Engineering a revolution

Joseph Affholter and Frances Arnold show how chemists are applying the rules of genetic engineering to breed catalysts for a biologically-based chemical industry.

Specialisation, order and diversity are the hallmarks of the biological world. We observe this in the vast complexity of organisms resident in any segment of the biosphere. We also observe this within a single species (or genus) as that species adapts to different environments. What is true at the macroscopic level is mirrored at the molecular level. We can now identify whole families among today's fantastic array of natural biocatalysts - the enzymes - which appear to share a molecular ancestry, but which have acquired distinctly different properties through the opposing demands of random change and natural selection. Scientists and engineers who want to redesign these same molecules are now beginning to follow Nature's lead. The result is a new catalyst improvement technology - directed molecular evolution (DME) - which may revolutionise the way that we do chemistry.

Nature's algorithm

The biological world uses a robust, versatile design algorithm to respond to new demands. When environmental opportunities or stresses change, rare sub-populations may enjoy a 'fitness' advantage that allows them to propagate faster than their neighbours. Previously minor trait(s) or gene(s) can eventually dominate the population. New traits can appear. This ubiquitous perturbation-response algorithm of 'natural selection' has helped organisms (and their individual protein catalysts) to adapt to an incredible range of environments, from the boiling acid of volcanoes to Antarctic ice fields.

Animal and plant breeders practise a kind of 'directed selection', using criteria other than survival as the measure of fitness. They influence the direction of change by choosing parents with increasingly enhanced traits of interest, for example better speed, strength, agility, docility, or growth under drought conditions. In a few hundred years, breeders have generated thousands of different varieties of dogs, cattle, sheep and crops.

'Natural' and 'directed' selection processes share a diversification strategy - they recruit...
and re-process the genetic diversity within single genomes (the entire complement of genes making up a particular individual), between parents with different genomes, and from random mutations. Together they define a fantastically successful algorithm that takes place at the molecular level, but does not rely on a molecular understanding of the structures and functions of genes, proteins or metabolic networks.

**Breeding catalysts**

Plants and animals are not the only things that might benefit from natural selection. Can we similarly adapt or breed catalysts for a biologically-based chemical industry? The emerging tools of DME promise to let us do just that. In a typical DME experiment (Fig 1) scientists introduce a library of diverse DNA sequences created by mutagenesis and/or recombination into microbial (bacteria or yeast) cells, such that each cell receives a single version of the genetic sequence. Within each colony, each cell transcribes and translates the diversified DNA into a protein biocatalyst. To screen for the activity of interest, we can inoculate the cells into microcultures (typically plates containing 96384 or even 1536 micro-wells). These microcultures are compatible with a growing array of high-throughput analytical devices for rapid assays of biocatalyst activity. We screen the individual colonies for appearance or enhancement of catalytic activities of interest under relevant chemical process conditions. The next step is to isolate genes from cells exhibiting improved enzymes in the desired property and use them to create a new library of DNA sequences for the next generation of prospective biocatalysts. DME uses mutation and selection to identify and accumulate beneficial changes in protein function. Unlike the virtually imperceptible process of natural selection occurring at the whole organism level, however, DME can direct, at higher frequency, sequence alterations into DNA encoding the specific enzyme or metabolic pathway of interest. As with classical breeding, ‘survival of the fittest’ – the direction of selection – is subject to the control of the experimenters. Here, high-throughput screening technologies are used to identify cells carrying improved genes. Repeating this cycle allows the rapid accumulation of beneficial changes. The result is that evolution in the laboratory has a timescale of weeks or months and can progress towards a defined goal.

