Directed Evolution of Sesquiterpene Synthesis in *Escherichia coli*

Ronald Tanaka[1], Syd Withers[2], Jay Keasling[3]

MarBEC Summer Undergraduate Research Fellowship
Marine Bioproducts Engineering Center
University of Hawaii at Manoa and University of California, Berkeley

Summer 2002

[1]University of Hawaii at Manoa (Biology/English)
[2]University of California, Berkeley
[3]Jay Keasling
Phone: (510) 642-4862
Fax: (510) 643-1228
Email: keasling@socrates.berkeley.edu

Abstract

Isoprenoid prevalence and diversity in the natural world is the remarkable result of terpene enzymes capable of incredible molecular specificity. The ability to manipulate terpene diversity and production using *in vitro* directed evolution techniques could have significant commercial value. The high sequence similarity between two sesquiterpene plant genes coding for the production of epi-cedrol and amorpha-4,11-diene made these genes ideal parents for DNA shuffling experiments. Previous family shuffling experiments gave five synthetic epi-cedrol overproducing genes with a 20 to 60-bp amorphadiene cyclase crossover at the 5' end. These five genes were randomly fragmented and then reassembled to create a hybrid gene library. The resulting gene products were transformed into *Escherichia coli* strain DH10B using the plasmid pTrc99A and screened for epi-cedrol production. When production failed to improve further, exon shuffling techniques were used to more deliberately create genetic diversity. Seven exons exist in both epi-cedrol and amorphadiene synthases. Each exon was individually PCR amplified using primers that code for intronic areas of the genes. The exon fragments were then randomly reassembled to make 128 possible gene combinations. Four hundred individually screened clones yielded seven epi-cedrol producers and no amorphadiene producers.

Introduction

Isoprene and its biosynthetic derivatives, terpenes, represent one of the largest and most diverse groups of natural products known with over 30,000 identified compounds (Lange et al., 2000). The molecular diversity inherent in these compounds, is largely due to terpene synthases capable of designating very specific regio- and stereochemical preferences in protein formation (Greenhagen et al., 2001). The wide-ranging applications of terpenes, including anti-microbial agents, pharmaceuticals, and fragrances, make isoprenoid research an important initial step towards eventual commercial production (Greenhagen and Chappell, 2001). The introduction of engineered genes into a heterologous host has made possible the harvesting of terpene quantities sufficient for commercial use where inadequate expression in the natural host organism has been the limiting factor (Martin et al., 2001).

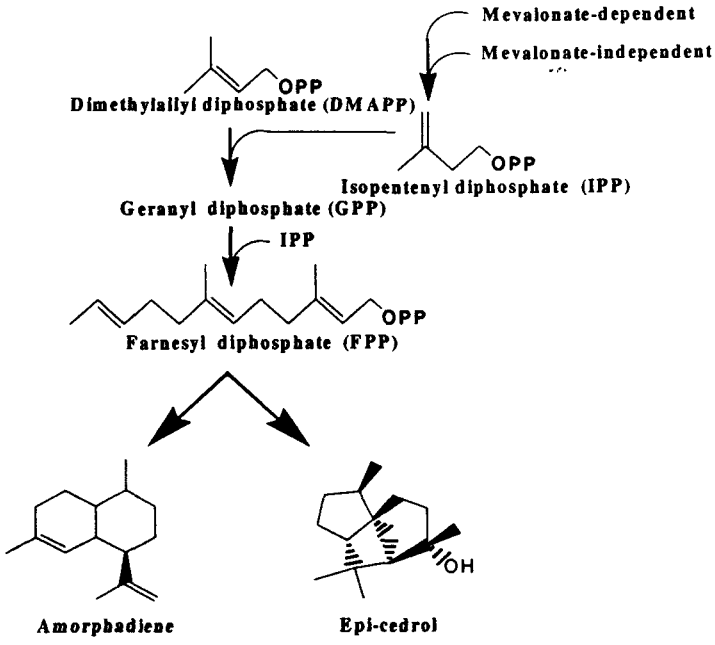


Figure 1. pathways leading to FPP production, the universal precursor molecule of sesquiterpenes. The MBIS gene codes for the mevalonate pathway.

