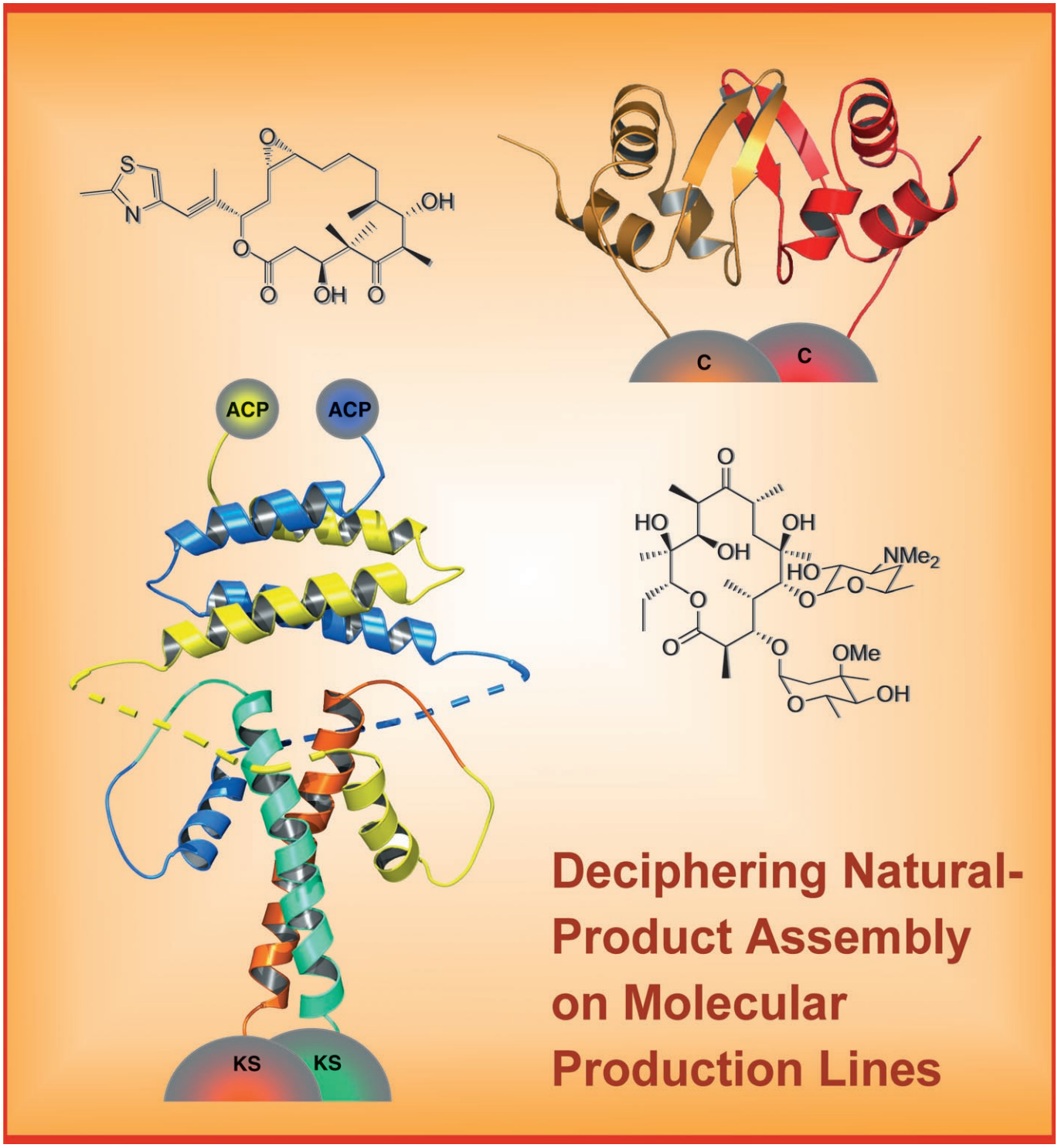
CHEMBIOCHEM

DOI: 10.1002/cbic.200700751

# Protein–Protein Interactions in Multienzyme Megasynthetases

Kira J. Weissman* and Rolf Müller*[a]



*The multienzyme polyketide synthases (PKSs), nonribosomal polypeptide synthetases (NRPSs), and their hybrids are responsible for the construction in bacteria of numerous natural products of clinical value. These systems generate high structural complexity by using a simple biosynthetic logic—that of the assembly line. Each of the individual steps in building the metabolites is designated to an independently folded domain within gigantic polypeptides. The domains are clustered into functional modules, and the modules are strung out along the proteins in the order in which they act. Every metabolite results, therefore, from the successive action of up to 100 individual catalysts. Despite the conceptual simplicity of this division-of-labor organization, we are only beginning to decipher the molecular details of the numerous protein–protein interactions that support assembly-line biosynthesis, and which are critical to attempts to re-engineer these systems as a tool in drug discovery. This review aims to summarize the state of knowledge about several aspects of protein–protein interactions, including current architectural models for PKS and NRPS systems, the central role of carrier proteins, and the structural basis for intersubunit recognition.*

## Introduction

Natural products and their derivatives continue to be an important source of drugs and drug leads in many therapeutic areas.[1] Prominent among them are the complex (or reduced) polyketide (PK) and nonribosomal polypeptide (NRP) metabolites, and their hybrids (Scheme 1).[2–5] These compounds exhibit a wide range of valuable pharmaceutical activities, from anticancer (epothilone, **1**) to immunosuppressive (cyclosporine, rapamycin (**2**), and FK506), to anti-infective (erythromycin (**3**), daptomycin (**4**), monensin (**5**), gramicidin (**6**)) properties. Although structural diversity is a hallmark of these metabolites, they are all generated by the linkage of simple building blocks—acyl-CoAs in the case of PK metabolites, and proteinogenic or modified amino acids in the case of NRPs—into long chains. The basic functional groups arrayed along the chains are further diversified by modification reactions, which include reduction, oxidation, methylation, acylation, and glycosylation.[3,6] Macrocyclization is also a fundamental feature of many pathways and introduces the conformational rigidity, which enables specific interactions with biological targets, and in the case of NRP-derived compounds, confers resistance to proteolytic degradation.[7]

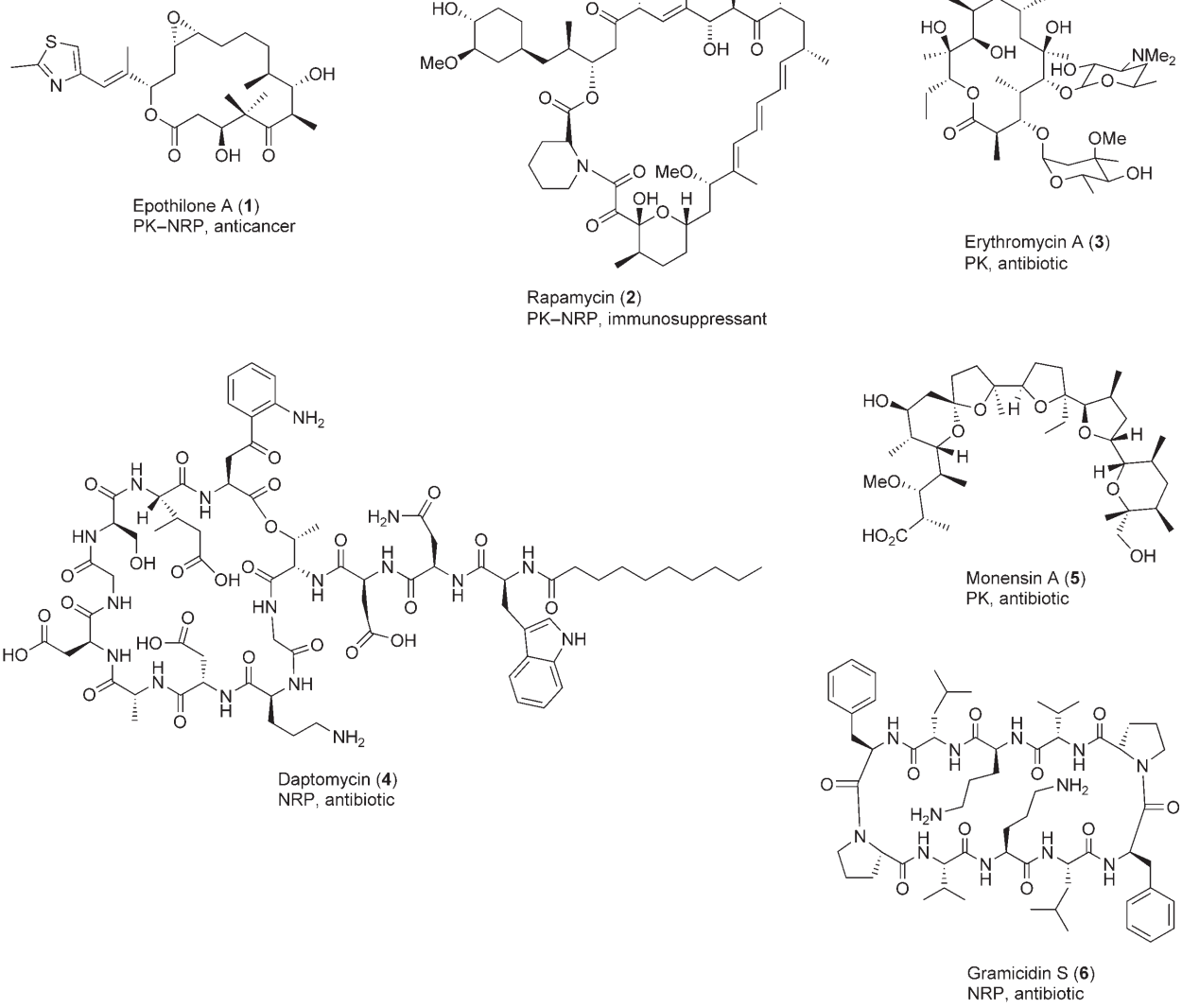
Nature has evolved a particularly elegant solution to the challenge of customizing a set of standard building blocks to generate structural complexity: each step in the pathway is designated to an independently folded enzymatic domain within large polypeptides. In accordance with their products, the multienzymes are known as polyketide synthases (PKSs), nonribosomal synthetases (NRPSs), and mixed PKS–NRPS systems (Figure 1). Within the gigantic protein subunits, sets of domains are organized into functional units called modules, and each module is responsible for a specific round of chain extension. The sequence of modules also directly correlates with the order in which they act. As this biosynthetic logic echoes the method of car manufacture invented by Henry Ford, these polypeptides have been dubbed "molecular assembly lines". In fact, the colinearity between the genetic organization and the sequence of enzymatic transformations in many systems is so strong that multiple features of the product structures can be predicted with confidence from inspection of the gene sequences alone.[8,9]

In these so-called "type I" PKSs (Figure 1), a functional module minimally comprises an acyl transferase (AT) domain, which recruits the building block from the cellular environment, and a ketosynthase (KS) domain, which accomplishes carbon–carbon bond formation using thioclaisen condensation chemistry. The extent of modification that occurs on the resulting β-keto group depends on the specific complement of reductive domains present in the module: inclusion of a ketoreductase (KR) domain results in a hydroxyl group, a DH–KR enzymatic pair in a double bond, and a complete DH–ER–KR tridomain in a fully reduced methylene group. In some systems, this set of domains is further augmented by *O*-, *N*-, and *C*-methyltransferase (MT) functions. PKS assembly lines also incorporate a module of domains to initiate the biosynthesis (a "loading module"), while the finished product is cleaved off the multienzyme by hydrolysis or, more typically, by macrocyclization by a dedicated thioesterase (TE) activity.

The analogous basic biosynthetic apparatus in the NRPS systems (Figure 1) comprises an adenylation (A) domain responsible for selection of a specific amino acid and activation as its aminoacyl adenylate, and a condensation (C) or heterocyclization (HC) domain, which catalyzes amide bond formation. HC domains are variant C domains, which additionally carry out cyclization of the thiol side chain of cysteine or the hydroxyl side chain of serine or threonine onto the newly formed peptide bond, resulting in the formation of thiazoline or oxazoline rings. The complement of specialized processing enzymes in NRPS modules can include epimerase (E), *N*- and *C*-methyltransferase, and oxidase (Ox) activities. Initiation of biosynthesis is also assigned to a specific set of domains, while TEs or alternatively reductase (R) functions are typically present at the end of the assembly lines to terminate chain extension. Following disconnection of PK, NRP, and hybrid intermediates from the multienzymes, the structures are often subject to various tailoring reactions, including acylations, oxidations, and glycosylations, which are usually required to elaborate the metabolites to their biologically active forms.[3]

[a] *Dr. K. J. Weissman, Prof. R. Müller*
*Pharmaceutical Biotechnology, Saarland University*
*P.O. Box 151150, 66041 Saarbrücken (Germany)*
*Fax: (+49) 681-302-5473*
*E-mail: k.weissman@mx.uni-saarland.de*
*rom@mx.uni-saarland.de*




**Scheme 1.** Structures and biological activities of representative polyketide (PK), nonribosomal polypeptide (NRP), and mixed PK–NRP natural products.

*Kira J. Weissman obtained her BS in Chemistry from Stanford University, California in 1995. Receipt of a Churchill Fellowship allowed her to relocate to the United Kingdom, where she carried out MPhil, PhD, and postdoctoral research at the University of Cambridge with Professors Jim Staunton and Peter Leadlay. Her work at Cambridge was supported by several fellowships, including a Royal Society Dorothy Hodgkin Research Fellowship. She is presently an Alexander von Humboldt Research Fellow at the University of Saarland, Saarbrücken, Germany, working with Prof. Rolf Müller. Kira's research interests include the mechanistic enzymology and structural biology of modular PKS and NRPS systems.*

*Rolf Müller studied pharmacy in Bonn and received his PhD with Professor Eckhard Leistner in 1994. He received a two-year research fellowship from the Deutsche Forschungsgemeinschaft to work in Professor Heinz G. Floss's laboratory in Seattle, USA. Starting in 1997, he studied myxobacteria at the German Research Center for Biotechnology as a junior group leader and received his Habilitation at the Technical University of Braunschweig in 2001. Rolf has received several research prizes, among them the BioFuture prize of the Bundesministerium für Bildung und Forschung. In October 2003, he was appointed Professor of Pharmaceutical Biotechnology at Saarland University. His current work focuses on the biosynthesis, regulation, and heterologous production of secondary metabolites from myxobacteria and actinomycetes. Rolf Müller is a member of the editorial advisory board of* ChemBioChem.

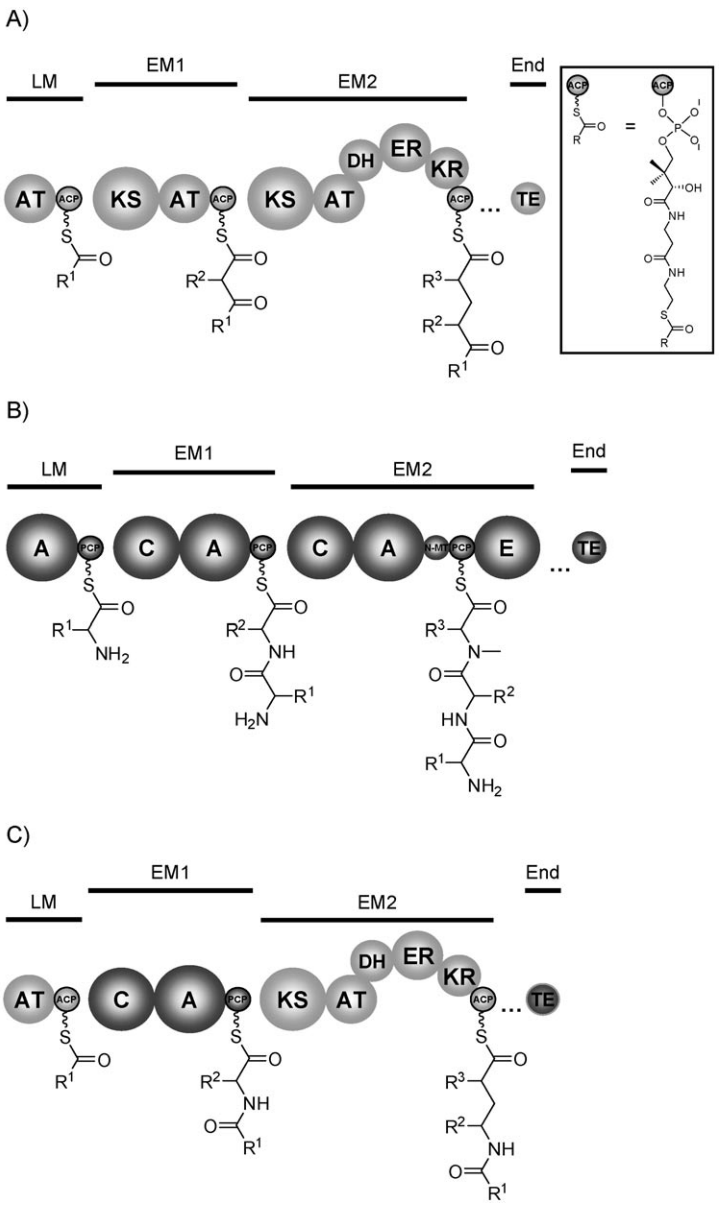



**Figure 1.** Biosynthetic logic of multienzyme assembly lines. A) Domain and modular organization of a generic PKS system, showing a typical loading module (LM), minimal extension module (EM1), and a second extension module (EM2), which incorporates a full "reductive loop". Biosynthesis is terminated by an integral thioesterase (TE) domain. B) Domain and modular organization of a generic NRPS system, showing a typical loading module (LM), minimal extension module (EM1), and a second extension module, which incorporates processing domains (EM2). C) Domain and modular organization of a generic mixed PKS–NRPS system. Abbreviations are: AT, acyltransferase; ACP, acyl carrier protein; KS, ketosynthase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; TE, thioesterase; A, adenylation; PCP, peptidyl carrier protein; C, condensation; N-MT, *N*-methyl transferase; E, epimerase.

