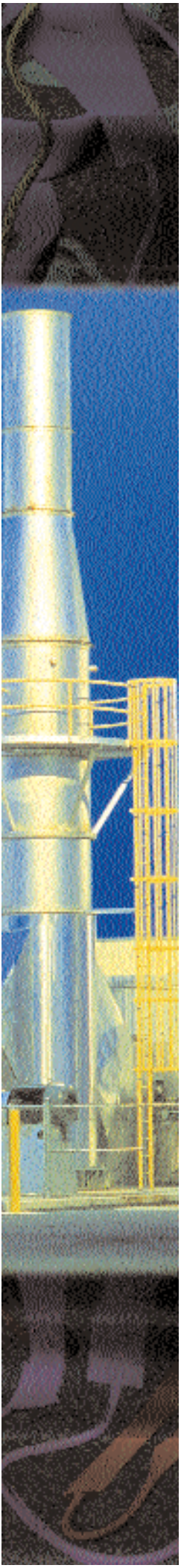

Modular enzymes

Chaitan Khosla & Pehr B. Harbury

Departments of Chemistry, Chemical Engineering and Biochemistry, Stanford University, Stanford, California 94305, USA (e-mail: ck@chemeng.stanford.edu)

Although modular macromolecular devices are encountered frequently in a variety of biological situations, their occurrence in biocatalysis has not been widely appreciated. Three general classes of modular biocatalysts can be identified: enzymes in which catalysis and substrate specificity are separable, multisubstrate enzymes in which binding sites for individual substrates are modular, and multienzyme systems that can catalyse programmable metabolic pathways. In the postgenomic era, the discovery of such systems can be expected to have a significant impact on the role of enzymes in synthetic and process chemistry.

A modular device is a multicomponent system in which individual components can be interchanged with functionally distinct analogues from related systems. Since the elucidation of the genetic code[1], the principle of modularity has been repeatedly uncovered in biological phenomena. Recent examples of modular mechanisms in biology have emerged from the analysis of signal transduction[2,3] and transcriptional activation[4,5]. In each case, the modular design of these systems has given the molecular biologist a powerful conceptual and technical base from which structure–function relationships can be probed.

In contrast to biological processes involving information transfer, metabolism is not regarded as a fertile hunting ground for modular systems. The reasons for this bias are understandable, given our deep admiration for two general properties of metabolic enzymes: their immense rate accelerations and their exquisite substrate selectivity. Because binding and catalysis are two sides of the same coin in the active site of an enzyme, the evolution of modular enzymes in which these two properties are localized in structurally distinct domains is often assumed to be incompatible with the physiological need for high reaction rates. Moreover, because most metabolic pathways involve diffusive transfer of intermediates from one enzyme to the next, maintenance of metabolic fidelity demands that an individual enzyme be able to discriminate sharply between its cognate substrate and related cellular metabolites. In this context, one might imagine that the additional requirement of modularity would present an unnecessary challenge for evolution.

Yet modular enzymes do occur in nature, although they are not common. Nonetheless, their potential utility to the chemist is enormous. Starting from a toolbox containing a reasonable number and diversity of modular enzymes, the ability to harness protein engineering to rapidly generate designer biocatalysts for any process would transform chemistry and chemical engineering. Where, then, in this postgenomic era should one look for modular enzymes? What clues can be regarded as useful indicators of the existence of modularity? Which features of enzymes can one expect to be modularized, and what mechanistic principles underlie modular operation? What are the constraints on modularity in nature's enzymes, and how might they be overcome? This review will attempt to address these questions.

Detection of modular enzymes

For an enzyme to have modular properties, it must minimally have a modular architecture. Typically this implies the existence of multiple domains (or subunits, in the case of an oligomeric enzyme). Formally, domains are defined as stable globular fragments of proteins that may refold autonomously and carry out specific functions[6]. In practice, they are identified typically by computer algorithms that search for segments (of about 50–500 residues) with sequence similarity within a group of larger, functionally distinct polypeptides[7–9]. There are two principal assumptions in this strategy for detecting a structurally modular enzyme. The first is that homologous domains have similar tertiary structures. Although exceptions to this rule have been observed[10], this generally seems to be a reasonable assumption. More demanding is the assumption that interdomain interactions are either absent or conserved within a family of modular enzymes. As discussed below, this represents a major limitation to the modularity of enzymes.

Whereas a domain-like architecture is suggestive of a modular enzyme, it does not guarantee functional modularity. Experimental evidence for modularity requires satisfaction of two criteria. First, distinct properties must be able to be assigned to each domain, identified by means of either proteolysis or protein engineering. Second, it should be possible to recombine these domains to generate functional chimaeras. Although very few enzyme families have been shown to satisfy both criteria for modularity, rapidly expanding sequence and structural databases continue to add to the list of potentially modular enzyme families. It should be noted that not every nominally modular

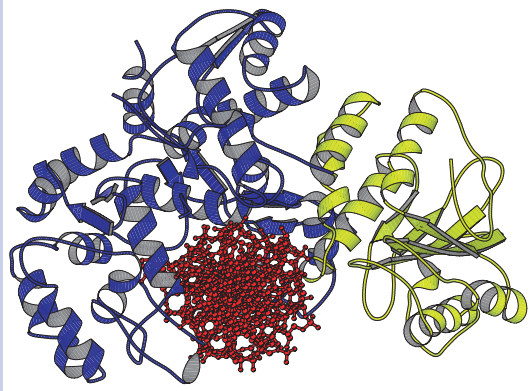

Figure 1 Separate catalytic and molecular recognition domains of the *Fok*I restriction endonuclease. The carboxy-terminal DNA-cleavage domain (yellow) piggy-backs on the amino-terminal DNA-binding domain (blue). The DNA double helix is coloured red. Figure generated from the protein data bank 1FOK coordinates using MOLSCRIPT[53].

