Chemistry & Biology, Vol. 11, 287–293, March, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2004.03.005

Previews

Optimized Chimeragenesis: Creating Diverse P450 Functions

A drawback to generating chimeric proteins by chimeragenesis, especially when the "parent" proteins share low sequence identity, is that unfolded proteins frequently result. In this issue of *Chemistry & Biology*, Arnold and coworkers report their use of the SCHEMA algorithm to effectively predict ideal hybrids of cytochromes P450 [10].

Protein evolution in vitro uses similar principles to those that nature has evolved over millions of years to create molecules with improved properties and capacity to adapt to the environment. Thus, techniques such as random mutagenesis and various recombination strategies have been successfully applied to obtain proteins and enzymes with improved stability, increased specific activities, or altered reactivity [1–5]. Recombination between most homologous proteins is, however, limited by the method employed. Both in vivo and in vitro recombination methods are dependent on correctly annealing DNA; therefore, successful recombination requires high homology between the proteins to be shuffled (normally 70% and higher), and crossovers are biased toward regions of high sequence identity [6–8]. A different problem is faced once the chimeric DNA constructs are made. It has been observed that very high amounts of unfolded and, thus, inactive proteins are generated in this way, rendering the screening needed to sort through these often dysfunctional mutants for those with improved properties a rather ineffective and tedious procedure. An innovative step toward overcoming the formation of unfolded proteins and subsequently improving the effectiveness of screening recombinant libraries was recently made by Frances Arnold and coworkers [8, 9]. This group devised an algorithm called SCHEMA, which calculates the potential for a given recombination to cause structural disruption based on three-dimensional data of the proteins to be recombined. The percentage of functional proteins in the library decreased as the calculated degree of disruption increased, illustrating that SCHEMA could predict crossover points for chimeragenesis that conserve structure and increase the number of folded and functional proteins.

In a paper published in this issue, the Arnold group [10] describes the application of this approach to use two parent bacterial enzymes, CYP102A1 and CYP102A2, to create hybrid cytochromes P450 with a wide range of thermostabilities, increased activity, and altered activity profiles formerly not present in either one of the parents.

Chimeragenesis has previously been successfully performed using P450s of the same subfamily (for review, see [11]), which show a high sequence homology (defined as greater than 55% homologous amino acids), with the goal of producing enzymes with altered substrate selectivity and activity and to identify individual amino acid residues involved in these functions. New activities have been observed in some of these chimeric mutants [12]. Chimeragenesis has also been applied to P450s that share only low sequence homology (around 20%) to solubilize membrane-bound P450s. Shimoji et al. [13] constructed a soluble chimera from ~50% bacterial CYP101 and ~50% mammalian membrane-bound CYP2C9, which was functional and catalyzed 4-chlorotoluene oxidation. A significant advance was achieved by Sieber et al. [14], who were able to create single-crossover hybrid proteins of the membrane-bound rat CYP1A2 and the heme domain of the bacterial CYP102. Approximately 2,000 variants out of 250,000 clones were preselected from the soluble, folded fraction of the library by chloramphenicol resistance due to CAT activity, which is only seen when CAT is expressed in a correct frame. After screening, two significantly more soluble chimeras were found that exhibit CYP1A2-type activity/chemistry. Another example of the creation and characterization of a highly chimeric library is a study by Abecassis et al. [15], who produced a library of CYP1A1 and CYP1A2 using a combination of PCR-based and in vivo recombination in yeast. However, in all cases where P450s with low homology were used, a considerable amount of the constructed chimeras as well as large parts of the library produced unstable mutant proteins.

The new direction taken by Arnold and colleagues reported in this issue [10] overcomes this shortcoming by applying the powerful structure-based algorithm SCHEMA to identify fragments of proteins that can be recombined to minimize disruptive interactions that would prevent protein folding. They used SCHEMA to design chimeras of CYP102A1 and CYP102A2, homologous proteins that share 63% amino acid sequence identity. Fourteen of the seventeen hybrid proteins constructed were folded correctly, as determined by carbon monoxide reduced difference spectroscopy. From this pool, 13 produced a band at 450 nm, indicating folded protein with heme correctly bound, whereas only one construct was inactive, as indicated by a band formation at 420 nm. This is an impressive fraction of structurally folded proteins. Interestingly, half of the folded chimeras had altered catalytic activities, while three mutants acquired affinity for a new substrate not hydroxylated by either parent. Thus, the method proposed by the authors is likely to be highly effective for developing libraries of mosaic P450s with changed activities and selectivities. However, an open question remains regarding the observation reported in this paper that protein stability does not correlate with predicted degree of structural disruption. It might be possible that this is due to the relatively low number of chimeric proteins investigated in this study, as the authors suggested, or to the effect of chimeragenesis on correct heme binding due to additional stabilizing effects by the prosthetic group (e.g., apo-P450 is, in general, less stable than the holo-protein [16]).

