J Ind Microbiol Biotechnol (2012) 39:227–241
DOI 10.1007/s10295-011-1044-2

REVIEW

Protein engineering towards natural product synthesis and diversification

Angelica O. Zabala · Ralph A. Cacho · Yi Tang

Received: 28 July 2011 / Accepted: 29 September 2011 / Published online: 18 October 2011
© Society for Industrial Microbiology 2011

Abstract A dazzling array of enzymes is used by nature in making structurally complex natural products. These enzymes constitute a molecular toolbox that may be used in the construction and fine-tuning of pharmaceutically active molecules. Aided by technological advancements in protein engineering, it is now possible to tailor the activities and specificities of these enzymes as biocatalysts in the production of both natural products and their unnatural derivatives. These efforts are crucial in drug discovery and development, where there is a continuous quest for more potent agents. Both rational and random evolution techniques have been utilized in engineering these enzymes. This review will highlight some examples from several large families of natural products.

The abundance of natural product-producing organisms, spanning all kingdoms of life, makes them a good resource of important molecules useful to man. This includes those with pharmacological relevance, like antibiotics, anti-cancer drugs, and cholesterol-lowering statins; as well as those that are used in everyday living, as in the case of cosmetics, fertilizers, food additives, and many others. Despite the structural and functional diversities observed with natural products, there is always room for further discovery or modification of these compounds driven by the need for more potent and effective agents [26]. This is especially true in drug development, where newer-generation therapeutics are continually needed to battle ever-evolving pathogens that develop resistance to currently used drugs. In addition, structural-activity-relationship studies on bioactive natural products have revealed that even subtle modifications can result in significant increases in activities [10, 23, 40, 64].

While organic synthesis remains a powerful tool in medicinal chemistry, the use of fermentative and biocatalytic approaches to generate and produce natural product derivatives have become more attractive, and in some cases, indispensable. The biosynthesis of natural products is accomplished by dedicated enzymes in different producing organisms. Typically, a collection of enzymes catalyzing diverse reactions are recruited into a biosynthetic pathway to assemble and decorate the complex natural products. Collectively, these enzymes can be considered nature's toolbox that can be used to generate structural and functional diversity. Because natural enzymes are typically only optimized under an evolutionary requirement towards a specific function, enzymes in natural product biosynthetic pathways are expected and have demonstrated to be highly evolvable towards utilizing unnatural substrates and gaining new functions in laboratory and industrial settings. With technological advancements in protein engineering and structural biology, it is now possible and even straightforward to remold biosynthetic enzymes towards the rational and combinatorial biosynthesis of new compounds. Some studies utilize a rational approach aided by structural information or bioinformatic analysis. On the other hand, some use directed evolution techniques, in combination with high-throughput screening to engineer

Angelica O. Zabala and Ralph A. Cacho contributed equally to this work.

A. O. Zabala · R. A. Cacho · Y. Tang (✉)
Department of Chemical and Biomolecular Engineering,
University of California, Los Angeles, CA, USA
e-mail: yitang@ucla.edu

Y. Tang
Department of Chemistry and Biochemistry,
University of California, Los Angeles, CA, USA

these natural enzymes to perform a novel function or to enhance their activity. This review highlights some of these studies that employ protein-engineering approaches for producing natural products and their derivatives. Specifically, it will cover the modifications of enzymes that are involved in the biosynthesis of the major classes of natural products—polyketides, nonribosomal peptides, terpenoids, and alkaloids. Domain shuffling and module rearrangement methods that have been employed in the past on collinear biosynthetic machineries will not be discussed in this review. The reader is directed to comprehensive reviews that cover these topics in detail [89].

Engineering polyketide synthases

Polyketide synthases (PKSs) are a family of enzymes responsible for the synthesis of a wide array of natural products, including the cholesterol-lowering lovastatin, the immunosuppressant rapamycin and the anti-bacterial tetracycline [34, 78]. The assembly of the polyketide backbone resembles that of fatty acid synthesis, with major differences in building block utilization, β-carbon reduction patterns and intramolecular cyclization [78]. The programming rules of the PKSs allow them to produce a far more diverse group of compounds relative to the fatty acid synthases (FASs). Understanding of rules governing the different types of PKSs has enabled the engineering of these enzymes towards generation of new compounds.

PKSs are often classified into three classes, loosely based on FAS nomenclature: a) type I PKSs are multi-domain megasynthases that act either modularly (as in some bacterial PKSs) [39] or iteratively (as in most fungal PKSs) [16] in order to catalyze successive polyketide chain extensions; b) type II PKSs, on the other hand, consist of discrete monofunctional proteins that are used repeatedly in producing the polyketide backbone [35]; and c) type III PKSs, most commonly found in plants, are homodimeric enzymes characterized by directly using coenzyme-A (CoA) as acyl carrier instead of a cognate carrier domain/protein and catalyzing the series of condensation reactions in a single multifunctional active site [6].

A minimum PKS unit is composed of a β-ketoacyl synthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP) (except in type III). The KS domain is responsible for the decarboxylative Claisen-like condensation between the growing polyketide and a carboxylated acyl building block to yield an elongated polyketide chain. In type II PKSs, the KS_α/KS_β heterodimer also dictates the length of the polyketide product [34]. The AT domain selects and transfers activated acyl group, typically from CoA thioesters, to the phosphopantetheinyl arm of the ACP; and therefore safeguards the identity of the polyketide building blocks [12]. In addition, other PKS enzymes or domains such as β-ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) are recruited in different combinations after each chain extension step to modify the β-keto positions [89]. In comparison to fatty acids, in which the β-positions are completely reduced to methylenes, polyketide backbones can be completely unreduced as in the case of aromatic polyketides [19], or only partially reduced in different permutations such as in the macrolides. An important strategy employed by different PKSs to generate structural diversity is via the incorporation of different starter and extender acyl units [12, 56]. This is attractive in creating structural analogues that are not observed in nature. Additionally, mutational studies aimed at altering the specificity of the chain elongation and reductive steps have also been successfully utilized to generate new derivatives of natural products, some of which have been found to have enhanced bioactivity.

Controlling chain length and substrate specificity

One of the key objectives of engineering PKS is to relax the machinery's stringency towards starter unit selection and to alter the product size, which is reflected in the polyketide chain length. Towards this end, Abe et al. performed mutagenesis studies on the active site residues of the type III PKS, aloesone synthase (ALS) from *Rheum palmatum*, based on homology modeling with the well-known chalcone synthase (CHS) [5]. Ferrer and his colleagues have previously solved the *Medicago sativa* CHS crystal structure, which guided the selection of ALS residues for mutation [24]. Despite high sequence similarity, these two enzymes are functionally distinct—ALS naturally selects an acetyl starter unit followed by six condensations with malonyl-CoA to produce the heptaketide aloesone, while CHS loads a bulkier 4-coumaroyl starter unit with three extensions to yield chalcone. Guided by the CHS model, some of the active-site residues in the ALS were mutated to the CHS counterparts, which resulted in production of new compounds with chalcone-like properties [5]. This study also provided insights into the respective roles of gatekeeping residues—Gly256 is important in the selection of the starter unit, Thr197 is involved in controlling the product chain length, while Ser38 is needed to guide the growing chain into the cavity.

Similarly, Abe et al. also mutated the *Aloe arborescens* pentaketide chromone PKS to produce aromatic polyketides SEK4 and SEK4b, which are octaketide shunt products in actinorhodin biosynthesis [4]. The crystal structures of the wild-type chromone synthase and the M207G mutant that synthesized the octaketides were solved, which showed numerous residues that control product chain length based on sterics [57]. Additional mutations of the bulky tyrosine and phenylalanine residues in the cavity

yielded a longer nonaketide naphthopyrone product (Fig. 1) [3]. Using a similar strategy, Klundt converted a benzophenone synthase (BHS) to a phenylpyrone synthase via the point mutation T135L. This mutation altered both the number of chain extension steps and the cyclization pattern of the resulting polyketide [42].

Among type II PKSs, which are constituted of discrete monofunctional enzymes, Tang et al. studied the Chain Length Factor (CLF) subunit in the KS/CLF heterodimer [80]. The KS/CLF catalyzes the iterative condensation of malonyl building blocks to yield an unreduced polyketide backbone. By aligning sequences of CLF involved in the synthesis of polyketides of different lengths (from C_{16} to C_{24}) and building a homology structure of the actinorhodin KS-CLF, four residues at the heterodimer interface were proposed to be involved in determining the chain length. These residues were noted to become progressively smaller in CLFs that are involved in the synthesis of longer polyketides, and were seen in the structure to define the size of the polyketide cavity. By changing bulkier residues (e.g., phenylalanine) found in C_{16}-specific CLF into smaller amino acids, C_{20} polyketides were preferentially produced, validating the role of CLF in controlling the number of decarboxylative condensations steps in type II PKS. The roles of these residues were subsequently confirmed by the X-ray crystal structure [38].

Incorporating different extender unit

The AT domains in PKSs have been engineered toward the biosynthesis of polyketides that contain alternative extender units. For example, Reeves et al. succeeded in changing the specificity of the fourth AT (AT4) in the type I modular PKS 6-deoxyerythronolide B synthase (DEBS) involved in erythromycin biosynthesis from its natural substrate methylmalonyl-CoA to malonyl-CoA [66]. By aligning sequences of ATs with different specificities and modeling to *Escherichia coli* FabD, three key regions in the active site cleft were identified to discriminate the different extender units. Site-directed mutagenesis on eight different residues in these regions, both individually and in combination, was performed, each time changing the methylmalonyl-specific residues into the malonyl-specific ones. The mutants were introduced into the heterologous production strain *Streptomyces lividans* and production of the expected 6-desmethyl analogue was observed. While AT swaps have been successfully employed on other DEBS modules to change the extender unit specificity [53], this strategy failed when performed on AT4.