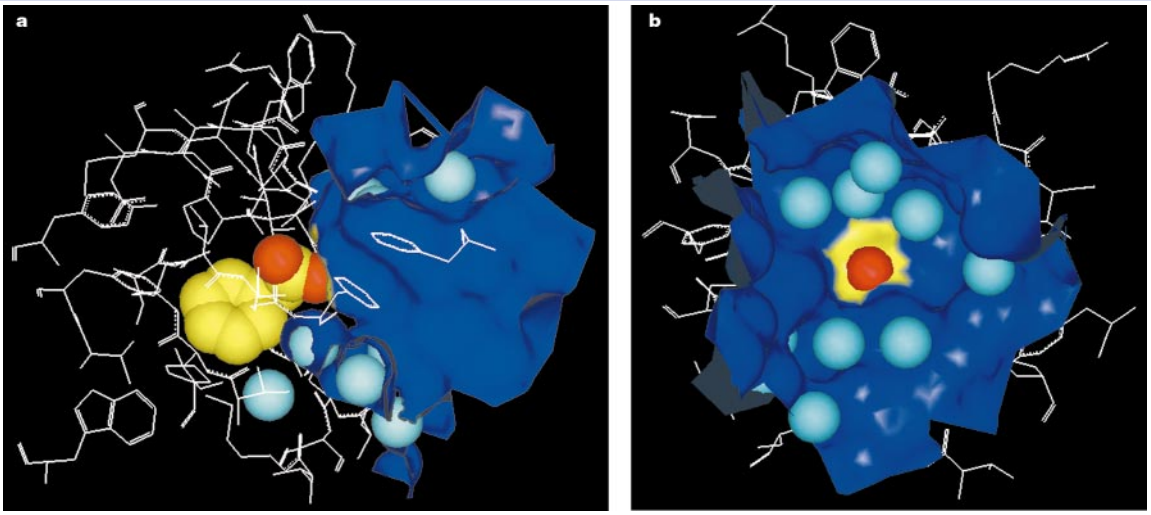

Figure 2 Recognition by penicillin G acylase of a modular chemical feature (the phenylacetyl group) in an otherwise generic substrate. Residues of penicillin acylase within 10 Å of the active site are depicted in white. The carbon and oxygen atoms of bound phenylacetic acid are coloured yellow and red respectively. A partial solvent-accessible surface of the enzyme is shown in blue. Ordered water molecules are rendered as cyan balls. **a**, Although the phenylacetyl moiety is surrounded by the enzyme, the distal oxygen atom of the acid (red) is directed into solvent. Generic amine and alcohol substrates of penicillin acylase are attached to the phenacetyl protecting group through this distal atom. **b**, A view of the same active site rotated by 90° around a vertical axis. The distal oxygen atom of phenylacetic acid (red) can be seen poking through the solvent-accessible surface. Figure generated from the 1PNL coordinates of the protein data bank using INSIGHTII (Molecular Simulations, San Diego).

biocatalyst is as perfectly modular as, for example, the machinery that catalyses ribosomal protein synthesis[1]. In most cases, although the properties of individual domains or subunits can be recombined in chimaeric enzymes, the resulting chimaeras display kinetic imperfections when compared with their parents.

One can define three general categories of modular catalysts in nature: (1) enzymes in which catalysis and substrate specificity are separable; (2) multisubstrate enzymes in which the binding sites for individual substrates are modular; and (3) multienzyme systems that can catalyse programmable metabolic pathways. Using selected examples, we will illustrate the properties of enzymes belonging to each of these categories. We also discuss the opportunities for engineering enzymes in each category, as well as their potential utility in the context of applied biocatalysis. As a frame of reference, we note that the prototypical modular enzyme — the ribosomal protein biosynthetic machinery — exhibits all three types of modularity. The catalyst (the ribosome) can be readily separated from the element that dictates substrate specificity (the messenger RNA template). Moreover, the two acylated transfer RNA substrates for the peptide bond-forming reaction bind to distinct sites in the ribosome (the A and P sites). And finally, the multistep pathway catalysed by this modular system can be reprogrammed by codon removal, addition or replacement. The same can be said about template-dependent DNA and RNA polymerases. Although such an extraordinary degree of modularity is clearly an exception rather than the norm, these modular features underlie the singular success of these systems as practically useful biocatalysts. Therefore a better understanding of the chemistry and biology of modular enzymes has important implications for biocatalysis.

Types of modular enzymes

Separation of catalysis from molecular recognition

Enzymes are prized by chemists for their regio- and stereoselectivity, yet their use is often restricted to a narrow spectrum of substrates. Paradoxically, an ideal catalyst for the chemist should be able to pinpoint a single functional group on a molecule, yet retain the ability to act on many different molecular species. If an enzyme's chemical reactivity could be separated in a modular fashion from its substrate recognition, it would be possible to create a catalyst (or family of

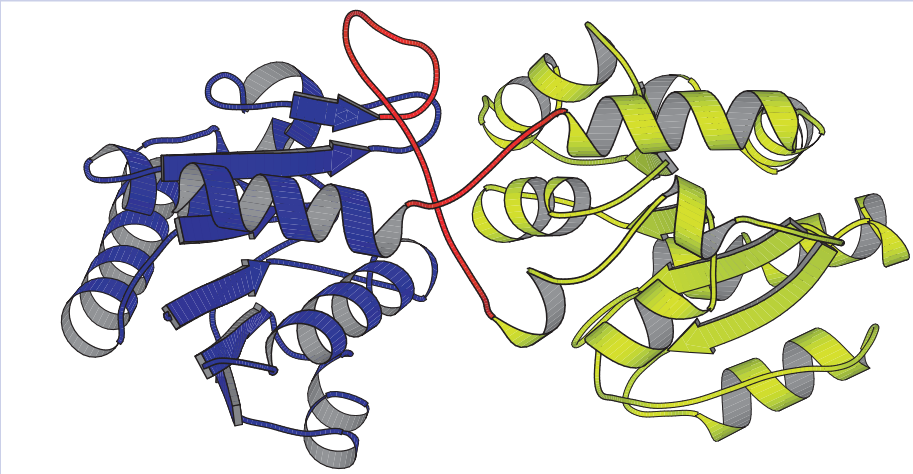

Figure 3 Separation of molecular-recognition features in modular multisubstrate enzymes. Crystal structure of the bi-domain MurG glycosyltransferase. The NDP-sugar and aglycone binding domains are coloured yellow and blue, respectively. The two loops that connect these two domains are coloured in red. The active-site residues lie at the interface of these two domains. For details, see ref. 36.

