DIRECTED EVOLUTION OF STEREOSELECTIVE BIOCATALYSTS

Reported by David M. Knapp                 March 13, 2008

INTRODUCTION

Stereoselective chemical reactions have become extremely important in both academic and industrial settings for the synthesis of useful molecules. Catalytic reactions are of particular importance due to their capacity to effect otherwise impossible transformations using substoichiometric quantities of the catalyst. Traditionally, transition metal catalysts have played a central role in enabling these transformations, and the last century has seen a tremendous amount of work in the research and development of organometallic catalysts based upon a wide variety of metal centers.

Another approach is to use nature’s own catalysts, enzymes, to catalyze chemical reactions. This approach has seen less implementation for a number of reasons. The number of enzymes available for use in reactions has, until recently, remained small. Also, the substrate specificity for many enzymes can be high, limiting the generality of such catalysts for use in a laboratory setting, and the enantioselectivity of available enzymes may be insufficient for practical use in a particular system. However, the substrate promiscuity of enzymes may have been underappreciated, and continuing advances in high-throughput combinatorics and molecular biology, combined with the technique of directed evolution, may soon make enzymatic biocatalysis a practical and economical solution to the problem of asymmetric synthesis.

Assuming that the issues of low stereoselectivity and substrate scope can be solved, there are a number of advantages associated with biocatalysts over traditional transition-metal-mediated catalysis. First, the size and complexity of proteins allows for extraordinarily high diversity in both structure and function. Ironically, this tremendous complexity largely arises from a relatively simple, modular construction of the primary peptide sequence. Harnessing the cellular machinery in bacteria via microbial engineering now allows the generation of large libraries of potential biocatalysts, something that is not true of small-molecule metal catalysts. Furthermore, reactions can be carried out in aqueous media at mild temperatures and pressures, which may facilitate the development of more environmentally friendly “green” chemistry.
Finally, protein-based catalysts may ultimately be more economical than those developed around rare and expensive metals.

Despite the benefits of biocatalysis, one important limitation that has prevented its more widespread use is that there is a relatively limited number of enzymes available that catalyze synthetically useful chemical reactions. Furthermore, the enzymes that are available may be inappropriate with respect to their activity or stereoselectivity in the reaction of interest. The technique of directed evolution (DE) currently stands as one of the most powerful means to expand the number of existing biocatalysts, as well as to tailor existing catalysts for a particular reaction.

**BASIC PRINCIPLES OF DE**

DE involves the application of selective Darwinian pressure to generate mutant enzymes of increasing fitness with respect to some desirable characteristic, such as activity or stereoselectivity. To begin, an enzyme is selected which displays unsatisfactory reactivity or selectivity. The gene for this enzyme is subjected to *in vitro* random mutagenesis, using one of a variety of available methods, to generate a library of mutants. This DNA library is then transformed into and expressed in a bacterial host, generating a library of enzymes that can then

![Directed Evolution Cycle](image)

*Figure 1. Directed Evolution Cycle*
be screened for the desired characteristic(s). The best mutants are subjected to a further round of mutagenesis. In this fashion, enzymes of increasing quality are generated. The process can be further iterated until a catalyst of satisfactory stereoselectivity or activity is obtained (Figure 1).

This sequence of events has two main challenges. The first problem is to select an appropriate mutagenesis protocol that will allow access to a sufficiently diverse library. The second is to design an effective and suitably high-throughput screening method. Some notable solutions to both of these problems will be discussed in the following sections.

**GENERAL STRATEGIES FOR DE**

**Mutagenesis Strategies**

In principle, standard polymerase chain reaction site-directed mutagenesis could be used to generate a library of mutant sequences. The problem with this approach is that it may not be clear which amino acids in a protein are important for activity. The location of the active site may not be known, and mutations distant from the active site may also prove effective at modulating enzyme activity. In contrast, non-selective mutagenesis has the benefit of requiring no knowledge whatsoever of the enzyme’s structure or catalytic mechanism. Several methods have been developed to perform both site-directed and non-selective random mutagenesis.

One of the first techniques to be developed was error-prone polymerase chain reaction (epPCR). PCR is a standard technique in molecular biology which involves the use of a polymerase to perform a replicative enhancement of a DNA sample. Generally in PCR, high replicative fidelity is desirable to maintain sequence integrity in the enhanced sample. In epPCR, a relatively low fidelity polymerase is chosen. Taq polymerase, from *Thermus aquaticus*, is commonly used, as it lacks the 3’→5’ exonuclease proofreading capability common to higher-fidelity polymerases. Additionally, adding Mn²⁺ to the reaction buffer further increases the polymerase error rate. Adjustment of the relative free nucleotide (dNTP) concentrations is also necessary to minimize Taq’s bias toward A→G and T→C mutations. By carefully selecting the protocol conditions, libraries of mutant sequences can be generated with an average number of one, two or more errors per sequence.

