration interaction (CHI) analysis, and normal mode dynamics analysis. Computer programs for implementing these techniques are readily available. See Rein et al., Computer-Assisted Modeling of Receptor-Ligand Interactions (Alan Liss, New York, 1989).

IX. Pharmaceutical Compositions

Polynucleotides, their expression products, and secondary metabolites whose formation is catalyzed by polypeptide products, generated by the above methods are optionally formulated as pharmaceutical compositions. Such a composition comprises one or more active agents, and a pharmaceutically acceptable carrier. A variety of aqueous carriers can be used, e.g., water, buffered water, phosphate-buffered saline (PBS), 0.4% saline, 0.3% glycine, human albumin solution and the like. These solutions are sterile and generally free of particulate matter. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions such as pH adjusting and buffering agents, toxicity adjusting agents and the like, for example, sodium acetate, sodium chloride, potassium chloride, calcium chloride and sodium is selected primarily based on fluid volumes, viscosities, and so forth, in accordance with the particular mode of administration selected.

EXAMPLES

Example 1

Novel Rhizobium Flaa Genes from Recombination of Rhizobium Lupini Flaa And Rhizobium Meliloti Flaa

Bacterial flagella have a helical filament, a proximal hook and a basal body with the flagellar motor (68). This basic design has been extensively examined in E. coli and S. typhimurium and is broadly applicable to many other bacteria as well as some archaea. The long helical filaments are polymers assembled from flagellin subunits, whose molecular weights range between 20,000 and 65,000, depending on the bacterial species (69). Two types of flagellar filaments, named plain and complex, have been distinguished by their electron microscopically determined surface structures (70). Plain filaments have a smooth surface with faint helical lines, whereas complex filaments exhibit a conspicuous helical pattern of alternating ridges and grooves. These characteristics of complex flagellar filaments are considered to be responsible for the brittle and (by implication) rigid structure that enables them to propel bacteria efficiently in viscous media (71–73). Whereas flagella with plain filaments can alternate between clockwise and counter clockwise rotation (68), all known flagella with complex filaments rotate only clockwise with intermittent stops (74). Since this latter navigation pattern is found throughout bacteria and archaea, it has been suggested that complex flagella may reflect the common background of an ancient, basic motility design (69).

Differing from plain bacterial flagella in the fine structure of their filaments dominated by conspicuous helical bands and in their fragility, the filaments are also resistant against heat decomposition (72). Schmitt et al. (75) showed that bacteriophage 7-7-1 specifically adsorbs to the complex flagella of R. lupini H 13-3 and requires motility for a productive infection of its host. Though the flagellins from R. meliloti and R. lupini are quite similar, bacteriophage 7-7-1 does not infect R. meliloti. Until now complex flagella have been observed in only three species of soil bacteria: Pseudomonas rhodos (73), R. meliloti (76), and R. lupini H113-3 (70, 72). Cells of R. lupini H113-3 possess 5 to 10 peripherically inserted complex flagella, which were first isolated and analyzed by high resolution electron microscopy and by optical diffraction (70).

Maruyama et al. (77) further found that a higher content of hydrophobic amino acid residues in the complex filament may be one of the main reasons for the unusual properties of complex flagella. By measuring mass per unit length and obtaining three-dimensional reconstruction from electron micrographs, Trachtenberg et al. (73, 78) suggested that the complex filaments of R. lupini are composed of functional dimers. FIG. 6 shows the comparison between the deduced amino acid sequence of the R. lupini H113-3 Flaa and the deduced amino acid sequence of the R. meliloti Flaa. Percent matches are indicated by vertical lines, and conservative exchanges are indicated by colons. The overall identity is 56%. The R. lupini Flaa and R. meliloti Flaa were subjected to in vitro heteroduplex formation followed by in vivo repair in order to create novel Flaa molecules and structures.

A. Methods

pRL20 containing R. lupini H113-3 Flaa gene and pRM40 containing R. meliloti Flaa gene are shown in FIGS. 6A and 6B. These plasmids were isolated from E. coli SCS110 (free from dam- and dcm-type methylation). About 3.0 pg of unmethylated pRL20 and pRM40 DNA were digested with Bam HI and Eco RI, respectively, at 37° C. for 1 hour. After agarose gel separation, the linearized DNA was purified with Wizard PCR Prep kit (Promega, Wis., USA). Equimolar concentrations (2.5 nM) of the linearized unmethylated pRL20 and pRM40 were mixed in 1xSSPE buffer (180 mM NaCl, 1 mM EDTA, 10 mM Na2HPO4, pH 7.4). After heating at 96° C. for 10 minutes, the reaction mixture was immediately cooled at 0° C. for 5 minutes. The mixture was incubated at 68° C. for 2 hour for heteroduplexes to form.

One microtiter of the reaction mixture was used to transform 50µl of E. coli ES130 mutS, E. coli SCS110 and E. coli JM109 competent cells. The transformation efficiency with E. coli JM109 competent cells was about seven times higher than that of E. coli SCS110 and ten times higher than that of E. coli ES130. mutant. Although the overall transformation efficiencies were 10–200 times lower than those of control transformations with the close, covalent and circular pUC19 plasmid.
Two clones were selected at random from the E. coli SCS110 transformants and two from E. coli ES1301 mutS transformants, and plasmid DNA was isolated from these four clones for further DNA sequencing analysis.

**B. Results**

FIG. 7 shows (a) the sequence of SCS01 (clone #1 from E. coli SCS110 transformant library), (b) the sequence of SCS02 (clone #2 from E. coli SCS110 transformant library), (c) the sequence of ES01 (clone #1 from E. coli ES1301 transformant library), and (d) the sequence of ES02 (clone #2 from E. coli ES1301 transformant library). All four sequences were different from wild-type R. lupini flaA and R. meliloti flaA sequences. Clones SCS02, ES01 and ES02 all contain a complete open-reading frame, but SCS01 was truncated. FIG. 8 shows that recombination mainly occurred in the loop regions (unmatched regions). The flaA mutant library generated from R. meliloti flaA and R. lupini flaA can be transformed into E. coli SCS110, ES1301, XL10-Gold and JM109, and transformants screened for functional flaA recombinants.

**Example 2**

**Directed Evolution of ECB Deacylase for Variants with Enhanced Specific Activity**

**Streptomyces** are among the most important industrial microorganisms due to their ability to produce numerous important secondary metabolites (including many antibiotics) as well as large amounts of enzymes. The approach described here can be used with little modification for directed evolution of native Streptomyces enzymes, some or all of the genes in a metabolic pathways, as well as other heterologous enzymes expressed in E. coli.

New antifungal agents are critically needed by the large and growing numbers of immune-compromised AIDS, organ transplant and cancer chemotherapy patients who suffer opportunistic infections. Echinocandin B (ECB), a lipopeptide produced by some species of Aspergillus, has been studied extensively as a potential antifungal. Various antifungal agents with significantly reduced toxicity have been generated by replacing the linoleic acid side chain of A. nidulans echinocandin B with different aryl side chains (79–83). The cyclic hexapeptide ECB nucleus precursor for the acylation reaction is obtained by enzymatic hydrolysis of ECB using *Actinoplanes utahensis* ECB deacylase. To maximize the conversion of ECB into intact nucleosides, this reaction is carried out at pH 5.5 with a small amount of miscible organic solvent to solubilize the ECB substrate. The product cyclic hexapeptide nucleus is unstable at pH above 5.5 during the long incubation required to fully deacylate ECB (84). The pH optimum of ECB deacylase, however, is 8.0–8.5 and its activity is reduced at pH 5.5 and in the presence of more than 2.5% ethanol (84). To improve production of ECB nucleus it is necessary to increase the activity of the ECB deacylase under these process-relevant conditions.

