

Applications of DNA shuffling to pharmaceuticals and vaccines

Phillip A Patten*, Russell J Howard† and Willem PC Stemmer‡

DNA shuffling is a practical process for directed molecular evolution which uses recombination to dramatically accelerate the rate at which one can evolve genes. Single and multigene traits that require many mutations for improved phenotypes can be evolved rapidly. DNA shuffling technology has been significantly enhanced in the past year, extending its range of applications to small molecule pharmaceuticals, pharmaceutical proteins, gene therapy vehicles and transgenes, vaccines and evolved viruses for vaccines, and laboratory animal models.

Addresses

Maxygen, Inc., 3410 Central Expressway, Santa Clara, CA 95051, USA

*e-mail: phil.patten@MAXYGEN.com

†e-mail: russell.howard@MAXYGEN.com

‡e-mail: pim_stemmer@MAXYGEN.com

Current Opinion in Biotechnology 1997, 8:724–733

<http://biomednet.com/elecref/0958166900800724>

© Current Biology Ltd ISSN 0958-1669

Abbreviations

IFN interferon
kb kilobases
MLV murine leukemia virus

Introduction

The sequence-design processes used by nature have yielded results far superior to those obtained so far by rational approaches to the design of biological structures and systems. The rational design of proteins, for example, is performed by computer modeling of individual changes, followed by construction of the corresponding DNA and expression of the recombinant protein for testing and evaluation. This approach seeks to design proteins for specific tasks by drawing correlations between amino acid sequence and specific shapes in an attempt to understand a complex system through its topography. At its root, our limited ability to rationally engineer complex biological systems stems from three main factors: limited structure/function knowledge of most proteins and of their complex interactions with expression machinery and other molecules within the cell; the computationally intensive and approximate nature of the modelling required; and the fact that the number of relevant mutants for which one would wish to make predictions in order to optimize a given function ranges from millions to numbers of astronomical proportions [1].

In contrast to rational design procedures, nature employs mutation, selection and recombination to evolve highly adapted individuals from the effectively infinite possibilities encoded implicitly in the genome. Recent

technological advances have demonstrated that it is now possible to mimic these natural evolutionary processes. Benchtop *in vitro* evolution techniques are used to construct libraries as large as 10^{15} molecules [2–5]. Researchers can mimic natural evolution by searching these libraries by, for example, affinity panning of phage-displayed or RNA ligands against pharmaceutical targets, for the best candidates for a specific task. Repeated rounds of selection and amplification of candidates has already produced improved enzymes and novel molecules capable of binding their targets with higher affinity than their natural counterparts. Unlike natural selection, in which multiple environmental forces select organisms with genomes that allow them to meet a variety of challenges, *in vitro* evolution exerts focused selection pressure on organisms in isolation, enabling the rapid development of variants with highly specialized traits.

Despite the enormous potential of these techniques, determination of the best strategy to exploit this diversity has been the topic of much debate. The most popular methods of creating combinatorial libraries are strategies that seek to evolve sequences that have individual point mutations or blocks of oligonucleotide encoded mutations. At present, most researchers use either repeated cycles of 'error-prone PCR' [6,7] or repeated oligonucleotide directed mutagenesis [8] to create these 'point mutation' libraries. Error-prone PCR employs a low fidelity replication step to introduce random point mutations at each round of amplification [6]. This method has the advantage of simplicity and ease of use. The power of these methods is limited, however, by one's ability to identify critical regions for mutagenesis and because, generally, only small regions of the genome can be mutagenized to saturation and be exhaustively sampled in screens or selections, due to limited library sizes relative to the size of the sequence spaces defined by exhaustive random searches [9].

Iterative cassette or point mutagenesis can overcome some of these limitations; however, as discussed below, DNA shuffling profoundly accelerates the process. In this review we summarize recent advances which extend the range of application of this technique to small molecule pharmaceuticals, pharmaceutical proteins, gene therapy vehicles and transgenes, vaccines, and evolved viruses for laboratory animal models of disease.

Recombination

A key limitation of random point mutagenesis and random cassette mutagenesis strategies can be traced to the fact that they introduce random 'noise' into the gene population at every cycle, and hence improvements are limited to small steps. If the noise level is too high relative to the library size and the selection stringency, the

message will gradually become riddled with deleterious mutations. This is analogous to the phenomenon of Muller's Ratchet [10] from population biology in which, in the absence of sexual recombination, deleterious mutations build up in a population over time. From decades of plant and animal breeding, classical breeders have learned to use recombination to rapidly evolve improved sequences for a specific task. Sexual replication in combination with directed selection can produce substantial improvements in a highly diverse genome within just a few cycles. The widespread prevalence in nature of sexual versus asexual reproduction has been the topic of much debate [11]. A gene favouring parthogenic rather than sexual reproduction would quickly take over in a population unless counterveiled by selection, the so-called twofold cost of sex. One hypothesized major force for the maintenance of sexual reproduction is that recombination of natural diversity allows for populations to rapidly evolve in order to adapt to changing physical environments or pathogens by combining beneficial mutations into single, more fit individuals and deleterious mutations into less fit individuals that are then selected out of the gene pool [11].

The technique of DNA shuffling comes the closest of any laboratory technique to mimicking natural recombination by allowing the *in vitro* homologous recombination of DNA [12,13]. This technique not only recombines DNA fragments, but also introduces point mutations at a very low, controlled rate [14], thus, combining recombination, point mutation, and selection techniques to create a general, parallel algorithm for evolving improved genes. This parallel search strategy is analogous to that used by massively parallel supercomputers [9], such as those used for climate modelling or to factor very large prime numbers. Figures 1 and 2 summarize the practical and theoretical advantages of DNA shuffling relative to existing recursive mutagenesis methods, such as error-prone PCR or recursive oligonucleotide directed mutagenesis. Recent progress with DNA shuffling has clearly demonstrated the utility of recombination for accelerating molecular evolution through simultaneously permuting both single mutations and large sequence blocks (Table 1). DNA shuffling combined with focused selection pressure in the laboratory will allow one to rapidly evolve genes for a wide variety of industrial applications: the optimization of enzymes, such as proteases, lipases, amylases and cellulases; the development of metabolic pathways specialized to synthesize large amounts of specialty chemicals, antibiotics, or pharmaceutical proteins; organisms designed for bioremediation; and plasmids or viruses for novel vaccines and gene therapy applications (Table 2). These emerging applications are discussed below.