Despite the obvious benefits from a programming standpoint, assembly-line biosynthesis is inherently inefficient, as it depends on the successive action of tens (and up to hundreds)[10] of individual enzymes. Indeed, measured rates for PK[11–13] and NRP[14,15] biosynthesis in vitro ($k_{cat}$ values are typically in the 0.1–10 $min^{-1}$ range) suggest that the multienzymes are among the more sluggish catalysts in nature.[16] This process would undoubtedly be slower were it not for "substrate channeling"[17]—a highly efficient method of substrate delivery employed by other multienzyme systems, including the 2-oxoacid dehydrogenase complexes and the biotin-dependent carboxylases.[18] This strategy is achieved by the inclusion in each PKS and NRPS module (Figure 1) of a small (80–100 residues), noncatalytic domain, called a "carrier protein" (CP; either acyl carrier proteins (ACPs), aryl carrier proteins (ArCPs), or peptidyl carrier proteins (PCPs), depending on the particular building block incorporated by the module).[19,20] Each CP is modified by post-translational addition of a phosphopantetheine (PPant) prosthetic arm derived from coenzyme A to a conserved Ser residue; this converts the domain from its inactive apo to its active holo form. This priming reaction is catalyzed by a recently identified superfamily of enzymes called the phosphopantetheinyl transferases (PPTases),[21] which are exemplified by the broad specificity catalyst Sfp from the surfactin NRPS pathway in *Bacillus subtilis*.[22]

Tethering of the growing chains to the terminal sulfhydryl of the PPant generates a thioester bond, which is activated towards both Claisen condensation and amide bond formation. An additional benefit of this "multiple thiotemplated" logic of natural-product assembly is that the substrates are sequestered away from competing cellular processes.[17] Such microcompartmentation is particularly important in these pathways, as many of the building blocks are diverted from primary metabolism, or are used by one of the many other PKS or NRPS assembly lines within the producing organisms.[23,24] Substrates covalently attached to CPs can additionally be protected from bulk solvent through interactions with the carrier protein itself or other domains; this sequestration provides an effective means to stabilize reactive intermediates.[17,25] Finally, attachment to a carrier protein means that the true substrate is not simply the growing PK or NRP chain, but the entire acylated domain; this feature introduces an additional level of molecular recognition, which can contribute to the specific programming required in these systems.[18]

The use of protein-bound molecules as substrates potentially places protein–protein interactions at the very heart of assembly-line biosynthesis of natural products. Dissection of a typical cycle of chain extension in PKS systems shows that the ACP (in either its holo or acylated form) must interact with every other domain within the module (Figure 2):[20,26] with the AT to load the building block; with the KS to accomplish chain elongation; with any reductive domains within the module to customize the resulting β-ketone; and with a KS or TE domain located immediately downstream to initiate the next round of chain extension, or alternatively, to terminate it. Similarly, in NRPSs, the PCP domain must communicate with multiple partners: with upstream A and C domains, any optional modifying enzymes present in the module, and with a downstream C or

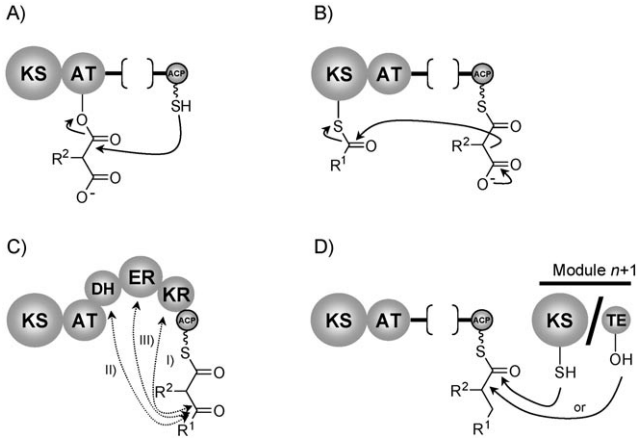



**Figure 2.** Suite of ACP-based interactions that occur during each round of chain extension on modular PKSs. A) The holo ACP cooperates with the AT during extender-unit selection. B) Condensation is accomplished by cooperation between the ACP and KS domains. (Although this is not accurately represented in the diagram, the KS and ACP domains are located on opposite subunits within the overall homodimeric PKS.) C) Reductive tailoring of the β-keto functionality occurs by delivery of the substrate to the reductive domains, KR, DH, and then ER. D) The fully processed intermediate is transferred to a domain in the downstream module, either a KS or a TE.

TE (Figure 3).[20] Another common feature of these systems is that the component modules are distributed among multiple protein subunits. Consequently, many ACP to KS intermodular transfers in PKSs occur between domains located on different polypeptides.[27] Similarly, peptide-bond formation catalyzed by a specific C domain might involve substrates tethered to PCP domains that are embedded in separate NRPS proteins. Therefore, forming the correct product requires a specific ordering of the multienzymes; each subunit must not only recognize its appropriate partner protein, but discriminate against all improper associations.[28] Finally, it has been revealed that for an increasing number of systems, essential transformations are accomplished by catalytic domains that are not located within the basic modules. For example, several PKSs have been identified that lack integral AT domains within the multienzymes (the "AT-less systems"[29,30]), and instead this function is typically provided by a discrete AT that delivers a common building block to each ACP, in trans. Similarly, discrete TE domains ("type II" TEs) in both PKS and NRPS systems perform crucial proof-reading functions by hydrolyzing aberrant substrates from the respective CP domains (Figure 4).[31–33]

This analysis clearly shows that efforts to understand these classes of biosynthetic enzymes must focus on illuminating the molecular basis for the many protein–protein recognition events that underlie the pathways. Questions of particular interest include the following: What is the structural arrangement of domains within each module that allows all of the interactions to occur? What is the role (if any) of the noncatalytic regions called "linkers" that separate the domains in maintaining the modular structure or facilitating interdomain communication? How is the order of interactions controlled when the carrier proteins have a "choice" of partners? What is the structural basis for the association between subunits, and how is interaction specificity assured? How do the many domains that operate in trans during chain assembly interact with their ACP-bound substrates? The answers to these questions are also likely to prove essential for optimizing the activity of genetically engineered assembly lines.

This review aims to summarize the current state of knowledge about these and other aspects of protein–protein interactions in PKS, NRPS, and mixed PKS–NRPS systems.

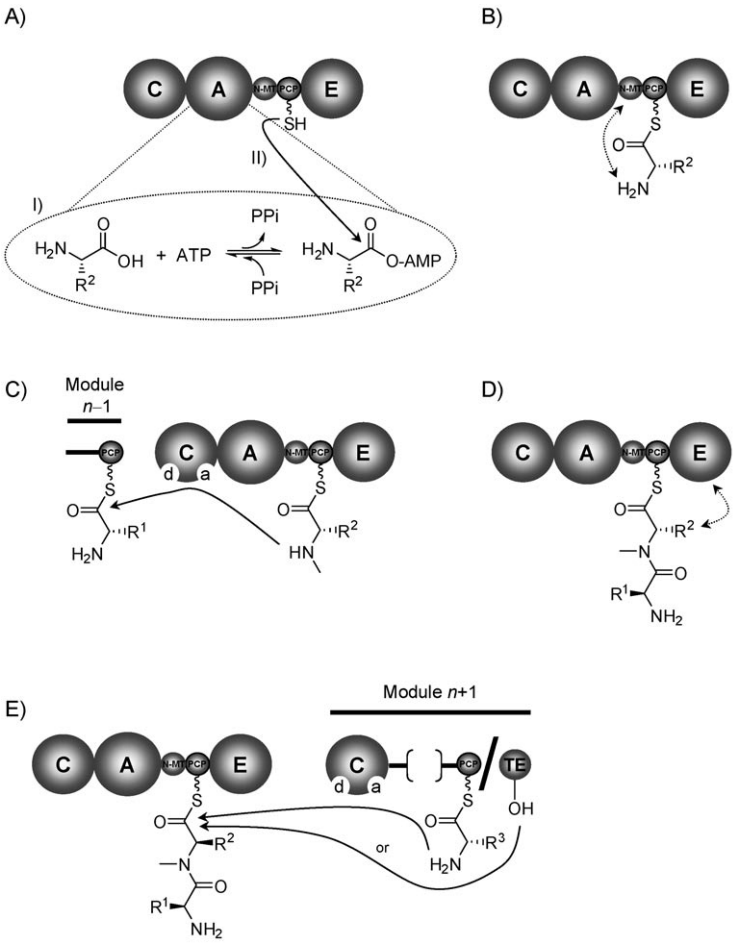



**Figure 3.** Set of PCP-based interactions that occur during each round of chain extension in NRPS systems. A) The amino acid to be added is activated as its adenylate by I) the A domain, and then II) transferred to the PCP domain. B) If an *N*-methyltransferase is present, it acts prior to the condensation reaction. C) The C domain catalyzes amide bond formation between the growing chain bound to the upstream PCP domain (d site), and the amino acid monomer tethered to the PCP (a site). D) Epimerization at the α center occurs following condensation, as it is energetically preferred. E) The processed chain is either used in chain extension by the downstream C domain, or transferred to the TE prior to release from the synthetase.

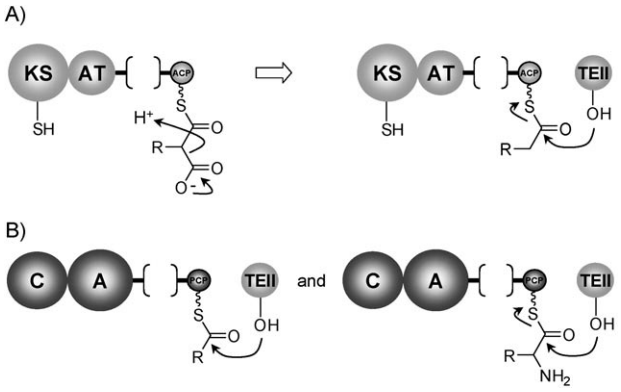



**Figure 4.** Roles of type II TE domains in megasynthase systems. A) In modular PKS, type II TEs are proposed to release from the ACP domains acyl groups that have been produced by aberrant decarboxylation of chain extender units. B) Type II TEs are proposed to play two functions in NRPS systems: removal of acyl groups added to the PCP domains during post-translational priming of the apo proteins, and release of amino acids that have been loaded by mistake.

## 1. Modular Architecture

In theory, a high-resolution crystal structure of an entire module or subunit derived from PKS and NRPS systems would at once clarify many protein–protein interaction issues, including the relative location of the domains within the complexes, and the function of the intervening linker regions. To date, however, such a structure has not appeared in the literature, although efforts towards this goal are almost certainly in progress. Encouragingly, crystallographic information has recently become available for the fatty acid synthase (FAS) of animals[34]—a multienzyme system the evolution of which is closely linked to that of PKSs—as well as for the FAS of fungi.[35]

### 1.1. The structure of animal FAS

The sequence of reactions performed by animal FAS parallels that in PK biosynthesis, and the analogous domains (KS, malonyl-CoA/acetyl-CoA transacylase (MAT), DH, ER, KR, ACP, and TE) occur in the same order in the primary sequence of the proteins (Figure 5). Nonetheless, an important general distinction between PKS and FAS is that FAS multienzymes function in an iterative manner, with the same set of domains used repeatedly to generate products of appropriate length. (In fact, iteration is a programmed feature of some bacterial PKS modules, such as in the pathways to borrelidin[36] and aureothin,[37] but the majority of modules are "single use".) Although the detailed logistics of chain extension also differ between PKS and FAS, the catalytic mechanisms employed by individual domains are closely similar (for an in depth comparison between PKS and FAS, see the recent excellent review by Smith and Tsai[38]). These considerations suggest that modular PKS and FAS multienzymes might have evolved from a common ancestor, a hypothesis that is supported by phylogenetic analysis.[35] Therefore, structural information on the overall architecture of FAS is likely to be relevant to modular PKS.

The crystal structure of porcine FAS was solved to 4.5 Å,[34] yielding an overall X-shaped dimer (Figure 5) that was in good agreement with an earlier 15 Å structure determined by electron microscopy.[39] Although it was not possible at this resolution to unambiguously trace the entire peptide backbone, by fitting the observed electron density to the atomic-resolution structures of homologous individual (type II) bacterial proteins, the authors were able to locate the majority of the functional domains within the structure. This analysis revealed that the large dimerization interface of FAS arises from self-association of three individual domains—the KS, the ER, and the pseudodimeric DH. In contrast, the MAT and KR domains do not oligomerize, but are positioned on the periphery of the structure through interactions with adjacent domains or linker regions. The structure was also notable for the absence of the ACP and TE, although the location of these domains directly adjacent to the KR was suggested by a blurred volume of electron density at the end of one arm of the dimer. The failure to unambiguously place these two domains within the structure likely arises from their high inherent mobility, a feature that could be facilitated by linker flexibility. This idea is bolstered by analysis of the distances between active sites in each of the two asymmetrical reaction chambers defined by the FAS structure (Figure 5); from these measurements it is clear that the ACP must be capable of significant motion in order to shuttle substrates among the deep-set active sites within its partner domains. Therefore, earlier models for substrate delivery by the swinging arm of the ACP must be broadened to include the concept of the ACP itself as a "swinging domain".[18]

Despite the good agreement between the structure and many earlier biochemical data, there is one troubling inconsistency.[40,41] Mutant-complementation studies have demonstrated that each ACP domain can interact with functional partners located both on the same and opposite subunit. Such extensive cross-talk is difficult to reconcile with the crystal structure, as the two reaction chambers appear to be physically isolated from one another. Thus, it is likely that fatty acid biosynthesis is accompanied by large-scale conformational transitions within the FAS multienzyme, and, therefore, that the solved structure represents only a static picture of a much more dynamic system.

### 1.2. The structure of fungal FAS

Although fungal FAS systems exhibit a type I organization, the component domains are ordered differently and distributed between two distinct subunits, $\alpha$ and $\beta$, which together form a dodecameric $\alpha_6\beta_6$ architecture.[42] However, it is relevant to consider emerging structural information on fungal FAS here, as they represent the only systems to date for which the motion of the ACP domains has been rationalized clearly.[34,43–45] Consideration of interdomain distances within the structures, as well as the constraints of the flexible linker regions that tether each ACP domain to two anchor points within the complex, suggests that the diffusive motion of the ACP is chan-

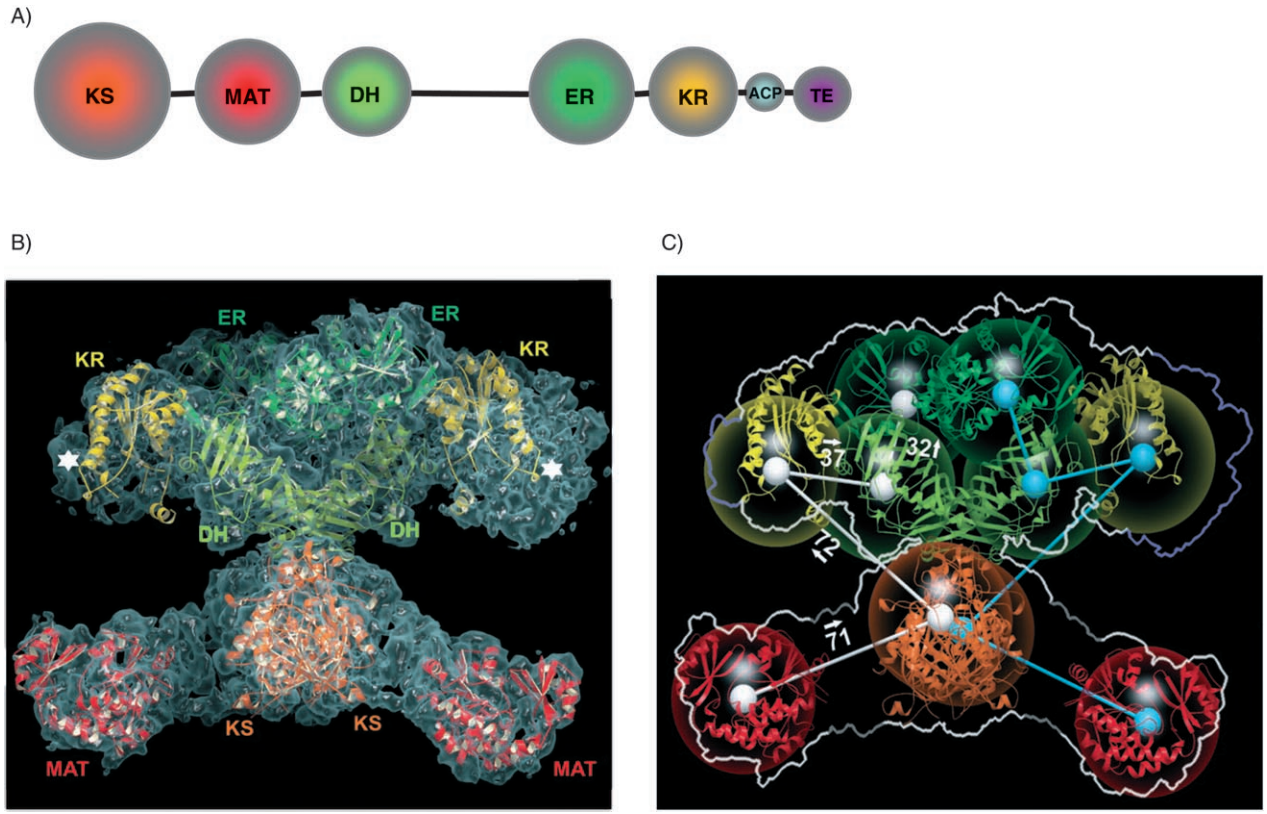



**Figure 5.** Domain organization and structure of porcine FAS. A) The linear domain organization of animal FAS. Abbreviations are: KS, ketosynthase; MAT, malonyl-CoA/acetyl-CoA-ACP transacylase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl carrier protein; TE, thioesterase. The linker regions between the domains are also included, and drawn approximately to scale. B) Semitransparent surface representation of the 4.5 Å crystal structure of the FAS, showing the fitted domains. The white stars indicate the suggested locations of the missing ACP and TE domains. Reprinted with permission from ref. [34]; copyright 2006, American Association for the Advancement of Science. C) Front view of FAS with ribbon representations of the fitted domains colored as in B). The hollow spheres in domain colors that surround the active sites denote the length of the phosphopantetheine arm, and reflect how close the ACP has to approach the domains during the catalytic cycle. The active sites are connected in order of the reaction sequence, with the distances between them indicated for the left-hand reaction chamber. Reprinted with permission from ref. [34]; copyright 2006, American Association for the Advancement of Science.