Relatively little is known about ECB deacylase. The enzyme is a heterodimer whose two subunits are derived by processing of a single precursor protein (83). The 19.9 kDa α-subunit is separated from the 60.4 kDa β-subunit by a 15-amino acid spacer peptide that is removed along with a signal peptide and another spacer peptide in the native organism. The polypeptide is also expressed and processed into functional enzyme in *Streptomyces lividans*, the organism used for large-scale conversion of ECB by recombinant ECB deacylase. The three-dimensional structure of the enzyme has not been determined, and its sequence shows so little similarity to other possibly related enzymes such as penicillin acylase that a structural model reliable enough to guide a rational effort to engineer the ECB deacylase will be difficult to build. We therefore decided to use directed evolution (85) to improve this important activity.

Protocols suitable for mutagenic PCR and random-priming recombination of the 2.4 kb ECB deacylase gene (75% G+C) have been described recently (86). Here, we further describe the use of heteroduplex recombination to generate new ECB deacylase with enhanced specific activity.

In this case, two *Actinoplanes utahensis* ECB deacylase mutants, M7-2 and M16, which show higher specific activity at pH 5.5 and in the presence of 10% McOH were recombined using technique of the in vitro heteroduplex formation and in vivo mismatch repair.

FIG. 12 shows the physical maps of plasmids pM7-2 and pM16 which contain the genes for the M7-2 and M16 ECB deacylase mutants. Mutant M7-2 was obtained through mutagenic PCR performed directly on whole *Streptomyces lividans* cells containing wild-type ECB deacylase gene, expressed from pM7-2 plasmid pSHP150–2*. ECB deacylase with pM7-2 show 1.5 times the specific activity of cells expressing the wild-type ECB deacylase (86). Clone pM16 was obtained using the random-priming recombination technique as described (86, 87). It shows 2.4 times specific activity of the wild-type ECB deacylase clone.

**A. Methods:**

M7-2 and M16 plasmid DNA (pM7-2 and pM16) (FIG. 9) were purified from E. coli SCS210 (in separate reactions). About 5.0 µg of unmethylated M7-2 and M16 DNA were digested with Xho I and Psh I, respectively, at 37°C for 1 hour (FIG. 10). After agarose gel separation, the linearized DNA was purified using a Wizard PCR Prep Kit (Promega, Wis., USA). Equimolar concentrations (2.0 nM) of the linearized unmethylated pM7-2 and pM16 DNA were mixed in 1×SSPE buffer (1×SSPE: 180 mM NaCl, 1.0 mM EDTA, 10 mM NaH₂PO₄, pH 7.4). After heating at 96°C for 10 minutes, the reaction mixture is immediately cooled at 0°C for 5 minutes. The mixture was incubated at 68°C for 3 hours to promote formation of heteroduplexes.

One microliter of the reaction mixture was used to transform 50 µl of E. coli ES1301 mutS, SCS110 and JM109 competent cells. All transformants from E. coli ES1301 mutS were pooled and E. coli SCS110 were pooled. A plasmid pool was isolated from each pooled library, and this pool was used to transform S. lividans TK23 protoplasts to form a mutant library for deacylase activity screening. Transformants from the S. lividans TK23 libraries were screened for ECB deacylase activity with an in situ plate assay. Treated protoplasts were allowed to regererate on R2YE agar plates for 24 hr at 30°C and to develop in the presence of thiostrepton for 48 hours. When the colonies grew to the proper size, 6 ml of 0.7% agarose solution containing 0.5 mg/ml ECB in 0.1 M sodium acetate buffer (pH 5.5) was poured on top of each R2YE-agar plate and allowed to develop for 18–24 hr at 30°C. Colonies surrounded by a clearing zone larger than that of a control colony containing wild-type plasmid pSHP150–2*, were selected for further characterization.

Selected transformants were inoculated into 20 ml medium containing thiostrepton and grown aerobically at 30°C for 48 hours, at which point they were analyzed for ECB deacylase activity using HPLC. 100 µl of whole broth was used for a reaction at 30°C for 30 minutes in 0.1 M NaAc buffer (pH 5.5) containing 10% (v/v) McOH and 200 µg/ml of ECB substrate. The reactions were stopped by
adding 2.5 volumes of methanol, and 20 μl of each sample were analyzed by HPLC on a 100×4.6 mm polyhydroxy-ethyl aspartamide column (PolyLC Inc., Columbia, Md., USA) at room temperature using a linear acetonitrile gradient starting with 50:50 of A:B (A=93% acetonitrile, 0.1% phosphoric acid; B=70% acetonitrile, 0.1% phosphoric acid) and ending with 30:70 of A:B in 22 min at a flow rate of 2.2 ml/min. The areas of the ECB and ECB nucleus peaks were calculated and subtracted from the areas of the corresponding peaks from a sample culture of S. lividans containing pPH702 in order to estimate the ECB deacylase activity.

B. Results

FIG. 11 shows that after one round of applying this heteroduplex repair technique on the mutant M7-2 and M16 genes, one mutant (M15) from about 500 original transformants was found to possess 3.1 times the specific activity of the wild-type. Wild-type and evolved M15 ECB deacylases were purified and their kinetic parameters for deacylation of ECB were determined by HPLC. The evolved deacylase sequence and the positions of the amino acid substitutions in the evolved variants M7-2, M16 and M15 are summarized in FIG. 12.

The heteroduplex recombination technique can recombine parent sequences to create novel progeny. Recombination of the M7-2 and M16 genes yielded M15, whose activity is higher than any of its parents (FIG. 13). Of the six base substitutions in M15, five (at positions α50, α71, β57, β129 and β340) were inherited from M7-2, and the other one (β30) came from M16.

This approach provides an alternative to existing methods of DNA recombination and is particularly useful in recombinating large genes or entire operons. This method can be used to create recombinant proteins to improve their properties or to study structure-function relationship.

Example 3

Novel Thermostable Bacillus Subtilis Subtilisin E Variants

This example demonstrates the use in vitro heteroduplex formation followed by in vivo repair for combining sequence information from two different sequences in order to improve the thermostability of Bacillus subtilis subtilisin E.

Genes RC1 and RC2 encode thermostable B. subtilis subtilisin E variants (88). The mutations at base positions 1107 in RC1 and 995 in RC2 (FIG. 14), giving rise to amino acid substitutions Asn218Ser (N218S) and Asn181Asp (N181D), lead to improvements in subtilisin E thermostability; the remaining mutations, both synonymous and nonsynonymous, have no detectable effects on thermostability. At 65°C, the single variants N181D and N218S have activities as a function of pH were measured for the purified ECB deacylases at 30°C. Each of the ECB deacylase mutants was used for activity assay reaction at 30°C. For 30 minutes in 0.1 M NaAc buffer (pH 5.5) containing 10% (v/v) MeOH and different concentrations of ECB substrate. Assays were performed in duplicate. The reactions were stopped by adding 2.5 volumes of methanol, and the HPLC assays were carried out as described above. The absorbance values were recorded, and the initial rates were calculated by least-squares regression of the time progress curves from which the Km and the kcat were calculated.

Activities as a function of pH were measured for the purified ECB deacylases at 30°C. Each of the ECB deacylase mutants was used for activity assay reaction at 30°C. For 30 minutes in 0.1 M NaAc buffer (pH 5.5) containing 10% (v/v) MeOH and different concentrations of ECB substrate. Assays were performed in duplicate. The reactions were stopped by adding 2.5 volumes of methanol, and the HPLC assays were carried out as described above. The absorbance values were recorded, and the initial rates were calculated by least-squares regression of the time progress curves from which the Km and the kcat were calculated.

