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, 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

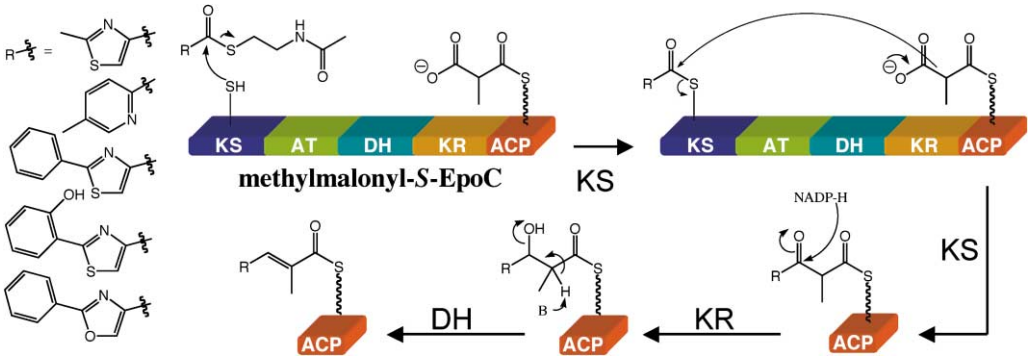

Figure 1. Proposed Mechanism for Monomer Extension by EpoC Module Showing Precursor Analogs Loaded onto EpoC

of the mechanisms of natural product synthesis may allow the rational design of new variants through genetic manipulation of the biosynthetic gene clusters [1, 2]. These enzymes assemble the core frameworks of these natural products from monomers that are incorporated into the growing chemical structures. Each module in the enzyme typically incorporates a single monomer and often contains domains to tailor the chemical functionality of the covalent intermediates. Most modules contain a single carrier protein domain that holds all covalently bound intermediates as thioesters to a phosphopantetheine cofactor. The carrier protein is initially loaded with a monomer, incorporates it into the growing natural product, and shuttles this covalently attached intermediate to any tailoring enzymes in the module before passing its cargo onto a subsequent module.

The mechanisms of many NRPS and PKS enzymes can be inferred from analysis of the structure of the natural product and sequence comparisons with related enzymes. However, methods for the direct interrogation of the status of individual modules within these multimodular enzymes are few. Much knowledge has been gained through the stoichiometric incorporation of radiolabeled precursors, cleavage of the intermediates, and identification through chromatographic comparison with authentic compounds or by mass spectrometry. A drawback to this approach is that some intermediates are unstable to the cleavage conditions, potentially biasing the results. Additionally, "resting" holoenzymes with unused phosphopantetheines are not quantified in such studies.

In this issue, the Kelleher and Walsh groups report the most comprehensive analysis to date of covalently bound megasynthase intermediates [3]. These groups have developed a semiquantitative method to measure and identify intermediates in epothilone biosynthesis using Fourier-transform mass spectrometry (FTMS). They interrogate the enzymology of EpoC, the second extender module in epothilone biosynthesis that activates a methylmalonate monomer for condensation, with a methylthiazolyl thioester produced by the starter and first extension modules of epothilone synthase (Figure 1). Their approach utilized a combination of experiments with reconstituted EpoA-ACP, EpoB, and EpoC modules [4] or analysis of EpoC directly loaded with acyl-SNAC substrate mimics in the presence or absence of the essential tailoring factor NADPH. In these assays, they used FTMS to resolve and quantify covalent intermediates on the carrier protein of EpoC after partial enzymatic digestion of the module. Interestingly, some of the unnatural analogs that were poor substrates for the synthase appeared to prevent the loading of methylmalonate, supporting the notion of communication between ketosynthase and acyltransferase domains within a module to prime loading of methylmalonate only when a suitable substrate is available. Additionally, they found that if the initial product of the condensation cannot be further processed, as is the case when NADPH is omitted, propionyl thioesters accumulate on the carrier protein domain, an unproductive intermediate formed from the net decarboxylation of methylmalonate. This intermediate can block the entire megasynthase until it is removed, allowing the assembly line to continue [5]. The authors propose that this may actually be the result of the enzyme catalyzing a retro-Claisen condensation, leaving propionate attached to the carrier protein.

The larger implications of this new technique for quantifying the biosynthetic intermediates in natural product biosynthesis lie in the possibilities for significantly deepening our understanding of the detailed enzymology underlying thiotemplate biosynthesis. Further, this approach could be used to identify intermediates on multimodular synthases where we do not know the structures of the biosynthetic intermediates, such as in the formation of pyrimidine rings in natural products like Bleomycin [6] or in the formation of the of the enediyne antibiotics [7]. As efforts toward re-engineering template-directed megasynthases are further developed, this experimental approach will help identify the bottlenecks in efficient enzymatic production of nonnatural analogs of these natural products by revealing the flux of intermediates during assembly of these compounds. The results reported in this issue of *Chemistry & Biology* highlight some advantages of high-performance mass spectrometry to probe the experimentally difficult enzymology of these megasynthases and will surely stimulate this field of research.