Family shuffling dramatically increases the rate of stepwise evolution

Forced hybridization between species, such as was performed in the cross-breeding of the plum and apricot

to yield the plumcot, was recognized last century by plant breeders as a highly effective method for generating novel, functional varieties with phenotypes differing dramatically from either parent [15]. The utility of this strategy has now been demonstrated at the level of a single gene. To evaluate whether recombining natural diversity accelerates the evolution process, the efficiency of obtaining a new substrate specificity from four homologous enzyme genes evolved separately was compared with that from a recombined pool of the four genes. The essential goal was to compare the rate of evolution of a single gene that is subjected to random mutation and selection to the rate of evolution of a library of genes created by recombining existing, functional genetic diversity present in a family of homologous genes. Since all of the recombinants are created from diversity that has proven functional in the context of its parental gene, the hope is that such libraries would be of much higher quality than random libraries. The results affirm this view. One cycle of single gene shuffling yielded eightfold improvements from each of the four separately evolved genes, versus a 270–540-fold improvement from the four genes shuffled together ([16••]; Figure 3a). This represents an approximately 50-fold increase in the rate of improvement per cycle. The best clone contained eight segments from three of the four genes as well as 33 amino acid point mutations. It is worth emphasizing that this evolved improvement relative to the initial gene pool was obtained in a single cycle of gene shuffling, rather than requiring many recursive cycles.

Thus, in contrast to classical breeding techniques, DNA shuffling allows one to readily recombine DNA derived from 'separate' species or genera. This results in a much more sparse sampling of sequence space (Figure 3b), in which the average similarity between library members is much lower than with other strategies. Sparse sampling yields mutants that, after a single cycle, are far more divergent from the parental genes than is possible with single gene shuffling or point mutation strategies. This recent experiment demonstrates that cross-species recombination is a remarkable accelerant of molecular evolution. Family shuffling will be widely applicable to the commercially important problems discussed below.

Protein pharmaceuticals

Recombinant pharmaceutical proteins form a multi-billion dollar sector of the pharmaceutical industry. This industry relies principally on cloning existing genes encoding cytokines, growth factors, and enzymes that are the products of millions of years of evolution. Many of these products have side effects that cause serious complications which limit or preclude their clinical use. Selective breeding using DNA shuffling provides a technology for rapidly improving pharmaceutical proteins through selective breeding for enhancement of desirable biological activities, while eliminating or reducing undesired activities (Table 2).

Table 1

Genes and operons evolved by DNA shuffling.

System	Improvement	Size	Cycles	Mutations	Comments	Reference
TEM-1 β -lactamase	Enzyme activity 32,000-fold	1 kb	3 + 2 [†]	6 aa	Selection MIC 0.02 μ g/ml to 640 μ g/ml In comparison, three cycles of mutagenic PCR and selection=16-fold	[12]
β -lactamase family	Enzyme activity 270–540-fold	1 kb	1	Chimerics	Selection 500-fold jump in fitness in one round of shuffling and screening 50,000 colonies	[16]
β -galactosidase	Fucosidase activity 66-fold Substrate specificity 1,000-fold	4 kb	7	6 aa 13 bp	Screen 10,000 colonies per round, best 20–40 chosen per cycle 66-fold fucosidase activity: 2–3-fold increase expression plus 20-fold increase activity 11/13 base mutations in coding sequence	[30**]
Green fluorescence protein	Protein folding 45-fold (<i>E. coli</i> and mammalian cells)	0.8 kb	3	3 aa 6 bp	Screen 10,000 colonies per round, best 20–40 chosen per cycle Protein folding improvement	[31]
Antibody (scFv)	Avidity >400-fold	0.8 kb	8 + 2 [†]	34 aa	Phage panning Naive human antibody library Stop codon modified to suppressible stop codon	[32]
Antibody (scFv)	Expression level 100-fold	0.8 kb		5	Phage panning Murine hybridoma library	[32]
Arsenate operon	Arsenate resistance 40-fold	2.3 kb	3	3 aa 13 bp	Selection Three genes; 5.5 kb plasmid shuffled 10,000 colonies per cycle All three mutations in <i>arsB</i> efflux pump, <i>arsC</i> was predicted to be rate limiting Arsenate reduction activity up 5–10-fold Growth in 0.5 M arsenate Plasmid integrated in round 3	[19]
Atrazine degradation	Atrazine degradation 80-fold	5.6 kb	4	8 aa [‡]	Screen Two genes; 8.5 kb plasmid shuffled Atrazine chlorohydrolase activity increased Wild-type inactive versus terbutylazine, evolved enz active Screen: insoluble atrazine to soluble product gives zone of clearing	
Alkyl transferase	DNA repair 10-fold	0.5 kb	6	7 aa	Selection Suicide enzyme: limited potential for improvement	FC Christians, G Dawes, WPC Stemmer, unpublished data
Benzyl esterase	Antibiotic deprotection 150-fold	1.5 kb	2	8 aa	Screen Four rounds of PCR mutagenesis and screening for improved variants prior to shuffling	[20]
tRNA synthetase	Charging of engineered tRNA 180-fold	2.0 kb	7	nd	Selection	[22]

[†]Additional backcross cycles. [‡]Rare mutation: 3 NA insertions in 5 codons restored frame, inserted 1 aa with 5 other aa changed atz gene lost 2 kb in final mutant, no loss of function. aa, amino acid changes; bp, base pair changes; nd, not determined.