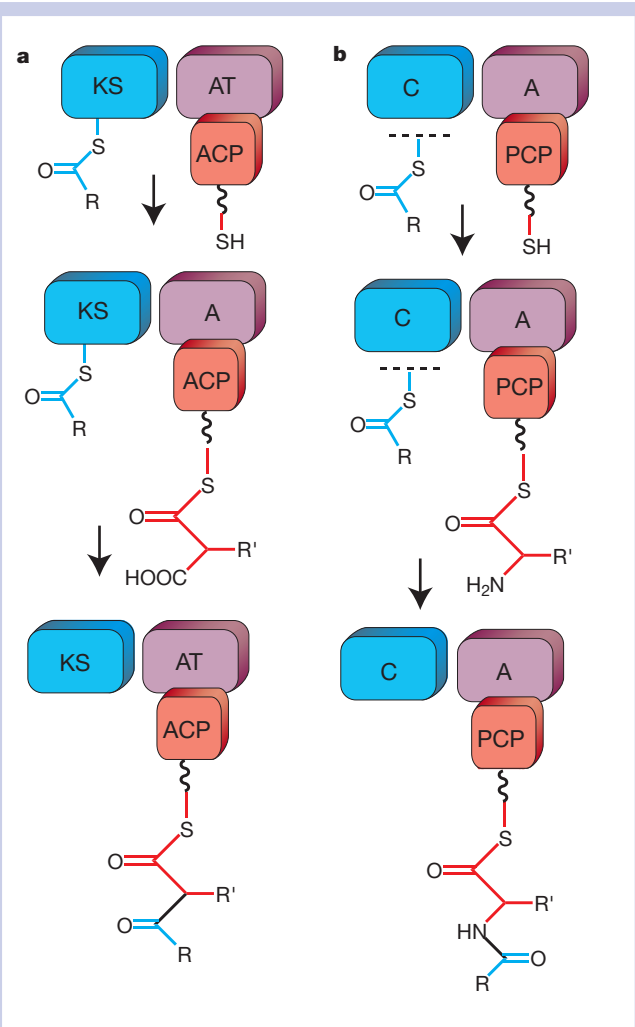

Figure 4 Separation of electrophile and nucleophile recognition in modules of polyketide synthases and non-ribosomal peptide synthetases. **a**, A polyketide synthase module is minimally comprised of three distinct domains: a condensation (KS, ketosynthase), an acyltransferase (AT), and an acyl carrier protein (ACP) domain. The electrophile (shown in blue) is recognized by and attached to the KS domain, whereas the nucleophile (shown in red) is recognized by the AT domain, which attaches this moiety to the 'swinging arm' of the ACP domain (wavy line). The C–C bond formed between the electrophile and the nucleophile is shown in black. **b**, Likewise, a non-ribosomal peptide synthetase module is also minimally comprised of three distinct domains: a condensation (C), an adenylation (A), and a peptidyl carrier protein (PCP) domain. The electrophile (shown in blue) is recognized by the C domain (although it is not believed to bind covalently to this domain, hence shown attached to the previous domain (dashed line)), whereas the nucleophile (shown in red) is recognized by the A domain, which attaches this moiety to the 'swinging arm' of the PCP domain (wavy line). The amide bond formed between the electrophile and the nucleophile is shown in black.

catalysts) that is both specific and general. This is a tall order, because substrate binding and catalytic activity are inextricably linked[11]. Binding interactions orient substrates with respect to enzyme active-site residues, and preferentially stabilize transition states relative to ground states. Nevertheless, enzymes exploit at least three different strategies for partially uncoupling substrate binding and catalysis in a modular fashion. They include the separation of catalytic and recognition functions in distinct domains, evolution of interchangeable substrate binding pockets, and recognition of a modular chemical feature in an otherwise generic substrate.

The *Fok*I restriction endonuclease exemplifies the division of labour between protein domains. It consists of two autonomously folded structures that can be separated by limited proteolysis (Fig. 1)[12]. The first domain binds a GGATG DNA-recognition motif, and the second domain catalyses DNA-strand scission nine and thirteen bases away. Fusion of the *Fok*I-cleavage domain to the DNA-binding domains of unrelated transcription factors produces chimaeras with new cleavage specificities[13,14]. However, the *Fok*I system has not proved to be perfectly modular. Unexpectedly, the binding domain sequesters and inactivates the cleavage domain when the endonuclease is nonspecifically associated with DNA[15]. Domain-swapped enzymes that lack these inhibitory interactions exhibit high levels of nonspecific DNA cleavage.

A consortium of bacterial hydrolases that break down plant cell walls also exhibits a partial domain division between substrate recognition and chemistry. The hydrolases include xylanases and cellulases that have one or two catalytic glycosidase domains, in addition to a series of carbohydrate-binding domains (CBDs). The catalytic domains can hydrolyse soluble oligosaccharides, but they require the CBD domains to break down the crystalline sugar polymers found in plant cell walls[16]. The CBDs act by increasing enzyme–substrate proximity[17]. Fusion of cellulose- and xylan-binding domains to heterologous glycosidase domains confers the ability to metabolize crystalline substrates.