Peter J. Belshaw
Departments of Chemistry and Biochemistry
University of Wisconsin, Madison
1101 University Avenue
Madison, Wisconsin 53706

Selected Reading

1. Cane, D.E., Walsh, C.T., and Khosla, C. (1998). Science *282*, 63–68.
2. Wilkinson, B., Kendrew, S.G., Sheridan, R.M., and Leadlay, P.F. (2003). Exp. Opin. Ther. Pathol. *13*, 1579–1606.
3. Hicks, L.M., O'Connor, S.E., Mazur, M.T., Walsh, C.T., and Kelleher, N.L. (2004). Chem. Biol. *11*, this issue, 327–335.
4. O'Connor, S.E., Chen, H.W., and Walsh, C.T. (2002). Biochemistry *41*, 5685–5694.
5. Heathcole, M.L., Staunton, J., and Leadlay, P.F. (2001). Chem. Biol. *8*, 207–220.
6. Du, L.C., Sanchez, C., Chen, M., Edwards, D.J., and Shen, B. (2000). Chem. Biol. *7*, 623–642.
7. Liu, W., Ahlert, J., Gao, Q.J., Wendt-Pienkowski, E., Shen, B., and Thorson, J.S. (2003). Proc. Natl. Acad. Sci. USA *100*, 11959–11963.

Chemistry & Biology, Vol. 11, March, 2004, DOI 10.1016/j.chembiol.2004.03.007

Profiling Natural Product Biosynthesis

Natural products are a rich source of therapeutics; however, artificially reengineering the biosynthetic pathways that generate these compounds could potentially generate "designer" drugs. Last month in *Chemistry & Biology*, Burkart and coworkers reported their technique to track and better understand the components of these pathways [1].

Many pharmacologically active natural products are of polyketide or polypeptide origin and are synthesized via secondary metabolism processes [2]. Although the structures of polyketides are myriad and diverse, natural products from this class are biosynthesized by a general mechanism involving the construction of poly-β-keto repeating chains in an "assembly line" fashion from the condensation of carboxylic acid (C_2) precursors by polyketide synthetases (PKS) [3–7]. Similarly, nonribosomally encoded peptide-derived natural products are assembled from amino acid precursors by the action of nonribosomal peptide synthetases (NRPS) [8–10].

Both PKSs and NRPSs are modular in construction, with separate enzymatic subdomains responsible for precursor recognition, activation, condensation, and postassembly modification. In preparation for polyketide synthesis and concomitant chain elongation, a serine residue of a carrier protein (CP) domain is phosphopantetheinylated by a 4′-phosphopantetheinyl transferase (PPTase) [11]. The terminal thiol of the phosphopantetheinyl arm is then acylated by an adjacent acyltransferase (AT) domain that utilizes acyl-Coenzyme A (acyl-CoA) as a substrate. The ketosynthase (KS) domain catalyzes the subsequent addition of the β-keto acid on a downstream CP domain to the monomer unit of an upstream CP domain. In any given module the resultant ketone may be functionalized by any combination of ketoreductase, dehydratase, and enoylreductase domains to yield the nascent linear product. Subsequent tailoring enzyme activities such as cyclization, epimerization, methylation, glycosylation, or hydroxylation are often employed during conversion of a linear precursor to the mature natural product.

In a similar fashion, NRPS enzymes consist of basic modules containing adenylation (A) domains, carrier protein (CP) domains, and condensation (C) domains. The NRPS CP domain is also 4′-phosphopantethienylated by a PPTase prior to loading with the activated amino acid (as an adenylate, the product of the A domain). The C domain then catalyzes transfer of the growing peptide chain of a downstream CP domain to the amino acid of an upstream CP domain via amide bond formation. Like PKS systems, NRPS modules may also contain any of a number of modification domains that contribute to the chemical diversity of the resultant product. In addition to isolated PK and NRP synthetases, hybrid PKS/NRPS systems have been observed for natural products such as bleomycin, epothilone, and others [12–16].

However, despite our current growing body of knowledge of the organization, mechanisms, and substrate specificity of NRPS and PKS assemblies, several formidable obstacles still preclude harnessing this machinery for metabolic engineering purposes. For example, there are significant challenges associated with performing genetic manipulations in many host producer microbial strains. There are also marked difficulties associated with expressing soluble active synthetases or subdomains in heterologous hosts. In part this is due to both the inherently large size of PK and NRP megasynthetases, impaired promotion of gene expression, codon usage differences between native and heterologous hosts, and improper folding of the recombinant proteins. In light of these issues, there is a pressing need for the development of alternative methods to identify, quantify, purify, and dissect the function of fully folded, active NRPS and PKS assemblies, both from host-producing organisms as well as heterologous hosts.

In last month's issue of *Chemistry & Biology,* University of California at San Diego assistant professor Michael Burkart, visiting scientist James La Clair, and coworkers described the development of novel activity-based proteomic tools for profiling PKS and NRPS activation [1]. This is the first example of activity-based protein profiling applied to natural product biosynthesis machinery. Burkart, La Clair, and colleagues described a method to covalently label CP domains of PK and