We have recently applied DNA shuffling to members of the human α -IFN gene family (85–97% pairwise amino acid identity). Greater than 10^{26} distinct recombinants can be generated from the natural diversity in this gene family (Figure 4a). While no foreseeable library technology will allow an 'exhaustive' sampling of this sequence space, DNA shuffling technology provides a design algorithm with which to selectively breed for IFNs with increased potency relative to naturally occurring IFNs. Typical chimeras produced by DNA shuffling are shown in Figure 4b. Generic high-throughput methods for α -IFN expression and biological assay as fusion proteins on phage have been developed and used for rapid parallel analysis of recombinant IFNs. Phage-displayed recombinants with improved potency on human and murine cells have been obtained (Figure 4).

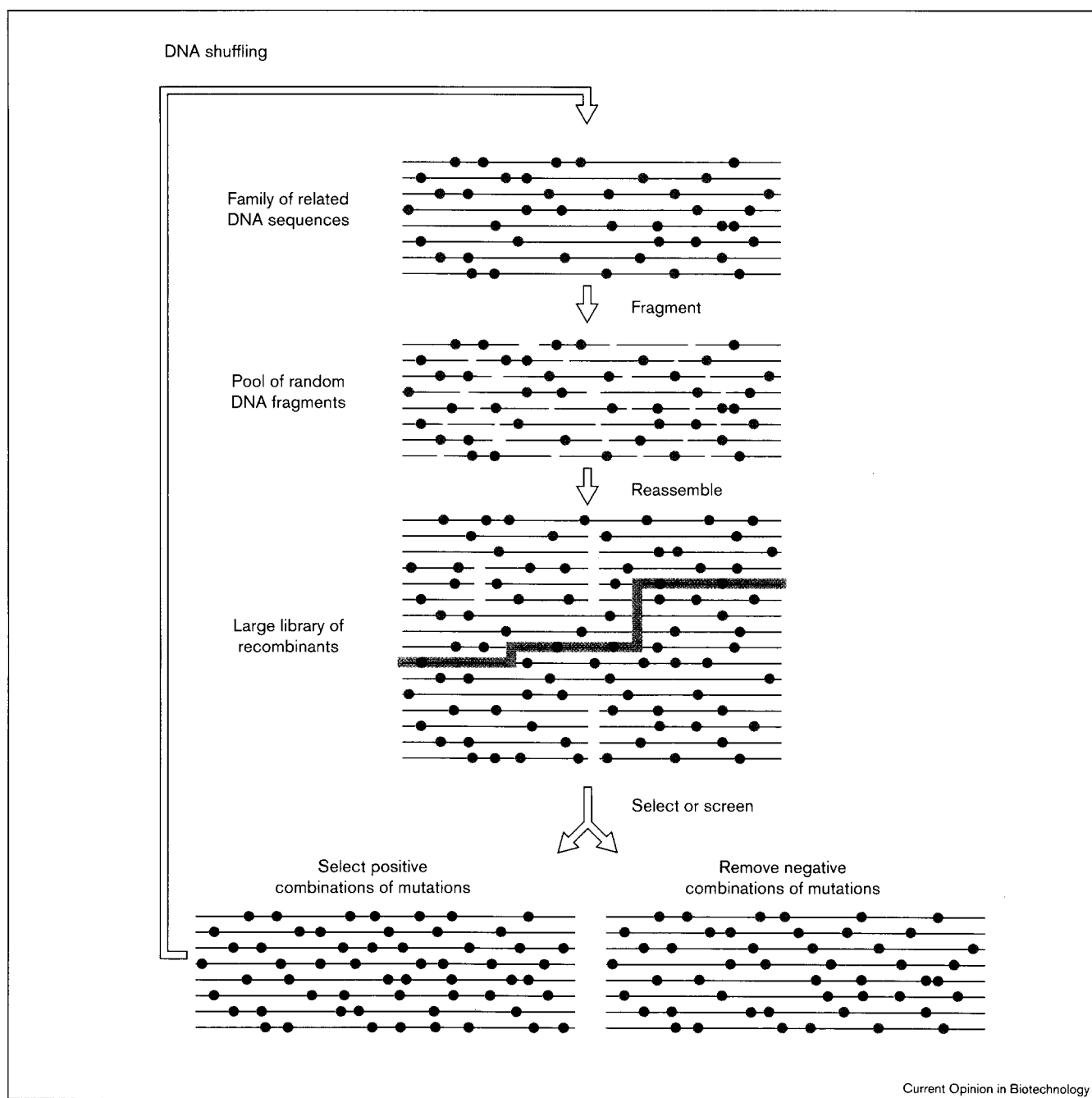
This generic approach can be applied to improve many pharmaceutical proteins. Proteins with novel activities have been created by directed chimerization of modules from existing pharmaceutical proteins [17,18], a strategy that is likely to be particularly effective with cytokines,

which typically act by dimerizing two or more receptor components. As with the rapid evolution of moxolactamase activity by shuffling a cephalosporinase gene family [16**], many new pharmaceutical activities may be discovered through breeding large libraries of chimeric pharmaceutical proteins. Selective breeding using DNA shuffling will allow rapid evolution of pharmaceutical proteins with potent activities from such recombinants, which initially have low levels of the desired activity. Backcrossing of these evolved variants with the wild type genes will allow one to remove functionally neutral changes, thus reducing the immunogenicity of the evolved proteins.