Among other mutagenic methods is site-saturation mutagenesis. In this strategy, residues of particular importance are chosen, and PCR is performed on the gene using a library of primers containing randomized nucleotides across the codons for the amino acids of interest. Assuming
enough clones are generated (usually 300-400 for one site), the resulting library will contain mutants with all 20 amino acids represented at the chosen site(s).\(^4\) This is an effective way of exhaustively probing the residue possibilities at crucial positions, but it requires prior knowledge of which positions are important for selectivity, as parallel site-saturation mutagenesis along the entire protein sequence is statistically infeasible. Nevertheless, this technique has been used as the primary means of mutagenesis to evolve the specificity of a β-glucuronidase.\(^5\) A third powerful method is DNA shuffling, which involves the recombinative assembly of mutants from multiple related proteins. By mixing the DNA encoding all of the proteins and subjecting the mixture to a restriction digest, a large pool of fragments is obtained. Subjecting this mixture to PCR-like conditions allows the fragments to prime each other, regenerating the full length sequences. However, homology between the separate original sequences allows cross-priming, ultimately resulting in a library of shuffled sequences.\(^6\) Each of these methods has been successfully utilized to direct the evolution of proteins. Some of the most dramatically successful examples involve the use of some combination of all three methods.

**Sorting Strategies**

The most effective mutagenesis strategies can generate libraries of thousands of mutants. With this comes the problem of efficiently identifying improved variants. Two basic strategies exist to tackle this problem, screening and selection. Screening involves individual examination of each library member in turn. Physically separating library members is not problematic, after transformation of the library DNA into bacteria, each resulting colony is derived from a single transformed cell, and therefore expresses only a single variant. Analyzing a variant is generally as simple as picking a colony, lysing the cells, and adding the substrate for the screening reaction to the cleared lysate. However, the need to examine thousands of variants makes this strategy extremely time and resource intensive. This strategy does benefit from simplicity of design. Reetz and coworkers have demonstrated the viability of screening among a library of mutant bacterial lipases from *Pseudomonas aeruginosa* for members that catalyze the hydrolytic kinetic resolution of ester 1 (Scheme 1). Cleavage of the substrate generates a p-nitrophenolate that absorbs strongly at 410 nm, allowing enzyme activity to be directly monitored via UV-Vis spectroscopy. Parallel screening of mutants with separately purified \(R\) and \(S\) substrates and calculation of the relative rates of the reactions allowed the determination of each variant’s
enantioselectivity. Performing the screen reactions in 96 well plates and using a UV-Vis plate-reader, Reetz and coworkers were able to process 500 mutants per day and enhanced the selectivity factor ($k_{rel}$) from 1.1 to 51, in favor of the formation of $S$-$\text{-2}$ (Scheme 1), corresponding to an increase in product enantiomeric ratio (e.r.) of 52:48 to 98:2.\(^7\) The use of robotic pipetting and colony picking was later utilized to increase the throughput of this system to 1500 variants per day.

\[
\begin{array}{c}
\text{rac-1 (R = } n\text{-C}_6\text{H}_{17}) \\
\text{Lipase Variants} \\
\text{H}_2\text{O} \\
\text{(S)-2} \\
\text{(R)-2} \\
3
\end{array}
\]

**Scheme 1. Hydrolytic cleavage of $p$-nitrophenylester**

In contrast, selection involves the development of a system in which an active mutant is essential for the survival of the expressing organism, or results in a phenotype that is distinguishable from non-active mutants. Thus, one does not need to examine each candidate separately and only a small fraction of the library needs to be examined more closely. However, the development of systems in which the activity of a catalyst is directly linked to bacterial survival, or a distinguishable phenotype, is not trivial. If the enzyme is part of an essential metabolic pathway within the bacterial host, the library may be expressed in a strain lacking the endogenous gene and plated onto media lacking the natural end product. Using this approach Taylor et al. selected for active variants of a chorismate mutase, which is involved in the *de novo* synthesis of tyrosine in bacteria.\(^8\) In an alternate approach, Reetz and coworkers have developed a selection system for the directed evolution of lipases, in which mutant-expressing colonies are grown on media containing an insoluble enzymatic substrate, glyceryl tributyrate.\(^9\) Colonies expressing active enzyme break down the substrate in water soluble products, creating clear spots in the cloudy media, and allowing active variants to be identified.