One class of isoprenoids, sesquiterpenes are formed from a universal precursor molecule, farnesyl diphosphate (FPP) (Figure 1). FPP is derived from the isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) isomers, which are, depending on the organism, products of the mevalonate or the recently elucidated non-mevalonate pathways (Rohdich et al., 2002). Sesquiterpene production from genes introduced into *Escherichia coli* was initially limited by low yields of FPP from the non-mevalonate pathway present in *E. coli*. To improve sesquiterpene expression in this heterologous host, the MBIS operon, which codes for the non-native mevalonate

pathway, was introduced via the pBBR1MCS-3 plasmid. The MBIS pathway allows *E. coli* to efficiently catalyze the synthesis of FPP from substrate mevalonate. The pBBRMBIS vector was a gift from Douglas Pitera.

Directed evolution techniques involving mutation, gene shuffling, and selection or screening (Figure 2), are powerful tools for creating libraries of hybrid or chimeric genes capable of novel protein expression (Crameri et al., 1998). Initially focusing on the shuffling of point mutations from error-prone PCR on a single gene, directed evolution has developed increasingly efficient techniques to produce molecular diversity (Stemmer, 1994). Family shuffling, for example, harnesses the power of natural evolution by encouraging functional crossovers between similar genes and exchanging entire blocks of gene sequences that have already selected against deleterious mutations while benefiting from productive ones (Crameri et al., 1998). Exon shuffling is more deliberate and requires amplification and then assembly of different exon combinations using carefully selected intron sequences as sites for novel recombinations (Kolkman et al., 2001). The catalytic power of enzymes can be significantly amplified or altered by the creation of this molecular diversity.

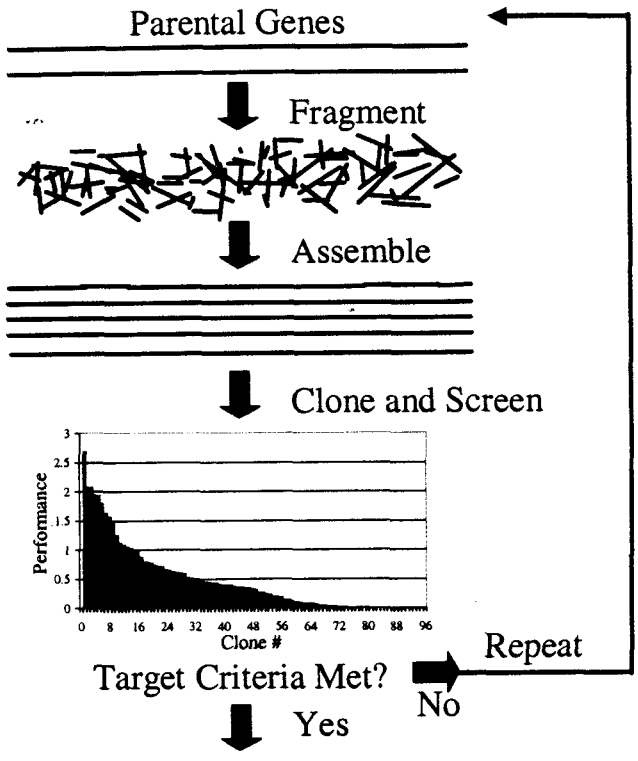


Figure 2. DNA shuffling of parental genes.

Amorpha-4,11-diene and epi-cedrol, identified in *Artemisia annua* L., are sesquiterpenes synthesized from FPP using two synthases of remarkable sequence similarity. Recent research has suggested that amorphadiene is a precursor to artemisinin, which is an essential ingredient of the alternative anti-malarial drug Coartem (Merke et al., 2000). Epi-cedrol is the first known alcohol synthesized from a sesquiterpene synthase (Merke et al., 1999). Our experiments used synthetic genes for

these sesquiterpenes that were designed to optimize codon usage in the heterologous host *E. coli* and to increase sequence similarity to 70% to promote crossover events between the two genes. Our goal in creating EPC-ADS gene libraries was to isolate genes with functional crossovers. These functional genes will then be used as parents in subsequent shuffling experiments with the goal of generating novel enzymes capable of producing unique compounds.

Our initial experiments used five epi-cedrol overproducing genes as parents. A previous family shuffling experiment had generated 20 to 60-bp amorphadiene cyclase crossovers at the 5' end of each of the genes in addition to various point mutations on each of the genes. While each of these genes had retained over 96% sequence similarity to the original synthetic EPC gene, epi-cedrol production was five to ten times higher than wild type levels. These 5 mutant genes were isolated from an ADS-EPC library in which 2000 colonies were screened and no new compounds or amorphadiene overproducers were found. When novel genes were not identified from the hybrid library created using these 5 parents, exon shuffling techniques were employed to more deliberately produce genetic diversity in the shuffling of the original ADS and EPC synthetic genes.