Analysis of covalent domain partnering in different genomes will probably yield other examples of enzymes with independent catalytic and binding domains[18]. But some mechanistic limitations of this class of enzyme should be noted. Most importantly, binding by the recognition domain does not orient the substrate for reaction. In fact, the inter-domain linker must be floppy, so that binding by the recognition domain does not prevent correct positioning of the substrate in the active site. The effective concentration of the bound substrate, held near the catalytic domain by the inter-domain linker, determines the selectivity of the enzyme. For example, if the effective concentration of a bound substrate at the active site is 10 mM, and unbound substrates are present at 1 mM, the maximum specificity exhibited by the enzyme for the bound substrate will be tenfold[19].

In addition to domain-based separation, catalysis can also be uncoupled from substrate recognition by the exchange of modular binding pockets that are located in the same structural domain as the active site. The *Tetrahymena* group I ribozyme exhibits this kind of modularity[20]. The internal guide sequence (IGS) of the ribozyme forms a duplex with the RNA substrate, positioning the substrate for cleavage. Watson–Crick base pairing between the substrate and the IGS governs substrate recognition. Base substitutions in the IGS produce ribozymes with specificity for RNA sequences bearing complementary substitutions. Likewise, the Src family of kinases provide an example of binding-pocket modularity in an enzyme with a non-polymeric substrate[21]. A single Ile→Gly substitution in the nucleotide binding pockets of Src or Fyn kinases allows the mutant proteins to efficiently use N^6-benzyl ATP. The wild-type Src kinase discriminates against the alkylated ATP substrate by a factor of more than 400-fold.

Although multiple sequence alignments in genome data together with the growth in protein structural data will undoubtedly offer new opportunities to identify and exploit substrate binding modularity, the existence of modular substrate recognition uncoupled from catalysis is likely to be relatively rare. Chimaeric enzymes with swapped binding sites often show affinity for new substrates, but this affinity is not correlated with the ability to convert substrate to product[22]. Gene shuffling of enzyme families[23] may be a useful tool to shed light on the generality of binding-pocket modularity.

A third class of enzymes that modularize catalysis and molecular recognition does so by recognizing a specific chemical feature in an otherwise generic substrate. The substrate can be practically anything as long as the recognition feature is present. Examples of this type of enzyme (and the functional group recognized) include lipases (straight-chain hydrocarbons), penicillin G acylase (phenylacetyl group), butyryl cholinesterase (choline), phthalyl amidase (phthalimide) and aryl acylamidases (acetanilide)[24–27]. Increasingly, these catalysts are being used to manipulate protecting groups on complex molecules,

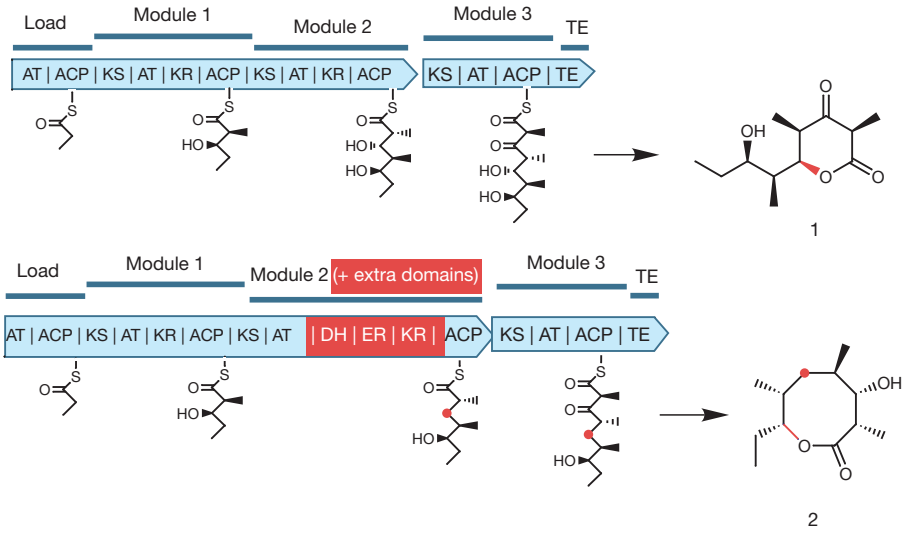

Figure 5 Introduction of auxiliary catalytic domains into the module of a polyketide synthase. Formation of the six-membered lactone 1 is catalysed by a trimodular derivative of the erythromycin polyketide synthase, comprised of a loading bi-domain (Load), three consecutive modules of catalytic domains (Module 1, Module 2 and Module 3), and a terminal thioesterase (TE) domain that catalyses release of 1 from module 3 via attack of the δ-OH on the thioester linkage. Domains designated as KS, AT and ACP are explained in Fig. 4. Ketoreductase (KR) domains in modules 1 and 2 are responsible for generating β-OH groups on the growing chains by reducing the ketone generated in the condensation reaction. By replacing the KR domain of module 2 with a KR–DH (dehydratase)–ER (enoylreductase) tridomain from the rapamycin synthase, the β-OH (highlighted in red) is eliminated by dehydration and enoylreduction. The extra catalytic domains introduced into module 2 are therefore able to compete kinetically with the downstream module 3. Moreover, the absence of the δ-OH group on the tetraketide product of module 3 forces the TE domain to form the eight-membered lactone 2. For details, see ref. 54.

which are sensitive to the harsh chemical reagents of traditional organic chemistry, or to produce chiral molecules[28–30]. For example, the *Serratia marcescens* lipase is used to manufacture a synthetic intermediate of the calcium antagonist diltiazem on the scale of 50 tons yr^{-1}.