Altering PKS reduction specificity

The reductive domains of PKSs, such as the KR and ER domains, have different regio and stereospecificities, thus further contributing to the structural diversity of partially-reduced polyketides. Engineering these domains through mutagenesis is therefore one strategy that can be utilized in the production of analogues with precise modifications in stereochemistry.

The KR domains of DEBS were modeled against the short chain dehydrogenase/reductase (SDR) superfamily by Reid et al. and a putative conserved triad of Tyr, Lys and Ser was identified [67]. Mutagenesis of these residues in KR6 of the truncated DEBS Module6 + TE (M6 + TE), followed by in vitro assay using a synthetic diketide precursor showed that the KR was completely inactivated. Subsequently, the same mutations were introduced into KR6 of the full-length DEBS expressed in *S. lividans*. Each of the mutations, which are K2664Q, S2686A and Y2699F, resulted in the production of the expected 3-deoxy-3-oxo-6-dEB, although only the Y2699F mutation fully inactivated the KR; both the K2664Q and S2686A mutations

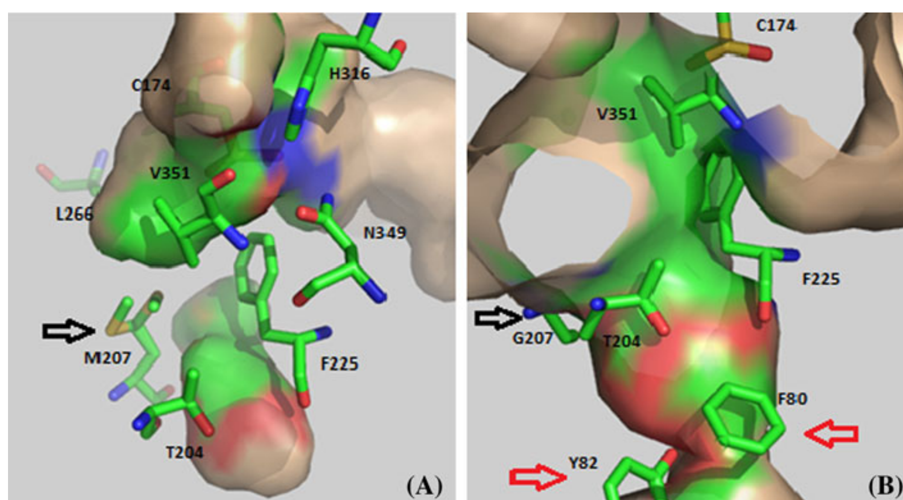

Fig. 1 Increasing type III PKS product chain length. **a** The active site cavity of the natural pentaketide chromone synthase (PDB 2D3M). It was shown that M207 dictates the cavity size; **b** the M207G mutation allowed entrance of the growing polyketide chain to the otherwise inaccessible cavity and led to the formation of the unnatural octaketide (PDB 2D52). Further mutations of the bulky Y82 and F80 residues found at the base of the cavity to alanine yielded the even larger nonaketide [3, 57]

resulted in mixed products. KR deletions have been performed in other DEBS modules that yielded the expected inactivation product [53]; however, in the case of KR6, it also affected the substrate specificity of the neighboring AT domain leaving the protein less efficient in producing the desired derivative. This is in contrast to the Y2699 mutation that exclusively produced the analogue and had better catalytic properties when tested in vitro.

Ding et al. used site-directed mutagenesis to inactivate the KR domain of a 6-methylsalicylic acid synthase (6-MSAS), converting it into a functional orsellinic acid synthase (OSAS) [21]. 6-MSAS differs from OSAS in that a functional KR domain is present. By homology modeling, a conserved tyrosine residue that is part of the catalytic triad was mutated to phenylalanine in the 6-MSAS ChlB1 from the spirotetronate pathway. The host strain bearing the Y1540F ChlB1 mutation was found to accumulate orsellinic acid, demonstrating inactivation of KR did not affect the remaining activities of the PKS. The engineered ChlB1 was reintroduced into the spirotetronate biosynthetic pathway, and the downstream enzymes were able to incorporate the orsellinate into a new spirotetronate analogue with similar bioactivity as the original antibacterial agent.

In addition to inactivation studies, the KR domains have also been engineered to produce analogues with altered stereochemistry. This was demonstrated by Leadlay and coworkers in successfully switching the stereospecificity of DEBS KR1 [7]. Using the synthetic substrate 2-methyl-3-oxopentanoyl-S-*N*-acetylcysteamine (SNAC), the wild-type KR1 was shown to have strict stereospecificity in the reduction to afford only the (*2S,3R*)-2-methyl-3-hydroxyl-pentanoyl product. Comparing KR1 to the downstream KR2 known to have the opposite stereospecificity, non-conserved residues that lie in the active sites were identified. In an accompanying study by O'Hare et al., saturation mutagenesis was performed on each of these residues and a library of KR1 mutants was constructed for product screening [61]. In order to select for active enzymes from this library and narrow down the number of mutants, a high-throughput spectrophotometric method that detected the reduction of *trans*-1 decalone was performed. The double mutant F141W/P144G was found to completely switch the stereospecificity and to produce the (*2S,3S*) product exclusively.

Concerning the stereospecificity of the ER domain, Kwan et al. identified a tyrosine residue that can possibly influence the stereochemistry of the α-methylated product, which originates from a methylmalonyl extender unit [46]. From surveying different ER domains with different stereospecificity, a correlation between the presence of a tyrosine residue in the active site and an *S* α-methyl-containing product was found. Substitution of the tyrosine with other residues, such as the prevalently found valine, resulted in a product with *R* configuration. To confirm the role of the tyrosine, module 4 ER of DEBS and module 13 ER of rapamycin synthase (RAPS), which have opposite stereospecificities, were studied. The tyrosine to valine mutation in DEBS ER4 switched the stereospecificity of enoylreduction and afforded the product in which the corresponding methyl is in the *R* configuration. However, the valine to tyrosine mutation in RAPS ER13 was not able to reverse the stereospecificity, suggesting additional residues may be involved. Probing deeper into this subject, Kwan performed mutagenesis on other active site residues in RAPS ER13 and examined individually and in combination the effect on the stereospecificity [45]. When the combination of mutations V46L, V47I, V52Y, and N52P were introduced to match those in DEBS ER4, the desired product with *S* stereochemistry was observed, albeit in low levels.

Producing polyketide analogues with enhanced bioactivity

KR inactivation proved useful in producing analogues of the antifungal polyketide amphotericin B, produced by a type I modular PKS. Using site-directed mutagenesis, the KR domains of module 12 and module 16 were both inactivated by site-directed mutagenesis of the conserved tyrosine residues (Y720F and Y6165F, respectively) [64]. The 16-descarboxyl-16-methyl-amphotericin produced by KR12 inactivation maintained the same antifungal activity as the amphotericin B, while the oxo-derivative from KR16-inactivated PKS was shown to have better activity, enhanced solubility and reduced hemolytic properties. Brautaset et al. performed numerous site-directed mutagenesis studies to investigate the structure–activity-toxicity correlations of nystatin analogs generated from the inactivation of a) *nysJ* DH15 responsible for the dehydration of C-10 hydroxyl bond; b) *nysN*, a P450 that putatively oxidizes the C16 methyl; c) *nysJ* KR16 and KR17 responsible for the reduction at C5 and C7, respectively; and d) some combinations of the above [10]. The nystatin analogue BSG020 (16-decarboxy-16-methyl-28,29-didehydronystatin) displayed enhanced activity over amphotericin B. More recently, the same group performed additional biosynthetic engineering of nystatin analogues to further elucidate structure-to-activity correlations, especially of the polyol region [9]. Double and triple inactivation experiments involving ER5, DH15, *nysN* and the other P450 monooxygenase *nsyL* were performed to generate four novel analogues. Notably, the authors found that the C16 methyl that resulted from *nysL* inactivation was beneficial in improving the activity and in lowering the hemolytic activity.

The same technique was utilized by Kim et al. in producing analogues of geldanamycin, an Hsp90 inhibitor that is a potential chemotherapeutic agent [40]. By inactivation

of DH1 via site-directed mutagenesis at the His2041 residue in the seven-module type I PKS, the expected C15-hydroxylated analogue was detected, along with other derivatives that resulted from an intramolecular reaction involving the C15 hydroxyl moiety. This demonstrated that the downstream modules and the tailoring enzymes were tolerant of this structural change originating from the first module. Further mutagenesis of tailoring enzymes, including the monooxygenase Gel7 and the *O*-carbamoyl transferase Gel8, yielded more analogues of geldanamycin, among which one compound exhibited a near fivefold improvement in Hsp90 ATPase activity.

Engineering non-ribosomal peptide synthetases

Nonribosomal peptide synthetases (NRPSs) are large multi-modular enzymes that are capable of catalyzing the formation of peptide bond using aminoacyl adenylates [73, 74]. NRPSs are responsible for the biosynthesis of important bioactive molecules such as penicillin, cyclosporins and vancomycin [70]. Each module of the NRPS has three basic domains: the adenylation domain (A), the thiolation domain (T) and the condensation domain (C) [73]. The adenylation domain activates the amino acid or carboxylic acid that is used as monomers of the NRPS product. The activated substrate is then tethered to the thiolation domain via a phosphopantetheinyl arm. The condensation domain subsequently catalyzes peptide bond formation between the tethered monomeric unit and the growing peptide chain.