Despite the attractive benefits of selection strategies, the difficulties in developing them have resulted in a general reliance on screening in applied examples of DE. When the product is UV-Vis active, screening by absorbance can be highly efficient. Similarly, fluorophoric systems have been utilized,\(^10\) but these strategies are relevant only to specialized reaction systems in
which the product of the catalyzed reaction is spectroscopically distinguishable from the substrate. Chiral chromatography is in principle a much more general strategy. High pressure liquid chromatography (HPLC) is perhaps the most general, but it is at best low throughput, being limited to several dozen analyses per day. Gas chromatography (GC), when fully automated, has been utilized to process up to 700 samples per day, but this method is restricted to products of low molecular weight. A variety of other methodologies for screening enantioselectivity have been developed which utilize capillary electrophoresis, infrared thermogenic imaging, circular dichroism and mass spectrometry.

SUCCESSFUL APPLICATIONS OF DE TO STEREOSELECTIVE CATALYSIS

Despite the aforementioned complexities, a number of successfully executed examples of DE strongly suggest its potential to enable biocatalysis. The example of Berry and coworkers highlights the potential usefulness of the DE methodology. Employing a combination of epPCR and site-saturation mutagenesis, the group evolved a pair of complementary stereoselective aldolases from N-acetyleneuraminic acid lyase (NAL). Wild-type NAL reversibly catalyses the aldol condensation of N-acetylmannosamine (ManNAc) and pyruvate to form the corresponding sialic acid (Scheme 2), one of a class of compounds that have found application in a number of commercially available anti-influenza drugs. Wild-type NAL is only moderately selective for the 4S-product, producing an e.r. of 74:26. A single round of epPCR generated a library of 2500 mutants which was screened by coupling the retro-aldol reaction of desired products to the reduction of NADH and pyruvate by lactate dehydrogenase. This led to the discovery of several

Scheme 2. Aldol condensation catalyzed by NAL
residues that were crucial for stereoselectivity, and subsequent site-saturation mutagenesis led to the discovery of 4S- and 4R-selective enzymes with selectivity factors of 50 and 48, respectively. The success in accessing both strongly S- and R-selective enzymes by this method is worth note, as achieving opposite stereoselectivity in biocatalysis is not as simple as it often is in transition metal based catalysis, i.e. simply using the opposite catalyst configuration.

The work of Reetz and coworkers on the biocatalysis of the Baeyer-Villiger oxidation is an excellent example of how DE can not only enable biocatalysis but fulfill an unmet need in organic chemistry. The Baeyer-Villiger oxidation involves the oxidative cleavage of a carbon-carbon bond adjacent to a carbonyl, and its asymmetric catalysis by traditional synthetic means remains an unsolved problem. Beginning with a cyclohexanone monooxygenase (CHMO), Reetz and coworkers were able to direct the evolution of a mutant capable of desymmetrizing rac-methoxycyclohexanone into lactone (R)-3, with an e.r. of 95:5 (Scheme 3); the wild type is only capable of achieving an e.r. of 55:45. Additionally, an S-selective variant was obtained that was even more selective (e.r. > 99:1) and that maintained enantioselectivity on a range of 4-substituted cyclohexanones. To date, no known transition metal catalyst has been demonstrated to match the performance of this enzyme.

![Scheme 3. Baeyer Villiger Oxidation catalyzed by CHMO](image)

**CONCLUSIONS**

Biocatalysis represents an emerging field which has the potential to provide a useful complement to traditional small molecule catalysis, and allow highly stereoselective transformations to be performed under greener reaction conditions. In this context, DE stands as an enabling strategy that could, in principle, allow chemists access to catalysts tailored for their particular synthetic needs. Currently, the biggest limitation preventing this is the challenge in
screening for e.r. in a both general and high-throughput fashion, and in developing widely applicable selection systems. The most general screening systems are relatively low-throughput, and the most high-throughput systems require either special circumstances with respect to the structure of the product (e.g. volatility or the presence of a chromophore/fluorophore). Additionally, the highest levels of throughput require significant robotic automation and substantial instrumental overhead, which may either preclude this methodology from being practical or prevent it from being maximally effective in small research laboratories. Despite these limitations, the high stereoselectivities that can be obtained without any need for “rational” catalyst design, especially for reactions which lack alternative efficient catalytic options, make this methodology quite promising. With further development of screening and selection methodology, DE of biocatalysts could become increasingly common in the arena of target-oriented organic synthesis.

REFERENCES