Enzymes with the unusual ability to recognize an isolated chemical module have so far been discovered empirically. The recent crystal structures of penicillin G acylase (Fig. 2) and of several lipases offer mechanistic insights into their mode of action. Both structures reveal shallow active sites, with variable regions of the substrate directed outwards into solvent[31–33]. Large portions of the substrate are not in intimate contact with the enzyme. Where might one look for additional examples of this class of catalyst? Bacterial lipases, phthalyl amidase and penicillin acylase are thought to function as carbon-source scavengers for the microorganisms that produce them. Aryl acylamidases in plants, rhizobacteria and soil microorganisms break down diverse acetanilide-derived pesticides (such as propanil), thereby conferring resistance. Butyryl cholinesterase exists in the liver and plasma of humans. It scavenges general choline esters and detoxifies a large number of drugs, including cocaine. The pattern of natural activities indicates that enzymes involved in nutrient recruitment and chemical warfare may have evolved under selective pressure to tolerate diverse substrates. Consequently, proteins involved in these processes might be expected to provide a rich source of undiscovered catalysts for enzymatic biotransformation.

Separation of molecular-recognition features in multisubstrate enzymes

Many enzymes catalyse reactions involving two (or more) substrates. Broadly speaking, these enzymes fall into two categories. Reactions in which all substrates bind to the enzyme before the first product is formed are called sequential, whereas those in which one or more products are released are called ping-pong[34]. The active sites of sequential enzymes can bind substrates in a random or a defined order. In ping-pong enzymes or in ordered sequential enzymes, the binding pockets for both substrates often overlap or interact with each other. But in random-binding, bi-substrate enzymes, the binding pockets are generally separated; indeed, in some cases they may even lie within architecturally distinct domains of the polypeptide backbone. A vivid example of such modular enzymes is methionine synthase, whose substrates — homocysteine, methyltetrahydrofolate, cobalamin and adenosylmethionine — bind to distinct regions of the protein[35]. Modularity of molecular-recognition features of multisubstrate enzymes represents a fertile starting point for protein engineering.

Sequence and structural analysis also indicates that many families of evolutionarily related bi-substrate enzymes have modular molecular-recognition features. For example, enzymes that use nucleoside diphosphate (NDP)-sugars, such as glycosyltransferases[36] and dehydrogenases[37], seem to have this property. NDP-glycosyltransferases in particular belong to a large family of enzymes that are known to possess relaxed specificity for both the sugar and the aglycones[38,39]. Recently the crystal structure of a prototype of this family has revealed a two-lobed architecture (Fig. 3)[36]. One domain binds to the NDP-sugar whereas the other domain binds to the aglycone. The two domains are separated by extended loops, and the active-site residues lie at the interface of the two domains. Sequence comparisons indicate that most members of this enzyme superfamily retain both the bi-domain architecture and the location of active-site residues[40]. Whereas functional evidence for modularity is lacking, it is plausible that the molecular-recognition features of these two domains have not only evolved independently but also been exchanged frequently without destroying the active-site geometry. If so, then the relaxed specificity of individual domains together with the biosynthetic importance of this enzyme family makes it a particularly attractive target for domain shuffling.

The individual modules of polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) represent another strategy that nature seems to have exploited for modularizing molecular recognition features in protein catalysts. Here, an electrophile and a nucleophile are attached covalently to two distinct domains (or subunits) of a multidomain (or multisubunit) enzyme (Fig. 4)[41]. Selectivity for the electrophile resides in the domain that catalyses bond formation between the two substrates (the condensing enzyme), whereas nucleophile selectivity is controlled by a transfer domain, which attaches the nucleophile onto the pantetheine arm of a carrier domain. Although direct structural evidence for the modularity of these three domains is lacking, numerous studies have shown that catalytic bond formation between the two substrates can be reconstituted by recombination of heterologous condensing, transfer and carrier domains (for reviews, see refs 42 and 43).

In contrast to the relative rarity of enzymes that modularize catalysis and molecular recognition, multisubstrate enzymes with modular recognition features are probably more common in nature.