Activities as a function of pH were measured for the purified ECB deacylases at 30°C. Each of the ECB deacylase mutants was used for activity assay reaction at 30°C. For 30 minutes in 0.1 M NaAc buffer (pH 5.5) containing 10% (v/v) MeOH and different concentrations of ECB substrate. Assays were performed in duplicate. The reactions were stopped by adding 2.5 volumes of methanol, and the HPLC assays were carried out as described above. The absorbance values were recorded, and the initial rates were calculated by least-squares regression of the time progress curves from which the Km and the kcat were calculated.

Activities as a function of pH were measured for the purified ECB deacylases at 30°C. Each of the ECB deacylase mutants was used for activity assay reaction at 30°C. For 30 minutes in 0.1 M NaAc buffer (pH 5.5) containing 10% (v/v) MeOH and different concentrations of ECB substrate. Assays were performed in duplicate. The reactions were stopped by adding 2.5 volumes of methanol, and the HPLC assays were carried out as described above. The absorbance values were recorded, and the initial rates were calculated by least-squares regression of the time progress curves from which the Km and the kcat were calculated.
were 250 times lower than that of the transformation with closed, covalent and circular control pUC19 plasmids. One microliter of the reaction mixture was used to transform 50 μl of E. coli ES 1381 mutS, E. coli SCS110 and E. coli HB101 competent cells.

The transformation efficiency with E. coli HB101 competent cells was about ten times higher than that of E. coli SCS110 and 15 times higher than that of E. coli ES1301 mutS. But in all these cases, the transformation efficiencies were 10–250 times lower than that of the transformation with closed, covalent and circular control pUC19 plasmids.

Five clones from E. coli SCS110 mutant library and five from E. coli ES1301 mutS library were randomly chosen, and plasmid DNA was isolated using a QIAprep spin plasmid miniprep kit for further DNA sequencing analysis.

About 2,000 random clones from E. coli HB101 mutant library were plated and total plasmid DNA was isolated using a QIAGEN-100 column. 0.5–4.0 μg of the isolated plasmid was used to transform Bacillus subtilis DB428 as described previously (88).

About 400 transformants from the Bacillus subtilis DB428 library were subjected to screening. Screening was performed using the assay described previously (88), on succinyl-Ala-Ala-Pro-Phe-p-nitroanilide. B. subtilis DB428 containing the plasmid library were grown on LB plates containing kanamycin (20 μg/ml) plates. After 18 hours at 37° C. single colonies were picked into 96-well plates containing 200 μl SG/kanamycin medium per well. These plates were incubated with shaking at 37° C. for 24 hours to let the cells to grow to saturation. The cells were spun down, and the supernatants were sampled for the thermostability assay.

Two replicates of 96-well assay plates were prepared for each growth plate by transferring 10 μl of supernatant into the replica plates. The subtilisin activities were then measured by adding 100 μl of assay activity solution (0.2 mM succinyl-Ala-Ala-Pro-Phe-p-nitroanilide, 100 mM Tris-HCl, 10 mM CaCl₂, pH 8.0, 37° C.). Reaction velocities were measured at 405 nm to over 1.0 min in a ThermoMax microplate reader (Molecular Devices, Sunnyvale Calif.). Activity measured at room temperature was used to calculate the fraction of active clones (clones with activity less than 10% of that of wild type were scored as inactive). Initial activity (A) was measured after incubating one assay plate at 65°C for 10 minutes by immediately adding 100 μl of prewarmed (37° C.) assay solution (0.2 mM succinyl-Ala-Ala-Pro-Phe-p-nitroanilide, 100 mM Tris-HCl, pH 8.0, 10 mM CaCl₂, pH 8.0) into each well. Residual activity (Ar) was measured after 40 minute incubation.

B. Results

In vitro heteroduplex formation and in vivo repair was carried out as described above. Five clones from E. coli SCS110 mutant library and five from E. coli ES1301 mutS libraries were selected at random and sequenced. FIG. 14 shows that four out of the ten clones were different from the parent genes. The frequency of occurrence of a particular point mutation from parent RC1 or RC2 in the resulting genes ranged from 0% to 50%, and the ten point mutations in the heteroduplex have been repaired without strong strand-specific preference.

Since none of the ten mutations locates within the dcm site, the mismatch repair appears generally done via the E. coli long-patch mismatch repair systems. The system repairs different mismatches in a strand-specific manner using the state of N6-methylation of adenine in GATC sequences as the major mechanism for determining the strand to be repaired. With heteroduplexes methylated at GATC sequences on only one DNA strand, repair was shown to be highly biased to the unmethylated strand, with the methylated strand serving as the template for correction. If neither strand was methylated, mismatch repair occurred, but showed little strand preference (23, 24). These results shows that it is preferable to demethylate the DNA to be recombinized to promote efficient and random repair of the heteroduplexes.

The rates of subtilisin E thermo-inactivation at 65° C. were estimated by analyzing the 400 random clones from the Bacillus subtilis DB428 library. The thermostabilities obtained from one 96-well plate are shown in FIG. 16, plotted in descending order. About 12.9% of the clones exhibited thermostability comparable to the mutant with the N181D and N218S double mutations. Since this rate is only half of that expected for random recombination of these two markers, it indicates that the two mismatches at positions 995 and 1107 within the heteroduplexes have been repaired with lower position randomness.

Sequence analysis of the clone exhibiting the highest thermostability among the screened 400 transformants from the E. coli SCS110 heteroduplex library confirmed the presence of both N181D and N218S mutations. Among the 400 transformants from the B.subtilis DB428 library that were screened, approximately 91% of the clones expressed N181D- and/or N218S-type enzyme stabilities, while about 80% of the transformants showed only wild-type subtilisin E stability.

Less than 1.0% inactive clone was found, indicating that few new point mutations were introduced in the recombinization process. This is consistent with the fact that no new point mutations were identified in the ten sequenced genes (FIG. 14). While point mutations may provide useful diversity for some in vitro evolution applications, they can also be problematic for recombination of beneficial mutations, especially when the mutation rate is high.

Example 4

Optimizing Conditions for the Heteroduplex Recombination.

We have found that the efficiency of heteroduplex recombination can differ considerably from gene to gene [17,57]. In this example, we investigate and optimize a variety of parameters that improve recombination efficiency. DNA substrates used in this example were site-directed mutants of green fluorescent protein from Aequorea victoria. The GFP mutants had a stop codon(s) introduced at different locations along the sequence that abolished their fluorescence. Fluorescent wild type protein could be only restored by recombination between two or more mutations. Fraction of fluorescent colonies was used as a measure of recombination efficiency.

A. Methods

About 2–4 μg of each parent plasmid was used in one recombination experiment. One parent plasmid was digested with Pst I endonuclease another parent with EcoRI. Linearized plasmids were mixed together and 20xSSPE buffer was added to the final concentration 1x (180 mM NaCl, 1 mM EDTA, 10 mM NaH₂PO₄, pH 7.4). The reaction mixture was heated at 96° C. for 4 minutes, immediately transferred on ice for 4 minutes and the incubation was continued for 2 hours at 68° C.

