© 1996 Nature Publishing Group http://www.nature.com/naturebiotechnology

PCR – DIRECTED EVOLUTION

DNA shuffling brightens prospects for GFP

Ichiro Matsumura and Andrew D. Ellington

Every so often, a technique revolutionizes biotechnology. DNA shuffling is such a technique, allowing researchers quickly and efficiently to direct protein evolution. In this issue of *Nature Biotechnology*, Willem Stemmer's group at Affymax (Palo Alto, CA) extend the horizons of DNA shuffling from selection to screening. Using an iterative approach, they produced a variant of the green fluorescent protein (GFP) — a reporter molecule that is fast becoming a workhorse in gene expression studies in molecular biology and biotechnology — that is an order of magnitude better than commercially available reporters[1].

Traditionally, random mutagenesis and selection have been used to produce and identify proteins with, for instance, altered substrate specificities[2] or increased affinities for a particular ligand[3]. However, the number of possible evolutionary pathways is dauntingly large and limits the usefulness of random techniques. For example, bacterial expression libraries generally contain fewer than 10^8 clones. Since the number of possible double mutants of a 500 amino acid protein exceeds 10^6, beneficial mutations must slowly be accreted in such a system. DNA shuffling vastly accelerates the acquisition of novel phenotypes by delimiting the most productive evolutionary pathways.

DNA shuffling is an *in vitro* recombination technique based on the polymerase chain reaction (PCR). Genes of similar nucleotide sequence are pooled and randomly nicked in controlled digestion reactions with deoxyribonuclease I. The fragments are reassembled in a thermocycled PCR-like reaction but without added primers. In this reaction, the randomly nicked DNAs denature and the single-stranded fragments re-anneal, often switching partners. The annealed fragments prime each other in polymerase catalyzed extension reactions that fill in the nonoverlapping regions. After many cycles, full length genes are regenerated. This process recombines all of the genes simultaneously, rather than pairwise, and also introduces random point mutations into the recombinant alleles. The recombinant products are then amplified in a conventional PCR with primers and subcloned into an expression vector for additional rounds of screening or selection[4].

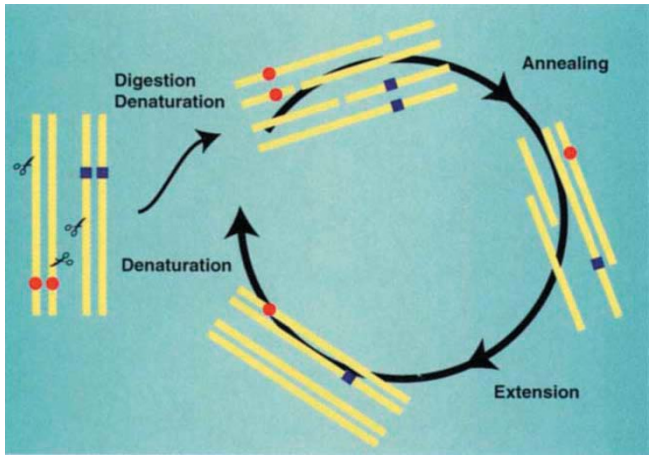

Figure 1. DNA shuffling after random nicking (scissors) of mixtures of alleles (blue squares, red circles). DNA shuffling can produce full length DNA strands carrying both mutations after one round of denaturation, annealing and extension. Shorter products require more than one round to attain full length.

The virtue of DNA shuffling is that it unites beneficial mutations isolated from a round of selection or screening in a single molecule, thereby obviating the need to independently regenerate them in future rounds of mutagenesis. Assuming that such mutations are either additive or synergistic, this permits optimization of a desired protein function in just a few cycles of mutagenesis and selection. For instance, three rounds of DNA shuffling and selection were sufficient to yield a variant of β-lactamase that imparted a 16,000-fold higher resistance to the antibiotic cefotaxime than the wild-type gene[5]. A comparable study using conventional mutagenesis and selection techniques improved TEM-1 catalyzed cefotaxime hydrolysis by only 16-fold[6].

The challenge for Stemmer and his coworkers with GFP was that, although the reporter protein was already proving very useful, the fluorescence levels at which it could be detected were still coarse compared with those of luciferase or other diagnostic enzymes. Their approach was straightforward. After DNA shuffling, they simply picked colonies that contained "brighter" variants of GFP (about one in every 250 clones) and recombined the advantageous substitutions. They were able to obtain a GFP that had an apparent emission intensity 42-fold higher than that of the wildtype molecule, and at least 3-fold higher than the best variant previously produced by conventional mutagenesis and selection techniques[1].

The mutations were scattered throughout the improved protein, and caused it to fold into a soluble form rather than being sequestered within unproductive inclusion bodies. This highlights one of the key advantages of DNA shuffling: that multiple properties (for instance, codon usage, folding, proteolytic stability, localization) of a protein can be simultaneously optimized for a particular function without regard for mechanism. Indeed, beyond the protein itself, improvements in gene expression (promoter strength) or the vector (copy number, dueling transcription units) will contribute to the desired function.

Future experiments should see the simultaneous optimization of multiple proteins, such as enzymes within metabolic pathways. Genetics is full of examples of simultaneous optimizations. However, if the architectures of proteins and metabolic pathways have "evolved to evolve" by linear recombination (for example, by lodging catalytic residues in distal portions of the primary sequence of a protein), then, by mimicking and accelerating these natural processes, DNA shuffling may in the end be the best method for improving function.

Ichiro Matsumura and Andrew D. Ellington are in the department of chemistry, Indiana University, Bloomington, IN. 47405. (Email: imatsuru@sunflower.bio.indiana.edu)

1. Crameri, A., Whitehorn, E.A., Tate E. and Stemmer, W.P.C. 1996. *Nature Biotechnology* **14:**315-319.
2. Osuna, J., Flores, H. and Sobéron, X. 1994. *Crit. Rev. Microbiol.* **20:**107-116.
3. Clackson, T. and Wells, J.A. 1994. *Trends Biotechnol.* **12:**173-184.
4. Stemmer, W.P.C. 1994. *Proc. Natl Acad. Sci. USA* **91:**10747-10751.
5. Stemmer, W.P.C. 1994. *Nature* **370:**389-391.
6. Palzkill, T. and Botstein, D.J. 1992. *J. Bact.* **170:**5237-5243.