Because of the collinear nature of an NRPS assembly line, inactivation of individual domains in the large NRPS can result in the predictive synthesis of unnatural analogues. For example, inactivating the methyltransferase (MT) domain of PchF of the pyochelin synthetase from *Pseudomonas aeruginosa* resulted in the formation of desmethyl-pyochelin [62]. Similarly, using an engineered *E. coli* strain that produced echinomycin [87], Watanabe and coworkers inactivated the MT domain of Ecm7 of the echinomycin NRPS by mutating the conserved glycine residues. When expressed in *E. coli,* the mutant produced the expected analogue des-*N*-tetramethyl-triostin A (TANDEM) (Fig. 2a) [88]. However, the yield of TANDEM was lower compared to that of triostin A by the wild-type enzyme.

Engineering adenylation domains

The adenylation domain serves as the “gate-keeper” of the NRPS by selectively activating a specific amino acid to form an aminoacyl adenylate using ATP [73]. As such, deciphering the selectivity-conferring code of adenylation domains can provide opportunities to generate novel bioactive molecules. A key step in achieving this goal is the

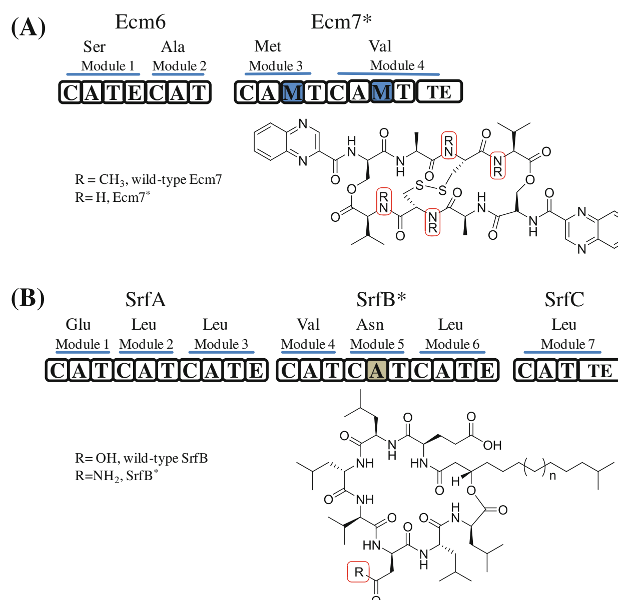

Fig. 2 Protein Engineering of NRPS to produce analogues. **a** Inactivation of the methyltransferase domain in Ecm7 led to the production of des-*N*-tetramethyl triostin A by the echinomycin NRPS complex [88]. **b** The H322E mutation in SrfB switched the specificity of the adenylation domain of Module 5 from L-aspartate to L-asparagine. The mutant surfactin NRPS complex was able to synthesize the corresponding surfactin analogue containing L-asparagine [22]

acquisition of the crystal structure of PheA, the phenylalanine-activating A domain of GrsA, complexed with phenylalanine and AMP [15]. Using the crystal structure and sequence alignment of the residues in A4-A5 core motifs of other adenylation domains, Stachelhaus et al. identified ten residues that are in the substrate-binding site of PheA. These ten residues, proposed to be the selectivity-conferring residues, are conserved among bacterial adenylation domains that activate the same amino acid [75]. Using the data obtained from the sequence alignment and crystal structure, they converted the selectivity of PheA from phenylalanine to leucine via two amino acid substitutions (T278M/A301G). In addition, they also successfully converted an aspartate-activating domain from the surfactin NRPS to an asparagine-activating domain [75].

Eppelmann and others were able to take this goal one step further by successfully biosynthesizing a full-length peptide from an NRPS with an adenylation domain that possess altered specificity [22]. First, they altered the specificity of the loading module of the surfactin synthetase of *Bacillus subtilis* from L-glutamate to L-glutamine. In addition, a novel lipoheptapeptide derivative of surfactin with an L-asparagine instead of L-aspartate in position 5 of the peptide was generated (Fig. 2b).

However, changing the amino-acid specificity of A domains seemed to have unexpected effects on overall NRPS function. Uguru and coworkers attempted to change

the specificity of the A domain of module 7 of the calcium-dependent antibiotics (CDA) peptide synthetase in *Streptomyces coelicolor* A3(2) from L-aspartate to L-asparagine [81]. While the mutant CDA peptide synthetase was able to synthesize the desired CDA derivative, the yield was considerably lower compared that of the wild-type. Additionally, a majority of the products of the mutant enzyme underwent hydrolysis to form a linear peptide instead, suggesting the terminal cyclization domain has unexpectedly high substrate specificity.

In light of the new knowledge from structural and mutational studies of PheA, other groups have attempted to further characterize the selectivity code of adenylation domain. Challis et al. performed an extensive analysis of most of the adenylation domains of known NRPS's to characterize the binding pockets [11]. Using PheA as the model for substrate-binding studies, they modeled the interactions of the specificity-conferring residues to different amino acid substrates. This study attempted to predict the selectivity of adenylation domains of unknown function, as well as the optimum mutations required to switch the specificity. From a computational design perspective, Lilien and coworkers applied the K* algorithm in redesigning the selectivity of PheA from phenylalanine to leucine [49]. An improved K* algorithm was later used to improve the selectivity of PheA towards amino acids with charged side chains, such as arginine, lysine, glutamate and aspartate [13].

Recently, directed evolution approaches were applied to AdmK, which incorporates a valine in the biosynthesis of andrimid, in order to create new andrimid analogues [23]. The mutant library was constructed via saturation mutagenesis of three residues in the A domain of AdmK. The mutants were then expressed in the native host *Pantoea agglomerans* that lacked the wild-type *admK* and were screened using LC-MS/MS. Antibiotic activities of the andrimid analogues, including substitutions containing isoleucine, leucine, alanine and phenylalanine, were determined against *E. coli*. Interestingly, a random R235K mutation introduced during PCR amplification resulted in higher activity and solubility of the NRPS. This residue is predicted to be located in the solvent-exposed portion of the adenylation domain. The authors suggested that residues outside of the selectivity-conferring region could be considered for future directed evolution studies for NRPSs.

Improving yield of chimeric NRPS

Domain swapping and module swapping were considered to be of high potential towards the combinatorial biosynthesis of novel NRPS compounds. In fact, derivatives of surfactin were obtained from the adenylation domain swapping [76], while module swapping was used to successfully synthesize daptomycin derivatives [60]. However, chimeric NRPSs typically have lower activities compared to their wild-type counterparts due to nonnative protein–protein interactions. Directed evolution was used to improve the activity of a chimeric EntF when the adenylation domain was swapped with a syringomycin synthetase adenylation domain [25]. After two rounds of directed evolution using mutagenic PCR, a mutant with ~26-fold improvement in product yield compared to the original chimeric EntF was isolated. The same approach was applied towards the andrimid hybrid NRPS-PKS, in which the adenylation domain of AdmK was replaced with the A domain of CytC1 from *Streptomyces* sp. RK95-74 that is specific for 2-aminobutyrate. The protein engineering efforts afforded a mutant with significantly increased activity.

Probing modular interactions

Communication-mediating (COM) domains bridge the interactions between the different modules of NRPS. Each donor COM domain recognizes a specific acceptor COM domain [32]. Hahn and coworkers investigated the residues that specify the interactions between the acceptor COM domain of TycC1 (COM^{A}_{TycC1}) with the donor COM domains of TycA (COM^{D}_{TycA}) and noncognate TycB3 (COM^{D}_{TycB3}) [33]. The K9D mutation in COM^{A}_{TycC} abolished its interaction with COM^{D}_{TycB3} while gaining the ability to interact with COM^{D}_{TycA}. In addition, the cognate interaction between COM^{D}_{TycA} and COM^{A}_{TycB1} was exploited to establish cross-talking between TycA and BacB2 from the bacitracin NRPS. BacB2 was in turn engineered to interact with the SrfAC via fusion with a C-terminal COM^{D}_{TycA}. This artificial NRPS linked by the Universal Communication System (UCS) was able to produce the expected L-Phe-D-Orn-L-Leu tripeptide.

In addition, Lai and coworkers mapped the residues that allow the interaction of the thiolation domain of EntB with its downstream interacting module EntF in the biosynthesis of enterobactin [47]. Based on structural homology, the T domain surface that interacts with EntB was predicted to be in helix 2 and loop 2/helix 3 region. Mutation of the residues in the region revealed that Phe264 and Ala268 found near the phosphopantetheinylated serine are the most essential residues in the interaction. Two adjacent residues, Met249 and Lys269, were also found to be of moderate importance to the interaction.

Engineering terpene synthases

Terpenes encompass a diverse array of compounds of medical and industrial importance such as cholesterol, the plant pigment lycopene, the antimalarial drug artemisinin

and the essential oil 4*S*-limonene [20, 84]. These compounds share five-carbon dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP) as precursors. DMAPP and IPP undergo condensation to form longer isoprenoid units such as geranyl (C_{10}) or farnesyl (C_{15}) diphosphate [20]. Typically, longer isoprenoid units undergo cyclization [30] with the notable exception being the linear C_{40} isoprenoid lycopene. The primary goal of protein engineering studies here is to change the cyclization regioselectivity of the enzymes and increasing the product yield.