Small molecule pharmaceuticals and industrial enzymes

Microorganisms are widely used for the production of pharmaceutical molecules, such as antibiotics, antifungals, anticancers and immunosuppressives. In many cases, the genes encoding the relevant biosynthetic enzymes are known, often occurring in operons or gene clusters. Rational engineering of these biosynthetic pathways to improve yield or generate analogs is difficult because,

Figure 1

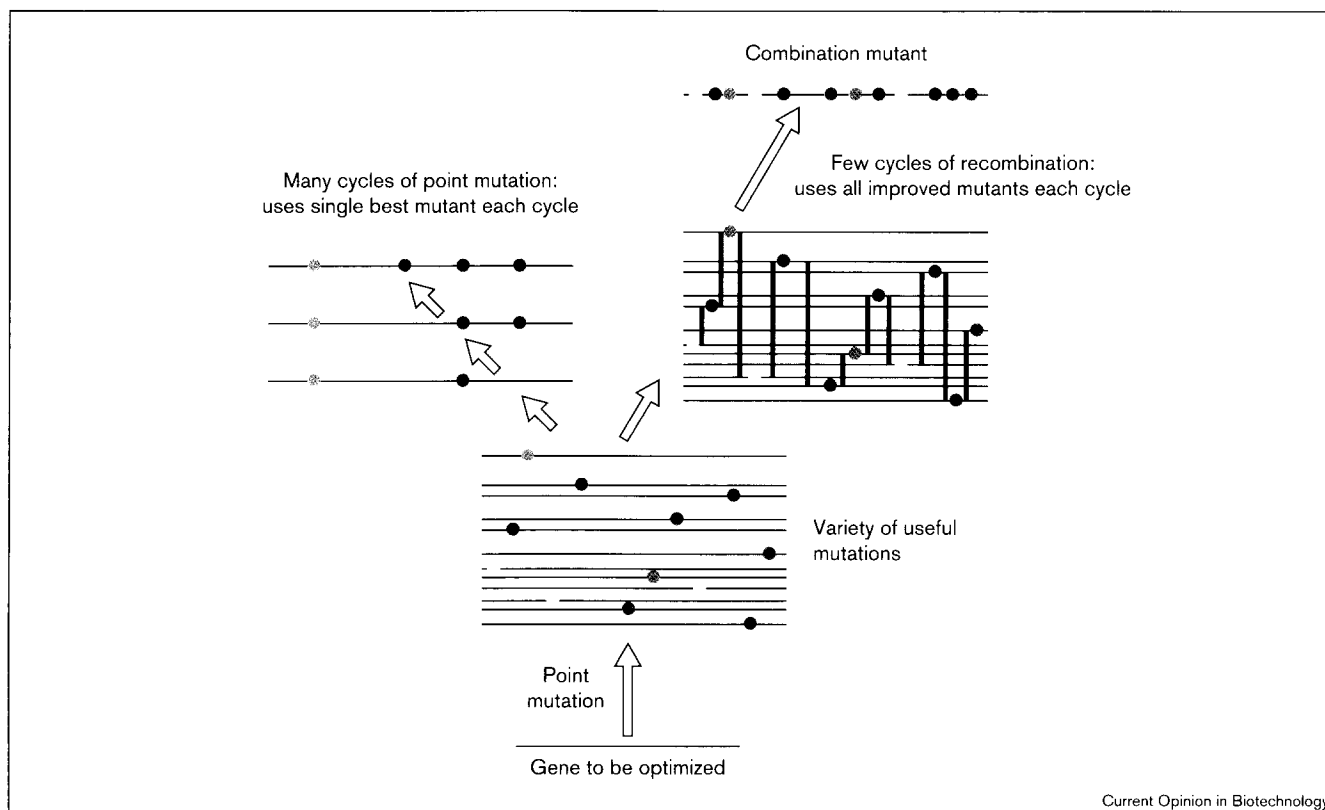


DNA shuffling methodology. The first step of this method is to randomly fragment a population of related genes using DNase I. This produces fragments of various lengths that, after denaturation, hybridize to form an equal mixture of 5' and 3' overhangs. Using PCR techniques, the 5' overhang fragments can be extended by Taq DNA polymerase—leaving the 3' overhang fragments unaffected. As a consequence of this extension, the average fragment length increases during each cycle. Recombination occurs when a fragment derived from one template primes a template with a different sequence [13]. Green dots represent beneficial mutations and red dots represent deleterious mutations. The coloured bars indicate recombinations of portions of three parents into recombinant progeny.

in addition to the difficulties of protein engineering, the determination of rate limiting steps in the pathway is laborious and uncertain. DNA shuffling is well suited to the optimization of such pathways because the entire pathway can be treated as the unit for evolution, with no requirements for knowledge of the rate limiting steps

or for detailed structure/function analysis of the proteins. Pathways for the detoxification of atrazine (J Minshall, personal communication) and arsenate [19•] have been improved using DNA shuffling (Table 1). Importantly, in these examples no *a priori* knowledge was needed to yield significant improvement. A benzylesterase used

Figure 2



The advantages of DNA shuffling as a sequence design algorithm for evolving complex new gene functions are shown schematically. The coloured dots represent beneficial mutations. The vertical direction represents a generalized measure of fitness (i.e. kcat/Km for an enzyme) with the fittest genes being at the top. Because the frequency of beneficial mutations is generally low relative to deleterious mutation, only single beneficial mutations are generally added in each cycle of random mutagenesis and screening or selection. Hence, procedures that use iterative point mutations must build up beneficial mutations one at a time through many rounds of selection, generally with only the best mutant from any given cycle being pursued. In contrast, DNA shuffling allows one to directly recombine all beneficial mutations from any given round into multi-step mutants with dramatically improved phenotypes.

industrially for deprotecting a precursor of the antibiotic loracarbef has been improved using DNA shuffling ([20]; Table 1). Recombination was shown to be superior to sequential error-prone point mutation for the evolution of this activity [21•]. DNA shuffling has also recently been used to modify the specificity of a tRNA charging enzyme [22], with the ultimate goal of evolving tRNA synthetases that can specifically charge tRNA's with unnatural amino acids incorporated at specific sites.