Target genes were amplified in a PCR reaction with primers corresponding to the vector sequence of pGFP
plasmid. Forward primer: 5'-CCGACTGGAAAGC
GGGCAGTG-3', reverse primer 5'-CGGGGCTGGCCTT
AACTATCGGG-3'. PCR products were mixed together and
purified using Qiagen PCR purification kit. Purified products
were mixed with 20×SSPE buffer and hybridized as
described above. Annealed products were precipitated with
ethanol or purified on Qiagen columns and digested with
EcoRI and PsII enzymes. Digested products were ligated
into PsII and EcoRI digested pGFP vector.

dUTP was added into PCR reaction at final concentrations
200 pM, 40 pM, 8 pM, 1.6 pM, 0.32 pM. PCR reaction and
subsequent cloning procedures were performed as described
above.

Recombinant plasmids were transformed into XL10 E.
coli strain by a modified chemical transformation method.
Cells were plated on ampicillin containing LB agar plates
and grown overnight at 37° C., followed by incubation at
room temperature or at 4° C. until fluorescence developed.

1. Effect of Ligation on Recombination Efficiency.

Two experiments have been performed to test the effect of
breaks in the DNA heteroduplex on the efficiency of recom-

In one experiment heteroduplex plasmid was treated with DNA ligase to close all existing single-strand
breaks and was transformed in identical conditions as an
unligated sample (see Table 1). The ligated samples show up
to 7-fold improvement in recombination efficiency over
unligated samples.

In another experiment, dUTP was added into PCR reaction
to introduce additional breaks into DNA upon repair by
uracil N-glycosylase in the host cells. Table 2 shows that
dUMP incorporation significantly suppressed recombi-
ation, the extent of suppression increasing with increased dUTP concentration.

2. Effect of Plasmid Size on the Efficiency of Heterodu-

Plasmid size was a significant factor affecting recombi-
nation efficiency. Two plasmids pGFP (3.3 kb) and a Bacili-

3. Efficiency of Recombination vs. Distance Between

A series of GFP variants was recombined pairwise to
study the effect of distance between mutations on the effi-
ciency of recombination. Parental genes were amplified by
PCR, annealed and ligated back into pGFP vector. Hetero-
duplexes were transformed into XL10 E. coli strain.

The first three columns in Table 3 show the results of three
independent experiments and demonstrate the dependence of recombinant efficiency on the distance between muta-
tions. As expected recombinant becomes less and less
efficient for very close mutations.

However, it is still remarkable that long-patch repair has
been able to recombine mutations separated by only 27 bp.


88. Zhao, H. and Arnold, F. H. 1997. Functional and nonfunctional mutations distinguished by random recombina-


SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 11
<210> SEQ ID NO 1
<211> LENGTH: 410
<212> TYPE: PRT
<213> ORGANISM: Rhizobium lupini
<220> FEATURE:
<223> OTHER INFORMATION: flagellin A (FlaA)

<400> SEQUENCE: 1
Met Ala Ser Val Leu Thr Asn Ile Asn Ala Met Ser Ala Leu Gln Thr
1 5 10 15
Leu Arg Ser Ile Ser Ser Ass Met Glu Asp Thr Gln Ser Arg Ile Ser 20 25 30
Ser Gly Met Arg Val Gly Ser Ala Ser Asp Ala Ala Tyr Trp Ser
35 40 45
Ile Ala Thr Thr Met Arg Ser Asp Ala Ser Leu Ser Ala Val Gln
50 55 60
Asp Ala Ile Gly Leu Ala Ala Lys Val Asp Thr Ala Ser Ala Gly
65 70 80
Met Asp Ala Val Ile Asp Val Val Lys Asp Arg Ile Ser Leu Val
85 90
Thr Ala Gln Glu Ser Ser Ala Asp Lys Thr Lys Ile Gln Gly Glu Val
100 105 110
Lys Gln Leu Gln Glu Leu Lys Gln Leu Val Asp Ser Ala Ser Phe
115 120 125
Ser Gly Glu Asn Trp Leu Lys Gly Asp Leu Ser Thr Thr Thr Lys
130 135 140
Ser Val Gly Ser Phe Val Arg Gly Gly Thr Val Ser Lys
145 150 155 160
Thr Ile Asp Tyr Ala Leu Asn Ala Ser Lys Val Leu Val Asp Thr Arg
165 170 175
Ala Thr Gly Thr Lys Thr Ile Leu Asp Thr Ala Tyr Gly Leu
180 185 190
Asn Ala Asn Thr Val Thr Asp Ile Asn Lys Gly Gly Val Ile Thr
195 200 205
Gln Ala Ser Val Arg Ala Tyr Ser Thr Asp Glu Met Leu Ser Leu Gly
210 215 220
Ala Lys Val Asp Gly Ala Asn Ser Val Ala Val Gly Gly Ser
225 230 235 240
Ala Ser Ser Arg Ser Thr Ala Ala Gly Leu Arg Val Ala Ser Thr Leu
245 250 255
Arg Pro Pro Ser Pro His Gln His Gln Ser Leu Ala Ser Leu Pro Pro
Leu Thr Pro Pro Leu Lys Leu Val Leu Gln Leu Leu Pro Val Thr Pro
275 280 290 295
Ser Ser Ser Thr Lys Pro Thr Ala Pro Val Gln Val Asn Leu Thr
295 300
Gln Ser Val Leu Thr Met Asp Val Ser Ser Met Ser Ser Thr Asp Val
305 310 315 320
Gly Ser Tyr Leu Thr Gly Val Gln Ala Leu Thr Ser Leu Thr Ser
325 330 335
Ala Gly Ala Glu Gly Ser Ile Lys Gln Arg Ile Asp Leu Gln Val
340 345 350
Asp Phe Ala Ser Lys Leu Gly Asp Ala Leu Ala Gln Asp Leu Arg
355 360 365
Leu Val Asp Ala Asp Met Asn Glu Ser Thr Lys Leu Lys Ala Leu
370 375 380
Gln Thr Gln Gin Gin Leu Ala Gin Ser Leu Ser Ile Ala Asn Ser
385 390 395 400
Asp Ser Gin Asn Ile Leu Ser Leu Phe Arg
405 410

SEQ ID NO 2
LENGTH: 394
TYPE: PRT
ORGANISM: Rhizobium meliloti
FEATURE: OTHER INFORMATION: flagellin A (FlaA)

SEQ: 2
Met Thr Ser Ile Leu Thr Asn Asn Ser Ala Met Ala Leu Ser Thr
1 5 10 15
Leu Arg Ser Ile Ser Ser Ser Met Glu Asp Thr Gin Ser Arg Ile Ser
20 25 30
Ser Gly Leu Arg Val Gly Ser Ala Ser Asp Ann Ala Ala Tyr Trp Ser
35 40 45
Ile Ala Thr Thr Met Arg Ser Asp Ann Gin Ala Leu Ser Ala Val Gin
55 60
Asp Ala Leu Gln Leu Gln Ala Lys Val Asp Thr Ala Tyr Ser Gly
65 70 75 80
Met Glu Ser Ala Ile Gln Val Val Lys Ile Lys Ala Lys Leu Val
85 90 95
Ala Ala Thr Gin Asp Gin Gin Val Gin Gin Gin Gin Gin
100 105 110
Thr Gin Leu Lys Asp Gin Leu Thr Ser Ile Ala Gln Ala Ala Ser Phe
115 120 125
Ser Gly Glu Asn Trp Leu Gin Ala Asp Leu Ser Gly Gin Glu ProVal Thr
130 135 140
Lys Ser Val Val Gln Gln Phe Val Arg Asp Ser Ser Gly Ala Val Ser
145 150 155 160
Val Lys Val Asp Tyr Ser Leu Ann Thr Asp Thr Val Leu Phe Asp
165 170 175
Thr Thr Gin Asn Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
180 185 190
Ala Ser Val Thr Leu Pro Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
195 200 205
Thr Val Gly Ala Tyr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
210 215 220
-continued