Changing the cyclization regioselectivity

Monoterpenes

Controlling the cyclization selectivity of monoterpene synthases by protein engineering was shown by Hyatt and others in changing the specificity of (–)-pinene synthase from *Abies grandis* to (–)-camphene synthase [36]. While mutations of twelve residues led to a significant decrease in product selectivity towards *β*-pinene, the mutants displayed promiscuity in the synthesis of other products such as *α*-pinene and camphene.

In order to obtain an enzyme with a better product selectivity, Kampranis and coworkers obtained the structure of Sf-CinS1, the 1,8-cineole synthase from *Salvia fruticosa* [37]. Using sequence alignment with other plant monoterpene synthase as well as comparison with the structure of bornyl pyrophosphate synthase [90] and domain swapping experiments between the two enzymes [63], they were able to identify the residues that contribute to the cyclization specificity of Sf-CinS1. Mutation of five residues in Sf-CinS1 completely abolished the production of 1,8-cineole while making sabinene the major product [37]. Asn338 in Sf-CinS1, in particular, contributes to the selectivity of Sf-Cin-S1 towards 1,8-cineole. The structural data from the study of Sf-CinS1 was also utilized by Kohzaki and coworkers to change the selectivity of sabinene synthase from rough lemons [43]. Based on the structural data of Sf-CinS1, Ile338 was mutated to asparagine, which resulted in the increased production of *β*-pinene.

Sesquiterpene

Sesquiterpenes synthases, in comparison to other terpenes synthases, have been the most frequent targets for protein engineering due to the availability of several X-ray structures such as the tobacco 5-*epi*-aristolochene synthase (TEAS) [77], aristolochene synthase from *Aspergillus terreus* [72], and *epi*-isozizaene synthase from *Streptomyces coelicolor* A3(2) [1]. TEAS, in particular, played an important role in modeling the mode of substrate binding for engineering the *Hyoscyamus muticus* premnaspirodiene synthase (HPS) [31], cotton (+)-*δ*-cadinene synthase [101], and *γ*-humulene synthase from *A. grandis* [100].

Aaron and coworkers, using the X-ray structure of *epi*-isozizaene synthase (EIZS), identified the residues necessary for controlling cyclization specificity [1]. The structure of EIZS bound with a substrate homolog revealed that the carbocation intermediate is stabilized by cation-*π* interaction with aromatic residues in the binding pocket. Mutation of these aromatic residues led to the production of alternatively cyclized sesquiterpenes. For instance, changing Phe198 to alanine resulted in a shift of selectivity towards *β*-farnesene and *β*-acoradiene. Meanwhile, the W203F mutation caused the production of *Z*-*γ*-bisabolene, a sesquiterpene not produced by the wild-type enzyme. Additionally, the F198A mutant led to a more promiscuous distribution of product [1].

In the absence of structural data from the enzyme, successful manipulation of product selectivity was still achieved for enzymes that use a homologous protein as a guide. For example, HPS was converted to *epi*-aristolochene synthase by mutation of eight residues predicted to be in the binding pocket using *epi*-aristolochene synthase as a model structure [31]. TEAS was also converted to premnaspirodiene synthase by nine amino acid substitutions, albeit with lower selectivity compared to wild-type HPS.

Another example of rational design based on homolog modeling is the engineering of *γ*-humulene synthase (GHS) from *A. grandis* [100]. Wild-type GHS, known to have an extremely promiscuous product profile, synthesizes 20 different sesquiterpene products, among them are *α*-longipinene, siberene and *γ*-humulene [79]. Thus, it is an attractive target to test whether a promiscuous terpene cyclase can be engineered to have more product selectivity. Using TEAS as the model, Yoshikuni and coworkers identified 19 putative "plasticity residues" that were targeted for saturation mutagenesis. Surprisingly, the mutation of one amino acid residue in GHS (W315P) was sufficient to dramatically increase the selectivity of the enzyme and to create an *E*-*β*-farnesene/*Z*-E-*α*-farnesene synthase [100]. Several other new sesquiterpene synthases were isolated using this approach that have a more selective product profile, such as *β*-bisabolene synthase, longifolene synthase and siberene synthase. Moreover, a mutant with more selectivity toward *γ*-humulene compared to the wild-type was also isolated. This study offered a glimpse of the potential for improving or completely altering the selectivity of promiscuous cyclases.

Another work by Yoshikuni and coworkers utilized both rational design and directed evolution to convert the

selectivity of (+)-δ-cadinene synthase towards germacrene D-4-ol (Fig. 3) [101]. Mutant synthases were fused to a C-terminal chloramphenicol acyltransferase (CAT) to select against mutations that cause protein insolubility. The preliminary selection eliminated a large number of inactive mutants. The remaining mutants were then screened using GC–MS for an increased germacrene-D-4-ol to (+)-δ-cadinene ratio. After two rounds of error-prone PCR, the double mutant N403K/L405S has 93% product selectivity towards germacrene-D-4-ol compared to ~0.1% in the wild-type enzyme. To explain the switching of product selectivity towards germacrene-D-4-ol, Yoshikuni and coworkers modeled the binding of farnesyl hydroxyl phosphate (FHP) in the enzyme and revealed Phe400, Asn403, Leu405 play important roles in protecting the substrate carbocation from solvent exposure.

Diterpenes

Most of the investigation in the cyclization mechanism of diterpene synthase, in particular those that utilize copalyl diphosphate as an intermediate, was done by the Peters group. Surprisingly, they were able to switch the cyclization pattern of copalyl diphosphate by mutating a threonine or serine residue in six different diterpene cyclases [58, 91, 97]. In all of these cases, the presence of a polar threonine or serine in the binding pocket stabilizes the cation and prevents further cyclization or rearrangement of the diterpene. OsKSL5i, an iso-kaur-15-ene synthase from *Oryza sativa*, was converted to an *ent*-primardiene synthase by the

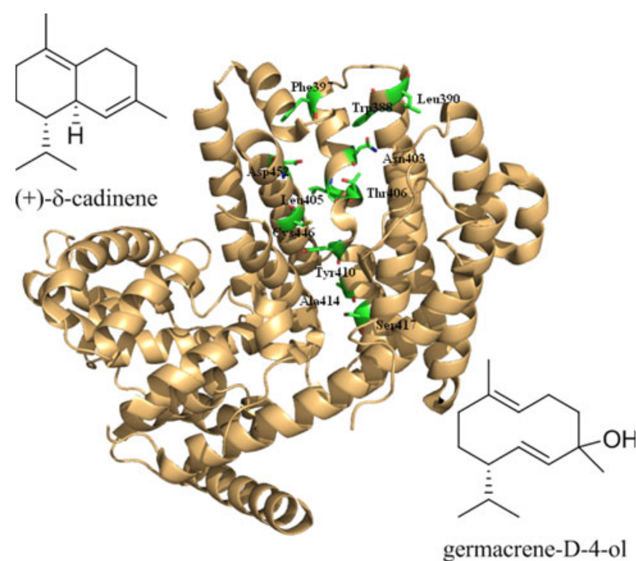

Fig. 3 Crystal structure of cotton (+)-δ-cadinene synthase (PDB 3G4F). Yoshikuni et al. re-engineered (+)-δ-cadinene synthase to germacrene-D-4-ol synthase [101]. The residues that were mutated over the course of directed evolution are labeled, most of which are located in the G-helix region between Phe397 and Ser417

T664I mutation. In addition, the *ent*-pimara-8(14),15-diene synthase OsKSL5j displayed a broader product profile upon T661I mutation and was able to produce isokaur-15-ene, atiser-16-ene and kaur-16-ene in a 50:37:13 ratio [97]. Similarly, switching Ala723 to serine in *A. grandis* abietadiene synthase (AgAs) resulted in the "short-circuiting" of the cyclization reaction, preventing the conversion of isopimar-15-en-8-yl cation to isopimar-8(14)-en-15-yl cation. Primardienes, the products that go through the isopimar-15-en-8-yl cation but not isopimar-8(14)-en-15-yl cation as an intermediate, are therefore the only diterpenes produced by the A723S AgAs mutant [91]. Moreover, the T696I mutation in *syn*-pimara-7,15-diene synthase (OsKSL4) from *Oryza sativa* resulted in the production of the four-ringed aphidicol-15-ene [58].

Sterols

Sterols, formed from the cyclization of triterpenes, are comprised of steroid hormones and cellular membrane components such as cholesterol. Studies done in changing the specificity of sterol cyclase include work by Dang and Prestwich, where they mutated the ^{376}DDTAVV^{381} motif in squalene-hopene cyclase (SHC) to ^{376}DCTAEA^{381} [18]. While the mutant was unable to convert squalene to hopene, it can cyclize 2,3-oxido-squalene to the pentacyclic hop-22(29)-en-3α-ol and hop-22(29)-en-3β-ol. Moreover, the mutant also produced the monocylic 3α-achilleol A and achilleol. Füll and others mutated the conserved tyrosine residues in the binding site of SHC of *Alicyclobacillus acidocaldarius* and terminated the later cyclization steps to form bicyclic triterpenes α-polypodatetraene and γ-polypodatetraene [27].