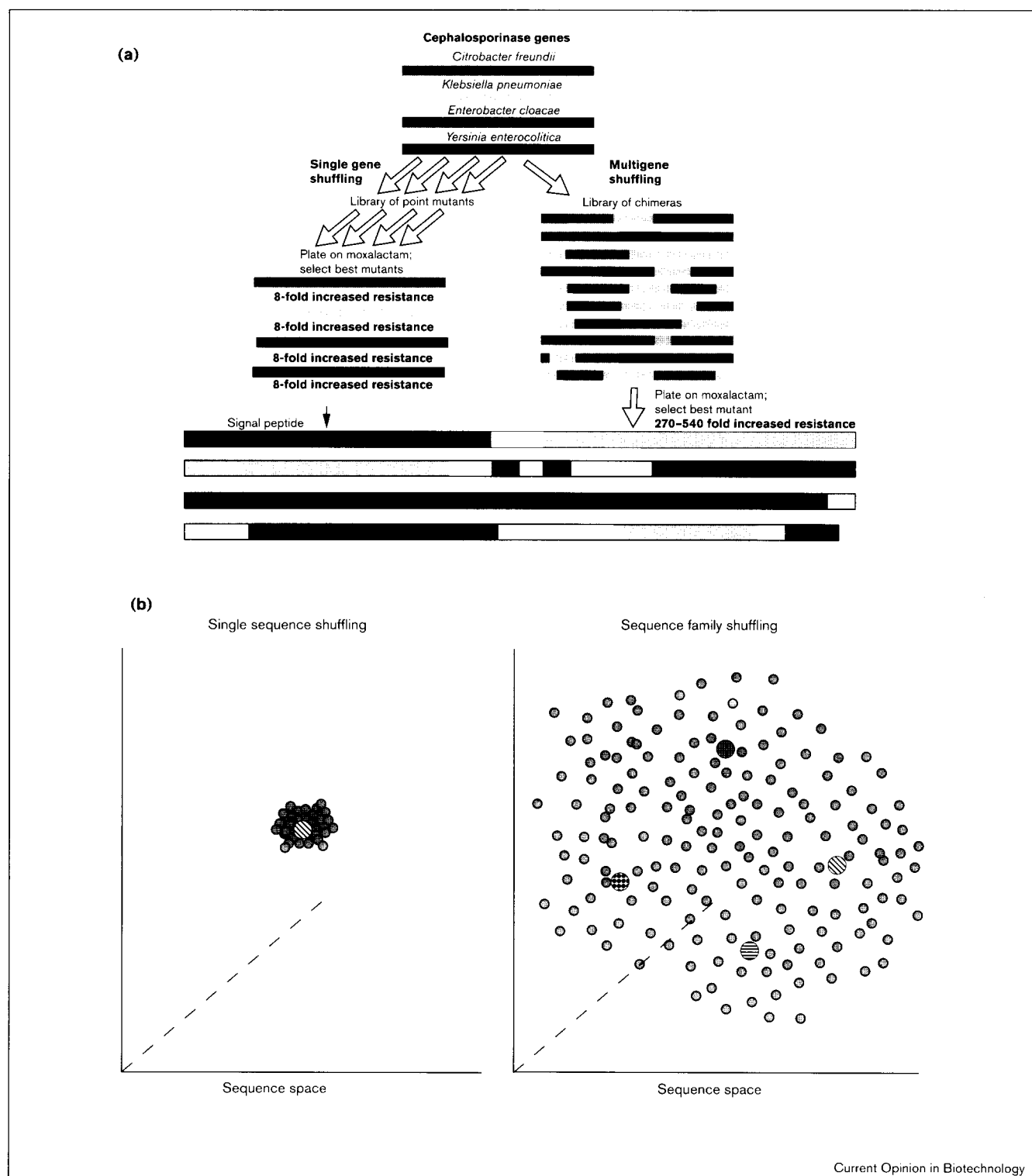
Evolved enzymes will find wide application in the replacement of multi-step chemical reactions required for manufacture of drugs or their precursors by an enzymatic conversion. Most naturally occurring enzymes capable of such valuable conversions require significant modification in activity, specificity, or expression level before they are suitable for large scale drug manufacture. DNA shuffling provides an important tool for the optimization of such enzymatic conversions.

Evolved viruses for pharmaceutical applications

The full length genomes of many viruses are in the range of 5–15 kilobases (kb), a size range that can be readily handled by current DNA shuffling methods (Table 1). Our ability to clone and sequence the wealth of natural viral isolates far outstrips our molecular understanding and our ability to rationally manipulate them. Three wild-type strains of human papilloma virus have been successfully shuffled (D Apt, personal communication). The biological properties of this library of recombinants are currently being investigated. This approach has potential for the evolution of human papilloma virus to overcome the blocks to growth in transformed fibroblasts, and thus be able to grow in readily manipulated tissue culture systems for drug screening.

Adenovirus is widely used as a gene therapy vehicle. Over 100 naturally occurring serotypes with differing

Figure 3



(a) The strategy and results from shuffling four homologous cephalosporinase genes are shown schematically. Single gene shuffling resulted in eightfold increased resistance to the antibiotic moxalactam, whereas shuffling the gene family gave a 270–540-fold increase in resistance in a single step [16]. **(b)** Evolution starting from a single gene is schematically contrasted with evolution based on shuffling a homologous gene family. The axes denote a generalized sequence space. Shaded dots indicate particular sequences present in a given library. Hatched dots represent sequences that are more fit than the best parental molecule. Greater distance indicates greater sequence divergence. Family shuffling results in a relatively 'sparse' sampling of sequence space with relatively few individuals that are highly similar to the parental molecules and many individuals that are very divergent from the parents.

Table 2**Development of novel human therapeutics through molecular breeding technology.**

Therapeutic area	Size of shuffled DNA	Novel properties for screening or selection
Protein pharmaceuticals	1–5 kb	Receptor selectivity Improve agonist/antagonist activity Novel agonist/antagonist activity Toxicity Optimal expression in desired host – shuffle the expression vector and/or protein gene Plasma half-life/protein and cell binding
Small molecule pharmaceuticals	1–5 kb	Manufacturing enzymes: desired properties for enzymes that replace chemical synthetic steps in drug manufacture (pH and temperature optimization, solvent tolerance, etc.)
	1–20 kb	Natural products: generation of drug analogs/new structures by shuffling entire pathways – screen for pharmaceutical activity Strain improvement
Gene therapy	1–20 kb	Vectors Transgenes
DNA vaccines	1–20 kb	Plasmid properties of promoter strength, tissue tropism and cell entry, plasmid stability Antigen shuffling to optimize immune response Shuffling to improve immunostimulatory sequences Qualitative nature of immune response (shuffling cytokine gene cassettes)
Recombinant protein vaccines	1–20 kb	Optimal expression in desired host – shuffle the expression vector and/or the protein gene while selecting for retention of recognition by protective mAbs Shuffling to generate family of immunogens for antigenically variant targets
Viral vaccines	1–20 kb	Enable viral growth/optimize yield in desired cell for manufacturing Select for retention of desired properties when viral families are shuffled – neutralization assay, antibody effects Select for attenuation (growth rate, temperature, retention of some properties)
Evolved viruses	1–20 kb	Select for optimal expression of recombinant antigen(s) genes in viral vaccine vectors Evolve to grow in animal models such as mice Evolve viruses to grow in easily manipulated tissue culture systems

