Biochemistry

 Volume 40, Number 8 February 27, 2001

Current Topics

Intermodular Communication in Polyketide Synthases: Comparing the Role of Protein−Protein Interactions to Those in Other Multidomain Proteins†

Stuart Y. Tsuji,‡ Nicholas Wu,§ and Chaitan Khosla*,‡,§,||

Departments of Chemical Engineering, Chemistry, and Biochemistry, Stanford University, Stanford, California 94305-5025

Received October 24, 2000; Revised Manuscript Received December 22, 2000

ABSTRACT: Although the role of protein−protein interactions in transducing signals within biological systems has been extensively explored, their relevance to the channeling of intermediates in metabolism is not widely appreciated. Polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) are two related families of modular megasynthases that channel covalently bound intermediates from one active site to the next. Recent biochemical studies have highlighted the importance of protein−protein interactions in these chain transfer processes. The information available on this subject is reviewed, and its possible mechanistic implications are placed in context by comparisons with selected well-studied multicomponent protein systems.

Many cellular processes rely upon the concerted action of multiple proteins. During such events as transcription activation and signal transduction, specific protein−protein interactions play an essential role. Because specificity in these processes has often emerged from evolutionarily related mechanisms, many of the constituent proteins consist of modularized domains, where the domains can comprise recognition features or enzymatic activities. The advantages of this strategy with respect to balancing the needs of selectivity and tolerance in a system as complex as a cell are clear. Although such modularity is relatively rare in the context of metabolic pathways, recent evidence suggests that certain polyketide synthases (PKSs) take advantage of modular protein−protein interfaces to facilitate substrate channeling. In this brief review the interactions of domains within modular PKSs and the related nonribosomal peptide synthetases (NRPSs) are compared to other well-characterized multidomain protein systems.

We start by reviewing the salient features of four diverse multicomponent systems: the Fos-Jun heterodimer, the Src kinase, methionine synthase, and pyruvate dehydrogenase. From structural and mechanistic studies on each of these well-characterized systems have emerged important principles of interdomain communication, which are likely to be relevant to future studies on PKSs and NRPSs. We then compare the properties of individual PKS modules to those of vertebrate fatty acid synthases (FASs). Although high-resolution structural insights into vertebrate FASs have not yet emerged, protein chemical and enzymological analyses of these systems have led to the establishment of working models for interdomain interactions. Given the close architectural and mechanistic relationships between a vertebrate FAS and a PKS module, these models are likely to also apply

† Research in the authors' laboratory on modular PKSs has been supported by grants from the National Institutes of Health (1 R01-CA66736), the National Science Foundation (BES 9806774), and the Strategic Research Fund of AstraZeneca, Inc. S.Y.T. is a recipient of an NSF Predoctoral Fellowship, and N.W. is a recipient of a Stanford−NIH Predoctoral Training Fellowship.

* Address correspondence to this author. Phone/Fax: 650-723-6538. E-mail: ck@chemeng.stanford.edu.

‡ Department of Chemical Engineering.

§ Department of Chemistry.

|| Department of Biochemistry.

10.1021/bi002462v CCC: $20.00
Published on Web 02/02/2001

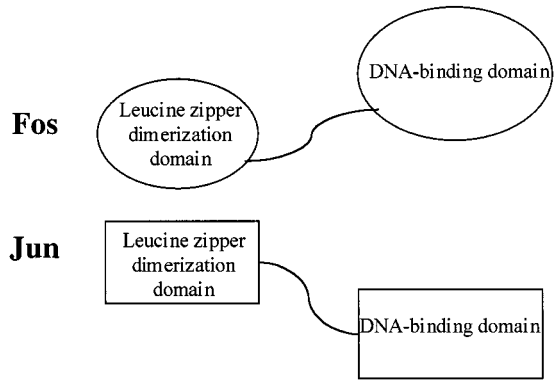

FIGURE 1: Domain organization of the Fos and Jun, which composes the core of the AP-1 transcription factor. The leucine zipper regions mediate dimer formation, a prerequisite for DNA binding by the distal domain of the proteins. The dimerization and DNA-binding domains are separated by a linker whose length has been shown to be critical for activity.

to PKS modules. Finally, we discuss the role of protein−protein interactions in the transfer of polyketide chains between adjacent PKS and NRPS modules.