neled into a two-dimensional, circular trajectory around the reaction chamber. Temporary docking of the ACP onto the individual domains is proposed to occur through complementary electrostatic interactions between a negatively charge recognition motif on the ACP domain, and a prominent positively charged patch near the active-site cavity of each partner.[43] Indeed, such an interface was observed between the KS and ACP domains in the crystal structures. The contact surface is not extensive, however, and involves only ten residues on the ACP; this is consistent with the need for the interaction with the KS to be transient.[43,45] Weaker docking against the remaining domains would explain the finding that the ACP is not randomly distributed among the active sites, but localized at the KS domain. While it remains to be demonstrated whether modular systems operate by similar principles, it is tempting to speculate that the respective CPs might also follow such economic trajectories during substrate translocation.

### 1.3. Structural information on modular PKS

It has been known for over a decade that modular PKS, like animal FAS, are homodimeric.[46,47] Several fragments (KS–AT didomains, TE) retain their dimeric character when released by limited proteolysis; this suggests that they contribute to the overall subunit association. Chemical cross-linking[46] and mutant-complementation studies analogous to those performed on animal FAS,[48,49] have also demonstrated that intersubunit cross-talk is a fundamental feature of chain assembly on PKS multienzymes. The strong parallels to animal FAS have been reinforced by the recent publication of high-resolution crystal or NMR structures for all of the domains, with the exception of the DH and ER activities.

The structures of KS and AT activities have been solved within the context of two KS–AT didomains derived from the PKS (DEBS) responsible for erythromycin biosynthesis (at 2.6 and 2.7 Å; Figure 6).[50,51] The overall structure of the didomains closely resembles that of the equivalent KS–MAT region of FAS, including an extensive dimerization interface formed by the KSs. As in the FAS structure, the active sites of KS and AT are separated by a distance (≈80 Å) that exceeds the reach of a static ACP; this is consistent with the requirement for a swinging ACP domain in PKS systems. In addition, the higher resolution of the PKS structures allowed clarification of the roles of various linker regions, which were suggested by the FAS struc-

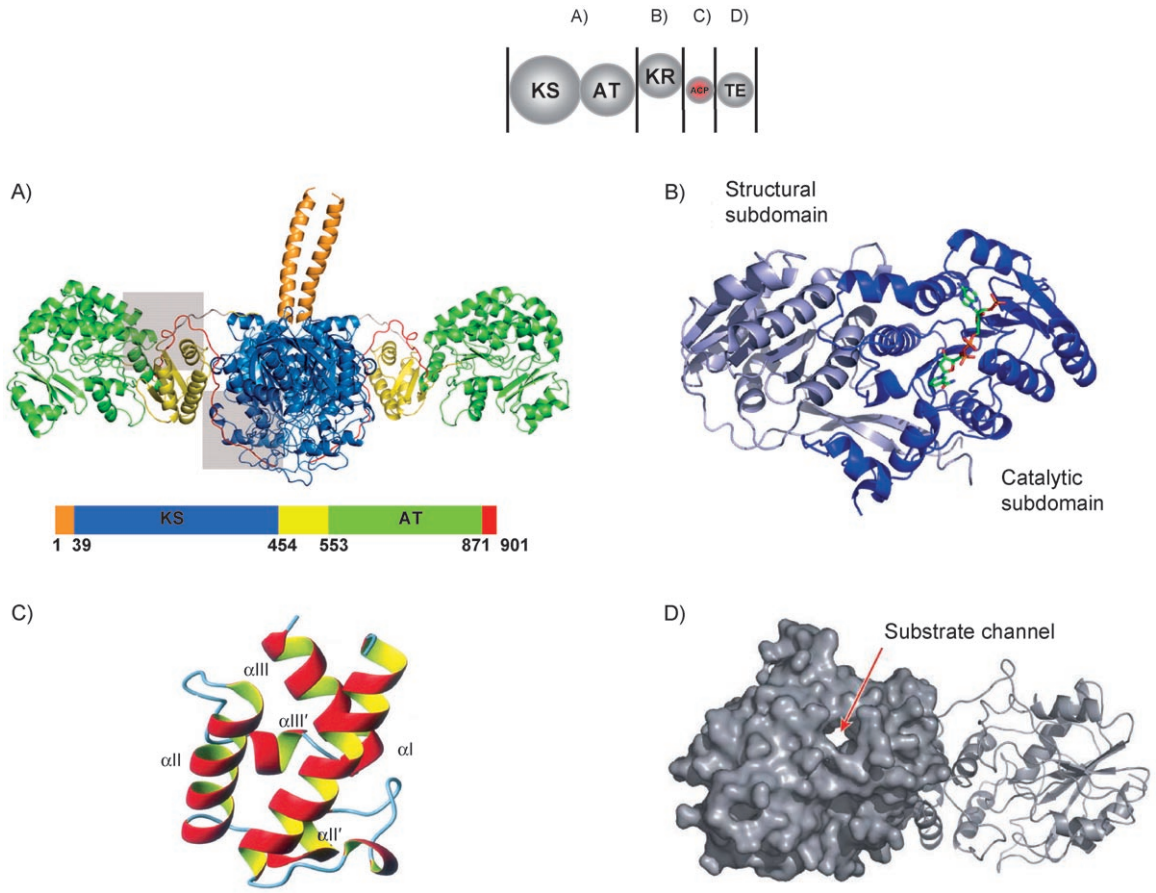



**Figure 6.** Structures of PKS domains derived from DEBS. A) Crystal structure of DEBS $KS_5$–$AT_5$ didomain at 2.7 Å resolution. The domains and linker regions are color coded, as follows: orange, N-terminal docking domain; blue, ketosynthase (KS); green, acyl transferase (AT); yellow, KS-to-AT linker; red, post-AT linker. Residues 458–465 (indicated in gray) lacked electron density and were therefore modeled manually. (Reprinted with permission from ref. [51]; copyright 2006, National Academy of Sciences, USA. B) Crystal structure of DEBS $KR_1$ at 1.79 Å resolution, showing the secondary structural elements. The AT-to-KR linker ("structural subdomain") is shown in light blue, and the catalytic domain in dark blue. The NADPH and catalytic tyrosine are represented as sticks. Reprinted with permission from ref. [67], copyright 2007, Annual Reviews. C) NMR solution structure of DEBS $ACP_2$; ribbon diagram of the minimized mean structure is shown. Reprinted with permission from ref. [55], copyright 2007, The Protein Society. D) Crystal structure of the DEBS TE solved at 2.8 Å resolution. One subunit in the monomer is shown in surface representation. The open substrate channel is indicated by the red arrow. Reprinted with permission from ref. [67]; copyright 2007, Annual Reviews.

ture.[34] The KS-to-AT linker region between the two domains interacts with the last α helix of the AT domain, to form an overall αβα fold. The linker C-terminal to the AT domain was observed to wrap back over both the AT domain and the KS-to-AT linker to interact with the KS, although this static configuration might have arisen through artifactual crystal contacts. Thus, both "linkers" appear to be structured adapter domains that serve to fix the relative positions of the KS and AT.[51]

A folded linker region was also discovered during structure elucidation of KR domains derived from the DEBS (Figure 6) and tylosin (TYL) systems.[52,53] In both cases, the crystallographic efforts were targeted at stable proteolytic fragments of the PKS subunits, and encompassed the catalytic KR domain and approximately 200 residues of the large upstream AT-to-KR linker region.[47] The structures, obtained at 1.79 and 1.95 Å resolution, respectively, showed that both the AT-to-KR linker region and the KR domain adopt similar folds that are characteristic of the short-chain dehydrogenase (SDR) family of enzymes. Although the linker–KR fragment is monomeric, which is consistent with the monomeric nature of the FAS KR, the AT-to-KR linker and KR domains interact with each other to form an overall pseudodimer. On the basis of this observation, the authors proposed that the unassigned electron density at the tip of the porcine FAS structure might arise not from the ACP–TE didomain, but from an analogous "structural subdomain" of the FAS catalytic KR.[52] However, there is no convincing sequence homology between the PKS AT-to-KR linker and the corresponding DH–ER interdomain sequence of animal FAS,[38,54] and so it cannot be assumed that the two regions will adopt a common fold. Finally, detailed insight into the structure of PKS ACP domains was recently provided by the publication of the NMR solution structure of a prototypical ACP from DEBS (Figure 6).[55] The ACP adopts a right-twisted helical bundle topology, which is a conserved fold previously observed for type I ACP from rat FAS,[56] and many type II ACPs from both FAS and PKS systems.[57–66]

The accumulated structural data have been combined to produce a model for PKS reductive loops[52] and a closely simi-

lar structure that additionally encompasses the KS and AT domains[67] in which the linker between the DH and ER domains plays the role of the KR structural subdomain (AT-to-KR linker; Figure 7). In both cases, the proposed organization bears a striking resemblance to the published FAS structure, although the authors took the extra step of locating the ACP domain in defined positions (at the C terminus of the KR[52] and on a KS docking site,[67] respectively). In support of the models, it has been shown that while the linker–KR fragment is monomeric, a construct that additionally encompasses the DH and ER is a homodimer;[52] this is consistent with the role of the DH and ER domains in promoting subunit self-association

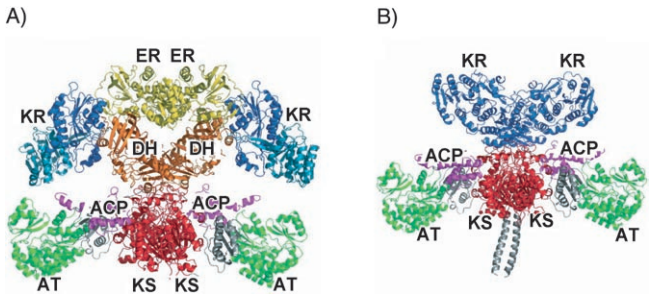



**Figure 7.** Structural models for type I PKS modules. A) Structural model of module 4 of DEBS, which incorporates a full reductive loop (DH–ER–KR). Reprinted with permission from ref. [67]; copyright 2007, Annual Reviews. B) Structural model of modules 3 or 5 of DEBS, which include only a KR domain. Reprinted with permission from ref. [67]; copyright 2007, Annual Reviews.

Despite the inherent appeal of models for PKS architecture that resemble that of FAS, these proposals are based largely on homology modeling and computational docking.[52,67] Indeed, there are multiple issues with deriving the structure of a PKS module by direct analogy with that of FAS. For example, many PKS modules lack DH and ER domains, and so it is not obvious in these cases how the overall stability of the PKS dimer would be maintained. A variant model has been suggested for PKS modules that include only a KR domain,[67] and in which the upstream AT-to-KR linker appears to substitute for the missing DH and ER dimerization elements; indeed this region can self-associate when expressed as a discrete protein.[54] However, in AT-less PKS modules even this portion of the multienzymes appears to be absent.[54] Another significant concern is that in modular PKSs, many ACP are joined directly to KS domains within the downstream module by short linker regions (15–25 residues). Thus, while FAS ACPs are tethered on their C-terminal ends to a monomeric TE domain—an arrangement that does not appear to hinder their mobility—many PKS ACPs are effectively attached to all the downstream modules within the polypeptide. A similar situation arises even for modules at the end of the assembly lines, as the solved structures of PKS TE domains have shown them to be homodimeric (Figure 6).[68–70] These architectural features would seem to impose very significant constraints on the ability of the ACP domain to traverse the large interdomain distances apparent in the FAS structure. As it seems premature to propose structural models based on animal FAS, further insight into the specific organization of domains within PKS modules (as well as the relative arrangement of the modules themselves) awaits real structural data obtained on larger portions of these systems.

### 1.4. Structural information on NRPS/mixed systems

As with PKS systems, structure elucidation efforts on NRPSs have not yet yielded overall structures for the multienzymes. However, high resolution crystallographic or NMR spectroscopy data are available for each type of domain, including the stand-alone C domain VibH from the vibriobactin pathway[71] (which is likely to be representative of HC and E activities), the phenylalanine-activating A domain from gramicidin synthetase,[72] an integral PCP from tyrocidine biosynthesis (TycC3–PCP),[59] and TE domains from the surfactin[73] and fengycin[74] NRPSs. These structures have identified accurate boundaries for the folded domains, and simultaneously of the intervening linker regions. Inspection of each type of linker sequence (refs. [57, 73, 75], and K.J.W. unpublished results) shows them to be uniformly short (from zero length to ~30 residues), and to incorporate a high proportion of proline, alanine, and charged residues. Thus, unlike many PKS interdomain regions that are likely to adopt specific folds, those present in NRPS systems appear to share characteristics with "classical" linkers, as exemplified by the interdomain sequences from the 2-oxo-acid dehydrogenase complexes.[18] Although such linkers are inherently flexible, a degree of rigidity is introduced through the presence of *trans* Ala–Pro peptide bonds.[76] These features serve to keep tethered domains apart, while allowing them to make essential interactions with other catalytic partners.

Several multidomain structures have also been reported, including the didomain EntB protein from the enterobactin NRPS (which contains functionally independent isochorismate lyase (ICL) and aryl carrier protein (ArCP) domains),[57] and a PCP–C didomain from the module 5–6 junction in the tyrocidine synthetase (referred to hereafter as $PCP_5$–$C_6$).[75] In the EntB structure, PCP protrudes from the dimeric ICL domain—an arrangement that presumably facilitates its interactions in trans with its partner adenylation domain EntE, and the downstream NRPS module EntF. Although the PCP and C domains are in intimate contact within the $PCP_5$–$C_6$ structure, their active sites are located ~50 Å apart—a distance that exceeds the reach of the 20 Å phosphopantetheinyl prosthetic group. Thus, this structure could represent a conformational state of the system prior to peptide-bond formation. The 18-residue linker region between PCP and C, while unstructured, participates in an intricate H-bonding network with both domains, and presumably stabilizes their respective orientations. However, the linker shows high inherent mobility, and should therefore accommodate the significant conformational adjustments and/or domain rearrangements necessary to allow the PCP and C domains to interact directly with each other.

Given the many analogies to PKS systems in terms of organization and reaction mechanism, it was anticipated that NRPSs would also operate as functional dimers.[77] Indeed, a shared quaternary architecture would neatly account for the ability of

NRPSs to form interfaces with PKS subunits in hybrid systems, as well as the presence of both NRPS and PKS modules within the same multienzymes.[78,79] However, analysis by the same methodologies used to demonstrate the homodimeric nature of FAS and PKS (size-exclusion chromatography, analytical ultracentrifugation, chemical cross-linking, and mutant complementation) failed to provide any evidence for physiologically relevant dimerization of subunits derived from the purely NRPS systems responsible for gramicidin (**6**), tyrocidine, and enterobactin (**7**) biosynthesis.[77,80] In addition, all individual NRPS domains whose structures have been solved, were purified as monomers.[71] On the other hand, analogous studies of the six-domain multienzyme VibF from the vibriobactin NRPS clearly showed it to be dimeric,[81] while gel-filtration studies of a mixed PKS–NRPS enzyme from the yersiniabactin pathway provided evidence for both monomeric and dimeric forms.[77] These findings have led to the proposal that NRPS subunits might exhibit a continuum of functional oligomerization states between monomers and dimers.[77]

## 2. CP-Based Interactions

Although the overall structure of PKS and NRPS modules remains unknown, the available data strongly suggest that mobile tethered elements (either the CP domains themselves and/or their attached "swinging arm" cofactors) will be a critical architectural feature. Another important conclusion from the ongoing structural work is that detailed insights into CP-based interactions are unlikely to come from single crystallographic snapshots of intact multienzymes. Such static images do not capture the dynamic aspects of chain extension (for example changes of domain position, and internal conformational motions), and furthermore, might not reveal the locations of the carrier proteins themselves.[34] One promising, though as yet untested idea, is to trap carrier proteins in the act of communicating with their partners by using specific mechanism-based inhibitors.[82,83] In the meantime, information on CP-based interactions is beginning to emerge from assays of specific interactions between the carrier proteins and their partners in solution, coupled with studies of dynamic domain behavior by solution-phase NMR spectroscopy.