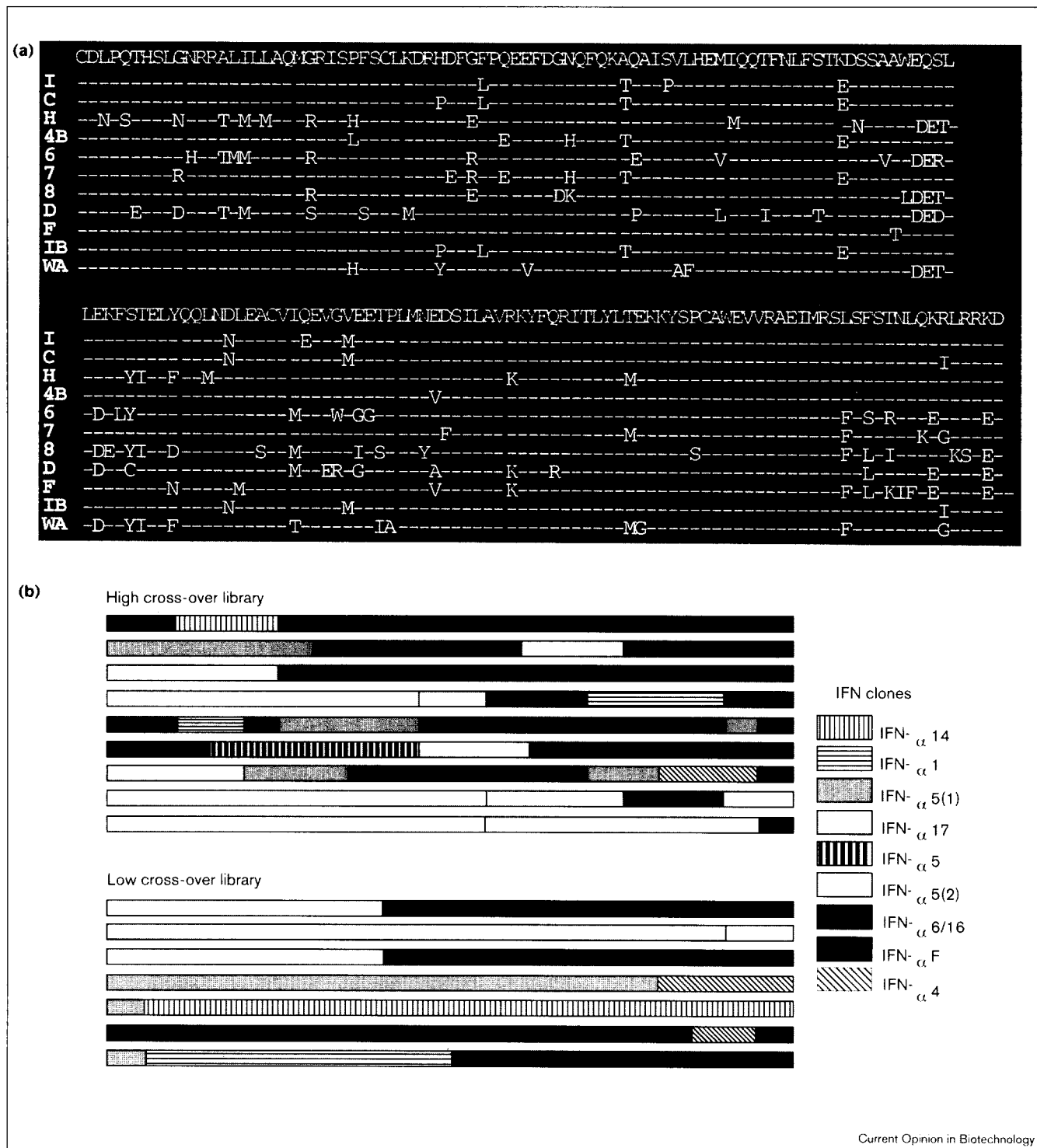
tropisms are known. The fiber and penton genes of adenovirus are the major determinants of tissue tropism, as they are responsible for cell adhesion. The pentons of adenoviruses interact with cellular integrins [23] and the fibers with cellular receptors, one of which has been newly identified [24]. Evolved adenoviral variants which could be selectively targeted to particular cell types would be of great utility for gene therapy and vaccine delivery vectors. The penton and fiber genes of various adenovirus serotypes have been shuffled (S Liu, personal communication) and this approach may allow one to evolve mutant adenoviral vectors with desired cell and tissue tropism.

Murine leukemia virus (MLV) is a retroviral vector that has received much attention as a gene therapy vehicle. As with adenovirus, there are <100 naturally occurring strains, whereas only a single MLV strain is being developed for gene therapy applications (Moloney MLV). The envelopes of 15 MLV strains have been shuffled and viruses are being selected for improved tropism, titer, stability and gene expression properties (N Soong, personal communication).