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>210</td>
<td>215</td>
<td>220</td>
</tr>
</tbody>
</table>

Phe Asp Gly Asp Tyr Ala Asn Val Gly Ala Gly Ala Leu Ala Gly Asp
225 230 235 240

Tyr Val Lys Val Gln Gly Ser Trp Val Lys Ala Val Asp Val Ala Ala
245 250 255

Thr Gly Glu Val Val Tyr Asp Gly Thr Thr Lys Trp Gly Val
260 265 270

Asp Thr Thr Val Thr Gly Ala Pro Ala Thr Asn Val Ala Ala Ala Pro Ala
275 280 285

Ser Ile Ala Thr Ile Asp Ile Thr Ile Ala Gln Ala Gly Asn Leu
290 295 300

Asp Ala Leu Ile Ala Gly Val Asp Glu Ala Leu Thr Asp Met Thr Ser
305 310 315 320

 Ala Ala Ala Ser Leu Gly Ser Ile Ser Ser Arg Ile Asp Leu Gln Ser
325 330 335

Asp Phe Val Asn Lys Leu Ser Asp Ser Gly Val Gly Arg
340 345 350

Leu Val Asp Ala Asp Met Asn Glu Ser Thr Arg Leu Lys Ala Leu
355 360 365

Gln Thr Gin Gin Leu Ala Ile Gin Ala Leu Ser Ile Ala Asn Ser
370 375 380

Asp Ser Gin Asn Val Leu Ser Leu Phe Arg
385 390

SEQ ID NO 3
LENGTH: 1201
TYPE: DNA
ORIGIN: Artificial Sequence
FEATURE: Description of Artificial Sequence:SCS01 mosaic flaA gene created by in vitro heteroduplex formation followed by in vivo repair

SEQUENCE: 3

atggcaagcg ttctcacaaa cattaacgca atgtctgctc ttcagacgct gcgttcgatt
tcttccaaca tggaagacac ccagagccgt atttccagcg gcatgcgcgt tggttcggct
tccgacaacg ccgcttattg gtctatcgcg accaccatgc gctcggacaa tgcctcgctt
tccgctgttc aggatgcaat tggcctcggt gccgccaagg tcgataccgc ttcggcgggt
atggatgcgg ttatcgatgt tgtaaagcag atcaagaaca aactggtcac tgccaccgaa
gacggcgtcg acaaggccaa gatcaagca atcaacgctc ttcggcgggt
agcgtcgaatc ccaggcctcc gtccgcgcct attcgctgga gtccctcacc
gacgaaacca acagcggcgc cggtgcaggt aaacctcacc cagtcggtcc tgaccatgga
tgtcagctcg atgagctcga cggatgtcgg cagctacctc acgggcgtgg aaaaggctct

60

120

180

240

300

360

420

480

540

600

660

720

780

840

900

960

<210> <211> <212> <213> <220> <223> <240>
caccagctg acgagctgtg gcgctgaact cggctctatc aaacagcgca tcgatctgca 1020
ggttgatttt gcttccaagc tgggcgacgc tctcgcaaaa ggtattggcc gtctcgttga 1080
tgctgacatg aatgaagagt ccactaagct taaggctctt cagacgcagc agcagctggc 1140
tatccagctg ototcatcg caaacagcga ctcgcagaac attctgtcgc tgttccgtta 1200
a 1201

<210> SEQ ID NO 4
<211> LENGTH: 1229
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:SCS02 mosaic flaA gene created by in vitro heteroduplex formation followed by in vivo repair
<400> SEQUENCE: 4
atgacgagca ttctcaccaa caactccgca atggccgcgc tttccggagt gcgctcgatc 60
tcttccagca tggaagacac gcagagccgc atctcctccg gccttcgcgt cggttcggcc 120
tggcgctgc agaacgctcc gcgcgcggcc gcgctcgggt ccgcggggtt gccggtggtt 180
tgctgacatg aatgaagagt ccactaagct taaggctctt cagacgcagc agcagctggc 240
atggaatcgg cgatcgaagt cgttaaggaa atcaagaaca aacgtgcatc tcgctgcaaa 300
tttgaccga caaaaaagga gattcggggt gagaatgctc tttctctctc gggtgttagt 360
ggcctggcc agaacgctcc gcgcgcggcc gcgctcgggt ccgcggggtt gcagaagcagc 420
tgtgaagcg cgatcgcggtc tgttccctgc ccggcgtctg ccggcgcggt ttgttgctgtg 480
tcggcagaagcg gatctcggcg cggcgcggcc gcgctcgggt ccgcggggtt tggctgctgg 540
ggcttggctt cggagctggat tcggagctgca gcggagctgc gccggtgctg ggctttcggt 600
ttcggtgctgc ctgctgctgc tggagctggat tcggagctgca gcggagctgc gccggtgctg 660
ccggagtgctg cggctgctgc cggctgctgc cggctgctgc cggctgctgc cggctgctgc 720
cggctgctgc ctgctgctgc cggctgctgc cggctgctgc cggctgctgc cggctgctgc 780
cgttccggcg cggctgctgc cggctgctgc cggctgctgc cggctgctgc cggctgctgc 840
ttcctctccgt ttcctctccgt ttcctctccgt ttcctctccgt ttcctctccgt ttcctctccgt 900
tccggctggc gcggagctggat tcggagctgca gcggagctgc gccggtgctg ggctttcggt 960
ttcggtgctgc ctgctgctgc tggagctggat tcggagctgca gcggagctgc gccggtgctg 1020
gtctggctgc gcggagctggat tcggagctgca gcggagctgc gccggtgctg ggctttcggt 1080
ttcggtgctgc ctgctgctgc tggagctggat tcggagctgca gcggagctgc gccggtgctg 1140
agggctggc gcggagctggat tcggagctgca gcggagctgc gccggtgctg ggctttcggt 1200
cggagctgc gcggagctggat tcggagctgca gcggagctgc gccggtgctg ggctttcggt 1229

<210> SEQ ID NO 5
<211> LENGTH: 1228
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:ES01 mosaic flaA gene created by in vitro heteroduplex formation followed by in vivo repair
<400> SEQUENCE: 5
atgacgagca ttctcaccaa caactccgca atggccgcgc tttccggagt gcgctcgatc 60
tttgaccga caaaaaagga gattcggggt gagaatgctc tttctctctc gggtgttagt 120
<210> SEQ ID NO 6
<211> LENGTH: 1209
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: ES02 mosaic SaA gene created by in vitro heteroduplex formation followed by in vivo repair