Most attributes of an enzyme – kinetic constants, stability, selectivity, inhibition profiles and so on – are ‘breedable’ properties. Given the huge database of genes now and soon-to-be available from natural sources, the ability to breed molecules in the laboratory makes vast regions of protein structural space available for generating new catalysts. While sharing conceptual similarities with traditional plant or animal breeding, molecular breeding also differs in significant ways. First, plants and animals breed only by successive recombination of two parent organisms. For breeding molecules, we use one or many parents. Secondly, traditional breeders produce only a few genetically unique offspring on each successive cross. Molecular breeding can generate virtually limitless combinations of offspring – the level of diversity is tailored to the throughput capacity of the screening process. Finally, plant and animal breeding is limited by strong phylogenetic barriers (only very closely related individuals can breed), and cycle times are very slow (usually 1–2 cycles per year). Molecules can be bred as rapidly as we can identify, isolate, recombine and rescreen improved genes. This can be as fast as 1–2 cycles per week for some gene–enzyme families.

**Genetic diversity**

Since directed evolution emerged as an approach to biocatalyst design in the mid-1990s, numerous cases of enzyme and pathway improvements have been reported. Table 1 lists just a few. The earliest attempts to evolve new

---

**Table 1. Examples of biocatalyst properties improved by DME**

<table>
<thead>
<tr>
<th>Altered property</th>
<th>Target enzyme(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased thermostability</td>
<td>subtilisins (proteases) p-nitrobenzyl esterase</td>
</tr>
<tr>
<td>Increased activity in organic solvents</td>
<td>p-nitrobenzyl esterase chloroperoxidase</td>
</tr>
<tr>
<td>Altered substrate specificity</td>
<td>β-galactosidase atrazine hydrolase thymidine kinase alky transferase aspartate aminotransferase dioxygenases</td>
</tr>
<tr>
<td>Increased enantioselectivity</td>
<td>lipase esterase transaminase</td>
</tr>
<tr>
<td>Increased activity</td>
<td>aminoacyl transferase atrazine degradation pathway arsenate resistance pathway p-nitrobenzyl esterase cytochrome P&lt;sub&gt;450&lt;/sub&gt;</td>
</tr>
<tr>
<td>Increased gene expression</td>
<td>green fluorescent protein subtilisin E horseradish peroxidase galactose oxidase</td>
</tr>
</tbody>
</table>

---

**Glossary**

- **Chimeric**: A hybrid gene (or protein) containing sequences from two or more parents.
- **Crossover**: Point at which sequences from different parents are recombined.
- **Mesophilic**: Enzymes that work best at temperatures between 20–45°C.
- **Phylogenetic**: To do with the genetic relationships between organisms.
- **Random point mutation**: Randomly induced mutation of a base in a DNA sequence.
- **Recombination**: Reassortment of genes or gene segments from different parents.
- **Thermophilic**: Enzymes that are adapted to high temperatures.
- **Wild-type enzyme**: Natural enzyme.

**Fig 2. Activity–temperature profiles of esterases evolved during eight sequential generations of primarily random mutagenesis**
enzymes in the laboratory invoked a relatively simple approach using only random point mutations – conversion of any arbitrary gene to another – to generate genetic diversity. These and many subsequent examples reveal that even this simple approach can yield significant improvements in biocatalyst performance, such as stabilising and activating enzymes in unusual environments, for example high temperatures or organic solvents. Figure 2 shows the improvements in esterase activity over sequential generations of random point mutagenesis and screening. The final eighth generation enzyme is stable to much higher temperatures than the naturally occurring wild type enzyme, and its activity is higher at all temperatures. Achieving this level of functional difference in nature usually takes a very long time, during which large numbers of mutations accumulate. As a result, related thermophilic and mesophilic enzymes can differ at hundreds of amino acids. However, the laboratory-evolved thermophilic esterase is a very close relative of its mesophilic starting point, differing at only 13 of its 490 amino acids.

In 1994 a more conservative, but very powerful mutagenic engine referred to as in vitro gene recombination or ‘gene shuffling’ was added to the molecular evolutionist’s repertoire by Pim Stemmer, then at the Affymax Research Institute in California. Recombination – the test-tube answer to sex – can speed up the evolutionary process by allowing beneficial mutations to accumulate and deleterious ones to drop out. Scientists have used combinations of point mutation and recombination widely in the past few years to improve individual enzymes and multi-enzyme pathways.