HIV-1 poses a major threat to human health which is increasing because of the growing viral load worldwide, the high rate of evolution of this pathogen, and the concomitant evolution of associated opportunistic pathogens

in HIV-1 infected individuals. There is currently no practical animal model in which to test the multitude of antiviral drugs or vaccine strategies [25]. Work is beginning on the genetically engineered animal models to support replication of HIV-1 (D Littman, M Goldsmith, personal communication), but no replication has yet been observed in these hosts. It is clear from viral phylogenetic trees that lentiviruses can evolve the ability to grow in new species and it is clear that recombination plays a major role in the natural high rate of evolution of lentiviruses [20]. DNA shuffling provides a powerful new tool with which to accelerate the adaption of viruses to grow in laboratory animal hosts. Recombination is believed to be of great importance for the naturally high rate of evolution of retroviruses [26]. Laboratory animals have been engineered with the human HIV-1 receptor and co-receptor genes (D Littman, personal communication). DNA shuffling is being used to recombine entire genomes and individual genes of natural HIV-1 isolates to accelerate the adaptation of HIV-1 to grow in these engineered animals (P Patten and N Landau, unpublished data). We anticipate that the adaptation of HIV-1 to replicate in a laboratory animal will open up many fertile avenues for drug and vaccine discovery on this important human pathogen. Shuffling of natural diversity to create large libraries of chimeric viruses is a general approach that can be applied to other viruses, such as the hepatitis B

Figure 4



(a) Human α -interferon diversity. The sequences of eleven natural human α -IFN sequences are shown [29]. Consensus α -IFN is given at the top. Dashes indicate identity to consensus. The number of distinct recombinants that can be generated by shuffling these eleven genes is 3×10^{26} . This number is calculated by multiplying the number of different amino acids observed at each polymorphic site ($2^6 \times 3^{15} \times 4^4 = 3 \times 10^{26}$). **(b)** Sixteen representative recombinant α -IFNs derived by shuffling eight natural human α -IFN genes are shown schematically (P Patten, unpublished data). The high crossover and low crossover libraries were generated by shuffling 20–50 bp or 50–100 bp fragments, respectively. The DNA shuffling was done essentially as described in [16]. The shuffled IFNs were expressed as fusions to gene III on bacteriophage M13 and screened for antiproliferative activity using Daudi cells. Phage displayed α IFN-MAX4 (high crossover library, fourth from the top) is 40-fold more active in a Daudi antiproliferation assay than IFN2 α and twofold more active than consensus 1 IFNs displayed on phage.

virus and hepatitis C virus, for the purpose of producing domesticated forms of these viruses (i.e., easier to handle in the laboratory) for vaccine development and drug screening on variants that can readily be grown in tissue culture systems.

Vaccines

Development of effective vaccine technologies has stimulated renewed government emphasis and interest from pharmaceutical researchers. DNA vaccines are particularly attractive because of their relatively low cost and the feasibility of rapid generation of diverse variant vaccines containing evolved promoters, immunostimulatory sequences, cytokines, etc., for comparison testing [27]. For example, assays for selection of DNA vaccines with improved promoter activity, immunostimulatory sequences, enhanced expression levels or cell tropism can be developed to produce second generation DNA vaccine plasmids. Viral vaccine vectors can be enhanced by DNA shuffling to give desired properties of tropism, stability and expression level. The promise of this new technology notwithstanding, existing human vaccines rely on live or killed whole organisms or components purified from whole organisms. We envision opportunities for DNA shuffling as a tool for increasing the efficiency and success rate of the development of novel whole organism, viral, bacterial and recombinant protein vaccines. Pathogenic viruses can be subjected to DNA shuffling followed by selection for desired attenuation properties while retaining the immunogenicity required for a vaccine. Viruses that would serve as excellent vaccine vectors or as vaccines but which cannot be manufactured to sufficient titer in manufacturing cell lines, can be shuffled and selected for improved titre to create new commercial opportunities. Recombinant proteins that are known to be excellent vaccine immunogens but which cannot be manufactured in appropriate yield or in suitable host systems can be shuffled and screened to solve such expression problems, while co-selecting for retention of necessary epitopes. These and other valuable opportunities for application of DNA shuffling technologies in vaccinology are summarized in Table 2.

The impact of genomics

The rapid rate of increase in the availability of known gene sequences and our ability to manipulate them in cloned form greatly exceeds our understanding of these genes and our ability to engineer these sequences based on rational models. Sequence information from informatics databases can readily be converted into functional DNA clones given tools such as PCR, synthetic DNA and methods for the rapid assembly of genes from synthetic DNA [28].

Shuffling of natural diversity to explore the sequence space defined by shuffled homologues is a demonstrably powerful strategy for accelerated evolution of biological molecules with novel activities ([16]; Figure 3). We expect the dramatic success seen with family shuffling

of cephalosporinase genes to be repeated in many other systems in which homologous genes are recombined. The explosion of DNA sequences available on the internet provides a rich, diverse and rapidly expanding supply of molecular breeding stock. DNA shuffling is a general and natural algorithm for functionally exploiting this natural diversity, with minimal requirements for *a priori* genetic or biochemical characterization to guide this exploitation.

Conclusions and future directions

In order to unlock desired biologically active sequences from the potential diversity present in an organism's gene pool, it is of great importance to understand which evolution algorithms are the most effective. Point mutation techniques, such as error-prone PCR and repeated oligonucleotide directed mutagenesis, search sequence space by creating libraries of randomly mutated molecules. In contrast, DNA shuffling exchanges large functional blocks of sequence containing previously selected mutations to search for the best candidate molecules, thus mimicking and accelerating the process of sexual recombination. We expect that, just as recombination has played a major role in the evolution of life, DNA shuffling will play a central role in the development of applied molecular evolution technologies and will prove indispensable for bringing existing biological diversity into the service of human health care, and agricultural and industrial chemical needs.