Materials and Methods

DNA shuffling of mutant overproducers. The 1.6kb EPC synthase genes of the 5 previously identified mutant epi-cedrol overproducers were PCR amplified using primers 5'-TGGAATTGTGAGCGGA-3' (forward) and 5'-CTGCAGGTCGACTCTA-3' (reverse). The 100 µL PCR reaction contained 10 µL of 10x taq buffer, 5 µL of 50 mM $MgCl_2$, 1 µL of 10 mM dNTPs, 100 µM of each primer, 0.2 µL taq, and 1 µL template. Cycling conditions were as follows: 94°C for 60 sec; 30 cycles of 94°C for 30 sec – 45 sec, 50°C – 55°C for 30 sec – 45 sec, and 72°C for 60 sec/kb DNA; 4°C thereafter.

The 5 PCR amplified EPC mutant genes were purified, and the DNA quantified and then pooled for digestion into fragments using DNase I. The 50 µL reaction contained 3.6 µg DNA (0.72 ug of each EPC mutant gene), 2.5 µL of 1 M Tris-HCl pH 8 buffer, and 2.5 µL of 200 mM $MnCl_2$. The mixture was incubated at 15°C for 30 min.

1.5 μL of a 1:20 dilution of DNase I was added. EDTA was added after the restriction reaction had continued for 2:20 min and when the fragments were 50 to 200 bp long. The QIAquick Gel Extraction Kit (QIAGEN, Valencia, CA) was used for fragment purification.

The 50 μL PCR reassembly mixture contained 5 μL of 10x taq buffer, 1.5 μL of 50 mM $MgCl_2$, 0.05 μL of 10 mM dNTPs, 0.5 μL taq, and 0.36 μg fragments.

Cloning of EPC gene into E. coli DH10B. Reassembled genes were recovered using a 100 μL PCR mixture of 10 μL of 10x taq buffer, 3 μL of 50 mM $MgCl_2$, 1 μL of 10 mM dNTPs, 100 μM of each primer, 0.2 μL taq, and 4 μL of assembled product. The forward and reverse primers were the same used in PCR amplification. The cycling conditions were identical to those used in the initial PCR amplification. Gel extraction recovered the gene at 1.6kb.

EPC genes were cloned into pTRC99a using the restriction sites NcoI and XmaI. The engineered vectors containing genes from the mutant DNA library were then transformed using electroporation into *E. coli* DH10B containing pBBRMBIS.

Expression/assay of mutant library using GC-MS. The bacteria were streaked and grown overnight at 37°C. Inoculations of individual colonies into 5 x 96-well plates containing 100 μL Luria Bertani (LB) medium containing 50 μg/mL carbenicillin and 10 μg/mL tetracycline were incubated overnight in a shaker at 37°C. The 5 x 96-well plates were pooled into 1 96-well plate for increased initial screening efficiency. Sesquiterpene overproduction, therefore, from a 5-clone sample could be explained by either 1 exceptional overproducer, or 4 or 5 moderate producers. A 5 mL, 2% inoculation of the pooled samples was incubated for 2 hours on a shaker at 37°C to optimize cell growth before terpene production was induced. Fifty μL of 1 M mevalonate and 5 μL of .5 M IPTG were added to induce sesquiterpene production and the cultures were incubated for 6 hrs. on a shaker at 37°C. IPTG induces expression of the sesquiterpene synthase gene and the genes of the MBIS operon. 700 μL samples were extracted with an equal volume of ethyl acetate in a glass vial. The screening of these samples by gas chromatography electron impact-mass spectrometry (GC-IE-MS) used a Hewlett Packard HP6890 gas chromatograph with a model 5973 mass spectrometer and chiral column.

In confirmation assays, individual colonies were subjected to the same conditions for terpene expression. Individual colonies were later assayed for production over time. Samples were taken every hour, and optical density (OD) and sesquiterpene concentration were measured. OD was measured with a spectrophotometer and was used to normalize sesquiterpene concentration to the number of cells. Restriction Fragment Length Polymorphism (RFLP) was used to identify crossovers of substantial length. The restriction enzymes Msp I and Alu I were used to identify crossovers in shuffled genes. Gel electrophoresis of the gene fragments provides a unique banding pattern for each hybrid gene. Small crossovers, however, usually do not affect fragment separation. RFLP, while providing only partial information, is an inexpensive and quick evaluation of crossover events before sequencing.