<400> SEQUENCE: 6

tcgcagcaca cgcgtactcg gtcgattgcg accaccatgc gctccgacaa ccaggccctt

tcggccgtcc aggacgccct cggcctcggc gccgccaagg ttgataccgc ctattccggt

atggaatcgg cgatcgaagt cgttaaggaa atcaaggcca agctcgtagc tgccaccgaa

180

gacggcgtcg acaaggccaa gatccaagaa gaaatcactc agctcaagga ccagctgacg

240

agcatcgccg acgcggcttc cttctccggt gagaactggc tgcaggcgga cctcagcggc

300

ggcgccgtca ccaagagcgt cgtcggctcg ttcgtccgtg acggaagcgg ttccgtagcc

360

gtcaagacca tgaataacgg tcaagggttc tgcagatagc ccgagcagc

420

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

480

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

540

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

600

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

660

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

720

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

780

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

840

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

900

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

960

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

1020

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

1080

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

1140

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

1200

gcggccgttc agcgcgttac gcgctccggt gataagctgcgc gcggccgttc gcggccgttc

1228
tggagtctg atgtcgcgtc gtcggtcgtc gcggcagctt gcggcagctt gcggcagctt
900
aacatcatca tgtcgacggt ggtgcagct gcgctcgatg ccctgatcag cggtgtcgac
960
gccgctttgc cagcatcgc cagctgtcgc gcgtcgtgc gcgtcatctc caggtcgatc
1020
gacgtcagc gcaattcgct ccaacatcgc tgcagatcgg gcgtcgcgtt
1080
tcgtcgcgc ccagcgcgtc cgccggtcgc gcgtcgcgc gcgtcgcgc gcgtcgcgc
1140
tcgtcgcgc ccagcgcgtc cgccggtcgc gcgtcgcgc gcgtcgcgc gcgtcgcgc
1200
tccgtcgc gcgtcgcgc gcgtcgcgc gcgtcgcgc gcgtcgcgc gcgtcgcgc
1260
<210> SEQ ID NO 7
<211> LENGTH: 4039
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> OTHER INFORMATION: Description of Artificial Sequence: Actinoplanes
utahensis echinocandin B (ECB) deacylase gene
mutant M-15 created by in vitro heteroduplex formation followed by in vivo repair
<221> NAME/KEY: CDS
<222> LOCATION: (1196),(3559)
<400> SEQUENCE: 7
tgcgcatcg tccagctgtc cgtgtgctgt atcgcggccg ggtgcacccg ggtgcagctt
60
gggtggcgc gcggcgcggc gcggcgcggc gcggcgcggc gcggcgcggc gcggcgcggc
120
tgcggttctg cagcatcgc cagctgtcgc gcgtcgtgc gcgtcatctc caggtcgatc
180
gacgtcagc gcaattcgct ccaacatcgc tgcagatcgg gcgtcgcgtt
240
aacatcatca tgtcgacggt ggtgcagct gcgctcgatg ccctgatcag cggtgtcgac
300
agccggtgga gcgtcatctc gcgtcgtgc gcgtcgtgc gcgtcgtgc gcgtcgtgc
360
tgcggttctg cagcatcgc cagctgtcgc gcgtcgtgc gcgtcatctc caggtcgatc
420
gggtggcgc gcggcgcggc gcggcgcggc gcggcgcggc gcggcgcggc gcggcgcggc
480
<400> SEQUENCE: 7
tgcgcatcg tccagctgtc cgtgtgctgt atcgcggccg ggtgcacccg ggtgcagctt
60
gggtggcgc gcggcgcggc gcggcgcggc gcggcgcggc gcggcgcggc gcggcgcggc
120
<221> NAME/KEY: CDS
<222> LOCATION: (1196),(3559)
<400> SEQUENCE: 7
tgcgcatcg tccagctgtc cgtgtgctgt atcgcggccg ggtgcacccg ggtgcagctt
60
gggtggcgc gcggcgcggc gcggcgcggc gcggcgcggc gcggcgcggc gcggcgcggc
120
tgcggttctg cagcatcgc cagctgtcgc gcgtcgtgc gcgtcatctc caggtcgatc
180
gacgtcagc gcaattcgct ccaacatcgc tgcagatcgg gcgtcgcgtt
240
aacatcatca tgtcgacggt ggtgcagct gcgctcgatg ccctgatcag cggtgtcgac
300
agccggtgga gcgtcatctc gcgtcgtgc gcgtcgtgc gcgtcgtgc gcgtcgtgc
360
tgcggttctg cagcatcgc cagctgtcgc gcgtcgtgc gcgtcatctc caggtcgatc
420
gggtggcgc gcggcgcggc gcggcgcggc gcggcgcggc gcggcgcggc gcggcgcggc
480
gac gcc ggc tat ggc gcc ctg atc gtc cgc cgg gcc tgc tac gcc gtc ccg
Asp Gly Gly Tyr Ala Ala Leu Ile Arg Arg Ala Ser Tyr Gly Val Pro
35 40 45

cac etc acc gcc gac gtc ttc ggg ctc ggt ttc ggc gtc ggg tac
His Ile Thr Ala Asp Phe Gly Ser Leu Gly Phe Gly Val Gly Tyr
50 55 60 65

gtg cag gcc gac acc ttc gtc etc ggc gac gac gtc gac
Val Gln Ala Ala Asp Ile Cys Val Ile Ala Glu Ser Val Val Thr
70 75 80

gcc acc ggt gac cgg tcc cgg tgc gtt ggc acc ggc cgc cgg gcc
Ala Asp Gly Glu Arg Ser Arg Thr Phe Gly Ala Thr Gly Pro Arg
85 90 95

gcc gat gtt cgc gac gcc ctc ttc cac cgc ctg cag gcc
Ala Asp Val Arg Ser Asp Leu Phe His Arg Ala Ile Asp Arg
100 105 110

gtc ggc gac gtc ctc gta gaa ggg ccc cgc gac ggc gtc gg ggc
Val Ala Glu Arg Leu Gly Pro Arg Asp Val Arg Ala Pro
115 120 125

tcg gac gac gtc cgg cag atg ctc ggg gag gtt ctc gcc gtc
Ser Asp Ala Val Arg Glu Arg Ser Met Ala Gly Tyr Asn
130 135 140 145

cac ttc cta cgc gcc aag gcc gtg cac gcc cgc
His Phe Leu Arg Arg Ala Gly Thr Arg Pro Ala Cys
150 155 160

cgc ggc aag gcc tgg tgc cgc ctc ttc gac gtc gat ctc gtc
Arg Gly Lys Ala Trp Val Arg Pro Leu Ser Glu Ile Asp Leu Trp Arg
165 170 175

cag tcg tgg acc agc atg gtc cgc ggc gac gcc ctc gcc gcc
Thr Ser Trp Asp Met Arg Ala Gly Pro Thr Ala Gly Pro Ala Ser Ala
180 185 190

ggc etc gtc gcc ggc acg cca cct aca gcc ggc cgg ccc ggc tca gcc
Gly Ile Val Ala Thr Pro Pro Thr Ala Gly Pro Ala Ser Ala
195 200 205

cgg gac gcc cgc gcc ggc gac gac gcc ccg cgg ggc tgg ggt ctc gcc
Pro Glu Ala Asp Ala Ala Ala Ala Leu Arg Gly Thr
210 215 220 225

gcg gac gcc gac gac gcc gtc gac gcc gcc gc ggc gac ggc gcc
Ser Ala Gly Ala Ala Ala Ala Ala Ala Ala Ala Leu Arg Gly Thr
230 235 240

gtg acc gcc aag gcc atg gtt ctc gcc acc ctc gcc cgc
Val Ala Gly Ser Met Leu Ala Asp His Pro Cys Thr Phe
245 250 255

ggc gcc gcc gcc ctc tact cgg atg cac ctc aag cgg ccc gcc gcc
Gly Ala Ala Arg Phe Tyr Met His Leu Lys Pro Gly Arg Tyr
260 265 270

gac gtc gac gcc ggc ggc ctc gtc gcc gcc ggc gac gcc
Aep Val Glu Gly Ala Ala Leu Ile Gly Asp Pro Ile Ile Gly
275 280 285

cac acc ggc gac gtc gcc tgc agc cac acc gtc tcc acc gcc cgc
His Asn Arg Thr Val Ala Trp Ser His Thr Val Ser Thr Ala Arg
290 295 300 305

ttc gtt gtc gac gcc ctc gcc cgc cgc gc cgg gcc gcc ctc ctc
Thr Phe Val Trp His Arg Leu Ser Leu Val Pro Gly Asp Pro Thr Ser Tyr
310 315 320

tac gtc gcc ggc cgc gcc ggc ctg cgc gcc gcc gtc aag gcc
Tyr Val Ala Gly Arg Pro Glu Arg Met Arg Ala Arg Thr Val Val
325 330 335

cgc acc gcc gcc ggc ctc gcc cgc ctc gcc gcc gcc
Gln Thr Gly Ser Pro Val Ser Arg Thr Phe His Arg Thr Arg
340 345 350
Gly Pro Val Ala Val Met Pro Gly Thr Phe Asp Thr Thr Pro Ala Thr 355 360 365