The Matsuda group utilized directed evolution to convert the cycloartenol synthase from *Arabidopsis thaliana* (*Ath*CAS1) to a lanosterol synthase [50, 54, 71]. The lanosterol synthase mutants were screened through the complementation of yeast with a non-functional lanosterol synthase [54]. To further improve the selectivity of the *Ath*CAS1 H477N mutant isolated from the directed evolution study, a H477N/I481V double mutant was constructed. The resulting mutant can cyclize 2,3-oxidosqualene to lanosterol with essentially no side products [50]. In addition to a lanosterol synthase, they also changed *Ath*CAS1 to a parkeol synthase through a H477Q mutation [71]. Similarly, protostadienol synthase from *Aspergillus fumigatus* was converted to a lanosterol synthase by conversion of ^{702}APPGGMR^{708} to ^{702}NKSCAIS^{708} [41]. As for the *Saccharomyces cerevisiae* oxidosqualene-lanosterol cyclase, a single amino acid mutation (F699T) was sufficient to convert the enzyme to a protosta-13(17),24-dien-3β-ol cyclase [94]. For more information regarding this subject, we recommend the review on triterpene cyclization by Abe [2].

Carotenoids

An extensive review has been published by Umeno and others on the studies in changing the product specificity of carotenoid biosynthetic enzymes [83]. Color assays of produced pigments were usually used to identify mutant enzymes. This enabled the high throughput screening of mutant libraries constructed by random mutagenesis [82, 86].

Umeno, using directed evolution, changed the number of desaturation steps in a C_{35} carotenoid pathway constructed in *E. coli* [82]. Using the same approach, phytoene desaturase from *Rhodobacter sphaeroides* was mutated to catalyze an additional desaturation step to form lycopene [85]. Another method of diversifying carotenoid production in heterologous host is by changing the number of cyclizations in the carotenoids. Using gene shuffling, Schmidt-Dannert and others constructed a chimeric phythoene desaturase that produced a linear 3,4,3′,4′-tetradehydrolycopene, as well as the cyclic torulene [69]. Moreover, an L488H mutation in *A. thaliana* lycopene ε-cyclase resulted in the formation of two epsilon rings [17].

Improving product yield

An approach to improve the production of terpenoid is by protein engineering of enzymes responsible for biosynthesis of common isoprenoid precursors such as geranylgeranyl diphosphate. Wang et al. used error-prone PCR and staggered extension process to create a mutant library of geranylgeranyl diphosphate synthase (GPPS) [86]. Using visual inspection and UV-Vis spectroscopy, mutants that have increased GGPPS activity and lycopene production levels were isolated. In contrast, Mijts and coworkers engineered the C_{30} carotenoid oxygenase from *Staphylococcus aureus* using directed evolution to increase the production of oxygenated linear C_{30} and C_{40} carotenoids [55].

While directed evolution is the preferred method to improve enzyme activity in terpenoid biosynthesis, Yoshikuni and coworkers developed a more direct approach to change the number of glycine and proline residues in 3-hydroxy-3-methylglutaryl-CoA reductase (tHMGR) to enhance the synthesis of mevalonate, a precursor to IPP and DMAPP [99]. By sequence alignment to related enzymes, they identified conserved glycine and proline residues and switched the non-conforming amino acids to match the consensus sequence. Similarly, the glycines and prolines that did not align were altered via saturation mutagenesis at that site. The resulting tHMGR-G9 with nine mutated residues displayed 2.5- to 3-fold increases in production of mevalonate compared to the wild-type tHMGR.

Leonard and others, meanwhile, used a combination of metabolic engineering and protein engineering to increase levopimaradiene production in an engineered *E. coli* strain [48]. The levopimaradiene synthase substrate-binding cavity was modeled based on the structure of paralogous LPS-type enzymes such as abietadiene synthase from *A. grandis* and *Picea abies* and isopimardiene synthase from *P. abies*. Moreover, the activities of GGPPS and levopimaradiene synthase were improved by 1.7- and 5-fold respectively using directed evolution [48].

Engineering alkaloid synthases

Monoterpenoid-derived indole alkaloids such as camptothecin and vinblastine are among the largest and most pharmacologically diverse alkaloid families in plants [44]. Strictosidine synthase, which catalyzes the Pictet-Spengler condensation between tryptamine and secologanin, is involved in the biosynthesis of important indole alkaloids such as reserpine, camptothecin and vinblastine [98]. The Stöckigt group solved the structures of both the native strictosidine synthase [52] and its substrate-bound derivative [51] in order to determine the key residues that control substrate specificity. Subsequent mutation of Val208 to alanine in *Rauvolfia serpentina* strictosidine synthase relaxed the specificity of the enzyme while maintaining its enantioselectivity. The mutant was able to convert substituted tryptamines such as 5-methyltryptamine, 5-fluorotryptamine, and 6-fluorotryptamine, together with secologanin into substituted strictosidine analogues. Bernhardt and coworkers developed a method to rapidly identify mutants with relaxed substrate selectivity in *Catharanthus roseus* strictosidine synthase [8]. They first developed a colorimetric assay that reported enzyme activities with alternate substrates. A N-terminal yeast mating α signal sequence was then fused to the strictosidine synthase to facilitate the transport of the enzyme to the culture media. With this in place, a mutant library generated from the saturation mutagenesis of Val171, Val214, Phe232, and Glu31 was screened, leading to the isolation of the V214M mutant that accepts 5-bromotyptamine (Fig. 4). In addition, they also found that a F232L mutant that utilizes (2′R)-tryptophanol as substrate. Chen and others modeled the binding of secologanin to *C. roseus* strictosidine synthase to predict the residues that control the selectivity of the enzyme to secologanin (Fig. 4) [14]. The D177A mutant was able to accept derivatives of secologanin, albeit with lower selectivity compared to the natural substrate.

Engineering tailoring enzymes

The diversity of natural products synthesized by the different classes of biosynthetic machineries is further increased by the action of tailoring enzymes. These

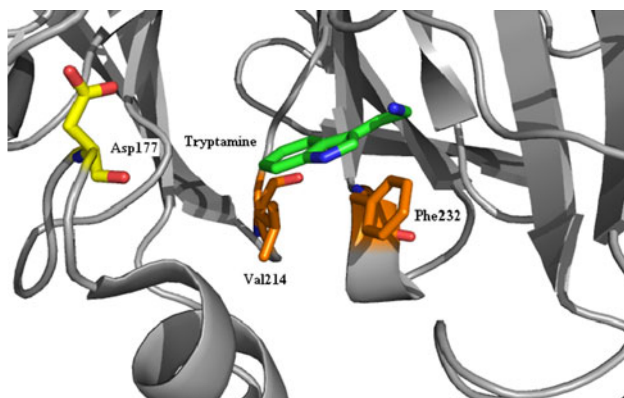

Fig. 4 Protein engineering of the *Catharanthus roseus* strictosidine synthase (PDB 2VAQ). Active site of the enzyme is shown here with the labeled residues mutated to increase substrate promiscuity. Phe232 and Val214 mutations enabled the enzyme to accept tryptamine derivatives [8], while the Asp177 mutation led to the acceptance of secologanin derivatives [14]

enzymes decorate the natural product scaffold with various functional groups that are often essential for the observed bioactivity of the final compounds. Enzymes such as acyltransferases, glycosyltransferases, and oxygenases are frequently found in various combinations in a biosynthetic pathway. Protein-engineering approaches have been applied to these enzymes to either improve the catalytic activities or more importantly, broaden the substrate specificity towards unnatural natural products [102].

Acyltransferases

Acyltransferases catalyze the transfer of an acyl group from a donor substrate to a nucleophilic site (for example, amine or hydroxyl) on an accepting substrate. Oftentimes, it is desirable to express and purify acyltransferases from a heterologous host in order to investigate their functionality. One common problem hampering these studies is the lack of soluble and functional protein, which can result from unfavorable interactions between surface-exposed residues and water. In studying the specificity of the plant 2-*O*-benzoyltransferase from *Taxus cuspidate* towards different acyl substrates, Nawaranthe et al. first improved the solubility of the protein, which was initially expressed as non-functional inclusion bodies [59]. By homology modeling to other functionally defined AT orthologoues and to the solved crystal structure of a related vinorine synthase, they identified two key residues, Glu19 and Asn23, that deviated from the consensus sequence and were also found to lie in the exposed regions of the protein. Using site-directed mutagenesis to incorporate the Q19P and N23K mutations, they obtained a fivefold improvement in its solubility, which enabled them to further probe the specificity of this acyltransferase.

Besides solubility, protein engineering has also been performed to improve the catalytic activity of acyltransferases, as in the case of LovD, an enzyme involved in lovastatin biosynthesis [29]. LovD was previously shown to mediate the transfer of the acyl group from the lovastatin diketide synthase LovF, as well as from the small molecule substrate α-dimethylbutyryl-*S*-methyl-mercaptopropionate, to monacolin J, and yielding either lovastatin or the semisynthetic simvastatin [96]. Gao and coworkers used multiple rounds of directed evolution via error-prone PCR and saturation mutagenesis to enhance its catalytic efficiency, thermal stability and solubility. They coupled these methods to a screen based on an agar-diffusion assay, where simvastatin production inhibited the growth of embedded spores of *Neurospora crassa.* After seven rounds of directed evolution, a mutant that displayed 11-fold increase in whole-cell activity in simvastatin synthesis was recovered. A fifth-round mutant was crystallized and the X-ray structure was compared to that of the wild-type LovD. It was proposed that mutations led to conformation changes of the enzyme and stabilized the active site in a configuration that is more favorable for catalysis.

