CHEMICAL
& Engineering News

Latest News

April 17, 2006
Volume 84, Number 16
p. 7

Computational Biology

Artificial P450 Enzymes Created

Computational method leads to library of enzymes that fold and function

Celia Arnaud

A family of nearly 3,000 artificial cytochrome P450 enzymes has been created by a California Institute of Technology team's efforts to recombine sections of three natural cytochrome P450s, a large family of oxidative enzymes that are widespread in nature (*PLoS Biol.*, published online April 11, dx.doi.org/10.1371/journal.pbio.0040112). In humans, P450s play a crucial role in the metabolism of drugs and other toxins.

Courtesy of Frances Arnold

Schemers Arnold **(left) and Otey let**

computation guide them to a new family of cytochrome P450s.

"I'm hoping that this new family will contain cytochrome P450s that people would want to use, for example, to make the human metabolites of drugs or to synthesize complex, biologically active compounds," says Frances H. Arnold, a chemical engineering professor at Caltech. "It's my dream to make a whole library of cytochrome P450s that could hydroxylate anything."

Arnold, graduate student Christopher R. Otey, and coworkers used a computational method called SCHEMA to guide the creation of the new protein sequences and increase the likelihood that they will be useful.

P450 enzymes are such a diverse family of enzymes that the usual method of random DNA shuffling would generate a set of new sequences in which most of the members would not fold, Arnold says. "You'd be screening garbage."

To generate P450 enzymes that catalyze new reactions, Arnold's team wanted the new proteins to be 70 to 100 amino acids different from the starting proteins, yet still fold properly. They made new proteins by recombining chunks of the P450 enzymes nature has provided, Arnold says. "It's a dual optimization problem. We recombine them to preserve as many structural interactions as possible, while at the same time making lots of mutations."

That's where SCHEMA comes in. SCHEMA's job is to improve the likelihood that a given sequence will fold by considering the structures of the parent proteins. The crystal structures of the paren proteins are encoded mathematically to make counting broken interactions between side chains a simple calculation. "SCHEMA penalizes you for every broken contact. It says, 'Thou shalt make a library such that most of the members have few broken interactions,' " Arnold says. "At the end, you get a design that penalizes you the least."

The Caltech team chopped each of the three original enzymes into eight pieces and recombined them, yielding 6,561 (3^8) possible sequences. Of those, nearly half fold into properly functioning cytochrome P450s that can catalyze a reaction.

But the nonfolding proteins serve a useful purpose, too. The team used a mathematical technique known as logistic regression analysis, which relies on having sequences that don't fold and function in addition to sequences that do, to glean information about why particular sequences fold. "If you're trying to understand what it is about a sequence of amino acids that makes it into a functional protein, it's nice to have ones that weren't successful with which to compare," Arnold says.

David R. Nelson, who maintains a database of cytochrome P450s at the University of Tennessee Health Science Center, Memphis, says the method provides a tool for looking at functional proteins that did not arise via evolution. "Evolution will not let you see the nonfunctional or poorly functional versions of a protein that can be very useful for understanding how a protein works," he says. "The paper showcases a method of structure-function analysis that should be applicable to many other proteins."
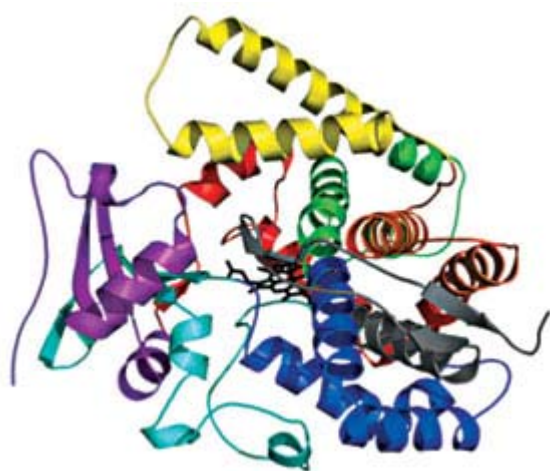
PLOS BIOLOGY © 2006

Protein Pieces
Recombining the pieces of
three cytochrome P450s
divided into eight sections
(indicated by the different

colors) results in a family
of nearly 3,000 new
enzymes that fold and
function properly.

Chemical & Engineering News
ISSN 0009-2347