### 2.1. ACP-based interactions

In analyzing ACP-based interactions, one useful starting point is to conceptualize the domain's movement as a "random walk" through the module. Unregulated encounters with the catalytic domains are possible because the enzymes can only operate when offered suitable substrates. For example, if the ACP were to present any of the reductive elements with a dicarboxylic extender unit, catalysis could not take place (Figure 8). After the chain is lengthened, the ACP could still engage in nonproductive interactions with the reductive loop activities without deleterious effects, as the DH domain can only operate after the KR has acted, and the ER can function only after catalysis by the DH. Therefore, the critical events in substrate processing are to position the intermediates within

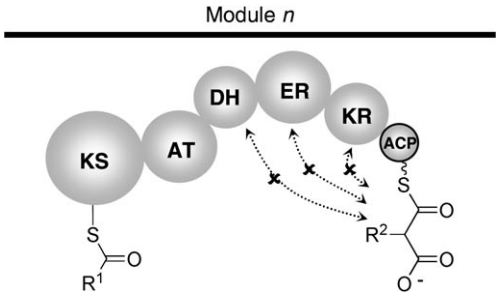



**Figure 8.** Analysis of ACP-based interactions. When the ACP is charged with carboxylated extender unit, its interactions (shown by black arrows) with the DH, ER, and KR domains are unproductive (indicated by an X).

reach of each reductive activity. In fact, the only programmed "choice" that each ACP has to make is to interact with domains within its own module, or to transfer the chain to a domain (KS or TE) located in the subsequent module. (The analogous decision for ACPs positioned upstream of NRPS modules in mixed systems, is when to participate in chain extension with the downstream C domain.)

This analysis leads to two possible mechanisms by which intramodular interactions can occur. In the first model, recognition by the catalytic domains is centered on the chain-extension intermediates (and possibly the Ppant tether), in the absence of a complex being formed with the ACP domain. The enforced proximity of the domains within the multienzymes obviates the need for specific interaction motifs between the ACP and its partners, as the high effective concentration of the substrates is sufficient to ensure catalytic efficiency. The decision to engage in an intra- or intermodular interaction would then be controlled by the substrate specificity of the downstream domain in the next module. Alternatively, the ACP and its partner domains, despite their covalent linkage, behave like discrete proteins. Complementary electrostatic docking sites, as postulated for fungal FAS,[43] could enable formation of ACP-catalytic domain complexes at each stage of the chain-extension cycle. Such interfaces are unlikely to be strong, however, as high affinity interactions would interfere with efficient substrate shuttling. Programmed changes in partnership could then be induced by subtle conformational changes on the ACP as it contacts the various forms of the chain-elongation intermediates, optimizing its surface features for recognition by the next catalytic domain. In this way, the ACP random walk could be constrained into a more economical series of protein–protein interactions with its partner activities.

#### 2.1.1. Studies of ACP-based interactions

On the basis of domain boundaries revealed by limited proteolysis[46,47,84] and the high-resolution structures (Figure 6),[50,51] module 3 from the DEBS PKS was dissected into its individual components ($KS_3$, $AT_3$, and $ACP_3$; the KR present in the module is inactive).[85] Critically, $KS_3$ and $AT_3$, when expressed as independent proteins retained their ability to self-acylate. Combining $ACP_3$ and acylated $AT_3$ resulted in transacylation to the

$ACP_3$ domain; this suggests that the domains share a defined interaction motif that enables reformation of an active chain-transfer complex. Addition of $KS_3$ was sufficient to reconstitute chain extension, although only in the presence of an $AT_3$ domain that incorporated its C-terminal linker region. In the KS–AT crystal structure, this linker appears to serve as a structural adaptor between the catalytic domains; this observation suggests that docking between the ACP and KS requires restoration of the native KS–AT didomain architecture. Furthermore, $KS_3$ exhibited preferences among heterologous ACPs derived from DEBS, as partners during chain extension.[84,86] Taken together, these results strongly implied the existence of specific molecular interfaces between the ACP and both its partner KS and AT domains.

Direct evidence for an interaction motif can be provided by identifying interface residues by using site-directed mutagenesis, as long as it can be demonstrated simultaneously that the mutations do not disrupt the structure or intrinsic catalytic properties of the domains. In the case of the KS–ACP interaction, attention was focused on helix αII by homology modeling[26] and computational docking by using solved structures.[50] Indeed, mutation of proposed interface residues on helix αII of DEBS $ACP_2$ to their counterparts in DEBS $ACP_6$, was sufficient to switch the condensation specificity of $ACP_2$ from DEBS $KS_3$ to $KS_6$.[55] Residues along the length of helix αII were also shown to lie at the interface between $ACP_6$ and the DEBS-specific PPTase, which acts on the domain, in trans.[26] Thus, helix αII is recognized by at least two of the ACP's partner domains. This mechanism is reminiscent of type II FAS systems, in which helix αII serves as a universal "recognition helix",[87–90] and highlights the fundamental mechanistic similarities between the two pathways.

An interesting question is whether this helix is also involved during chain transfer between ACP and KS domains located on successive modules. It cannot be assumed that the chain transfer and condensation complexes between ACP and KS domains are identical, given the very different connectivities between the participating proteins ($ACP_{n-1}+KS_n$ and $KS_n+ACP_n$ respectively). In fact, the existence of an alternative docking interface for chain transfer has been suggested by a study of modular "skipping". This phenomenon was observed in an engineered version of DEBS, in which an introduced module was simply ignored, resulting in omission of the expected chain-extension step.[91] In this case, the chain-extension intermediate was shown to pass through the interpolated module by direct ACP-to-ACP transfer.[92] This transesterification process requires that the thiol groups on the ACPs come together in close proximity, an arrangement that would arise naturally from simultaneous docking of two ACPs onto a single KS domain (Figure 9). If a conformational switch within the upstream ACP is required to interact with this alternative site on the KS domain, it could form the basis for substrate-induced programming at intermodular junctions.

The basis for communication between $ACP_6$ and the adjacent TE domain in DEBS has recently been investigated by using a combination of functional and binding assays.[93] These experiments did not provide any evidence for a strong interface between the TE domain and the apo ACP. An interaction was observed between the TE and the holo ACP, and the affinity of binding increased further by the presence of a model substrate, butyrate. These data suggest that molecular recognition occurs between the TE and prosthetic group (as well as attached substrates), and not with the ACP itself.[19] Similar conclusions have recently been reached for several recombinant KR domains from DEBS, which showed specificity for their β-ketoacylthioester substrates, but not for the ACP domains to which the substrates were tethered, or for the KS domains that synthesized the intermediates.[85]

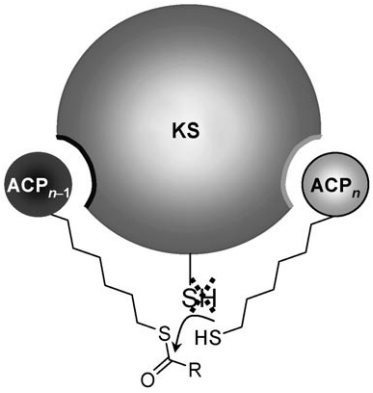



**Figure 9.** A two-site docking model for KS–ACP interactions. The existence of distinct docking sites on the KS for the upstream ($n-1$) and downstream ($n$) ACP domains explains how ACP-to-ACP chain transfer can occur in the absence of a functional KS domain.

Taken together, these experiments show that control of chain extension within PKS modules probably arises from a combination of the two mechanisms outlined above. Matched recognition motifs are likely to play a role within a subset of interactions (KS + ACP and AT + ACP), and these might be modulated by the internal conformational dynamics of the ACP domains.[55,84,86] In the case of the KS + ACP interaction, a higher affinity, longer-lived complex is probably required to satisfy the mechanistic and stereochemical constraints of the condensation reaction.[43,45] The remaining intramodular interactions (ACP + reductive activities), appear to be fostered by physical approximation of the domains within the multienzyme subunits. Intermodular communication between the ACP and the N-terminal domain of the downstream module within the same polypeptide (a KS or TE in PKS systems, or a C domain in mixed PKS–NRPS) should be facile as the domains are tethered directly through short linker sequences.[94] These linker regions show no evidence for sequence conservation within and among PKS systems (K.J.W., unpublished observations) and therefore are unlikely to serve as a shared mechanism for controlling the timing of intermodular interactions. Instead, changes in the conformational state of the ACP domain and/or the inherent substrate preferences of the downstream catalytic activities could suppress premature chain transfer between the modules.

### 2.2. PCP-based interactions

By analogy with PKSs, several mechanistic schemes can be proposed to account for the control of chain extension within NRPSs. In these systems, tailoring of the constituent amino acids can occur both before (for example, *N*-methylation) and after (for example, epimerization, *C*-methylation) the condensation reaction in each module.[95] In the simplest model, recognition is focused on the chain-extension intermediates. The inherent substrate specificities of the C domains at both their donor (upstream module) and acceptor (same module) sites would then control the directionality of chain elongation,[96] while the timing of modification would be regulated by the reaction chemistries (for example, epimerization, which is most favorable following condensation).[97] Alternatively, the sequence of PCP-based interactions (for example, with the *N*-methyltransferase, upstream C domain, epimerase, and then downstream C domain) could be regulated through formation of successive PCP-catalytic domain complexes. Similar considerations apply to mixed PKS–NRPS systems, in which NRPS modules are fused directly to PKS modules within the same polypeptide. In such cases, however, the PCP must choose to present its substrate to the domains within its own module, or to transfer the intermediate to the immediate downstream KS domain.

#### 2.2.1. Studies of PCP-based interactions

A particularly notable report[98] arose from re-examination of the original TycC3–PCP NMR solution structure.[59] Detailed analysis of the NMR spectra revealed the existence of two distinct conformational states for both the apo and holo forms of the domain (Figure 10). One of the conformations (the A/H state) is shared by both forms (despite the absence of the phosphopantetheine arm in the apo domain), and resembles the classical four α-helix bundle structure.[57–66] In the alternative A state of the apo protein, the lengths of helices αI, αII, and αIV are significantly reduced, and helix αIII is missing entirely; the corresponding loop region is buried within the core of the structure between helices αII and αIV. The major difference between the alternative H conformation of the holo protein and the A/H state is that helix αIII unravels and becomes extended, inducing a significant relocation of helix αII and loop elements, including the region that contains the active site Ser and its attached PPant. Together, these changes cause the cofactor to migrate some 100 Å across the face of the PCP, repositioning the terminal thiol group by ~16 Å. These data therefore provide the first direct evidence that the PPant arm "swings" during catalysis on NRPS multienzymes.

Using NMR titration experiments, the authors further demonstrated that the PPTase enzyme selectively interacts with the apo protein A state, and contacts some 25 residues along helix αII and the preceding loop, which are not fully accessible in the A/H state. Thus, despite the almost complete absence of sequence homology between ACP and PCP domains,[99] both exploit helix αII as an interaction element. The type II TE associated with the pathway exclusively recognizes the H state of the holo domain, through a contiguous interaction surface formed by the N-terminal part of helix αII and the loop II and III regions that flank it. Critically, in the A/H state, the corresponding amino acids are not adjacent to each other, which explains the selectivity of the TE for the H conformation. Although these experiments with trans-acting domains did not directly probe the control of chain extension, they bolster the idea that conformational changes within the PCP domain could contribute to programming within NRPSs.

Further support for the role of helix αII in partner recognition has been provided by genetic engineering. A hybrid TycC3 PCP domain was constructed by swapping helix αII with the equivalent region from a type II ACP of bacterial FAS.[99] Al-

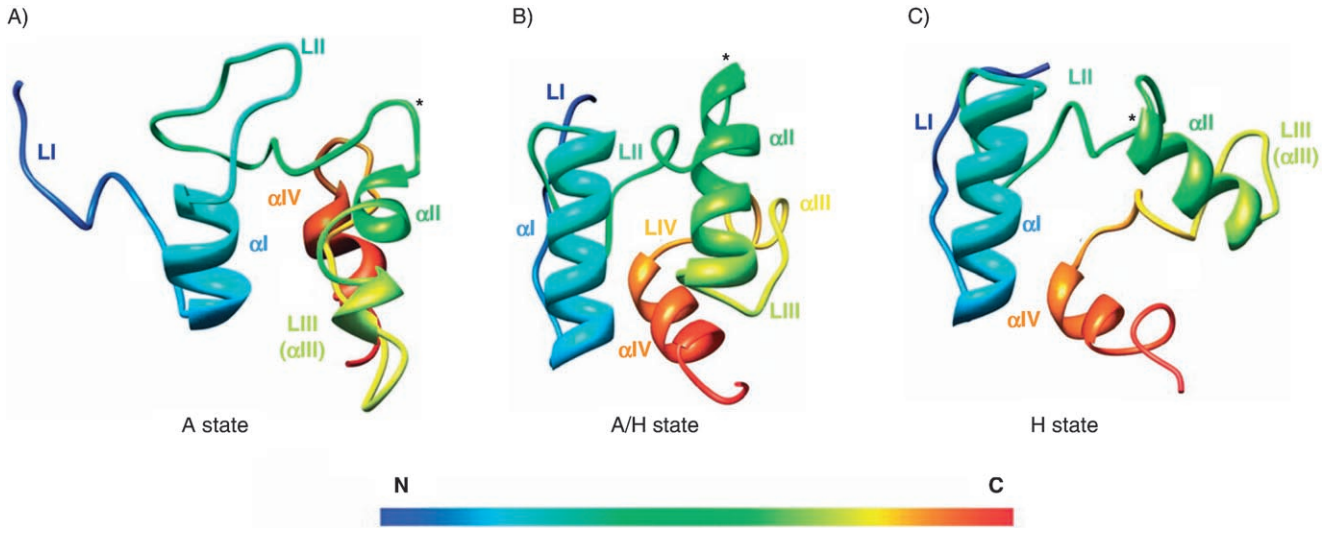



**Figure 10.** Ribbon diagrams of the average NMR solution structures of the TycC3 PCP domain. A) A state adopted by the apo protein. The PPTase Sfp interacts exclusively with this conformation. B) A/H state shared by both the apo and holo domains. This state most closely resembles the classical four α-helix bundle structure. C) H state adopted only by the holo protein. This state is recognized by the type II TE. In each case, the position of the active Ser is indicated by an asterisk. Reprinted with permission from ref. [98]; copyright 2006, American Association for the Advancement of Science.

though wild-type PCP was not recognized by the ACP-specific PPTase AcpS, the chimaeric PCP was efficiently phosphopantetheinylated both in vivo and in vitro. The hybrid PCP failed to interact with the neighboring E domain, suggesting that helix αII also participates in this interaction. In contrast, no effect was observed on the rate of in cis aminoacylation by the A domain directly upstream. Therefore, either the A domain recognizes an alternative motif on the PCP, or a specific interface is not required for interdomain communication. Targeted mutagenesis within helix αII further demonstrated that a single-point mutation was sufficient to alter the specificity of phosphopantetheinylation. Such "hot-spot" residues are typical of protein–protein interaction surfaces.[100]

A residue on helix αII was also identified at the interface between PCP and the downstream C domain in the Tyc $PCP_5$–$C_6$ crystal structure,[75] in which the A/H conformation of the PCP suggests that it is primed for interaction with the upstream A or C domains within its own module. In support of this hypothesis, mutation of several residues in helix αII of the TycB1 PCP compromised its ability to participate in chain extension with the upstream C domain.[101] However, mutation of helix αII within the PCP of TycA also impaired the ability of the domain to cooperate with a downstream C activity located on the next subunit, TycB.[99] One possible explanation for this result is that interactions between the TycA PCP domain and its aminoacyl substrate induced a switch between its A/H and H conformational states,[98] presenting residues on helix αII to the downstream C domain.

Additional interaction motifs for NRPS carrier proteins have been revealed by a series of studies on the synthetase responsible for biosynthesis of enterobactin (**7**) in *Escherichia coli* (Figure 11).[102–104] Enterobactin is a cyclic trilactone siderophore comprised of three *N*-(2,3-dihydroxybenzoyl)-serine (DHB–Ser) residues. Assembly occurs on an NRPS that consists of two modules, which are distributed among three proteins EntE, EntB, and EntF (Figure 11). EntE is a DHB-specific A domain, which delivers its cargo to the ArCP of the bidomain ICL–ArCP, EntB. ArCP-bound DHB is then condensed with serine by the C domain of the NRPS module EntF, to yield DHB–Ser. The products of three such elongation cycles are themselves condensed by the terminal TE domain of EntF, which then releases the final product enterobactin by macrolactonization. Both carrier proteins (ArCP and the integral PCP of EntF) are phosphopantetheinylated by the PPTase EntD. As the ArCP is a standalone protein, all of its interactions (with EntE, EntD, and EntF) occur in trans. Conversely, the EntF PCP cooperates in cis with partners within its own subunit (the C, A, and TE domains). Thus, the two contexts for carrier-protein recognition, in cis and in trans, can be probed within a single system. Furthermore, *E. coli* are dependent on the production of enterobactin for growth on iron-deficient media, which enabled a high-throughput, selection-based method for investigating recognition determinants.