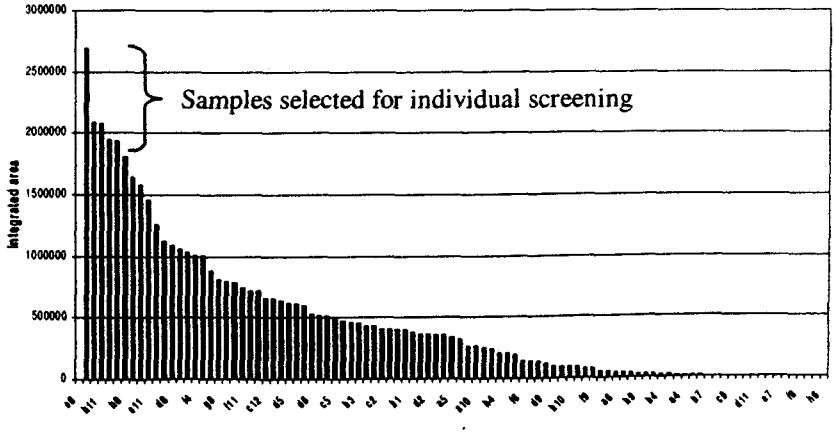
Exon shuffling of synthetic ADS and EPC genes. Exon shuffling techniques modified only slightly the methods used in family shuffling. Each of the 7 exons of both the ADS and EPC genes was PCR amplified in 2 separate reactions that prime at an exon sequence. The first reaction primed one end of an exon to anneal to the next ADS exon while the second reaction primed one end of the same exon to anneal to the next EPC exon, both with a 20-bp overlap. ADS exon #2, for example, would be amplified in 2 different ways so that in the assembly reaction, it could anneal to either an ADS #3 exon or an EPC #3 exon. There are 128 possible combinations of full gene assemblies using this method. 28 100μL-PCR reactions each contained 10 μL of 10x cloned pfu buffer, 1 μL of 10 mM dNTPs, 100 μM of each primer, 2 μL pfu, and 1 μL template. "Turbo" pfu was used to decrease random mutagenesis inherent in PCR using taq polymerase.

The 28 amplified exons were then combined into an equimolar mixture. The 100 μL PCR reassembly contained 10 μL of 10x cloned pfu buffer, 1 μL of 10 mM dNTPs, 2 μL pfu, and 15 μL

fragments.

Results

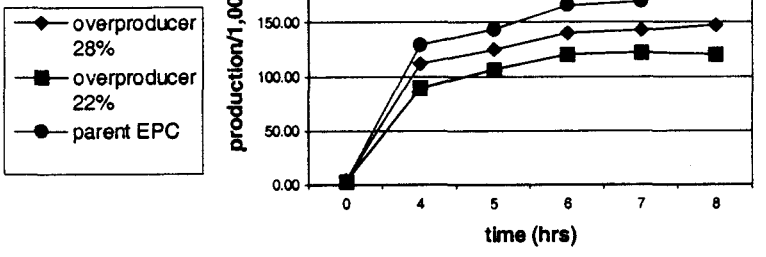
Figure 3. Epi-cedrol production of pooled samples



Initial screening of 2 96-set samples of the library made from the mutant epi-cedrol overproducers resulted in an extensive range of epi-cedrol production. Each sample contained 5 pooled colonies and exceptional production of a sample was determined relative to other samples in its set since a parent control sample was not used. Out of each set, 5 or 6 samples were exceptional producers (Figure 3). Clones from samples with exceptional epi-cedrol production were individually assayed for terpene overproduction. Screening of individual clones against an EPC parental control revealed 2 colonies producing epi-cedrol at 28% and 22% higher levels than the previous best clone. However, when three replicates of each clone were screened and averaged over an 8-hour period, epi-cedrol production for the two colonies did not exceed the parental (Figure 4).

Exon shuffling experiments assayed 400 individual colonies from which 25 samples appeared to not only produce amorphadiene and epi-cedrol, but also a number of unidentified compounds. However, the chromatograph peaks of these new samples indicated relatively low abundance of these compounds. These 25 samples were assayed a second time to confirm sesquiterpene production and only 6 epi-cedrol producers were identified. Amorphadiene producers were absent from this second set of assayed samples and no new compounds were detected.