Glycosyltransferase

The ligation of deoxysugars to the natural product aglycon is an essential step in introducing bioactivity to the final molecules. Therefore the corresponding glycosyltransferases are valuable in the synthesis of analogues containing altered glycosylation patterns. Most of the studies are focused on expanding the substrate specificities of the enzymes to be more promiscuous. A prime example is the expansion of both the donor and acceptor specificities of the oleandomycin glycosyltransferase, OleD, that naturally transfers uridine diphosphoglucose (UDP-Glc) to the oleandomycin aglycon to form the glucosylated product [93]. By performing several rounds of directed evolution, a "universal glycosyltransferase" that can transfer fifteen different deoxysugars onto eight different acceptor molecules, including the natural substrate oleandomycin, 4-methylumbelliferone and novobiocic acid, etc., was isolated (Fig. 5). This paved the way for glycorandomization of several natural products via pairing of different deoxysugars and aglycons. The various monosaccharide substrates that were accepted by the P67T/A242V/S132F triple mutant are shown in Fig. 5b. Gantt and coworkers further probed the aglycon specificity of these OleD mutants by testing 137 different drug-like acceptors, more than half of which were glycosylated [28].

In a more focused approach, Williams and his coworkers isolated an OleD mutant that specifically and efficiently glycosylates novobiocic acid to produce novobiocin [92]. Using "hot spot" saturation mutagenesis on the sequence

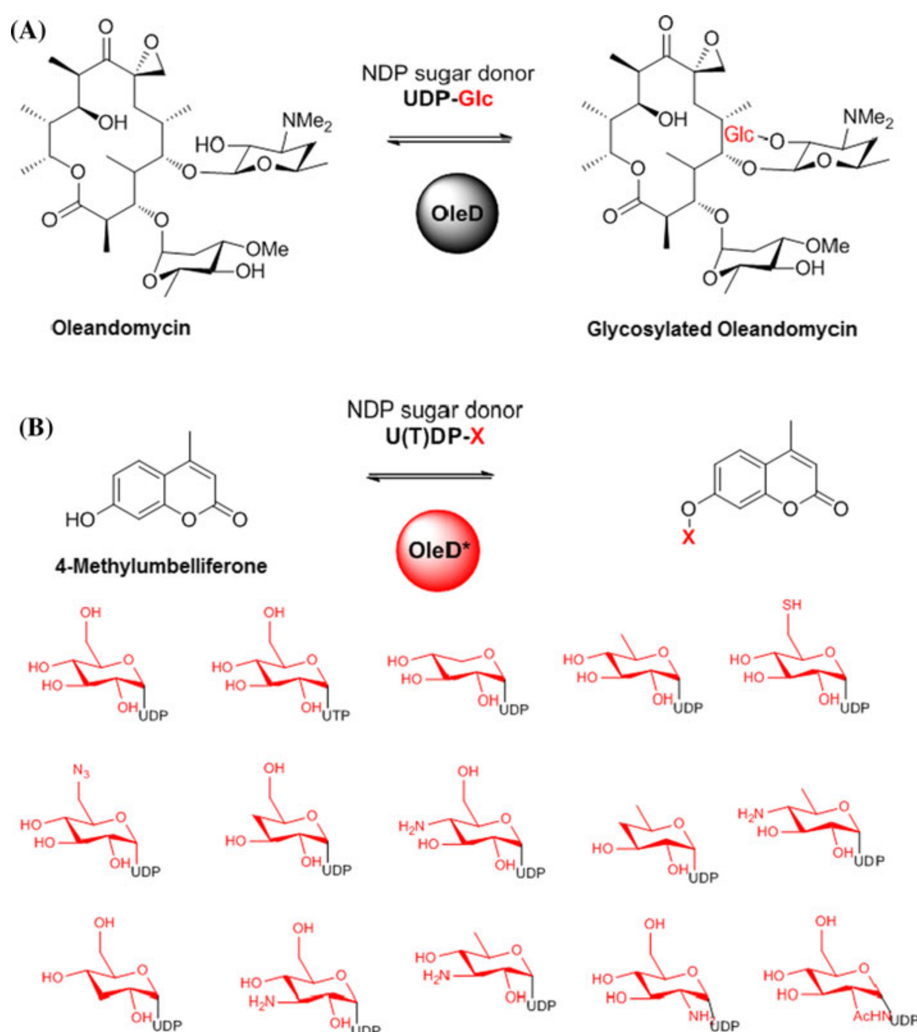

Fig. 5 Protein engineering of a "universal" glycosyltransferase. **a** The wild-type OleD catalyzes the transfer of glucose from UDP-Glu to oleandomycin. **b** A triple mutant P67T/S132F/A242V, OleD*, isolated via directed evolution using the fluorescent 4-methylumbelliferone as the aglycon acceptor was able to transfer a wide array of sugar substrates (denoted as U(T)DP-X) [93]

previously identified to be involved in the acceptor promiscuity, they obtained a glycosyltransferase mutant with several 100-fold improvement in catalytic activity. In a similar manner, Ramos and coworkers mutated several residues in the elloramycin glycosyltransferase, a moderately promiscuous enzyme, in order to make it more specific towards glycosylation of the important antitumor drug, elloramycin [65].

Oxygenases

Oxidative modification on natural products is another important tailoring step, and can lead to drastic changes in product polarity and structure, as well as installing sites for further enzymatic modifications such as acylation and glycosylation. Nature uses different types of oxygenases, including flavin-dependent, cytochrome P450, and α-keto-glutarate/iron-dependent enzymes, to catalyze a variety of redox modifications, such as hydroxylation, epoxidation and Baeyer–Villiger oxidative cleavages, etc. [68]. Several cytochrome P450 oxygenases have been engineered to date, the main goals of which are to process a wider array of substrates, to enhance the catalytic activity and to alter the enantioselectivity. Xiang et al. succeeded in relaxing the substrate specificity of the EryF P450 by mutation of Ala245 to the highly conserved threonine found among its homologs [95]. The conserved threonine residue among this family P450s is believed to participate in the O–O bond scission step during catalysis, and its absence in EryF is compensated by the 5-hydroxl group in the substrate. With the A245T mutation, this strict substrate constraint is eliminated and conferred EryF promiscuity towards a much wider array of substrates. For example, the mutant catalyzed the efficient oxidation of testosterone in the 1-, 11α, 12- and 16α positions, activities that were not observed with the wild-type.

Zocher and his coworkers solved the X-ray crystal structures of four isoforms of the P450 AurH involved in the synthesis of aureothin [103]. AurH catalyzes the two-step reaction including an initial hydroxylation at C7,

followed by heterocyclization to yield the tetrahydrofuran ring. Structural insights guided mutational studies to characterize residues involved in the stereospecific oxidation reaction. By analyzing the "open" and "closed" (inhibitor-bound) conformations of the enzyme, as well as modeling substrate docking, Ser66 and Asn91 were predicted and confirmed to undergo conformational changes to accommodate the binding of both deoxy and hydroxyaureothin substrates to the active site. Interestingly, in an effort to understand the roles of other residues lining the binding site, two mutants, F89W and T239F were found to have altered regiospecificity. The mutants selectively oxidized the C9a methyl instead of the natural C7 methyl into a carboxylic acid.

Future outlook

Protein engineering has been an extremely powerful tool in the field of metabolic engineering and biocatalysis. Its utility in natural product biosynthesis towards production of unnatural compounds is emerging and will undoubtedly expand. With the accumulating wealth of genomic, structural and biochemical information on natural product biosynthetic pathways from different organisms, the use of protein engineering, in combination with other tools such as genome mining and synthetic biology, will improve the prospects of combinatorial biosynthesis in green chemistry and drug discovery.

Acknowledgments Due to space constraints, we apologize for any work that we were not able to cite. Related projects in our laboratory are supported by the NIH, David and Lucile Packard, Alfred Sloan and Dreyfus Foundations. R. Cacho and A. Zabala are supported by the NIH predoctoral training grants.