Since P450s generally share very low sequence identity, often <20% at the amino acid level, and thus established recombination methods are not applicable, the method described in Otey et al. [10] could be a powerful

tool for engineering improved enzyme activity as well as changing the substrate specificity or regioselectivity of this interesting group of enzymes. To this end, it is reassuring to consider that, as far as we know today, the members of the cytochrome P450 family share high structural similarity despite their low sequence identity, which could make the construction of appropriate fragments for recombination less complicated. Moreover, the method will not only be of extraordinary value for P450 engineering for biotechnological applications, but will also enable deeper insight into the structural prerequisites for P450 activity, substrate specificity and selectivity. Previously, researchers showed that replacing only one or very few amino acids close to the active site was sufficient to lead to new selectivities in substrate conversion [17-19]. In contrast, Otey et al. [10] demonstrate that residues located far from the active site exert long-range effects on enzyme activity and substrate selectivity. Thus, combining site-directed mutagenesis and random chimeragenesis should allow one to gain a deeper insight into the interplay of near- and long-range effects of mutation.

Future work in this area must demonstrate that the SCHEMA algorithm can be used to generate a library of mosaic P450 mutants containing a high percentage of folded proteins from P450 parents that have low sequence homology. The authors of this recent study have predicted, *in silico,* that such a library resulting from hybrids created from CYP102A1 and CYP102A2 (which share 63% sequence identity) would give 75% folded proteins, whereas a library obtained by random recombination may contain as little as 9%, and on average 42% of folded proteins. If this holds true when tested experimentally, it would establish the SCHEMA algorithm as general tool for predicting functional and diverse libraries of chimeric P450s and would open new and exciting opportunities for engineering P450s.

Moreover, SCHEMA may be generally applicable for predicting the outcomes of recombining other proteins with low homology, creating a methodology for recombination of very distantly related proteins. One prerequisite for this strategy to be successful, however, might be the availability of related crossover sites in the parent proteins that do not interfere with the overall protein structure in order to give chimeras of the correct length (without insertions or deletions).

Rita Bernhardt
Department of Biochemistry
P.O. Box 15 11 50
Saarland University
D-66041 Saarbrücken
Germany

Selected Reading

1. Crameri, A., Dawes, G., Rodriguez, E., Silver, S., and Stemmer, W. (1997). Nat. Biotechnol. *15*, 436–438.
2. Li, Q.S., Schwaneberg, U., Fischer, P., and Schmid, R.D. (2000). Chemistry *6*, 1531–1536.
3. Glieder, A., Farinas, E.T., and Arnold, F.H. (2002). Nat. Biotechnol. *20*, 1135–1139.
4. Lingen, B., Grotzinger, J., Kolter, D., Kula, M.R., and Pohl, M. (2002). Protein Eng. *15*, 585–593.
5. Baik, S.H., Ide, T., Yoshida, H., Kagami, O., and Harayama, S. (2003). Appl. Microbiol. Biotechnol. *61*, 329–335.
6. Ness, J.E., Del Cardayre, S.B., Minshull, J., and Stemmer, W.P. (2000). Adv. Protein Chem. *55*, 261–292.
7. Lutz, S., Ostermeier, M., Moore, G.L., Maranas, C.D., and Benkovic, S.J. (2001). Proc. Natl. Acad. Sci. USA *98*, 11248–11253.
8. Voigt, C.A., Martinez, C., Wang, Z.G., Mayo, S.L., and Arnold, F.H. (2002). Nat. Struct. Biol. *9*, 553–558.
9. Meyer, M.M., Silberg, J.J., Voigt, C.A., Endelman, J.B., Mayo, S.L., Wang, Z.G., and Arnold, F.H. (2003). Protein Sci. *12*, 1686–1693.
10. Otey, C.R., Silberg, J.J., Voigt, C.A., Endelman, J.B., Bandara, G., and Arnold, F.H. (2004). Chem. Biol. *11*, this issue, 309–318.
11. Domanski, T.L., and Halpert, J.R. (2001). Curr. Drug Metab. *2*, 117–137.
12. Straub, P., Lloyd, M., Johnson, E.F., and Kemper, B. (1994). Biochemistry *33*, 8029–8034.
13. Shimoji, M., Yin, H., Higgins, L., and Jones, J.P. (1998). Biochemistry *37*, 8848–8852.
14. Sieber, V., Martinez, C.A., and Arnold, F.H. (2001). Nat. Biotechnol. *19*, 456–460.
15. Abecassis, V., Pompon, D., and Truan, G. (2000). Nucleic Acids Res. *28*, E88.
16. Pfeil, W., Nölting, B.O., and Jung, C. (1993). Biochemistry *32*, 8856–8862.
17. Lindberg, R.L., and Negishi, M. (1989). Nature *339*, 632–634.
18. Böttner, B., Schrauber, H., and Bernhardt, R. (1996). J. Biol. Chem. *271*, 8028–8033.
19. Böttner, B., Denner, K., and Bernhardt, R. (1998). Eur. J. Biochem. *252*, 458–466.

Chemistry & Biology, Vol. 11, March, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2004.03.008

Quantifying Intermediates in Template-Directed Natural Product Biosynthesis

High-performance mass spectrometry is providing new experimental windows into the enzymology of natural product biosynthesis. The first quantitative assessments of covalently attached biosynthetic intermediates promise to shine new light on template-directed biosynthesis.

Nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) are multimodular megasynthase enzymes that catalyze the template-directed biosynthesis of natural products. The mechanisms of natural product biosynthesis are of great interest from both a basic science and a technological perspective, as elucidation