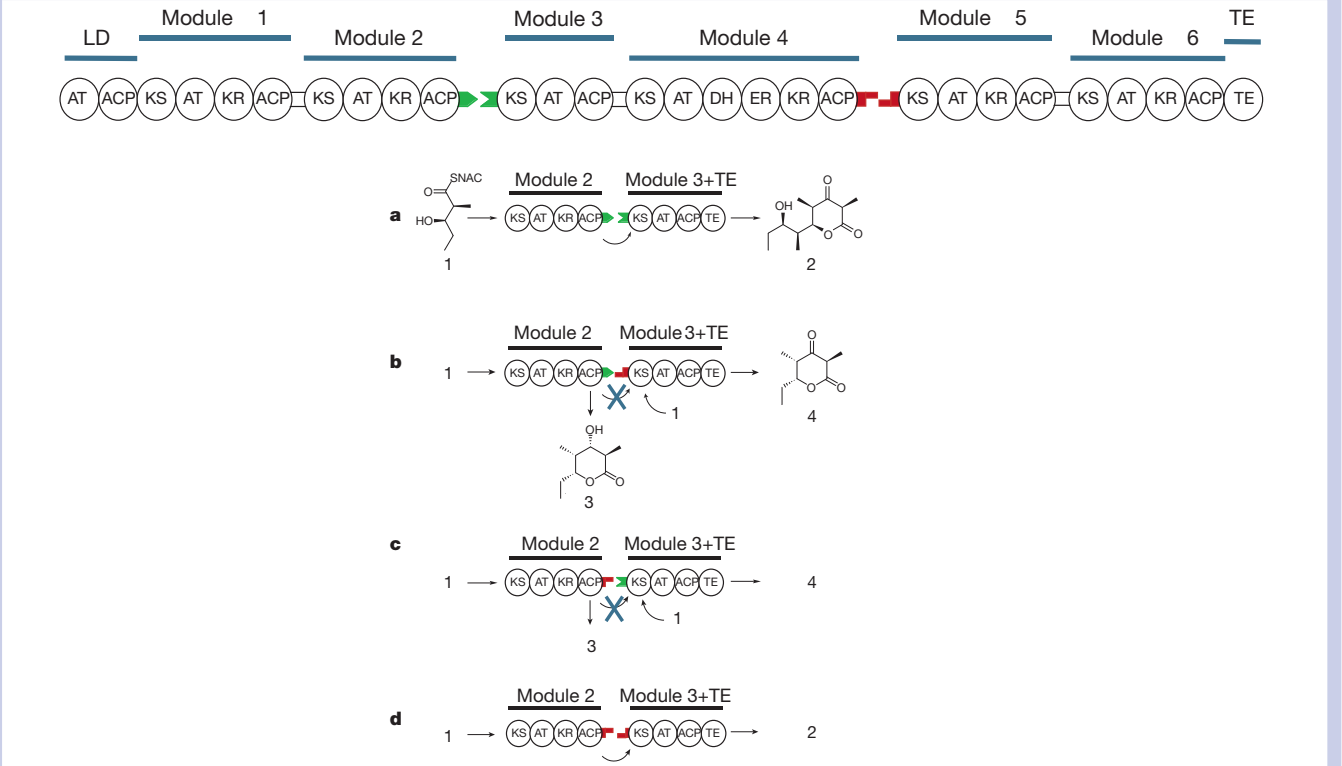

Figure 6 Modular protein–protein interactions in the selective channelling of intermediates between successive catalysts in a multistep metabolic pathway. The erythromycin polyketide synthase (top) is comprised of three very large proteins (relative molecular mass > 300,000), each possessing two modules (defined in Fig. 4)[55,56]. The triketide synthesized by module 2 is channelled selectively into module 3 (to the exclusion of module 5), whereas the pentaketide product of module 4 is channelled selectively into module 5 (to the exclusion of module 3). As illustrated in schemes **a**–**d**, this selectivity is determined largely by short 'linker pairs' (coloured in green and red) at the C and N termini of the polypeptides. Each linker is 25–50 residues in length, lacks significant sequence similarity to other linkers or other sequences in the database, but shows propensity to assume a coiled-coil structure. Matched linker pairs can effectively mediate intermediate transfer between modules, as illustrated in schemes **a** and **d**, whereas mismatched linkers abolish this channelling, as illustrated in schemes **b** and **c**. Thus, by appropriate engineering of linkers, substrates can be channelled between heterologous polyketide synthase modules, giving rise to new polyketide products.

In particular they might be expected to be abundant in secondary metabolism, where extraordinarily high metabolic fluxes are not required, and the driving force to mix and match enzymatic function is perhaps greater. For example, the turnover numbers of enzymes that synthesize the aglycone substrates of most of the glycosyltransferases shown in Fig. 2 are in the range of 1 min^{-1}; consequently an otherwise unimpressive k_{cat} of 1 min^{-1} is reasonable for these glycosyltransferases. This lack of evolutionary pressure to maximize rate constants attenuates the limitations associated with modular substrate-recognition features.

What are the limits to the modularity of the substrate-recognition features of catalysts such as glycosyltransferases, PKS modules or NRPS modules? Analysis of individual PKS and NRPS modules has demonstrated that the kinetics of catalytic bond formation is influenced by the selection of both the electrophile and the nucleophile substrates[44,45]. Although these 'imperfections' in modularity do not seem to present qualitative barriers to the predictive design and biosynthetic engineering of new 'unnatural' natural products, they do affect the productivity of hybrid multifunctional catalysts in fermentation processes[46]. A combination of random mutagenesis and structure-based approaches may be useful in ameliorating these bottlenecks.

Channelling of intermediates in multienzyme systems

Systems such as PKSs and NRPSs illustrate yet another principle for modular catalysis that, if generalized, would have significant implications for the development of one-pot biocatalytic processes. Instead of relying strictly on diffusion, they have evolved a highly modular strategy to channel intermediates from one active site to the next. Channelling can be defined as the direct transfer of an intermediate between consecutive enzymes in a metabolic sequence without equilibrating with the bulk fluid[47]. It has been demonstrated most vividly in the cases of enzymes such as tryptophan synthase and certain aminotransferases, where channels 10–100 Å wide have been identified crystallographically that restrict the diffusion of indole and ammonia, respectively, from one reaction centre to the next[48–50]. Channelling can be useful in cases where extremely reactive intermediates (such as ammonia) need to be transferred between active sites. It can also reduce the need for high specificity in the recipient enzyme for its cognate substrate.

Multienzyme assemblies such as α-keto-dehydrogenases, fatty acid synthases, PKSs and NRPSs have 'swinging arms' (lipoamide in the case of dehydrogenases[51], and phosphopantetheine in the cases of fatty acid synthases, PKSs and NRPSs[41]). These swinging arms are flexible, long (~10–15 Å) tethers that channel covalently bound intermediates between successive active sites (see Fig. 4). Two examples illustrate possible mechanisms by which a swinging arm can be combined with other structural features to make a metabolic pathway modular. First, individual PKS modules have been altered to expand their repertoire of catalytic functions. Gain-of-function mutagenesis involves grafting auxiliary catalytic domains into a core PKS module, and can be used to introduce new chemistry into the reaction sequence catalysed by the module (Fig. 5). Not only does this highlight the extraordinary structural plasticity of a module, but it also suggests that the presence of a swinging arm provides an effective mechanism for newly grafted domains to compete for potential substrates before they are transferred from one module to the next. Second, in combination with selective protein–protein interactions, swinging arm-mediated chemistry can facilitate transfer of natural and unnatural intermediates between mod-

ules by attenuating the role of protein–substrate interactions (Fig. 6; S. Y. Tsuji, D. E. Cane and C.K., unpublished results). Here metabolism seems to have borrowed a chapter directly from signal-transduction mechanisms by incorporating selectivity into matched pairs of short 'linkers' — as exemplified by the case of Fos–Jun interactions, where modular coiled-coil segments stabilize the heterodimer preferentially over either homodimer[52].