Ala Tyr Ala Ile Thr Asp Val Asn Ala Gly Asn Arg Ala Phe Asp 370 375 380 385

Gly Thr Leu Arg Met Gly Gln Ala Lys Asp Val Arg Ala Leu Lys Ala 390 395 400

Gly Pro Val Ala Val Met Pro Gly Thr Phe Asp Thr Thr Pro Ala Thr 405 410 415

Asp Ala Arg Gly Glu Ala Leu Tyr Gly Asp Ser Arg Ser Asp 420 425 430

Gly Trp Leu Ala Ala Ala Cys Ile Pro Ala Pro Phe Gin Pro 435 440 445

Val Thr Gly Ala Leu Ala Ala Ala Cys Ile Pro Ala Pro Phe Gin Pro 450 455 460 465

Asp Ala Arg Gly Glu Ala Leu Tyr Gly Asp Ser Arg Ser Asp 470 475 480

Pro Ala Ser Leu Pro Val Arg Phe Arg Asp Tyr Val Thr Pro Ala Thr 485 490 495

Asn Asp Ser His Thr Leu Ala Ser Pro Ala Ala Pro Leu Glu Gly Phe 500 505 510

Gly Ile Leu Gly Asn Arg Asp Ser Ser Leu Leu Arg Thr Arg Thr Arg 515 520 525

Leu Gly Leu Asp Gin Ile Gin Gin Arg Leu Ala Gly Thr Arg Leu Leu 530 535 540 545

Arg Leu Thr Ala Leu Ala Pro Arg Leu Pro Leu Gly Thr Phe Thr Thr 550 555 560

Leu Ala Gly Ala Leu Arg Leu Arg Asp Ala Ala Pro Ala Ala Pro Ala 565 570 575

Leu Thr Ala Cys Thr Ala Ala Gly Thr Leu Leu Arg Phe Arg Ala Asp 580 585 590

Leu Thr Ala Cys Thr Ala Ala Gly Thr Leu Leu Arg Phe Arg Ala Asp 595 600 605

Leu Thr Ala Cys Thr Ala Ala Gly Thr Leu Leu Arg Phe Arg Ala Asp 610 615 620 625

Gly Ile Arg Phe Ala Asp Thr Phe Gly Val Thr Asp Pro Val Arg Thr 630 635 640

Gly Ile Arg Phe Ala Asp Thr Phe Gly Val Thr Asp Pro Val Arg Thr 640 645 650 655

Amp Ala Val Gin Arg Leu Ala Gly Ile Pro Leu Asp Ala Aas Asp 660 665 670 675

Gly Pro Val Ala Val Met Pro Gly Thr Phe Asp Thr Thr Pro Ala Thr 680 685 690 695

Gly Thr Leu Arg Met Gly Gln Ala Lys Asp Val Arg Ala Leu Lys Ala 700 705 710
Val Thr Ser Ser Tyr Met Arg Leu Lys Ala Ala Ala Ile Ala Phe Gly
1  5  10  15
Val Ile Val Ala Thr Ala Ala Val Pro Ser Pro Ala Ser Gly Arg Glu
20  25  30
His Asp Gly Gly Tyr Ala Leu Arg Arg Arg Ser Tyr Gly Val
35  40  45
Pro His Ile Thr Ala Asp Phe Gly Ser Leu Gly Phe Gly Val Gly
50  55  60
Tyr Val Gin Ala Glu Asp Ile Cys Val Ile Ala Glu Ser Val Val
65  70  75  80
Thr Ala Asn Gly Glu Arg Ser Arg Trp Phe Gly Ala Thr Gly Pro Asp
85  90  95
Phe Pro Arg Ile Leu Gly Asn Glu Arg Thr Pro Arg Ser Leu Arg Thr 515 520 525
Arg Leu Gly Leu Asp Gin Ile Gin Gin Arg Leu Ala Gly Thr Asp Gly 530 535 540
Leu Pro Gly Lys Gly Phe Thr Thr Ala Arg Leu Trp Gin Val Met Phe 545 550 555 560
Gly Asn Arg Met His Gly Ala Glu Leu Ala Arg Asp Asp Leu Val Ala 565 570 575
Leu Cys Arg Arg Gin Pro Thr Ala Thr Ala Ser Asn Gly Ala Ile Val 580 585 590
Asp Leu Thr Ala Ala Cys Thr Ala Leu Ser Arg Phe Asp Gin Arg Ala 595 600 605
Asp Leu Asp Ser Arg Gin Asp Phe Gin Thr Glu Phe Ala Leu Ala 610 615 620
Gly Gin Ile Gin Gin Asp Thr Phe Gin Val Thr Gin Gin Gin Gin Gin Gin 625 630 635 640
Thr Pro Arg Arg Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin GinGly Gly Arg 785

<210> SEQ ID NO 9
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: OTHER INFORMATION: Description of Artificial Sequence: forward primer corresponding to the vector sequence of pOFF plasmid (Aequorea victoria green fluorescent protein)

<400> SEQUENCE: 9
cgcacctggaagagggactagt g 21

<210> SEQ ID NO 10
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: OTHER INFORMATION: Description of Artificial Sequence: reverse primer corresponding to the vector sequence of pOFF plasmid (Aequorea victoria green fluorescent protein)
What is claimed is:
1. A method for evolving a polynucleotide toward acquisition of a desired functional property, comprising
   (a) incubating a population of parental polynucleotide variants having sufficient diversity that recombination between the parental polynucleotide variants can generate more recombinant-polynucleotides than there are parental polynucleotide variants under conditions to generate annealed polynucleotides comprising heteroduplexes;
   (b) exposing the heteroduplexes to one or more enzymes of a DNA repair system in vitro to convert the heteroduplexes to parental polynucleotide variants or recombinant polynucleotide variants; and
   (c) screening or selecting the recombinant polynucleotide variants for the desired functional property.

2. The method of claim 1, wherein the DNA repair system comprises cellular extracts.

3. The method of claim 1, wherein the cells are bacterial cells.

4. The method of claim 1 further comprising introducing the products of step (b) into cells.

5. The method of claim 4, wherein the introducing step selects for transformed cells receiving recombinant polynucleotides resulting from resolution of heteroduplexes in step (b) relative to transformed cells receiving polynucleotides resulting from resolution of homoduplexes in step (b).

6. A method for evolving a polynucleotide toward acquisition of a desired functional property, comprising
   (a) incubating a population of parental polynucleotide variants having sufficient diversity that recombination between the parental polynucleotide variants can generate more recombinant polynucleotides than there are parental polynucleotide variants under conditions to generate annealed polynucleotides comprising heteroduplexes;
   (b) introducing the annealed polynucleotides into cells having a DNA repair system and propagating the cells under conditions to select for cells receiving heteroduplexes relative to cells receiving homoduplexes, and to convert the heteroduplexes to parental polynucleotide variants or recombinant polynucleotide variants; and
   (c) screening or selecting the recombinant polynucleotide variants for the desired functional property.