In total, approximately 80% of the residues on the ArCP surface have been mutated, including helices αI, αII, αIII, and the intervening loops. The majority of amino acids were tolerant to mutation, suggesting that the recognition motifs on the ArCP are more localized than the interface regions on the Tyc PCP.[98] Another possibility, however, is that the particular selection methodology employed in these experiments was unsuited to revealing more subtle determinants of interdomain communication. The EntD–ArCP and ArCP–EntF in trans interactions are mediated, at least in part, by a small cluster of strongly conserved residues on either side of the phosphopantetheinylated Ser (Figure 11). In the case of EntD (PPTase), the interface is formed by loop II, while two of the three residues in the EntF interaction motif are provided by helix αIII, and the third by helix αII. In an independent study, the interface between the ArCP and EntE (A domain) was explored by rational site-directed mutagenesis,[57] and shown to involve residues on loop II and helices αII and αIII. Surface-exposed residues on helix αIII also make a critical contribution to the in cis interaction between the EntF PCP and the downstream TE.[104] Like helix αII, the structure of helix αIII is substantially altered during the PCP A/H-to-H transition.[98] Therefore, these results add further weight to a model for NRPS operation in which substrate-induced conformational switching in the PCP is used to program alternative interactions with the domain's up- or downstream partners.[20]

A module-dissection approach has also been applied to EntF,[105] as well as to the NRPS subunit EpoB from the epothilone mixed PKS–NRPS.[106] The HC domain, when excised from EpoB, was able to recognize the remainder of its module (A–Ox–PCP), but with reduced catalytic efficiency relative to the intact protein. In the case of EntF, the isolated C domain was completely incapable of interacting with the remaining three domains (A–PCP–TE) in the subunit, while the A domain within a C–A fragment was incompetent for transfer of seryl-AMP to the severed carrier-protein partner, PCP–TE. Evidently, the affinity of interaction between partner domains in these systems is insufficient to reconstitute the active complexes from their fragments. These data reinforce the idea that for both NRPS and PKS systems, covalent tethering of catalytic and carrier-protein domains into single multienzymes also plays a critical role in optimizing their interactions.

## 3. Intersubunit Recognition

As a consequence of the multi-multienzyme architecture of PKS, NRPS, and hybrid systems, intermodular interfaces are formed by domains that lie within the same polypeptide, but in addition, by domains located on discrete subunits (Figure 12). Therefore, in order to generate the correct product, the individual proteins within the assembly lines must organize themselves appropriately. This positioning requires not only that each polypeptide recognize its specific partner, but that improper associations are actively prevented. In principle, each subunit may not only have to discriminate among multienzymes within its own assembly line, but all other modular PKS and NRPS proteins that are present within the cell.[23,107] The structural basis for correct end-to-end docking is therefore one of the critical issues in understanding catalysis on modular megaenzymes.

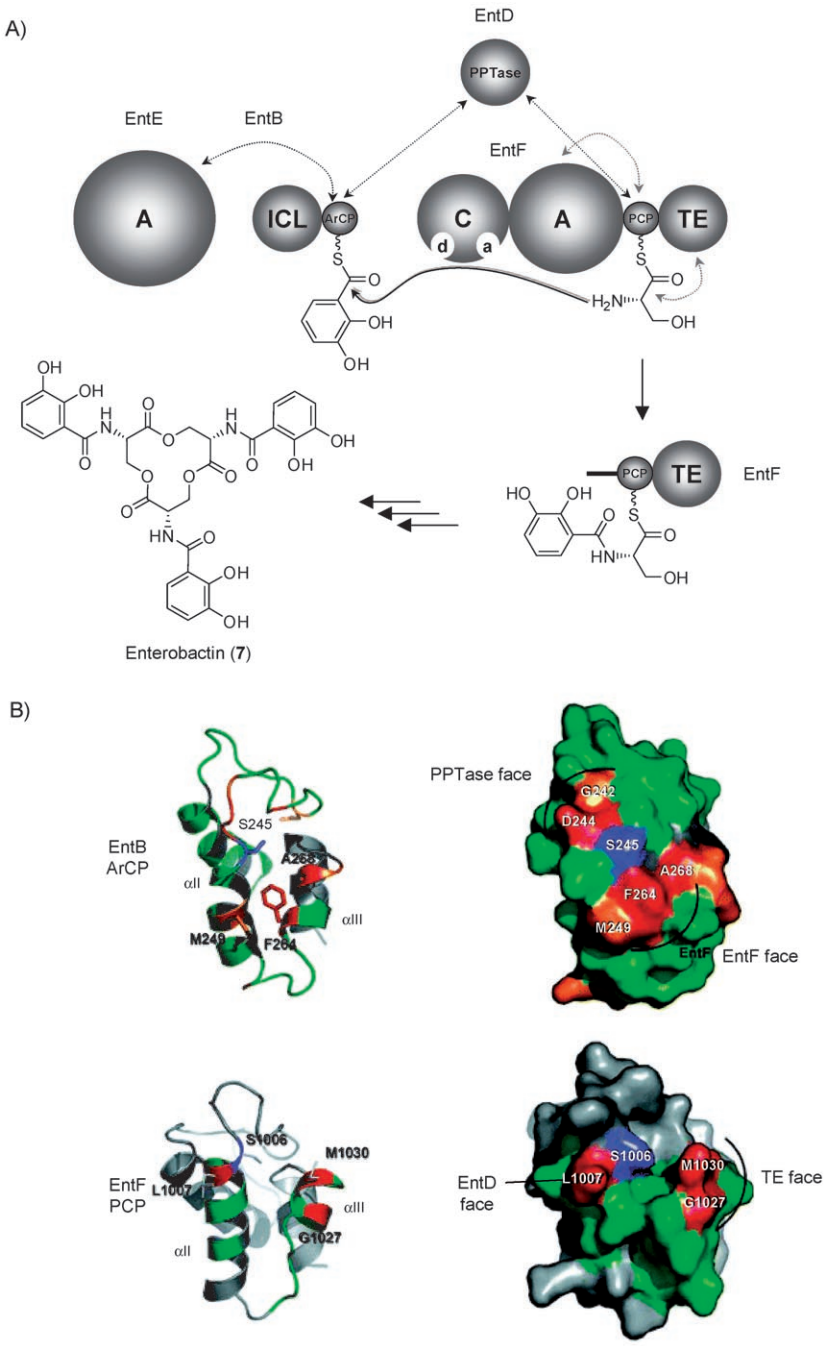



**Figure 11.** Protein–protein interactions in enterobactin biosynthesis. A) Biosynthetic scheme for enterobactin (**7**) assembly. Protein–protein interactions are indicated by double-headed, dashed arrows (black: in trans; gray: in cis). The ArCP of EntB must interact in trans with the PPTase EntD, the standalone A domain EntE, and the C domain of EntF, while the PCP of EntF interacts in cis with all of its modular partners, A, C, and TE. B) Localized recognition surfaces on the EntB ArCP and EntF PCP domains identified by combinatorial mutagenesis. The primary determinants of the interactions are shown in red, with secondary contributions from amino acids in orange. Reprinted with permission from ref. [20]; copyright 2006, American Chemical Society.

### 3.1. Multienzyme docking in modular PKS

The first hint as to how such discrimination might be accomplished in modular PKS came from recognition that the N termini of multienzymes, which house extender modules (i.e., the regions N-terminal of the conserved KS domains), contain regularities in their amino-acid sequences typical of amphipathic α-helical coiled coils.[108,109] This observation led to the proposal that these N termini are involved in specific coiled-coil interactions,[110] which stabilize the overall PKS homodimers. Shortly after, studies on DEBS demonstrated that these N-terminal "linker" regions make specific interactions with partner "linkers" at the extreme C termini of the previous PKS multienzyme (i.e., the sequences C-terminal to the conserved ACP domains).[27] These regions will be referred to hereafter as "docking domains", as this nomenclature reflects the fact that they adopt three-dimensional folds (see below).[28] Furthermore, matched pairs of docking domains could be substituted by other such partners without impairing the biological function of the hybrid PKSs,[111,112] and they could also mediate chain transfer between domains and modules that do not normally cooperate with each other.[113–115] Together, these results suggested that docking domains can function independently of their parent subunits—that is, that they are portable—and might adopt specific structures. However, it was not immediately obvious how to study their interactions directly, because PKS multienzymes bind to even their correct partners weakly, at least in vitro.[46,116,117]

This issue was addressed by solving the NMR solution structure of a complex of docking domains from the DEBS system, by fusing the elements together through their respective C and N termini.[28] Attempts to study the individual domains were hampered by their sensitivity to limited thrombolysis during purification,[28] and tendency to aggregate (K.J.W., unpublished results). Critically, an identical fusion approach had already been applied in vivo with DEBS and other PKS systems, with no apparent ill effects on the biosynthetic activity.[36,118,119] The solved structure (Figure 13) revealed that docking domains not only mediate specific intersubunit interactions, but, as proposed earlier,[108] promote associ-

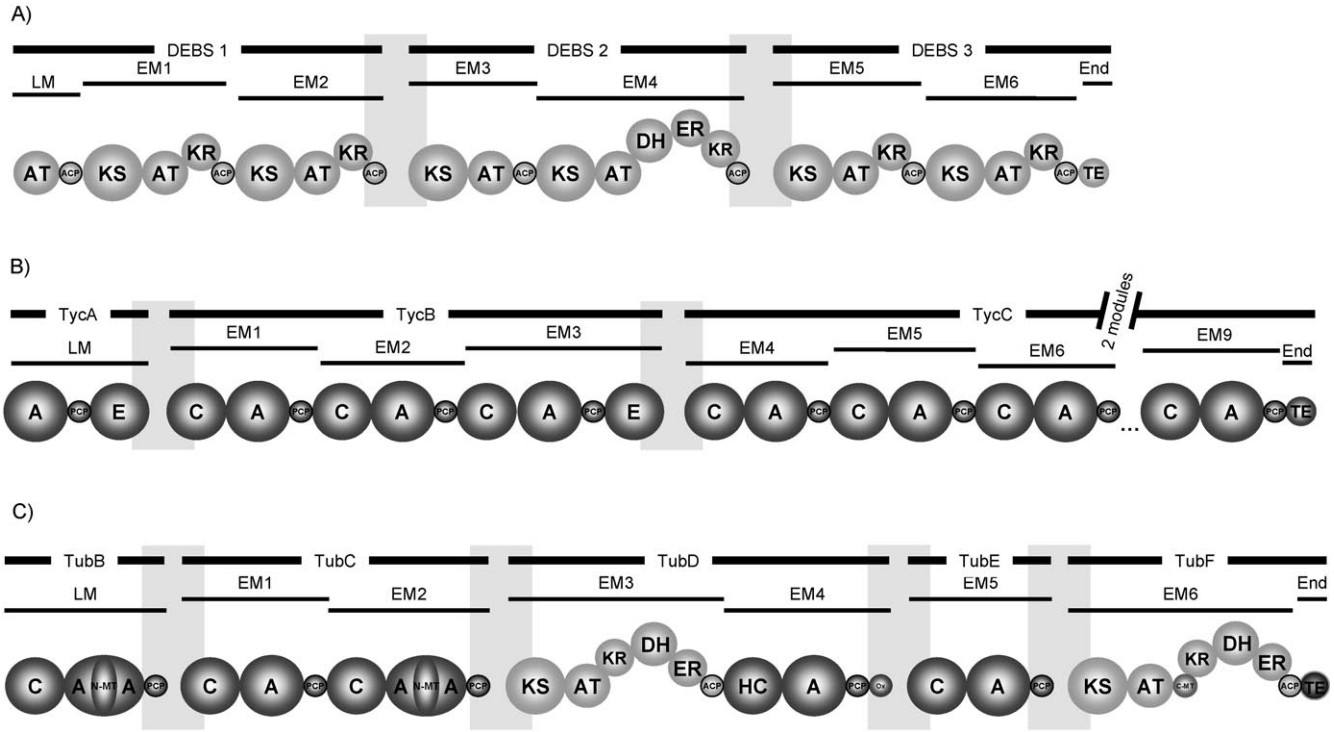



**Figure 12.** Multi-multienzyme organization of PKS, NRPS, and mixed PKS–NRPS systems. A) The 6-deoxyerythronolide B synthase (DEBS) responsible for erythromycin biosynthesis incorporates three subunits, DEBS 1, 2, and 3. Two intermodular junctions are formed across intersubunit interfaces (boxed gray regions). B) NRPS subunits TycA, TycB, and TycC cooperate to assemble the polypeptide tyrocidine. Chain extension occurs across two intersubunit interfaces (boxed gray regions). C) Tubulysin is assembled by a hybrid PKS–NRPS that incorporates three NRPS subunits (TubB, TubC, and TubE), a PKS subunit (TubF), and a mixed PKS–NRPS subunit (TubD). Both chain extension (NRPS–NRPS) and chain transfer (NRPS–PKS) are accomplished by domains located at intersubunit junctions (boxed gray regions). Abbreviations are: Ox, oxidase; C-MT, *C*-methyltransferase.

ation of each homodimer. The structure incorporates two separate dimerization elements, an unusual intertwined four α-helix bundle[120] formed by the first two helices of the C-terminal docking domain, and a coiled coil formed by the N-terminal docking domain. Reassuringly, the coiled-coil motif was later observed in the solved structure of the parent KS domain (Figure 6).[51] The docking interaction is localized to a second four α-helix bundle of different topology, formed when the third helix of the C-terminal docking domain wraps around the coiled-coil structure. A set of hydrophobic residues at the interface likely contributes to the stability of the docked complex. The structure also revealed several salt bridges at critical positions within the helical bundle (Figure 13); this suggests that association ($k_{on}$) between correct subunits is driven by favorable electrostatic interactions,[121] while misdocking is discouraged by charge–charge repulsion. Differences in stability between the resulting docked complexes ($k_{off}$) could contribute an additional layer of specificity. Replacement of particular helical segments within the docking complex by genetic engineering, yielded results in a minibiosynthetic system that were entirely consistent with this proposed docking model.[122] Taken together, these data suggest that this artificially docked complex captured the native interaction. Nonetheless, given the size of the overall structure, it is difficult to imagine how the docking domain region can be functionally equivalent to intraprotein linkers of only 20–30 residues, which mediate chain transfer between ACP and KS domains within the same PKS multienzyme.[28]

The subunit termini from many other PKSs share a high level of sequence homology with the DEBS-derived docking domains, making it likely that they adopt similar three-dimensional folds. However, sequence analysis reveals a second group of putative recognition elements, the structures of which might differ.[28] Thus, specificity appears to arise in many PKSs through the inclusion of sets of docking domains that are structurally incompatible. Nevertheless, in many systems, multiple docking interactions occur through DEBS-type docking domains;[28] for example, in the monensin (MON) PKS, the DEBS class of docking domain is present at all seven intersubunit junctions. Furthermore, analysis of all key residue positions (both hydrophobic and charged within the docking four α-helix bundle) shows that at least two docking domain pairs in the MON PKS are functionally identical. Thus, contacts between the docking domains alone are insufficient to insure interaction fidelity.[28,123] In such cases, interactions between surface residues on the flanking ACP and/or KS domains are highly likely to be involved in discriminating between correct and incorrect docking partners.[124] In support of this hypothesis, mutation of single residues on the surface of the DEBS docking complex significantly altered docking affinity.[125] This result demonstrates that the targeted amino acids are unlikely to be solvent exposed, but instead to lie within an interprotein interface that

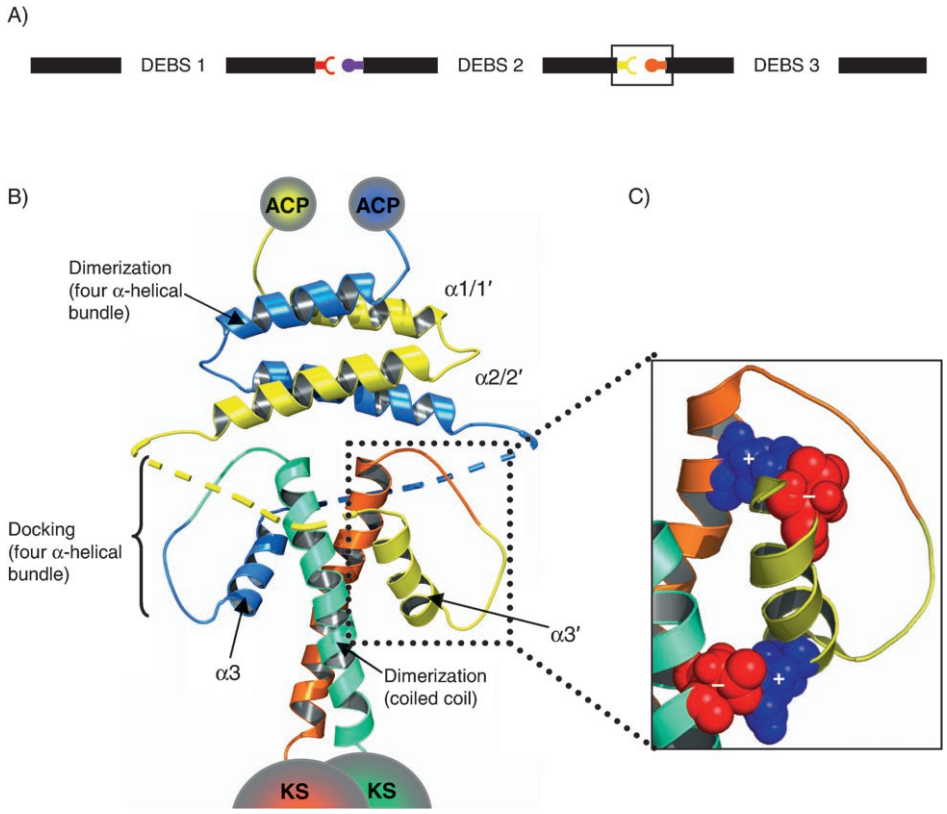



**Figure 13.** The structural basis for docking in modular PKS. A) PKS docking domains are located at the extreme C and N termini of the subunits. The complex of docking domains solved by NMR spectroscopy models the junction between polypeptides DEBS 2 and DEBS 3 in the erythromycin PKS. B) NMR solution structure of the DEBS docking complex. The dimeric C-terminal docking domain is shown in blue and yellow (three helices), while the dimeric N-terminal docking domain is shown in green and orange. Two dimerization elements are present, an intertwined four α-helical bundle formed by helices 1, 1′, 2, and 2′, and a coiled-coil motif formed by the N-terminal docking domain. Docking between the two domains, results in formation of a second four α-helical bundle, as indicated. The linker region between helices 2 and 3 is highly mobile, and therefore is represented as a dashed line in the structure. C) Charged residues located at critical positions in the interface (see box in B) are likely to contribute to the specificity of docking.

includes either one or both of the ACP and KS domains. In fact, it has been estimated that the adjacent domains make an equal contribution to docking specificity.[114] The precise way in which the flanking domains interact with each other and/or with the docking elements, remains to be determined.