What limits more widespread exploitation of modular channelling mechanisms in multistep bioconversion processes? First, these mechanisms seem to be restricted to systems that rely extensively on covalent catalysis, presumably because of the need for a swinging arm. Second, even where selective linkers can be engineered to direct intermediates between designated catalytic modules, in themselves they can at most provide the advantage of intramolecularity. For some catalytic reactions, approximation can provide a huge rate enhancement; in other cases it is only of relatively modest value. Finally, not all protein–protein interactions are modular. For example, in a PKS module, if interactions between the donor carrier domain and the acceptor condensing enzyme (ACP_2 and KS_3 in Fig. 6) are important to chain transfer, these properties will vary with module ultrastructure and will be difficult to categorize universally.

Conclusions

The brief history of molecular and cellular biology has demonstrated repeatedly that modularity in biological macromolecules can be exploited by both evolution and engineering. In many such enzymes, the structural and mechanistic basis for modularity is only now being elucidated. Yet notwithstanding our rudimentary knowledge about these enzymes, their utility in practical biocatalysis has been well established, and their attractiveness as targets for protein engineering is becoming apparent. As our understanding of these remarkable catalysts advances, and as the tools of molecular biology and knowledge-based protein design improve, one can expect to see more such engineered enzymes making the transition from the proof-of-principle stage to industrially useful biocatalysts with respectable space-time yields. At the same time, as sequence and structural databases continue to grow, new families of modular enzymes will surely emerge to expand the repertoire of chemistry that is accessible to modular biocatalysts. □