Acknowledgements

We thank Doris Apt, Scott Liu, Nai Wei Soong, Juha Punnonen and Jeremy Minshull for helpful discussions and for sharing unpublished results.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
 - of outstanding interest
1. Kauffman SA: *The Origins of Order*. New York: Oxford University Press; 1993.
 2. Joyce GF: **Directed molecular evolution**. *Sci Am* 1992, **267**:90-97.
 3. Bartell DP, Szostak JW: **Isolation of new ribozymes from a large pool of random sequences**. *Science* 1993, **261**:1411-1418.
 4. Tuerk C, Gold L: **Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase**. *Science* 1990, **249**:505-510.
 5. Breaker R, Joyce GF: **Emergence of a replicating species from an in vitro RNA evolution reaction**. *Proc Natl Acad Sci USA* 1994, **91**:6093-6097.
 6. Caldwell RC, Joyce GF: **Randomization of genes by PCR mutagenesis**. *PCR Methods Applic* 1992, **2**:28-33.
 7. You L, Arnold FH: **Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide**. *Protein Eng* 1994, **9**:77-83.
 8. Reidhaar-Olson J, Bowie J, Breyer RM, Hu JC, Knight KL, Lim WA, Mossing MC, Parsell DA, Shoemaker KR, Sauer RT: **Random mutagenesis of protein sequences using oligonucleotide cassettes**. *Methods Enzymol* 1991, **208**:564-586.
 9. Stemmer WPC: **Searching sequence space using recombination to search more efficiently and thoroughly**

- instead of making bigger combinatorial libraries. *Bio-Technology* 1995, **13**:549-555.
10. Muller HJ: **The relation of recombination to mutational advance.** *Mutat Res* 1964, **1**:2-9.
 11. Maynard Smith J: **The evolution of sex.** In *The Evolution of Sex*. Edited by Bellig R, Stevens G. San Francisco: Barper and Row; 1988:3-20.
 12. Stemmer WPC: **Rapid evolution of a protein *in vitro* by DNA shuffling.** *Nature* 1994, **370**:389-391.
 13. Stemmer WPC: **DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution.** *Proc Natl Acad Sci USA* 1994, **91**:10747-10751.
 14. Zhao H, Arnold FH: **Optimization of DNA shuffling for high fidelity recombination.** *Nucleic Acids Res* 1997, **25**:1307-1308.
 15. Burbank L: **Short-cuts into the centuries to come: better plants secured by hurrying evolution.** In *Luther Burbank His Methods and Discoveries: Their Practical Application*, vol 1. Edited by Whitson J, Williams RUHS. New York: Luther Burbank Press; 1914:176-210.
 16. Cramer AC, Raillard S, Stemmer WPC: **DNA shuffling of a family of genes from diverse species accelerates directed evolution.** *Nature* 1997, in press.
- This paper describes the use of homologous gene family shuffling to obtain large jumps in fitness in a single cycle of DNA shuffling. This is the first paper to show the utility of family shuffling for rapid evolution of a gene and it is a technique which is expected to open up approaches to many new problems by directed evolution of gene families.
17. Wolfson AJ, Kanaoka M, Lau F, Ringe D, Yound P, Lee J, Blumenthal J: **Modularity of protein function: chimeric interleukin 1 betas containing specific protease inhibitor loops retain function of both molecules.** *Biochemistry* 1993, **32**:5327-5331.
 18. Campbell RK, Bergert ER, Wang Y, Morris JC, Moyle WR: **Chimeric proteins can exceed the sum of their parts: Implications for evolution and protein design.** *Nat Biotechnol* 1997, **15**:439-443.
 19. Cramer A, Dawes G, Rodriguez E, Silver S, Stemmer WPC: **Molecular evolution of an arsenate detoxification pathway by DNA shuffling.** *Nat Biotechnol* 1997, **15**:436-438.
- This is the first paper to describe the use of DNA shuffling to evolve a multi-gene operon or gene cluster.
20. Moore JC, Arnold FH: **Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents.** *Nat Biotechnol* 1996, **14**:458-467.
 21. Arnold FH, Moore J: **Optimizing industrial enzymes by directed evolution.** *Adv Biochem Eng Biotechnol* 1997, **58**: 2-14.
• This paper describes the evolution of an industrial enzyme using both asexual random point mutation and DNA shuffling strategies.
 22. Liu DR, Magliery TJ, Pastnak M, Schultz PG: **Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*.** *Proc Natl Acad Sci USA* 1997, in press.
 23. Mathias P, Wickham T, Moore M, Nemerow G: **Multiple adenovirus serotypes use αv integrins for infection.** *J Virol* 1994, **68**:6811-6814.
 24. Bergelson JM, Cunningham JA, Droguett G, Kurt-Jones EA, Krithivas A, Hong JS, Horwitz MS, Crowell RL, Finberg RW: **Isolation of a common receptor for Coxsackie B viruses and adenoviruses a and 5.** *Science* 1997, **275**:1320-1323.
 25. Stott J, Almond N: **Assessing animal models of AIDS.** *Nat Med* 1995, **1**:295-297.
 26. Wei-Shau H, Temin HM: **Retroviral recombination and reverse transcription.** *Science* 1990, **250**:1227-1233.
 27. Ulmer JB, Donnelly JJ, Liu MA: **Toward the development of DNA vaccines.** *Curr Opin Biotechnol* 1996, **7**:653-658.
 28. Stemmer WPC, Cramer A, Ha KD, Brennan TM, Heyneker H: **Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides.** *Gene* 1995, **164**:49-53.
 29. Henco K, Brosius J, Fujisawa A, Haynes JR, Hochstadt J, Kovacic T, Pasek M, Schambeck A, Schmid J, Todokoro K *et al.*: **Structural relationship of human interferon alpha genes and pseudogenes.** *J Mol Biol* 1985, **185**:227-260.
 30. Zhang J, Dawes G, Stemmer WPC: **Evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening.** *Proc Natl Acad Sci USA* 1997, **94**:4504-4509.
• This paper illustrates how visually screening a modest number of colonies (rather than using selection), combined with DNA shuffling, can give substantial improvements in a relatively large enzyme within a small number of cycles.
 31. Cramer A, Whitehorn E, Tate E, Stemmer WPC: **Improved green fluorescent protein by molecular evolution using DNA shuffling.** *Nat Biotechnol* 1996, **14**:315-319.
 32. Cramer A, Cwirla S, Stemmer WPC: **Construction and evolution of antibody-phage libraries by DNA shuffling.** *Nat Med* 1996, **2**:100-102.