References

1. Aaron J, Lin X, Cane DE, Christianson DW (2010) Structure of *epi*-isozizaene synthase from *Streptomyces coelicolor* A3(2), a platform for new terpenoid cyclization templates. Biochemistry 49(8):1787–1797
2. Abe I (2007) Enzymatic synthesis of cyclic triterpenes. Nat Prod Rep 24(6):1311–1331
3. Abe I, Morita H, Oguro S, Noma H, Wanibuchi K, Kawahara N, Goda Y, Noguchi H, Kohno T (2007) Structure-based engineering of a plant type III polyketide synthase: formation of an unnatural nonaketide naphthopyrone. J Am Chem Soc 129(18):5976–5980
4. Abe I, Utsumi Y, Oguro S, Morita H, Sano Y, Noguchi H (2005) A plant type III polyketide synthase that produces pentaketide chromone. J Am Chem Soc 127(5):1362–1363
5. Abe I, Watanabe T, Lou W, Noguchi H (2006) Active site residues governing substrate selectivity and polyketide chain length in aloesone synthase. FEBS J 273(1):208–218
6. Austin MB, Noel JP (2003) The chalcone synthase superfamily of type III polyketide synthases. Nat Prod Rep 20(1):79–110
7. Baerga-Ortiz A, Popovic B, Siskos AP, O'Hare HM, Spiteller D, Williams MG, Campillo N, Spencer JB, Leadlay PF (2006) Directed mutagenesis alters the stereochemistry of catalysis by isolated ketoreductase domains from the erythromycin polyketide synthase. Chem Biol 13(3):277–285
8. Bernhardt P, McCoy E, O'Connor SE (2007) Rapid identification of enzyme variants for re-engineered alkaloid biosynthesis in periwinkle. Chem Biol 14(8):888–897
9. Brautaset T, Sletta H, Degnes KF, Sekurova ON, Bakke I, Volokhan O, Andreassen T, Ellingsen TE, Zotchev SB (2011) New nystatin-related antifungal polyene macrolides with altered polyol region generated via biosynthetic engineering of *Streptomyces noursei*. Appl Environ Microbiol. doi:10.1128/AEM.05780-05711
10. Brautaset T, Sletta H, Nedal A, Borgos SE, Degnes KF, Bakke I, Volokhan O, Sekurova ON, Treshalin ID, Mirchink EP, Dikiy A, Ellingsen TE, Zotchev SB (2008) Improved antifungal polyene macrolides via engineering of the nystatin biosynthetic genes in *Streptomyces noursei*. Chem Biol 15(11):1198–1206
11. Challis G, Ravel J, Townsend CA (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem Biol 7(3):211–224
12. Chan YA, Podevels AM, Kevany BM, Thomas MG (2009) Biosynthesis of polyketide synthase extender units. Nat Prod Rep 26(1):90–114
13. Chen C-Y, Georgiev I, Anderson AC, Donald BR (2009) Computational structure-based redesign of enzyme activity. Proc Natl Acad Sci USA 106(10):3764–3769
14. Chen S, Galan MC, Coltharp C, O'Connor SE (2006) Redesign of central enzyme in alkaloid biosynthesis. Chem Biol 13(11):1137–1141
15. Conti E, Stachelhaus T, Marahiel MA (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin. EMBO J 16(14):4174–4183
16. Cox RJ (2007) Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. Org Biomol Chem 5(13):2010–2026
17. Cunningham F, Gantt E (2000) One ring or two? Determination of ring number in carotenoids by lycopene ε-cyclases. Proc Natl Acad Sci USA 98(5):2905–2910
18. Dang T, Prestwich GD (2000) Site-directed mutagenesis of squalene-hopene cyclase: altered substrate specificity and product distribution. Chem and Biol 7(8):643–649
19. Das A, Khosla C (2009) Biosynthesis of aromatic polyketides in bacteria. Acc Chem Res 42(5):631–639
20. Dewick PM (2002) The biosynthesis of C5–C25 terpenoid compounds. Nat Prod Rep 19(2):181–222
21. Ding W, Lei C, He Q, Zhang Q, Bi Y, Liu W (2010) Insights into bacterial 6-methylsalicylic acid synthase and its engineering to orsellinic acid synthase for spirotetronate generation. Chem Biol 17(5):495–503
22. Eppelmann K, Stachelhaus T, Marahiel MA (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. Biochemistry 41(30):9718–9726
23. Evans B, Chen Y, Metcalf WW, Zhao H, Kelleher NL (2011) Directed evolution of nonribosomal peptide synthetase AdmK generates new andrimid derivatives in vivo. Chem Biol 18(5):601–607
24. Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP (1999) Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nat Struct Biol 6(8):775–784
25. Fischbach M, Lai JR, Roche ED, Walsh CT, Liu DR (2007) Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. Proc Natl Acad Sci USA 104(29):11951–11956

26. Fischbach MA, Walsh CT (2009) Antibiotics for emerging pathogens. Science 325(5944):1089–1093
27. Füll C (2001) Bicyclic triterpenes as new main products of squalene-hopene cyclase by mutation at conserved tyrosine residues. FEBS Lett 509(3):361–364
28. Gantt RW, Goff RD, Williams GJ, Thorson JS (2008) Probing the aglycon promiscuity of an engineered glycosyltransferase. Angew Chem Int Ed Engl 47(46):8889–8892
29. Gao X, Xie X, Pashkov I, Sawaya MR, Laidman J, Zhang W, Cacho R, Yeates TO, Tang Y (2009) Directed evolution and structural characterization of a simvastatin synthase. Chem Biol 16(10):1064–1074
30. Greenhagen B, Chappell J (2001) Molecular scaffolds for chemical wizardry: learning nature's rules for terpene cyclases. Proc Natl Acad Sci USA 98(24):13479–13481
31. Greenhagen B, O'Maille PE, Noel JP, Chappell J (2006) Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. Proc Natl Acad Sci USA 103(26):9826–9831
32. Hahn M, Stachelhaus T (2004) Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. Proc Natl Acad Sci USA 101(44):15585–15590
33. Hahn M, Stachelhaus T (2006) Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. Proc Natl Acad Sci USA 103(2):275–280
34. Hertweck C (2009) The biosynthetic logic of polyketide diversity. Angew Chem Int Ed Engl 48(26):4688–4716
35. Hertweck C, Luzhetskyy A, Rebets Y, Bechthold A (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. Nat Prod Rep 24(1):162–190
36. Hyatt D, Croteau R (2005) Mutational analysis of a monoterpene synthase reaction: Altered catalysis through directed mutagenesis of (-)-pinene synthase from *Abies grandis*. Arch Biochem Biophys 439(2):222–233
37. Kampranis S, Ioannidis D, Purvis A, Mahrez W, Ninga E, Katerelos NA, Anssour S, Dunwell JM, Degenhardt J, Makris A, Goodenough PW, Johnson CB (2007) Rational conversion of substrate and product specificity in a *Salvia* monoterpene synthase: Structural insights into the evolution of terpene synthase function. Plant Cell 19(6):1994–2005
38. Keatinge-Clay AT, Maltby DA, Medzihradszky KF, Khosla C, Stroud RM (2004) An antibiotic factory caught in action. Nat Struct Mol Biol 11(9):888–893
39. Khosla C, Kapur S, Cane DE (2009) Revisiting the modularity of modular polyketide synthases. Curr Opin Chem Biol 13(2):135–143
40. Kim W, Lee D, Hong SS, Na Z, Shin JC, Roh SH, Wu CZ, Choi O, Lee K, Shen YM, Paik SG, Lee JJ, Hong YS (2009) Rational biosynthetic engineering for optimization of geldanamycin analogues. Chembiochem 10(7):1243–1251
41. Kimura M, Kushiro T, Shibuya M, Ebizuka Y, Abe I (2010) Protostadienol synthase from *Aspergillus fumigatus*: Functional conversion into lanosterol synthase. Biochem Biophys Res Commun 391(1):899–902
42. Klundt T, Bocola M, Lutge M, Beuerle T, Liu B, Beerhues L (2009) A single amino acid substitution converts benzophenone synthase into phenylpyrone synthase. J Biol Chem 284(45): 30957–30964
43. Kohzaki K, Gomi K, Yamasaki-Kokudo Y, Ozawa R, Takabayashi J, Akimitsu K (2009) Characterization of a sabinene synthase gene from rough lemon (*Citrus jambhiri*). J Plant Physiol 166(15):1700–1704
44. Kutchan TM (1993) Strictosidine: from alkaloid to enzyme to gene. Phytochemistry 32(3):493–506
45. Kwan DH, Leadlay PF (2010) Mutagenesis of a modular polyketide synthase enoylreductase domain reveals insights into catalysis and stereospecificity. ACS Chem Biol 5(9):829–838
46. Kwan DH, Sun Y, Schulz F, Hong H, Popovic B, Sim-Stark JC, Haydock SF, Leadlay PF (2008) Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases. Chem Biol 15(11):1231–1240
47. Lai JR, Fischbach MA, Liu DR, Walsh CT (2006) Localized protein interaction surfaces on the EntB carrier protein revealed by combinatorial mutagenesis and selection. J Am Chem Soc 128(34):11002–11003
48. Leonard E, Ajikumar PK, Thayer K, Xiao W-H, Mo JD, Tidor B, Stephanopoulos G, Prather KLJ (2010) Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. Proc Natl Acad Sci USA 107(31):13654–13659
49. Lilien R, Stevens BW, Anderson AC, Donald BR (2005) A novel ensemble-based scoring and search algorithm for protein redesign and its application to modify the substrate specificity of the gramicidin synthetase A phenylalanine adenylation enzyme. J Comput Biol 12(6):740–761
50. Lodeiro S, Schulz-Gasch T, Matsuda SP (2005) Enzyme redesign: two mutations cooperate to convert cycloartenol synthase into an accurate lanosterol synthase. J Am Chem Soc 127(41):14132–14133
51. Loris E, Panjikar S, Ruppert M, Barleben L, Unger M, Schübel H (2007) Structure-based engineering of strictosidine synthase: auxiliary for alkaloid libraries. Chem Biol 14(9):979–985
52. Ma X, Panjikar S, Koepke J, Loris E, Stöckigt J (2006) The structure of *Rauvolfia serpentina* strictosidine synthase is a novel six-bladed β-propeller fold in plant proteins. Plant Cell 18(4):907–920
53. McDaniel R, Thamchaipenet A, Gustafsson C, Fu H, Betlach M, Ashley G (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. Proc Natl Acad Sci USA 96(5):1846–1851
54. Meyer M, Xu R, Matsuda SPT (2002) Directed evolution to generate cycloartenol synthase mutants that produce lanosterol. Org Lett 4(8):1395–1398
55. Mijts B, Lee PC, Schmidt-Danner C (2006) Identification of a carotenoid oxygenase synthesizing acyclic xanthophylls: combinatorial biosynthesis and directed evolution. Chem Biol 12(4):453–460
56. Moore BS, Hertweck C (2002) Biosynthesis and attachment of novel bacterial polyketide synthase starter units. Nat Prod Rep 19(1):70–99
57. Morita H, Kondo S, Oguro S, Noguchi H, Sugio S, Abe I, Kohno T (2007) Structural insight into chain-length control and product specificity of pentaketide chromone synthase from *Aloe arborescens*. Chem Biol 14(4):359–369
58. Morrone D, Xu M, Fulton DB, Determan MK, Peters RJ (2008) Increasing complexity of a diterpene synthase reaction with a single residue switch. J Am Chem Soc 130(16):5400–5401
59. Nawarathne IN, Walker KD (2010) Point mutations (Q19P and N23K) increase the operational solubility of a 2alpha-O-benzoyltransferase that conveys various acyl groups from CoA to a taxane acceptor. J Nat Prod 73(2):151–159
60. Nguyen KT, Ritz D, Gu JQ, Alexander D, Chu M, Miao V, Brian P, Baltz RH (2006) Combinatorial biosynthesis of novel antibiotics related to daptomycin. Proc Natl Acad Sci USA 103(46):17462–17467
61. O'Hare HM, Baerga-Ortiz A, Popovic B, Spencer JB, Leadlay PF (2006) High-throughput mutagenesis to evaluate models of stereochemical control in ketoreductase domains from the erythromycin polyketide synthase. Chem Biol 13(3):287–296