A particularly interesting recent demonstration of in vitro gene recombination technology is the directed evolution of a family of enzymes from different sources (often different species). In this ‘family shuffling’ approach, genetic sequences are recombined in the test tube to create a molecular library of unique chimeric sequences containing information from any or all of the parents (Fig 3). The process can be repeated for several cycles until the catalyst or pathway exhibits the desired performance.

Scientists at Maxygen, a molecular evolution platform technology company in Santa Clara, California, started with four β-lactamase enzymes derived from four unrelated organisms. After recombining these sequences in vitro and selecting for bacterial growth in the presence of an antibiotic, they isolated a clone whose activity towards that antibiotic was more than 250 times higher than any of its parents. Detailed characterization revealed that this new enzyme contained sequence information from three of the four: starting genes (seven crossovers) and shared no more than 50 per cent sequence identity with any one of its three parents.

Recently, in another example of family shuffling, Maxygen scientists recombined over 20 parental genes encoding an important family of industrial proteases and found enzymes improved in each of five key performance parameters. The ability of family shuffling to access new and very different functional sequences bodes well for the evolution of functions that are highly desirable, but not found (or needed) in nature.

**Fig 3. Creating new genetic diversity by ‘family shuffling’**

**Breeding molecules**

The ability of DME to generate novel biocatalysts and to improve them progressively for industrial applications has long been demonstrated. Groups in several laboratories, including those at the California Institute of Technology (Caltech) and Maxygen, are now focusing on evolving the enzymes and pathways that will be important in chemical production. Clean, highly selective oxidations and reductions, for example, are extremely important, but difficult to perform chemically. The enzymes that catalyse these reactions at room temperature with virtually no waste products are wonderful models for industrial catalysis. But can we go beyond thinking of them merely as models and evolve oxygenases that are active and stable enough to use in manufacturing? The Caltech group is trying to do just that. With support from BP Amoco, among others, researchers are targeting cytochrome P450 mono-oxygenases, a large ‘super-family’ of haem enzymes that catalyse oxygen insertion, hydroxylation and epoxidation (Fig 4). The unique substrate preferences and reaction specificities of each of the hundreds of P450 mono-oxygenase sequences identified to date reflect the flexibility of this catalytic framework. Just as natural evolution has added so many different versions of this enzyme, DME should be able to make them more useful for these difficult transformations by increasing their activities, stabilities and expression levels, and fine-tuning their substrate and reaction specificities.

The Caltech group is de-
Developing powerful new screening technologies to search very large libraries of mono-oxygenases for interesting catalysts. E. coli cells expressing an active P₄₅₀ and a peroxidase enzyme, for example, convert aromatic substrates into fluorescent compounds. This allows us to search tens of thousands of P₄₅₀ mutants using fluorescence digital imaging (Fig 5). The Caltech DME project has already generated P₄₅₀ mono-oxygenases that have increased activity towards a variety of substrates and show altered regiospecificity of hydroxylation.

In addition researchers will continue to apply DME to improve already widely used hydrolases (hydantoinases, esterases, lipases, proteases, and dehalogenases) and many other industrial enzymes. Attractive short-term goals for pharmaceutical and specialty chemical applications include increasing productivities and altering the regio- and stereoselectivities of a number of important biotransformations. Work at Maxygen has shown that the DME process – so effective for tuning single protein biocatalysts – is also highly effective for improving multi-enzyme–protein pathways. The Maxygen group, for example, has developed a dual-enzyme cellular pathway capable of converting atrazine and other triazines to a variety of commercially interesting intermediates. This opens exciting possibilities for developing the kinds of industrial single-pot multi-step chemical transformations that will tip the whole capital investment scenario of the chemical industry in favour of biotechnology.

Joseph Affholter is vice-president of biocatalysts and chemicals at Maxygen, Santa Clara, CA 95051, US (website: http://www.maxygen.com or e-mail: Joe_Affholter@maxygen.com) and Frances H Arnold is professor of chemical engineering at the California Institute of Technology, Pasadena, CA 91125, US (website: http://www.che.caltech.edu/groupfha or e-mail: frances@chem.caltech.edu).

Further reading