These structural data together with considerations of molecular symmetry, support a model for assembly-line biosynthesis on PKS multienzymes[46] in which the proteins align themselves one after another, so that the crucial contacts occur between the ends of the subunits (Figure 13). A recently proposed alternative architecture for DEBS in which the successive polypeptides are stacked in antiparallel fashion,[67] is more difficult to reconcile with these data.

### 3.2. Multienzyme docking in NRPS

It has also been tacitly assumed that docking between successive enzymes in NRPS systems occurs at the subunit ends. While the N-terminal domain of NRPS polypeptides is most often a C domain (or its HC analogue), the C-terminal domain can be a PCP or, more commonly, an E domain. A role was proposed for the E domains in ordering the subunits within the synthetases through recognition of the downstream C domain.[126,127] However, further experiments demonstrated that the interaction motif is actually located C-terminal to the catalytic E domain, at the very end of the protein, and that the N termini of NRPS subunits contain corresponding recognition features. These regions have been dubbed "communication-mediating" (COM) domains.[128] COM domains are significantly shorter than the respective C- and N-terminal docking domains of PKSs, approximately 20–30 and 15–25 amino acids, respectively, and exhibit little mutual sequence conservation.[128,129] As in PKS systems, swapping of matched pairs of COM domains can also foster communication between noncognate multienzymes, demonstrating that portability is a shared feature of these elements.[128–130] Strikingly, the decisive factor in establishing a productive interface in these experiments was a matched set of COM domains. This result suggests that the flanking domains make only a minor contribution to the interactions, although a role for specific PCP–C domain recognition has been postulated.[130]

A structural model for COM-based interactions has been proposed on the basis of sequence analysis, site-directed mutagenesis, and structural data.[129] Although the overall sequence homology among COM domains is low, the amino-acid composition is relatively uniform. The C-terminal COM domains incorporate a higher than average proportion of acidic amino acids, while N-terminal COM domains are biased towards polar residues (Figure 14). This observation suggests that, as in PKS systems, specificity is mediated largely by polar and/or electrostatic interactions between key residues. Indeed, reversal of the charge of a single amino acid of a Tyc COM domain was apparently sufficient to switch its partner specificity.[129] Based on the crystal structure of the VibH C domain[71] and secondary structure prediction, both C- and N-terminal COM domains are believed to adopt α-helical structures. Therefore, docking between compatible pairs is proposed to result in the formation of an antiparallel leucine zipper. Crucially, the five putative specificity determinants on each COM domain occur at approximately every third position, and so should be displayed on the

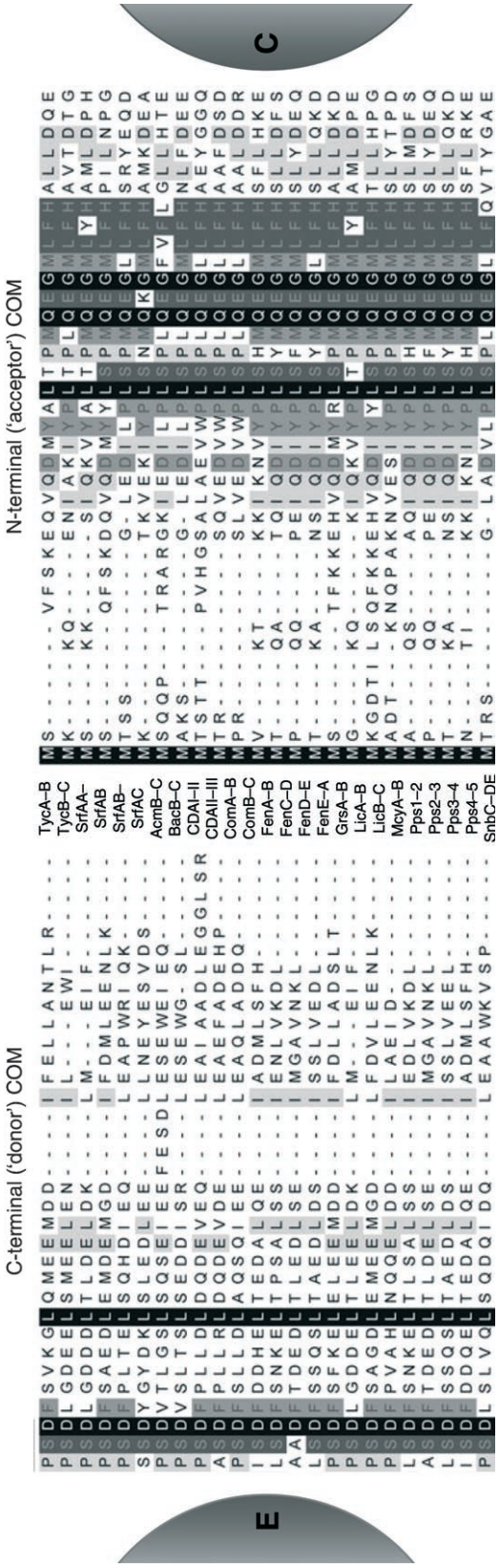



**Figure 14.** Sequences of proposed sets of C- and N-terminal COM domains from twelve biosynthesis systems. The subunits from which the COM domains are derived are indicated. Abbreviations: Tyc, tyrocidine; Srf, surfactin; Acm, actinomycin; Bac, bacitracin; CDA, calcium-dependent antibiotic; Com, complestatin; Fen, fengycin; Grs, gramicidin; Lic, lichenysin; Mcy, microcystin; Pps, pliplastatin; Snb, pristinamycin. Reprinted and adapted with permission from ref. [129]; copyright 2006, National Academy of Sciences, USA.

same side of the respective helices. Confirmation of this proposal from high-resolution structural information is eagerly anticipated.

### 3.3. Multienzyme docking in mixed PKS–NRPS

Mixed PKS–NRPS systems can incorporate as many as four different types of intersubunit interfaces: PKS–PKS, PKS–NRPS, NRPS–PKS, and NRPS–NRPS. As analogous PKS–PKS and NRPS–NRPS junctions occur in purely PKS and NRPS systems, in principle, the mechanisms of intersubunit recognition could also be shared. However, it is immediately obvious from primary sequence analysis that this is unlikely to be the case. Putative docking elements from mixed systems do not exhibit discernible homology to PKS docking domains or NRPS COM domains, despite the high level of sequence similarity between the corresponding catalytic activities.[94,131] Nevertheless, both C- and N-terminal docking elements show obvious mutual sequence similarities, and typically cluster according to their module of origin (either PKS or NRPS).[131] The exception is a group of PKS C-terminal docking domains that operate at PKS–NRPS interfaces, which align instead with their counterparts from NRPS modules. This analysis suggests that docking is likely to occur in a common fashion at PKS–PKS junctions in many mixed systems, and by a second type of mechanism at PKS–NRPS and NRPS–NRPS interfaces (Figure 15). However, many putative docking domains from these systems fall outside of the identified sequence groups (for example, those that operate at NRPS–PKS junctions), and so further types of interfaces are possible. Thus, as in purely PKS systems, one probable determinant of interaction specificity at intersubunit junctions is the presence of architecturally orthogonal sets of docking domains.

Similarly, sequence analysis suggests that multiple interfaces within mixed assembly lines are formed by docking domains that adopt the same overall fold (for example, between subunits MelC–D, MelD–E, and MelG–F in the melithiazol system[131]). Insight into the molecular basis for specificity at such junctions was recently provided by the NMR solution structure of a representative N-terminal docking domain, called TubCdd (Figure 15).[131] TubCdd operates at the intersection between NRPS subunits TubB and TubC in the tubulysin mixed megasynthetase, but members of its domain family are also present at PKS–NRPS interfaces. The solution structure revealed that TubCdd is a homodimer. This result suggests that while subunits in purely NRPS systems can be monomeric,[77,80] their counterparts in hybrid PKS–NRPS are likely to oligomerize in order to communicate effectively with their homodimeric PKS partners, and that the docking elements actively contribute to subunit self-association. This result argues against a proposed structure for hybrid enzymes that consists of a dimeric PKS core with monomeric NRPS loops.[77] The study also revealed that TubCdd adopts a novel $\alpha\beta\beta\alpha\alpha$ protein fold that features an exposed $\beta$-hairpin, which serves as the binding site for the C-terminal docking domain of the partner polypeptide. The pattern of charged residues on the surface of the $\beta$-hairpin appears to define an electrostatic "code" for docking at this type

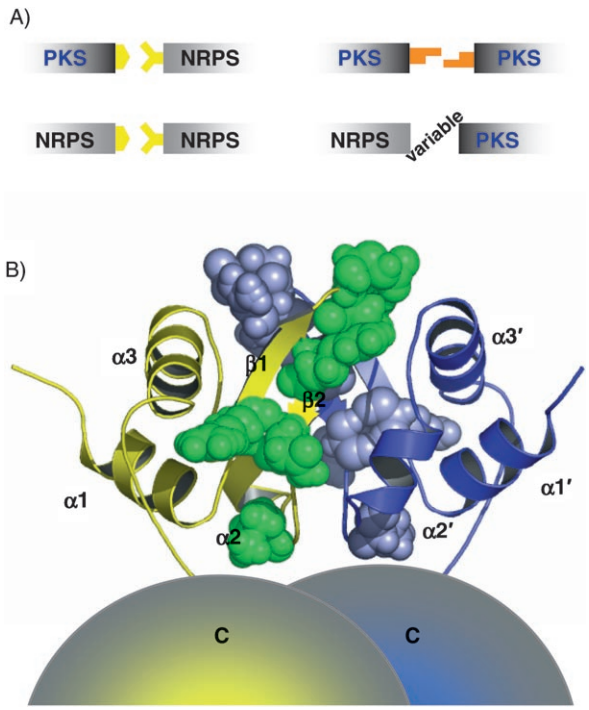



**Figure 15.** Intersubunit docking in mixed PKS–NRPS systems. A) Sequence analysis predicts that docking occurs by a common mechanism at PKS–NRPS and NRPS–NRPS junctions in mixed systems, while an alternative recognition architecture is operative at PKS–PKS interfaces. Docking elements at NRPS–PKS intersections exhibit no apparent homology to other docking domains, impling the existence of additional interface types.[131] B) NMR solution structure of the N-terminal docking domain from the NRPS subunit TubC (TubCdd). Each monomer of the overall homodimeric complex exhibits a novel $\alpha\beta\beta\alpha\alpha$ fold. Residues on either side of the dimeric complex that contact the partner docking domain of TubB, are indicated.

of interface. Although the C-terminal partner of TubCdd, TubBdd, is unstructured when expressed as a discrete protein, it might fold through interaction with TubCdd. In either case, recognition features within TubBdd are likely to be centered on the short, acidic-charge cluster at its extreme C terminus,[94,131] as only this region is conserved throughout the docking-domain group. In future, it will be important to test features of this docking model by genetic engineering of intact multienzymes, as performed with the DEBS system.

Studies with the epothilone mixed PKS–NRPS system have demonstrated that several of its constituent docking domains are portable, and can mediate communication between both cognate and noncognate partners.[132,133] However, the chain-elongation efficiency of the resulting hybrids (at both PKS–NRPS and NRPS–PKS junctions) was compromised,[132] even with native substrates. This result implies that, as in PKS systems, intersubunit communication involves additional molecular-recognition features that possibly include matched interaction interfaces between the flanking domains. A straightforward mechanism to introduce this added level of discrimination would be to optimize ACP or PCP domains to interact only with a downstream C or KS domain, respectively; such a feature would at least define the specific type of interpolypeptide junction. This alternative specificity should, in principle, be reflected in systematic sequence differences between domains in purely PKS or NRPS systems, and their counterparts in mixed systems (for example, C-domain-specific ACPs and their KS-preferring counterparts). However, neither ACPs nor PCPs from mixed assembly lines exhibit any distinguishing sequence features relative to the corresponding domains from entirely PKS and NRPS systems.[94] And while analysis of KS domains that function directly downstream of NRPS modules shows that they form a unique branch within the KS phylogenetic tree, the residues that differ have been implicated in adapting their substrate specificity to accommodate peptidyl intermediates.[94] Thus, evaluating whether such recognition motifs exist awaits high-resolution structural information on intersubunit junctions, or direct interrogation of putative interface residues by site-directed mutagenesis.

It is interesting that mixed assembly lines use an alternative docking strategy to those in purely PKS and NRPS systems, despite the conservation of catalytic functions. Such divergence could reflect a different evolutionary origin of the docking elements in hybrid systems. Alternatively, the change in oligomerization state of NRPS subunits on transition from purely NRPS to mixed systems could have driven the development of alternative forms of docking at all relevant junctions (NRPS–PKS, PKS–NRPS, and NRPS–NRPS). The factors favoring evolution of a second docking solution for PKS–PKS intersections are less obvious.

## 4. Domains that Operate in trans

This section describes several further protein–protein interaction issues in megaenzyme systems, about which relatively little is known. Every PKS, NRPS, or hybrid system depends on the in trans action of a PPTase enzyme or enzymes for post-translational activation of the constituent carrier proteins. For both ACPs and PCPs, recognition appears to be mediated by residues in the loop II and helix $\alpha$II regions, which lie directly adjacent in the primary sequence to the active Ser.[26,98] Many assembly lines also incorporate a type II TE editing function.[31–33] In the case of PCP domains at least, the PCP–TE interface is also formed by residues on helix $\alpha$II, and adjacent loops.[98]

A large and growing number of pathways feature further trans-acting domains. Notable among them are the group of "AT-less" PKS, which lack integral AT domains within each module (Figure 16).[134] Instead, one or several ATs are encoded elsewhere in the cluster, as monodomains,[29,135,136] tandem ATs,[137,138] or as genetic fusions with proteins of unrelated function, such as oxidase domains.[30,139,140] Each AT interacts iteratively with every carrier protein within the synthase to deliver a common substrate, usually malonate. It has been proposed that the sequences between the KS and KR domains within the modules in many systems, which are likely to be relics of functional ATs, serve as "AT docking domains" and restore the normal positioning of the AT within the complexes.[141] However, it is unclear why such a vestigial domain should interact with an active copy of the AT,[140] particularly given the monomeric nature of the ATs observed in the KS–AT crystal structures.[50,51] A reasonable alternative is that the trans AT domains