Figure 4. Average epi-cedrol production of triplicate overproducer confirmation screen



A control experiment was performed to determine if exon reassembly of amorphadiene was possible. ADS and EPC exon fragments were PCR reassembled separately from each other, cloned, and 20 samples from each library assayed. Twenty out of 20 ADS clones functionally produced while 18 out of 20 EPC clones were functional.

RFLPs of the 6 epi-cedrol producers using Msp I and Alu I resulted in 4 banding patterns that matched the parental EPC control. The remaining 2 epi-cedrol producers had banding patterns which were identical but different from the EPC parent gene.

Discussion

The relatively unchanged production of epi-cedrol subsequent to the shuffling of the mutant parents suggests that the N-terminal amorphadiene crossover that all 5 parental genes experienced as a result of prior family shuffling was the important event in overproduction. This was not unexpected since the parent gene sequences were nearly identical to each other. 20 to 60-bp amorphadiene crossovers in addition to various other mutations initially resulted in parental overproduction. The subsequent shuffling of the point mutations between these parent genes did not yield further improvement. Thus it was hypothesized that significant progress in the overproduction of epi-cedrol or the production of new terpenes would require more deliberate methods of shuffling that would be capable of consistently exchanging entire sequences as opposed to single point mutations. Exon shuffling techniques were therefore used to create greater molecular diversity.

The goal of the exon shuffling experiments was to use crossover events to create molecular diversity in genes that functionally produced epi-cedrol or amorphadiene. Those genes would then be used as parents in future shuffling experiments. The four genes that had banding patterns identical to the EPC parent in RFLP experiments suggested that the reassembled exons had simply reformed the parent gene. The banding patterns of the two genes that were different from the EPC parent but identical to each other had probably experienced some type of similar functional crossover event. Future sequencing of these genes is necessary to make any definite conclusions. The molecular diversity in these functional genes will be important in determining their use in future sequencing experiments.

The absence of amorphadiene producers from the screened colonies initially led us to believe that there may be something wrong in the way the amorphadiene exons were reassembling. However, the functional reassembly of the amorphadiene gene in

control experiments indicated that amorphadiene producers in the exon shuffled libraries exist but had avoided detection. Future screening of exon shuffled colonies will therefore pool multiple colonies into single samples to more efficiently isolate functional producers to be used as parents in future shuffling experiments.

Acknowledgements

This work was supported primarily by the ERC Program of the National Science Foundation under Award Number EEC-9731725.

References

Crameri A, Raillard S, Bermudez E, Stemmer WPC. 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature. 391:288-91.

Greenhagen B, Chappell J. 2001. Molecular scaffolds for chemical wizardry: Learning nature's rules for terpene cyclases. PNAS. 98:13479-13481.

Kolkman J, Stemmer WPC. 2001. Directed evolution of proteins by exon shuffling. Nature Biotechnology. 19:423-8.

Joern JM, Meinhold P, Arnold FH. 2002. Analysis of shuffled gene libraries. Jour Mol Bio. 316:643-656.

Lange BM, Rujan T, Martin W, Croteau R. 2000. Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. PNAS. 97:13172-7.

Martin VJ, Yoshikuni Y, Keasling JD. 2002. The in vivo synthess of plant sesquiterpenes by *Escherichia coli*. Biotechnology and Bioengineering. 75:497-503.

Merke P, Bengtsson M, Bouwmeester HJ, Posthumus MA, Brodelius PE. 2000. Molecular cloning, expression, and characterization of amorpha-4,11-diene synthase, a key enzyme of artemisinin biosynthesis in *Artemisia annua* L. Archives of Biochemistry and Biophysics. 381:173-80.

Mercke P, Crock J, Croteau R, Brodelius PE. 1999. Cloning, expression and characterization of epi-cedrol synthase, a sesquiterpene cyclase from *Artemisia annua* L. Archives of Biochemistry and Biophysics. 369:213-22.

Rohdich F, Hecht S, Gartner K, Adam P, Krieger C. Amslinger S, Arigoni D, Bacher A, Eisenreich, W. 2002. Studies on the nonmevalonate terpene biosynthetic pathway: Metabolic role of IspH (LytB) protein. PNAS. 99:1158-63.

Stemmer WPC. 1994. DNA shufflingby random fragmentation and reassembly: In vitro recombination for molecular evolution. PNAS. 91:10747-51.

Yano T, Kagamiyama H. 2001. Directed evolution of ampicillin-resistant activity from a functionally unrelated DNA fragment: A laboratory model of molecular evolution. PNAS. 98:903-7.