1. Crick, F. H., Barnett, I., Brenner, S. & Watts-Tobin, R. General nature of the genetic code for proteins. *Nature* **192,** 1227–1232 (1961).
2. Cohen, G. B., Sen, R. & Baltimore, D. Modular binding proteins in signal transduction proteins. *Cell* **80,** 237–248 (1995).
3. Sudol, M. From Src homology domains to other signaling modules: proposal of the protein recognition code. *Oncogene* **17,** 1469–1474 (1998).
4. Frankel, A. D. & Kim, P. S. Modular structure of transcription factors: implications for gene regulation. *Cell* **65,** 717–719 (1991).
5. Ptashne, M. & Gann, A. Imposing specificity by localization: mechanism and evolvability. *Curr. Biol.* **8,** R897–R904 (1998).
6. Rashin, A. A. Location of domains in globular proteins. *Nature* **291,** 85–87 (1981).
7. Orengo, C. A., Jones, D. T. & Thornton, J. M. Protein superfamilies and domain superfolds. *Nature* **372,** 631–634 (1994).
8. Henikoff, S. *et al.* Gene families: the taxonomy of protein paralogs and chimeras. *Science* **278,** 609–614 (1997).
9. Riley, M. & Labedan, B. Protein evolution viewed through *Escherichia coli* protein sequences: introducing the notion of a structural segment of homology, the module. *J. Mol. Biol.* **268,** 857–868 (1997).
10. Dalal, S., Balasubramanian, S. & Regan, L. Protein alchemy: changing β-sheet into α-helix. *Nature Struct. Biol.* **4,** 548–552 (1997).
11. Jencks, W. P. *Catalysis in Chemistry and Enzymology* 615–806 (Dover, New York, 1987).
12. Li, L., Wu, L. P. & Chandrasegaran, S. Functional domains in Fok-I restriction endonuclease. *Proc. Natl Acad. Sci. USA* **89,** 4275–4279 (1992).
13. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes—zinc finger fusions to Fok I cleavage domain. *Proc. Natl Acad. Sci. USA* **93,** 1156–1160 (1996).
14. Huang, B. H., Schaeffer, C. J., Li, Q. H. & Tsai, M. D. Splase—a new class I zinc-finger restriction endonuclease with specificity for Sp1 binding sites. *J. Protein Chem.* **15,** 481–489 (1996).
15. Aggarwal, A. K. & Wah, D. A. Novel site-specific DNA endonucleases. *Curr. Opin. Struct. Biol.* **8,** 19–25 (1998).
16. Fernandes, A. C. *et al.* Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. *Biochem. J.* **342,** 105–110 (1999).
17. Bolam, D. N. *et al. Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem. J.* **331,** 775–781 (1998).
18. Marcotte, E. M. *et al.* Detecting protein function and protein-protein interactions from genome sequences. *Science* **285,** 751–753 (1999).
19. Jencks, W. P. On the attribution and additivity of binding energies. *Proc. Natl Acad. Sci. USA* **78,** 4046–4050 (1981).
20. Murphy, F. L. & Cech, T. R. Alteration of substrate specificity for the endoribonucleolytic cleavage of RNA by the *Tetrahymena* ribozyme. *Proc. Natl Acad. Sci. USA* **86,** 9218–9222 (1989).
21. Liu, Y., Shah, K., Yang, F., Witucki, L. & Shokat, K. M. Engineering Src family protein kinases with unnatural nucleotide specificity. *Chem. Biol.* **5,** 91–101 (1998).
22. Hedstrom, L., Farrjones, S., Kettner, C. A. & Rutter, W. J. Converting trypsin to chymotrypsin—ground-state binding does not determine substrate specificity. *Biochemistry* **33,** 8764–8769 (1994).
23. Stemmer, W. P. C. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370,** 389–391 (1994).
24. Jaeger, K. E., Dijkstra, B. W. & Reetz, M. T. Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases. *Annu. Rev. Microbiol.* **53,** 327–351 (1999).
25. Pathak, T. & Waldmann, H. Enzymes and protecting group chemistry. *Curr. Opin. Chem. Biol.* **2,** 112–120 (1998).
26. Costello, C. A., Kreuzman, A. J. & Zmijewski, M. J. Selective deprotection of phthalyl protected amines. *Tetrahedron Lett.* **37,** 7469–7472 (1996).
27. Villarreal, D. T., Turco, R. F. & Konopka, A. A structure-activity study with aryl acylamidases. *Appl. Env. Microbiol.* **60,** 3939–3944 (1994).
28. Roberts, S. M. Preparative biotransformations: the employment of enzymes and whole-cells in synthetic organic chemistry. *J. Chem. Soc. Perkin Trans.* **1,** 157–169 (1998).
29. Roberts, S. M. Preparative biotransformations. *J. Chem. Soc. Perkin Trans.* **1,** 1–21 (1999).
30. Roberts, S. M. Preparative biotransformations. *J. Chem. Soc. Perkin Trans.* **1,** 611–633 (2000).
31. Duggleby, H. J. *et al.* Penicillin acylase has a single amino-acid catalytic centre. *Nature* **373,** 264–268 (1995).
32. Done, S. H., Brannigan, J. A., Moody, P. C. E. & Hubbard, R. E. Ligand-induced conformational change in penicillin acylase. *J. Mol. Biol.* **284,** 463–475 (1998).
33. Lang, D. A., Mannesse, M. L. M., DeHaas, G. H., Verheij, H. M. & Dijkstra, B. W. Structural basis of the chiral selectivity of *Pseudomonas cepacia* lipase. *Eur. J. Biochem.* **254,** 333–340 (1998).
34. Fersht, A. *Enzyme Structure and Mechanism* (Freeman, 1985).
35. Goulding, C. W., Postigo, D. & Matthews, R. G. Cobalamin-dependent methionine synthase is a modular protein with distinct regions for binding homocysteine, methyltetrahydrofolate, cobalamin, and adenosylmethionine. *Biochemistry* **36,** 8082–8091 (1997).
36. Ha, S., Walker, D., Shi, Y. & Walker, S. The 1.9 A crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.* **9,** 1045–1052 (2000).
37. Campbell, R. E., Mosimann, S. C., van de Rijn, I., Tanner, M. E. & Strynadka, N. C. The first structure of UDP-glucose dehydrogenase reveals the catalytic residues necessary for the two-fold oxidation. *Biochemistry* **39,** 7012–7023 (2000).
38. Borisova, S. A., Zhao, L., Sherman, D. H. & Liu, H. W. Biosynthesis of desosamine: construction of a new macrolide carrying a genetically designed sugar moiety. *Org. Lett.* **1,** 133–136 (1999).
39. Olano, C., Lomovskaya, N., Fonstein, L., Roll, J. T. & Hutchinson, C. R. A two-plasmid system for the glycosylation of polyketide antibiotics: bioconversion of epsilon-rhodomycinone to rhodomycin D. *Chem. Biol.* **6,** 845–855 (1999).
40. Kapitonov, D. & Yu, R. K. Conserved domains of glycosyltransferases. *Glycobiology* **9,** 961–978 (1999).
41. Cane, D. E., Walsh, C. T. & Khosla, C. Harnessing the biosynthetic code. Combinations, permutations, mutations. *Science* **282,** 63–68 (1998).
42. Khosla, C., Gokhale, R., Jacobsen, J. R. & Cane, D. E. Tolerance and specificity of polyketide synthases. *Annu. Rev. Biochem.* **68,** 219–253 (1999).
43. Mootz, H. D. & Marahiel, M. A. Design and application of multimodular peptide synthetases. *Curr. Opin. Biotechnol.* **10,** 341–348 (1999).
44. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284,** 486–489 (1999).
45. Wu, N., Kudo, F., Cane, D. E. & Khosla, C. Analysis of the molecular recognition features of individual modules derived from the erythromycin polyketide synthase. *J. Am. Chem. Soc.* **122,** 4847–4852 (2000).
46. McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl Acad. Sci. USA* **96,** 1846–1851 (1999).
47. Geck, M. K. & Kirsch, J. F. A novel, definitive test for substrate channeling illustrated with the aspartate aminotransferase/malate dehydrogenase system. *Biochemistry* **38,** 8032–8037 (1999).
48. Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W. & Davies, D. R. Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. *J. Biol. Chem.* **263,** 17857–17871 (1988).
49. Krahn, J. M. *et al.* Coupled formation of an amidotransferase interdomain ammonia channel and a phosphoribosyltransferase active site. *Biochemistry* **36,** 11061–11068 (1997).
50. Thoden, J. B., Holden, H. M., Wesenberg, G., Raushel, F. M. & Rayment, I. Structure of carbamoyl phosphate synthetase: a journey of 96 Å from substrate to product. *Biochemistry* **36,** 6305–6316 (1997).
51. Perham, R. N. & Reche, P. A. Swinging arms in multifunctional enzymes and the specificity of post-translational modification. *Biochem. Soc. Trans.* **26,** 299–303 (1998).
52. O'Shea, E. K., Rutkowski, R., Stafford, W. F. III & Kim, P. S. Preferential heterodimer formation by isolated leucine zippers from Fos and Jun. *Science* **245,** 646–648 (1989).
53. Kraulis, P. J. Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
54. Kao, C. M. *et al.* Gain of function mutagenesis of a modular polyketide synthase II. Engineered biosynthesis of an eight-membered ring tetraketide lactone. *J. Am. Chem. Soc.* **119,** 11339–11340 (1997).
55. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348,** 176–178 (1990).
56. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252,** 675–679 (1991).

Acknowledgements

Research on modular enzymes in C.K.'s laboratory is supported by grants from the National Science Foundation and the National Institutes of Health. P.B.H. is a Terman Fellow, a Searle Scholar and a Burroughs–Wellcome Young Investigator in the Pharmacological Sciences. We thank S. Walker for helpful discussions regarding glycosyltransferases.