PROTEIN BIOCHEMISTRY

Efficient evolution

A new approach to directed evolution provides a way of searching through numerous variants to identify the most desirable mutations.

Nature is very good at optimizing protein function, but sometimes it just is not good enough. For such cases researchers figured out that artificial mutagenesis and selection could be used to direct the evolution of a protein toward a desired end point.

Scientists have developed many such directed-evolution approaches, but sometimes even these are not adequate. This turned out to be the case for Gjalt Huisman, Richard Fox and colleagues at Codexis, Inc. before they used a new approach for analyzing these variants (Fox *et al.*, 2007).

Many of the Codexis employees once worked with Willem Stemmer, one of the pioneers of directed evolution, so they consider themselves at the forefront of this technology. But they found themselves unable to make the progress they needed in optimizing an enzyme called halohydrin dehalogenase (**Fig. 1**). Huisman says, "We had a number of options. Either stop the program, brute force it and see whether just by dedicating more resources we could force the enzyme to get better, or we could apply new technology." They decided on the last option.

Richard Fox had been working on the technology from a different angle than other people. "The question was whether there were more efficient ways of searching through a given set of mutations," Fox remarks. He had written some theory papers on this subject but had never tried it on a real evolution project—this was a perfect opportunity.

In traditional directed evolution approaches, mutants are generated and screened for activity without sequencing until variants with desirable activities are found. In contrast, Fox's method uses protein sequence activity relationships (ProSAR). Briefly, all the protein variants are assayed, as is typically done, but the genes of a small, diverse subset are sequenced. Activity and sequence information are combined in a statistical analysis to determine the impact of each mutation on protein activity. The power of the approach rests in its ability to identify beneficial and detrimental mutations in a background of other mutations that conceal their effects. This requires about three sequence-activ-
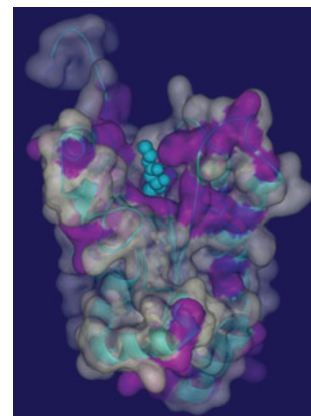

Figure 1 | Structural model of the optimized halohydrin dehalogenase enzyme. Residues that were selectively changed as a result of the ProSAR-based directed evolution are highlighted in magenta. Image courtesy of G. Huisman.

ity measurements for every mutation in a suitably diverse library.

The method used to introduce mutations is unimportant, provided that it yields a suitably rich source of activity variation. "Our diversity comes from any source we can grab it—either random mutagenesis, rational or semi-rational hypotheses, scanning, anything that can go into the bucket is fully in play, but where we are taking it to the next level is trying to rapidly sift through that diversity once you have it in hand," says Fox.

This approach is most valuable for proteins that require several assays to fully characterize the activity or when more than one property is being selected for. "We need the enzyme to be highly active, stable in the process, and tolerant of high concentrations of organics and product. We are interested in whether this enzyme is highly active when the reaction has already reached 90% conversion at 1 M substrate concentration," says Huisman.

Is this method suitable for all directed evolution applications? "If you have something like phage display and a very rapid way of generating diversity, [ProSAR] may not be necessary. But it's certainly applicable to any directed evolution project where you have a certain number of mutations and you want to search through them as efficiently as possible," says Fox.

Daniel Evanko

RESEARCH PAPERS

Fox, R.J. *et al.* Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.* **25**, 338–344 (2007).