62. Patel H, Walsh CT (2001) In vitro reconstitution of the *Pseudomonas aeruginosa* nonribosomal peptide synthesis of Pyochelin: characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities. Biochemistry 40(30):9023–9031
63. Peters RJ, Croteau RB (2003) Alternative termination chemistries utilized by monoterpene cyclases: chimeric analysis of bornyl diphosphate, 1, 8-cineole, and sabinene synthases. Arch Biochem Biophys 417(2):203–211
64. Power P, Dunne T, Murphy B, Nic Lochlainn L, Rai D, Borissow C, Rawlings B, Caffrey P (2008) Engineered synthesis of 7-oxo- and 15-deoxy-15-oxo-amphotericins: insights into structure-activity relationships in polyene antibiotics. Chem Biol 15(1):78–86
65. Ramos A, Olano C, Brana AF, Mendez C, Salas JA (2009) Modulation of deoxysugar transfer by the elloramycin glycosyltransferase ElmGT through site-directed mutagenesis. J Bacteriol 191(8):2871–2875
66. Reeves CD, Murli S, Ashley GW, Piagentini M, Hutchinson CR, McDaniel R (2001) Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. Biochemistry 40(51):15464–15470
67. Reid R, Piagentini M, Rodriguez E, Ashley G, Viswanathan N, Carney J, Santi DV, Hutchinson CR, McDaniel R (2003) A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. Biochemistry 42(1):72–79
68. Rix U, Fischer C, Remsing LL, Rohr J (2002) Modification of post-PKS tailoring steps through combinatorial biosynthesis. Nat Prod Rep 19(5):542–580
69. Schmidt-Dannert C, Umeno D, Arnold FH (2000) Molecular breeding of carotenoid biosynthetic pathways. Nat Biotechnol 18(7):750–753
70. Schwarzer D, Finking R, Marahiel MA (2003) Nonribosomal peptides: from genes to products. Nat Prod Rep 20(3):275–287
71. Segura M, Lodiero S, Meyer MM, Patel AJ, Matsuda SPT (2002) Directed evolution experiments reveal mutations at cycloartenol synthase residue His477 that dramatically alter catalysis. Org Lett 4(25):4459–4462
72. Shishova E, Di Constanzo L, Cane DE, Christianson DW (2007) X-ray crystal structure of aristolochene synthase from *Aspergillus terreus* and evolution of templates for the cyclization of farnesyl diphosphate. Biochemistry 46(7):1941–1951
73. Sieber SA, Marahiel MA (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. Chem Rev 105(2):715–738
74. Stachelhaus T, Marahiel MA (1995) Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. J Biol Chem 270(11):6163–6169
75. Stachelhaus T, Mootz HD, Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol 6(8):493–505
76. Stachelhaus T, Schneider A, Marahiel MA (1995) Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. Science 269(5220):69–72
77. Starks C, Back K, Chappell J, Noel JP (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-*epi*-Aristolochene synthase. Nature 277(5333):1815–1820
78. Staunton J, Weissman KJ (2001) Polyketide biosynthesis: a millennium review. Nat Prod Rep 18(4):380–416
79. Steele CL, Crock J, Bohlmann J, Croteau R (1998) Sesquiterpene synthases from grand fir (*Abies grandis*) - Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization and bacterial expression of delta-selinene synthase and gamma-humulene synthase. J Biol Chem 273(4): 2078–2089
80. Tang Y, Tsai SC, Khosla C (2003) Polyketide chain length control by chain length factor. J Am Chem Soc 125(42):12708–12709
81. Uguru G, Milne C, Borg M, Flett F, Smith CP, Micklefield J (2004) Active-site modifications of adenylation domains lead to hydrolysis of upstream nonribosomal peptidyl thioester intermediates. J Am Chem Soc 126(16):5032–5033
82. Umeno D, Arnold FH (2004) Evolution of a pathway to novel long-chain carotenoids. J Bacteriol 186(5):1531–1536
83. Umeno D, Tobias AV, Arnold FH (2005) Diversifying carotenoid biosynthetic pathways by directed evolution. Microbiol Mol Biol R 69(1):51–78
84. Walter MH, Strack D (2011) Carotenoids and their cleavage products: biosynthesis and functions. Nat Prod Rep 28(4):663–692
85. Wang C-W, Liao JC (2001) Alteration of product specificity of *Rhodobacter sphaeroides* phytoene desaturase by directed evolution. J Biol Chem 276(44):41161–41164
86. Wang C-W, Oh M-K, Liao JC (2000) Directed evolution of metabolically engineered *Escherichia coli* for carotenoid production. Biotechnol Prog 16(6):922–926
87. Watanabe K, Hotta K, Praseuth AP, Koketsu K, Migita A, Boddy CN, Wang CC, Oguri H, Oikawa H (2006) Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. Nat Chem Biol 2(8):423–428
88. Watanabe K, Hotta K, Praseuth AP, Searcey M, Wang CCC, Oguri H, Oikawa H (2009) Rationally engineered total biosynthesis of a synthetic analogue of a natural quinomycin depsipeptide in *Escherichia coli*. ChemBioChem 10(12):1965–1968
89. Weissman KJ, Leadlay PF (2005) Combinatorial biosynthesis of reduced polyketides. Nat Rev Microbiol 3(12):925–936
90. Whittington DA, Wise ML, Urbansky M, Coates RM, Croteau RB, Christianson DW (2002) Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci USA 99(24):15375–15380
91. Wilderman PR, Peters RJ (2007) A single residue switch converts abietadiene synthase into a pimaradiene specific cyclase. J Am Chem Soc 129(51):15736–15737
92. Williams GJ, Goff RD, Zhang C, Thorson JS (2008) Optimizing glycosyltransferase specificity via “hot spot” saturation mutagenesis presents a catalyst for novobiocin glycorandomization. Chem Biol 15(4):393–401
93. Williams GJ, Zhang C, Thorson JS (2007) Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. Nat Chem Biol 3(10):657–662
94. Wu T-K, Wen HY, Chang C-H, Liu Y-T (2008) Protein plasticity: a single amino acid substitution in the Saccharomyces cerevisiae oxidosqualene-lanosterol cyclase generates Protosta-13(17), 24-dien-3b-ol, a rearrangement product. Org Lett 10(12):2529–2532
95. Xiang H, Tschirret-Guth RA, Ortiz De Montellano PR (2000) An A245T mutation conveys on cytochrome P450eryF the ability to oxidize alternative substrates. J Biol Chem 275(46):35999–36006
96. Xie XK, Watanabe K, Wojcicki WA, Wang CCC, Tang Y (2006) Biosynthesis of lovastatin analogs with a broadly specific acyltransferase. Chem Biol 13:1161–1169
97. Xu M, Wilderman PR, Peters RJ (2007) Following evolution’s lead to a single residue switch for diterpene synthase product outcome. Proc Natl Acad Sci USA 104(18):7397–7401
98. Yang L, Stöckigt J (2010) Trends for diverse production strategies of plant medicinal alkaloids. Nat Prod Rep 27(10):1469–1479
99. Yoshikuni Y, Dietrich FA, Nowroozi FF, Babbitt PC, Keasling JD (2008) Redesigning enzymes based on adaptive evolution for

optimal function in synthetic metabolic pathways. Chem Biol 15(6):607–618
100. Yoshikuni Y, Ferrin TE, Keasling JD (2006) Designed divergent evolution of enzyme function. Nature 440(7087):1078–1082
101. Yoshikuni Y, Martin VJJ, Ferrin TE, Keasling JD (2006) Engineering cotton (+)-d-cadinene synthase to an altered function: germacrene D-4-ol synthase. Chem Biol 13(1):91–98
102. Zhou H, Xie X, Tang Y (2008) Engineering natural products using combinatorial biosynthesis and biocatalysis. Curr Opin Biotechnol 19(6):590–596
103. Zocher G, Richter ME, Mueller U, Hertweck C (2011) Structural fine-tuning of a multifunctional cytochrome P450 monooxygenase. J Am Chem Soc 133(7):2292–2302