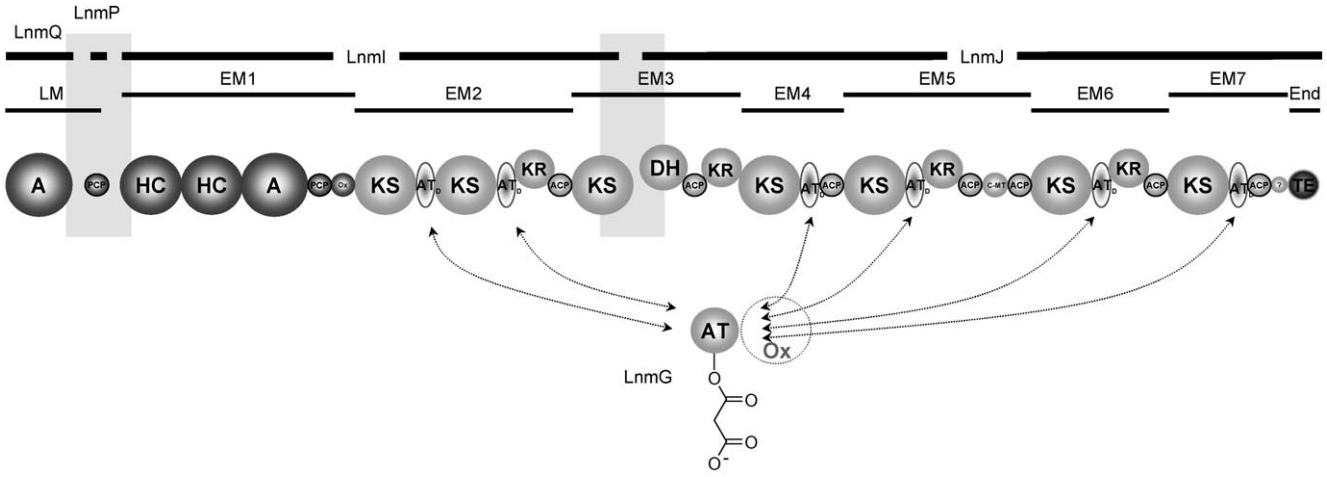
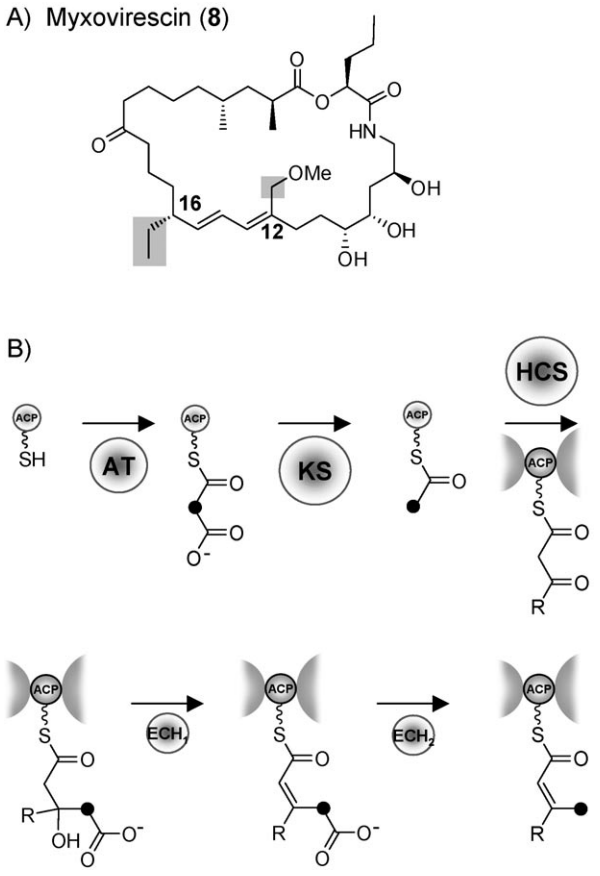



**Figure 16.** Architecture of the "AT-less" leinamycin hybrid PKS–NRPS, showing the location of the proposed "AT docking domains" ($AT_D$). The discrete AT (located on the didomain LnmG), has been proposed to dock onto the $AT_D$ domains, in order to deliver the common substrate malonate to the ACP domains.

dock directly with the ACP domains through recognition of a common surface feature. This hypothesis is supported by the finding that exogenous MAT from *Streptomyces coelicolor* restored the function of a DEBS module in which the native AT domain had been disabled by active-site directed mutagenesis;[142] in this case, the complemented module lacked a putative AT docking domain.

The proposed role of the AT-docking domain has been explored directly within the context of the mycosubtilin assembly line.[143] Surprisingly, inclusion of the upstream AT-docking domain in a recombinant ACP construct significantly reduced the rate of malonate transfer by the discrete AT, FenF. The authors postulated that FenF might form a long-lived or catalytically incompetent complex with the docking domain, in order to control the timing of malonyl transfer. The need for such a regulatory function is unclear, however, given that transfer of malonate to the ACP can only occur when the active site is unoccupied. FenF was also unable to discriminate significantly between the two possible ACP substrates within the assembly line, suggesting a role for additional recognition determinants in the interaction. Taken together, these data indicate that the residual fragments of AT domains found in AT-less systems are not involved in AT docking, but further exploration of this issue is clearly warranted.

Another modification that occurs in trans is the addition of alkyl functionalities to the β-position of nascent chains, as found in myxovirescin (**8**; Scheme 2), bacillaene, mupirocin, and curacin, among other metabolites.[137,138,144–148] Introduction of such β branches requires a cassette of five proteins, comprising a 3-hydroxy-3-methylglutaryl-CoA synthase homologue (HCS), a decarboxylase-type KS domain (which incorporates an active-site Cys to Ser substitution), a discrete ACP, and two enoyl-CoA hydratase homologues (ECH). The modification cycle is proposed to begin with loading of the alkyl precursor (either malonate or methylmalonate) onto the ACP, followed by KS-catalyzed decarboxylation; alternatively the corresponding alkyl chains are loaded directly onto the KS. Condensation




**Scheme 2.** Modification by β-alkylation. A) The mixed PK–NRP metabolite myxovirescin (**8**) contains two alkyl groups (at C12 and C16, highlighted in gray), which are introduced by the action of a 5-member "HMG gene cassette". B) Mechanism of β-methylation. A stand-alone ACP is acylated with malonate, presumably by the action of a discrete AT. The malonate is then decarboxylated by a free-standing KS domain. The resulting acetate unit is then added to an ACP-bound chain-extension intermediate by HMG–CoA synthase (HCS), followed, sequentially, by elimination of water and decarboxylation, which are catalyzed by two enoyl–CoA hydratase (ECH) homologues. Recognition between the HCS, ECHs, and ACP might increase the efficiency of the condensation, dehydration, and decarboxylation steps.

of the resulting alkyl unit (as its enolate nucleophile) with the β-keto group of the multienzyme-bound intermediate is accomplished by the HCS. Dehydration and decarboxylation catalyzed by the ECH enzymes then furnish the final alkyl functionality. The timing of β-alkylation critically depends on selection of the correct chain-extension intermediate by the HCS enzyme, and must occur before the chain is processed further. In principle, the HCS could be specific for a particular polyketide intermediate. However, the demonstrated promiscuity of HCS enzymes towards alternative acceptor substrates[135,146] suggests that additional recognition features are required, and likely include the modular ACP to which the substrate is tethered. Indeed, data have been obtained that support the formation of a specific complex between the HCS and the discrete ACP bearing the alkyl functionality.[145] Analogous interaction motifs might also facilitate the efficient processing of the alkylated intermediate by the ECHs. Additional studies will be required to fully elucidate the recognition requirements for β-branch incorporating enzymes.

A final feature of interest in several mixed PKS–NRPS systems is the in trans operation of domains that normally function in cis. This situation arises with the so-called "split modules" found primarily in myxobacterial systems, in which the normal complement of domains that comprises a module is shared between two different proteins (Figure 17).[135,138–140,149,150] Module splitting can occur at multiple sites (for example, following the KS,[135,139] AT,[150] DH,[140] or KR[135,135,149] domains) and is taken to an extreme in the andrimid PKS–NRPS, in which 15 domains are distributed over nine proteins.[151,152] It has been proposed, in some cases, that the splitting observed on the genetic level might not be realized in the protein structure, due to translational by-passing.[139] Alternatively, however, specific interactions between modular components (which likely include the KS and ACP domains as these are always located on the separate proteins of split modules) or supramolecular interactions among multiple polypeptides as proposed for the bacillaene PKS–NRPS,[153] are sufficient to reform the active complexes. Direct studies of split modules in vitro will be required to understand this phenomenon.

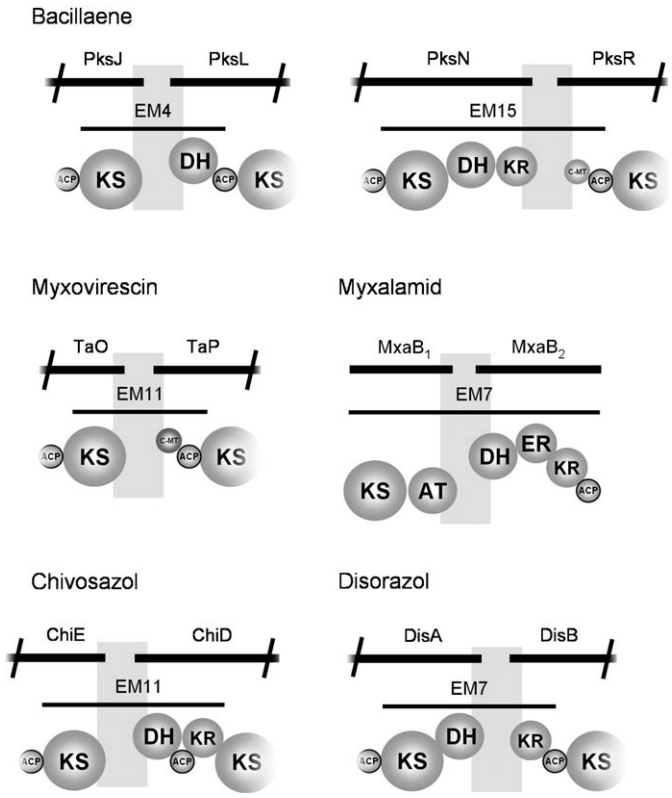



**Figure 17.** Examples of biosynthetic systems that incorporate "split"-extension modules. In each case, the subunit names containing the split module are indicated.

## 5. Summary and Outlook

In type II FAS and PKS systems in which all domains are present as individual components, interactions between most if not all of the proteins must arise through compatible molecular interfaces.[19] In contrast, within the modules of PKS, NRPS, and their hybrids, the majority of CP-based interactions occur with domains to which they are physically attached, albeit in many cases indirectly. In principle, this arrangement allows the number of essential protein–protein interactions to be reduced relative to type II systems, with the exception of those with PPTase enzymes, and possibly TEII functions. Nevertheless, at least for several interactions within modular megasynthetases, specific interfaces do play a role. In PKS systems, evidence supports the formation of KS–ACP complexes,[55,84,86] both during the condensation reaction and to maintain the fidelity of intersubunit transfer. Correspondingly, C–PCP interfaces appear to be operative in NRPSs.[99,101] The participation of such recognition motifs likely allows the systems to program one of the most critical events in the pathways—the switch between intra- and intermodular communication. Commonalities are also evident in the particular surface features characterized to date, which are used by the CP domains to recognize their partners. Although it is logical that such contact loci should center around the active Ser residues, both ACPs and PCPs exploit residues in helices αII and αIII and the surrounding loops.[19] It is unlikely to be coincidence that these regions of the proteins show the greatest structural flexibility.[58,62,98] This observation supports the idea that conformational switching within the CPs, possibly provoked by interaction with chain-extension intermediates, allow the proteins to adapt structurally to one or the other of their specific partners, giving rise to a programmed series of interactions at intermodular junctions.

Further commonalities are observed in the strategies to achieve docking specificity between subunits in PKS, NRPS, and mixed systems, although the particular structural solution used in each case differs. For both PKSs and PKS–NRPSs, more than one type of docking element type may be present, which would clearly tend to suppress mispairing of individual multienzyme subunits at those interfaces. However, in all three systems, multiple interfaces are formed by docking domains that adopt the same overall fold, and which therefore might be expected to compete for binding to "wrong" partners. Here, the contribution of the docking domains to binding specificity ap-

pears to depend upon precise interactions between a small set of polar or charged residues that occupy critical positions in the interfaces.[28,128,129,131] In PKS and mixed systems, at least, the flanking catalytic domains are also likely to make a decisive contribution. The underlying logic governing intersubunit recognition is therefore very similar. The reliance on charged rather than hydrophobic residues at intersubunit junctions is consistent with the overall weak affinity of intersubunit interactions. This feature of the pathways might allow these systems to cope with the mistakes in translation or protein folding that likely arise with proteins of this size, as defective proteins can be easily exchanged for functional alternatives.

Many of the protein–protein interaction models discussed in this review await direct experimental evaluation. In addition, the majority of CP-based interactions within PKS and NRPS modules have yet to be probed comprehensively. Such experiments should reveal whether further interdomain interactions involve the formation of specific protein–protein complexes, or whether proximity is also the primary determinant of interaction efficiency in these cases. Research in the field is also likely to focus on obtaining high-resolution crystallographic information on larger, multidomain portions of the complexes, including entire modules and subunits, and using NMR spectroscopy to probe the dynamic aspects of these gigantic protein complexes. In future, deciphering in detail the rules for recognition of (acyl)carrier proteins, subunits, and trans-acting domains, will undoubtedly improve our ability to productively engineer megasynthetase systems towards the production of novel drug candidates.

## Acknowledgements

*The authors gratefully acknowledge Profs. Peter Leadlay and Jim Staunton for critical reading of this manuscript. We would also like to thank the students and postdoctoral researchers whose work contributed to this review. Their names appear in the reference list. Research in R.M.'s laboratory is supported by the Deutsche Forschungsgemeinschaft (DFG) and the Bundesministerium für Bildung und Forschung (BMB+F). K.J.W's work on protein–protein interactions has been additionally supported by the Royal Society, the Biotechnology and Biological Sciences Research Council (BBSRC), and the Alexander von Humboldt Foundation.*

**Keywords:** biosynthesis · natural products · nonribosomal polypeptide synthetase · polyketide synthase · protein–protein interactions

[1] J. Clardy, C. Walsh, *Nature* **2004**, *432*, 829–837.
[2] C. T. Walsh, *Acc. Chem. Res.* **2008**, *41*, 4–10.
[3] C. T. Walsh, *Science* **2004**, *303*, 1805–1810.
[4] J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380–416.
[5] J. Grunewald, M. A. Marahiel, *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 121–146.
[6] U. Rix, C. Fischer, L. L. Remsing, J. Rohr, *Nat. Prod. Rep.* **2002**, *19*, 542–580.
[7] F. Kopp, M. A. Marahiel, *Nat. Prod. Rep.* **2007**, *24*, 735–749.
[8] G. L. Challis, J. Ravel, *FEMS Microbiol. Lett.* **2000**, *187*, 111–114.
[9] S. Lautru, R. J. Deeth, L. M. Bailey, G. L. Challis, *Nat. Chem. Biol.* **2005**, *1*, 265–269.
[10] T. P. Stinear, A. Mve-Obiang, P. L. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. Johnson, J. K. Davies, R. E. Lee, S. Adusumilli, T. Garnier, S. F. Haydock, P. F. Leadlay, S. T. Cole, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1345–1349.
[11] R. Pieper, S. Ebert-Khosla, D. Cane, C. Khosla, *Biochemistry* **1996**, *35*, 2054–2060.
[12] R. Pieper, R. S. Gokhale, G. Luo, D. E. Cane, C. Khosla, *Biochemistry* **1997**, *36*, 1846–1851.
[13] M. Bycroft, K. J. Weissman, J. Staunton, P. F. Leadlay, *Eur. J. Biochem.* **2000**, *267*, 520–526.
[14] D. B. Stein, U. Linne, M. Hahn, M. A. Marahiel, *ChemBioChem* **2006**, *7*, 1807–1814.
[15] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5848–5853.
[16] A. R. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman Company, New York **1998**.
[17] H. O. Spivey, J. Ovadi, *Methods* **1999**, *19*, 306–321.
[18] R. N. Perham, *Annu. Rev. Biochem.* **2000**, *69*, 961–1004.
[19] A. C. Mercer, M. D. Burkart, *Nat. Prod. Rep.* **2007**, *24*, 750–773.
[20] J. R. Lai, A. Koglin, C. T. Walsh, *Biochemistry* **2006**, *45*, 14869–14879.
[21] R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla, C. T. Walsh, *Chem. Biol.* **1996**, *3*, 923–936.
[22] L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, C. T. Walsh, *Biochemistry* **1998**, *37*, 1585–1595.
[23] M. Oliynyk, M. Samborskyy, J. B. Lester, T. Mironenko, N. Scott, S. Dickens, S. F. Haydock, P. F. Leadlay, *Nat. Biotechnol.* **2007**, *25*, 447–453.
[24] H. Ikeda, J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, S. Omura, *Nat. Biotechnol.* **2003**, *21*, 526–531.
[25] R. Castonguay, W. He, A. Y. Chen, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2007**, *129*, 13758–13769.
[26] K. J. Weissman, H. Hong, B. Popovic, F. Meersman, *Chem. Biol.* **2006**, *13*, 625–636.
[27] R. S. Gokhale, C. Khosla, *Curr. Opin. Chem. Biol.* **2000**, *4*, 22–27.
[28] R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay, K. J. Weissman, *Chem. Biol.* **2003**, *10*, 723–731.
[29] J. Piel, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14002–14007.
[30] Y. Q. Cheng, G. L. Tang, B. Shen, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3149–3154.
[31] M. L. Heathcote, J. Staunton, P. F. Leadlay, *Chem. Biol.* **2001**, *8*, 207–220.
[32] D. Schwarzer, H. D. Mootz, U. Linne, M. A. Marahiel, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14083–14088.
[33] E. Yeh, R. M. Kohli, S. D. Bruner, C. T. Walsh, *ChemBioChem* **2004**, *5*, 1290–1293.
[34] S. Jenni, M. Leibundgut, T. Maier, N. Ban, *Science* **2006**, *311*, 1263–1267.
[35] H. Jenke-Kodama, A. Sandmann, R. Müller, E. Dittmann, *Mol. Biol. Evol.* **2005**, *22*, 2027–2039.
[36] C. Olano, B. Wilkinson, S. J. Moss, A. F. Brana, C. Mendez, P. F. Leadlay, J. A. Salas, *Chem. Commun.* **2003**, 2780–2782.
[37] J. He, C. Hertweck, *Chem. Biol.* **2003**, *10*, 1225–1232.
[38] S. Smith, S. C. Tsai, *Nat. Prod. Rep.* **2007**, *24*, 1041–1072.
[39] F. J. Asturias, J. Z. Chadick, I. K. Cheung, H. Stark, A. Witkowski, A. K. Joshi, S. Smith, *Nat. Struct. Mol. Biol.* **2005**, *12*, 225–232.
[40] S. Smith, *Science* **2006**, *311*, 1251–1252.
[41] C. A. Townsend, J. M. Crawford, T. Bililign, *Chem. Biol.* **2006**, *13*, 349–351.
[42] E. Schweizer, J. Hofmann, *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 501–517.
[43] I. B. Lomakin, Y. Xiong, T. A. Steitz, *Cell* **2007**, *129*, 319–332.
[44] S. Jenni, M. Leibundgut, D. Boehringer, C. Frick, B. Mikolasek, N. Ban, *Science* **2007**, *316*, 254–261.
[45] M. Leibundgut, S. Jenni, C. Frick, N. Ban, *Science* **2007**, *316*, 288–290.
[46] J. Staunton, P. Caffrey, J. F. Aparicio, G. A. Roberts, S. S. Bethell, P. F. Leadlay, *Nat. Struct. Biol.* **1996**, *3*, 188–192.
[47] J. F. Aparicio, P. Caffrey, A. F. Marsden, J. Staunton, P. F. Leadlay, *J. Biol. Chem.* **1994**, *269*, 8524–8528.
[48] C. M. Kao, R. Pieper, D. E. Cane, C. Khosla, *Biochemistry* **1996**, *35*, 12363–12368.