7. The method of claim 6, wherein the heteroduplexes are exposed to the cellular DNA repair system in vitro.

8. A method for evolving a polynucleotide toward acquisition of a desired functional property, comprising
   (a) incubating first and second pools of parental polynucleotide variants having sufficient diversity that recombination between the parental polynucleotide variants can generate more recombinant polynucleotides than there are parental polynucleotide variants under conditions whereby a strand from any polynucleotide variant in the first pool can anneal with a strand from any polynucleotide in the second pool to generate annealed polynucleotides comprising heteroduplexes;
   (b) exposing the heteroduplexes to a DNA repair system to convert the heteroduplexes to parental polynucleotide variants or recombinant polynucleotide variants; and
   (c) screening or selecting the recombinant polynucleotide variants for the desired functional property.

9. The method of claim 8, further comprising introducing the heteroduplexes into cells, whereby the heteroduplexes are exposed to the DNA repair system of the cells in vivo.

10. The method of claim 9, wherein the annealed polynucleotides further comprise homoduplexes and the introducing step selects for transformed cells receiving heteroduplexes relative to transformed cells receiving homoduplexes.

11. The method of claim 10, 6, or 5, wherein a first polynucleotide variant is provided as a component of a first vector, and a second polynucleotide variant is provided as a component of a second vector, and the method further comprises converting the first and second vectors to linearized forms in which the first and second polynucleotide variants occur at opposite ends, whereby in the incubating step single-stranded forms of the first linearized vector reanneal with each other to form linear first vector, single-stranded forms of the second linearized vector reanneal with each other to form linear second vector, and single-stranded linearized forms of the first and second vectors anneal with each to form a circular heteroduplex bearing a nick in each strand, and the introducing step selects for transformed cells...
receiving the circular heteroduplexes or recombinant polynucleotides derived therefrom relative to the linear first and second vector.

12. The method of claim 11, wherein the first and second vectors are converted to linearized forms by PCR.

13. The method of claim 11, wherein the first and second vectors are converted to linearized forms by digestion with first and second restriction enzymes.

14. The method of claim 10, 6 or 5, wherein the population of polynucleotides comprises first and second polynucleotides provided in double stranded form, and the method further comprises incorporating the first and second polynucleotides as components of first and second vectors, whereby the first and second polynucleotides occupy opposite ends of the first and second vectors, whereby in the incubating step single-stranded forms of the first linearized vector reanneal with each other to form linear first vector, single-stranded forms of the second linearized vector reanneal with each other to form linear second vector, and single-stranded linearized forms of the first and second vectors anneal with each to form a circular heteroduplex bearing a nick in each strand, and the introducing step selects for transformed cells receiving the circular heteroduplexes before the annealing step.

15. The method of claim 10, 6 or 5, further comprising sealing nicks in the heteroduplexes to form covalently-closed circular heteroduplexes before the introducing step.

16. The method of claim 1, 6 or 8, wherein the population of polynucleotide variants are provided in double stranded form, and the method further comprising converting the double stranded polynucleotides to single stranded polynucleotides before the annealing step.

17. The method of claim 1, 6 or 8 wherein the converting step comprises:

conducting asymmetric amplification of the first and second double stranded polynucleotide variants to amplify a first strand of the first polynucleotide variant, and a second strand of the second polynucleotide variant, whereby the first and second strands anneal in the incubating step to form a heteroduplex.

18. The method of claim 17, wherein the first and second double-stranded polynucleotide variants are provided in vector-free form, and the method further comprises incorporating the heteroduplex into a vector.

19. The method of claim 18, wherein the first and second polynucleotides are from chromosomal DNA.

20. The method of claim 1, 6 or 8, further comprising repeating steps (a)-(c) whereby the incubating step in a subsequent cycle is performed on recombinant variants from a previous cycle.

21. The method of claim 1, 6 or 8, wherein the polynucleotide variants encode a polypeptide.

22. The method of claim 1, 6 or 8, wherein the population of polynucleotide variants comprises at least 20 variants.

23. The method of claim 1, 6 or 8, wherein the population of polynucleotide variants are at least 10 kb in length.

24. The method of claim 1, 6 or 8, wherein the population of polynucleotide variants comprises natural variants.

25. The method of claim 1, 6 or 8, wherein the population of polynucleotides comprises variants generated by mutagenic PCR.

26. The method of claim 1, 6 or 8, wherein the population of polynucleotide variants comprises variants generated by site directed mutagenesis.

27. The method of claim 1, 6 or 8, further comprising at least partially demethylating the population of variant polynucleotides.

28. The method of claim 27, whether the at least partially demethylating step is performed by PCR amplification of the population of variant polynucleotides.

29. The method of claim 27, wherein the at least partially demethylating step is performed by amplification of the population of variant polynucleotides in host cells.

30. The method of claim 29, wherein the host cells are defective in a gene encoding a methylase enzyme.

31. The method of claim 27, wherein the population of variant polynucleotides are double stranded polynucleotides and only one strand of each polynucleotide is at least partially demethylated.

32. The method of claim 1, 6 or 8, wherein the population of variant polynucleotide variants comprises at least 5 polynucleotides having at least 90% sequence identity with one another.

33. The method of claim 1, 6 or 8, further comprising isolating a screened recombinant variant.

34. The method of claim 33, further comprising expressing a screened recombinant variant to produce a recombinant protein.

35. The method of claim 34, further comprising formulating the recombinant protein with a carrier to form a pharmaceutical composition.

36. The method of claim 1, 6 or 8, wherein the polynucleotide variants encode enzymes selected from the group consisting of proteases, lipases, amylases, cutinases, cellulases, amylases, oxidases, peroxidases and phytases.

37. The method of claim 1, 6 or 8, wherein the polynucleotide variants encode a polypeptide selected from the group consisting of insulin, ACTH, glucagon, somatostatin, somatotropin, thyrotropin, parathyroid hormone, pigmentary hormones, somatomedin, erythropoietin, luteinizing hormone, chorionic gonadotropin, hyperthalnic releasing factors, antidiuretic hormones, and relaxin, interferon, thrombopoietin (TPO), and prolactin.

38. The method of claim 1, 6 or 8, wherein the polynucleotide variants encode a plurality of enzymes forming a metabolic pathway.

39. The method of claim 1, 6 or 8, wherein the polynucleotide variants are in concatemeric form.

40. The method of claim 39, wherein the functional property is an enzymatic activity.

41. The method of claim 1, 6 or 8, wherein the at least two polynucleotide variants differ at between 0.1–25% of positions.

42. The method of claim 1, 6 or 8, wherein the functional property is an enzymatic activity.

* * * * *
UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,537,746 B2
DATED : March 25, 2003
INVENTOR(S) : Frances Arnold, Zhixin Shao and Alexander Volkov

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title page,
Item [73], Assignee: should be -- California Institute of Technology --

Signed and Sealed this
Sixteenth Day of September, 2003

JAMES E. ROGAN
Director of the United States Patent and Trademark Office
UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,537,746 B2
DATED : March 25, 2003
INVENTOR(S) : Arnold et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 51,
Line 32, replace “recombinated” with -- recombined --.

Column 53,
Line 12, replace “fist” with -- first --.

Signed and Sealed this Sixth Day of April, 2004

[Signature]

JON W. DUDAS
Acting Director of the United States Patent and Trademark Office