[49] R. S. Gokhale, J. Lau, D. E. Cane, C. Khosla, *Biochemistry* **1998**, *37*, 2524–2528.
[50] Y. Tang, A. Y. Chen, C. Y. Kim, D. E. Cane, C. Khosla, *Chem. Biol.* **2007**, *14*, 931–943.
[51] Y. Tang, C. Y. Kim, I. I. Mathews, D. E. Cane, C. Khosla, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11124–11129.
[52] A. T. Keatinge-Clay, R. M. Stroud, *Structure* **2006**, *14*, 737–748.
[53] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898–908.
[54] C. D. Richter, D. A. Stanmore, R. N. Miguel, M. C. Moncrieffe, L. Tran, S. Brewerton, F. Meersman, R. W. Broadhurst, K. J. Weissman, *FEBS J.* **2007**, *274*, 2196–2209.
[55] V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi, C. Khosla, *Protein Sci.* **2007**, *16*, 2093–2107.
[56] M. A. Reed, M. Schweizer, A. E. Szafranska, C. Arthur, T. P. Nicholson, R. J. Cox, J. Crosby, M. P. Crump, T. J. Simpson, *Org. Biomol. Chem.* **2003**, *1*, 463–471.
[57] E. J. Drake, D. A. Nicolai, A. M. Gulick, *Chem. Biol.* **2006**, *13*, 409–419.
[58] S. C. Findlow, C. Winsor, T. J. Simpson, J. Crosby, M. P. Crump, *Biochemistry* **2003**, *42*, 8423–8433.
[59] T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, *Structure* **2000**, *8*, 407–418.
[60] T. A. Holak, S. K. Kearsley, Y. Kim, J. H. Prestegard, *Biochemistry* **1988**, *27*, 6135–6142.
[61] Y. Kim, J. H. Prestegard, *Biochemistry* **1989**, *28*, 8792–8797.
[62] Q. Li, C. Khosla, J. D. Puglisi, C. W. Liu, *Biochemistry* **2003**, *42*, 4648–4657.
[63] A. K. Sharma, S. K. Sharma, A. Surolia, N. Surolia, S. P. Sarma, *Biochemistry* **2006**, *45*, 6904–6916.
[64] B. F. Volkman, Q. Zhang, D. V. Debabov, E. Rivera, G. C. Kresheck, F. C. Neuhaus, *Biochemistry* **2001**, *40*, 7964–7972.
[65] G. Y. Xu, A. Tam, L. Lin, J. Hixon, C. C. Fritz, R. Powers, *Structure* **2001**, *9*, 277–287.
[66] H. C. Wong, G. Liu, Y. M. Zhang, C. O. Rock, J. Zheng, *J. Biol. Chem.* **2002**, *277*, 15874–15880.
[67] C. Khosla, Y. Tang, A. Y. Chen, N. A. Schnarr, D. E. Cane, *Annu. Rev. Biochem.* **2007**, *76*, 195–221.
[68] S. C. Tsai, L. J. Miercke, J. Krucinski, R. Gokhale, J. C. Chen, P. G. Foster, D. E. Cane, C. Khosla, R. M. Stroud, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14808–14813.
[69] S. C. Tsai, H. X. Lu, D. E. Cane, C. Khosla, R. M. Stroud, *Biochemistry* **2002**, *41*, 12598–12606.
[70] D. L. Akey, J. D. Kittendorf, J. W. Giraldes, R. A. Fecik, D. H. Sherman, J. L. Smith, *Nat. Chem. Biol.* **2006**, *2*, 537–542.
[71] T. A. Keating, C. G. Marshall, C. T. Walsh, A. E. Keating, *Nat. Struct. Biol.* **2002**, *9*, 522–526.
[72] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J.* **1997**, *16*, 4174–4183.
[73] S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh, M. T. Stubbs, *Structure* **2002**, *10*, 301–310.
[74] S. A. Samel, B. Wagner, M. A. Marahiel, L. O. Essen, *J. Mol. Biol.* **2006**, *359*, 876–889.
[75] S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L. O. Essen, *Structure* **2007**, *15*, 781–792.
[76] R. N. Perham, *Biochemistry* **1991**, *30*, 8501–8512.
[77] S. A. Sieber, U. Linne, N. J. Hillson, E. Roche, C. T. Walsh, M. A. Marahiel, *Chem. Biol.* **2002**, *9*, 997–1008.
[78] B. Silakowski, H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blöcker, G. Höfle, S. Beyer, R. Müller, *J. Biol. Chem.* **1999**, *274*, 37391–37399.
[79] S. Weinig, H. J. Hecht, T. Mahmud, R. Müller, *Chem. Biol.* **2003**, *10*, 939–952.
[80] S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **2005**, *105*, 715–738.
[81] N. J. Hillson, C. T. Walsh, *Biochemistry* **2003**, *42*, 766–775.
[82] B. Wilkinson, J. Micklefield, *Nat. Chem. Biol.* **2007**, *3*, 379–386.
[83] A. S. Worthington, H. Rivera, J. W. Torpey, M. D. Alexander, M. D. Burkart, *ACS Chem. Biol.* **2006**, *1*, 687–691.
[84] C. Y. Kim, V. Y. Alekseyev, A. Y. Chen, Y. Tang, D. E. Cane, C. Khosla, *Biochemistry* **2004**, *43*, 13892–13898.
[85] A. Y. Chen, D. E. Cane, C. Khosla, *Chem. Biol.* **2007**, *14*, 784–792.
[86] A. Y. Chen, N. A. Schnarr, C. Y. Kim, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2006**, *128*, 3067–3074.
[87] L. M. Worsham, L. Earls, C. Jolly, K. G. Langston, M. S. Trent, M. L. Ernst-Fonberg, *Biochemistry* **2003**, *42*, 167–176.
[88] Y. M. Zhang, M. S. Rao, R. J. Heath, A. C. Price, A. J. Olson, C. O. Rock, S. W. White, *J. Biol. Chem.* **2001**, *276*, 8231–8238.
[89] Y. M. Zhang, B. Wu, J. Zheng, C. O. Rock, *J. Biol. Chem.* **2003**, *278*, 52935–52943.
[90] Y. M. Zhang, H. Marrakchi, S. W. White, C. O. Rock, *J. Lipid Res.* **2003**, *44*, 1–10.
[91] C. Rowe, I. Böhm, I. Thomas, B. Wilkinson, B. Rudd, G. Foster, A. Blackaby, P. Sidebottom, Y. Roddis, A. Buss, J. Staunton, P. F. Leadlay, *Chem. Biol.* **2001**, *8*, 475–485.
[92] I. Thomas, C. J. Martin, C. J. Wilkinson, J. Staunton, P. F. Leadlay, *Chem. Biol.* **2002**, *9*, 781–787.
[93] L. Tran, M. Tosin, J. B. Spencer, P. F. Leadlay, K. J. Weissman, *ChemBioChem* **2008**, *9*, 905–915.
[94] L. H. Du, C. Sanchez, B. Shen, *Metab. Eng.* **2001**, *3*, 78–95.
[95] C. T. Walsh, H. W. Chen, T. A. Keating, B. K. Hubbard, H. C. Losey, L. S. Luo, C. G. Marshall, D. A. Miller, H. M. Patel, *Curr. Opin. Chem. Biol.* **2001**, *5*, 525–534.
[96] S. Lautru, G. L. Challis, *Microbiology* **2004**, *150*, 1629–1636.
[97] U. Linne, M. A. Marahiel, *Biochemistry* **2000**, *39*, 10439–10447.
[98] A. Koglin, M. R. Mofid, F. Lohr, B. Schafer, V. V. Rogov, M. M. Blum, T. Mittag, M. A. Marahiel, F. Bernhard, V. Dotsch, *Science* **2006**, *312*, 273–276.
[99] M. R. Mofid, R. Finking, M. A. Marahiel, *J. Biol. Chem.* **2002**, *277*, 17023–17031.
[100] D. Reichmann, O. Rahat, M. Cohen, H. Neuvirth, G. Schreiber, *Curr. Opin. Struct. Biol.* **2007**, *17*, 67–76.
[101] R. Finking, M. R. Mofid, M. Marahiel, *Biochemistry* **2004**, *43*, 8946–8956.
[102] J. R. Lai, M. A. Fischbach, D. R. Liu, C. T. Walsh, *J. Am. Chem. Soc.* **2006**, *128*, 11002–11003.
[103] J. R. Lai, M. A. Fischbach, D. R. Liu, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5314–5319.
[104] Z. Zhou, J. R. Lai, C. T. Walsh, *Chem. Biol.* **2006**, *13*, 869–879.
[105] D. E. Ehmann, C. A. Shaw-Reid, H. C. Losey, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2509–2514.
[106] W. L. Kelly, N. J. Hillson, C. T. Walsh, *Biochemistry* **2005**, *44*, 13385–13393.
[107] S. Omura, H. Ikeda, J. Ishikawa, A. Hanamoto, C. Takahashi, M. Shinose, Y. Takahashi, H. Horikawa, H. Nakazawa, T. Osonoe, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12215–12220.
[108] J. F. Aparicio, I. Molnár, T. Schwecke, A. König, S. F. Haydock, L. E. Khaw, J. Staunton, P. F. Leadlay, *Gene* **1996**, *169*, 9–16.
[109] L. Tang, Y. J. Yoon, C. Y. Choi, C. R. Hutchinson, *Gene* **1998**, *216*, 255–265.
[110] A. Lupas, *Curr. Opin. Struct. Biol.* **1997**, *7*, 388–393.
[111] S. Y. Tsuji, D. E. Cane, C. Khosla, *Biochemistry* **2001**, *40*, 2326–2331.
[112] N. Wu, S. Y. Tsuji, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2001**, *123*, 6465–6474.
[113] R. S. Gokhale, S. Y. Tsuji, D. E. Cane, C. Khosla, *Science* **1999**, *284*, 482–485.
[114] N. Wu, D. E. Cane, C. Khosla, *Biochemistry* **2002**, *41*, 5056–5066.
[115] C. D. Reeves, S. L. Ward, W. P. Revill, H. Suzuki, M. Marcus, O. V. Petrakovsky, S. Marquez, H. Fu, S. D. Dong, L. Katz, *Chem. Biol.* **2004**, *11*, 1465–1472.
[116] P. Caffrey, D. J. Bevitt, J. Staunton, P. F. Leadlay, *FEBS Lett.* **1992**, *304*, 225–228.
[117] R. S. Gokhale, D. Hunziker, D. E. Cane, C. Khosla, *Chem. Biol.* **1999**, *6*, 117–125.
[118] R. McDaniel, C. M. Kao, S. J. Hwang, C. Khosla, *Chem. Biol.* **1997**, *4*, 667–674.
[119] C. M. Squire, R. J. Goss, H. Hong, P. F. Leadlay, J. Staunton, *ChemBioChem* **2003**, *4*, 1225–1228.
[120] R. B. Rose, J. A. Endrizzi, J. D. Cronk, J. Holton, T. Alber, *Biochemistry* **2000**, *39*, 15062–15070.
[121] G. Schreiber, *Curr. Opin. Struct. Biol.* **2002**, *12*, 41–47.
[122] K. J. Weissman, *ChemBioChem* **2006**, *7*, 485–494.
[123] M. Thattai, Y. Burak, B. I. Shraiman, *PLoS Comput. Biol.* **2007**, *3*, 1827–1835.

[124] A. Ranganathan, M. Timoney, M. Bycroft, J. Cortés, I. P. Thomas, B. Wilkinson, L. Kellenberger, U. Hanefeld, I. S. Galloway, J. Staunton, P. F. Leadlay, *Chem. Biol.* **1999**, *6*, 731–741.
[125] K. J. Weissman, *ChemBioChem* **2006**, *7*, 1334–1342.
[126] U. Linne, D. B. Stein, H. D. Mootz, M. A. Marahiel, *Biochemistry* **2003**, *42*, 5114–5124.
[127] T. Stachelhaus, H. D. Mootz, V. Bergendahl, M. A. Marahiel, *J. Biol. Chem.* **1998**, *273*, 22773–22781.
[128] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15585–15590.
[129] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 275–280.
[130] C. Chiocchini, U. Linne, T. Stachelhaus, *Chem. Biol.* **2006**, *13*, 899–908.
[131] C. D. Richter, D. Nietlispach, R. W. Broadhurst, K. J. Weissman, *Nat. Chem. Biol.* **2008**, *4*, 75–81.
[132] S. E. O'Connor, C. T. Walsh, F. Liu, *Angew. Chem.* **2003**, *115*, 4047–4051; *Angew. Chem. Int. Ed.* **2003**, *42*, 3917–3921.
[133] F. Liu, S. Garneau, C. T. Walsh, *Chem. Biol.* **2004**, *11*, 1533–1542.
[134] S. C. Wenzel, R. Müller, *Curr. Opin. Chem. Biol.* **2005**, *9*, 447–458.
[135] R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straight, R. Kolter, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1506–1509.
[136] M. Royer, L. Costet, E. Vivien, M. Bes, A. Cousin, A. Damais, I. Pieretti, A. Savin, S. Megessier, M. Viard, R. Frutos, D. W. Gabriel, P. C. Rott, *Mol. Plant-Microbe Interact.* **2004**, *17*, 414–427.
[137] A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson, C. M. Thomas, *Chem. Biol.* **2003**, *10*, 419–430.
[138] V. Simunovic, J. Zapp, S. Rachid, D. Krug, P. Meiser, R. Müller, *ChemBioChem* **2006**, *7*, 1206–1220.
[139] O. Perlova, K. Gerth, A. Hans, O. Kaiser, R. Müller, *J. Biotechnol.* **2006**, *121*, 174–191.
[140] M. Kopp, H. Irschik, S. Pradella, R. Müller, *ChemBioChem* **2005**, *6*, 1277–1286.
[141] G. L. Tang, Y. Q. Cheng, B. Shen, *Chem. Biol.* **2004**, *11*, 33–45.
[142] P. Kumar, A. T. Koppisch, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2003**, *125*, 14307–14312.
[143] Z. D. Aron, P. D. Fortin, C. T. Calderone, C. T. Walsh, *ChemBioChem* **2007**, *8*, 613–616.
[144] V. Simunovic, R. Müller, *ChemBioChem* **2007**, *8*, 497–500.
[145] V. Simunovic, R. Müller, *ChemBioChem* **2007**, *8*, 1273–1280.
[146] C. T. Calderone, W. E. Kowtoniuk, N. L. Kelleher, C. T. Walsh, P. C. Dorrestein, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8977–8982.
[147] C. T. Calderone, D. F. Iwig, P. C. Dorrestein, N. L. Kelleher, C. T. Walsh, *Chem. Biol.* **2007**, *14*, 835–846.
[148] Z. Chang, N. Sitachitta, J. V. Rossi, M. A. Roberts, P. Flatt, J. Jia, D. H. Sherman, W. H. Gerwick, *J. Nat. Prod.* **2004**, *67*, 1356–1367.
[149] B. Julien, Z. Q. Tian, R. Reid, C. D. Reeves, *Chem. Biol.* **2006**, *13*, 1277–1286.
[150] B. Silakowski, G. Nordsiek, B. Kunze, H. Blöcker, R. Müller, *Chem. Biol.* **2001**, *8*, 59–69.
[151] M. Jin, M. A. Fischbach, J. Clardy, *J. Am. Chem. Soc.* **2006**, *128*, 10660–10661.
[152] P. D. Fortin, C. T. Walsh, N. A. Magarvey, *Nature* **2007**, *448*, 824–827.
[153] P. D. Straight, M. A. Fischbach, C. T. Walsh, D. Z. Rudner, R. Kolter, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 305–310.

Received: December 12, 2007
Published online on March 20, 2008