The Fos-Jun Heterodimer

The products of the oncogenes *fos* and *jun* comprise the activator protein-1 (AP-1), a transcription factor that induces transcription of several genes, including human metallothionein IIa (MTIIa), collagenase, and interleukin 2 (IL2). Transcription of AP-1 can be induced by the addition of 12-*O*-tetradecanoylphorbol-13-acetate (TPA), a tumor promoter which activates protein kinase C (PKC). Although it had been thought that AP-1 binding is necessary and sufficient for transcriptional activation after activation of PKC, the AP-1-controlled genes also display significant levels of basal transcription (*1*). The high level of basal transcription reflects the mechanism of activation of AP-1, which relies on the interactions of the Fos and Jun proteins.

Experiments where Jun expression plasmids and an appropriate reporter plasmid were cotransfected into carcinoma stem cells indicated that Jun was able to activate transcription. Given the sequence similarity (44% identity) between Jun and the DNA-binding domain of GCN4, the yeast transcription factor, this result was not surprising. On the other hand, Fos, which resides in the nucleus like Jun, did not demonstrate any sequence-specific DNA binding when expressed alone. It was shown that Fos mediated only weak expression of AP-1-dependent reporters and that Fos, unlike Jun, cannot form homodimers, thereby explaining its inability to bind DNA. However, it was discovered that Fos and Jun associate to form the core of the AP-1 complex, which suggested that Fos is able to mediate activation through the DNA-binding ability of Jun. More specifically, it was found that Fos forms heterodimers with Jun that bind to DNA more strongly than their Jun-Jun homodimeric counterparts (Figure 1) (*1*). Therefore, since Jun and its mRNA transcript persist, Jun likely causes the basal level of AP-1-dependent expression. However, strong induction requires the formation of the Fos-Jun heterodimer. The nature of the dimeric interaction was revealed by direct biochemical and biophysical analyses, as described below.

Inspection of the amino acid sequences of Fos and Jun revealed a pattern of repeated leucines characteristic of leucine zippers, also homologous to a region of GCN4. Although initially thought to be a novel secondary structure, leucine zippers can be considered as a subgroup of the coiled-coil motif. These structures typically contain repeated groups of seven residues, $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$, where the a and d positions contain hydrophobic residues and the remaining amino acids are generally polar (*2*). The hydrophobic interactions between the a and d residues on partner strands (or more, depending on the number of strands in the coil) mediate dimerization (or multimerization) of the individual chains. An isolated α-helical coil contains 3.6 residues per turn, thus seven residues would be just short of two full turns. Therefore, to maintain the hydrophobic contacts at the coiled-coil interface, the bundle assumes a superhelical twist, thereby yielding the observed coiled coil.

Although the hydrophobic residues are well conserved in coiled coils, they alone do not dictate dimerization. In fact, residues outside the a and d positions must allow for selectivity of interactions. The preference of Fos and Jun to form heterodimers was investigated through biophysical experiments of the isolated leucine zipper regions of the proteins (*3*). By determining the parallel orientation of the helices, the juxtaposition of the interhelical ion pairs at the e and g positions become more apparent. Thus, not only did Fos and Jun have a general electrostatic attraction due to the overall negative charge of Fos and positive charge on Jun, but the positions of the charged residues suggested that their specific patterns could dictate the selectivity of association.

Finally, although the dimerization domains of Fos and Jun can be functionally isolated, the linker joining the leucine zippers to their respective DNA-binding domains is also important. The identities of amino acids of the linker may not be crucial, but its approximate length is, as evidenced by the insertion of five additional residues in the linker eliminating all DNA-binding activity of Fos and Jun (*1*). The importance of less-conserved regions such as these will be further established below.

Src Kinase

Although the earliest identified signal transduction pathways involved communication via small molecule second messengers such as cAMP, the present understanding of signaling acknowledges the importance of more specific, weaker signals channeled through the interactions and activities of proteins (*4*). Many of these proteins consist of modular domains that can either mediate specific interactions or transduce the message via catalytic activity or conformational changes. A prominent example is the Src-family kinases (Rous sarcoma virus oncogene). Src kinase is a nonreceptor tyrosine kinase whose activity (which resides in the Src-homology-1 or "SH1" domain) is regulated by its constituent SH2 and SH3 domains (Figure 2) (*5*). The SH2 domain is a representative of a group of domains that recognizes phosphorylated tyrosine residues in specific sequence contexts. However, different SH2 domains interact with different sequences; thus there is a common feature (phosphotyrosine) and then a finer set of features for specificity. The SH3 domain is another separable module that binds to a set of peptides, those containing a characteristic PxxP motif that typically adopts the polyproline type

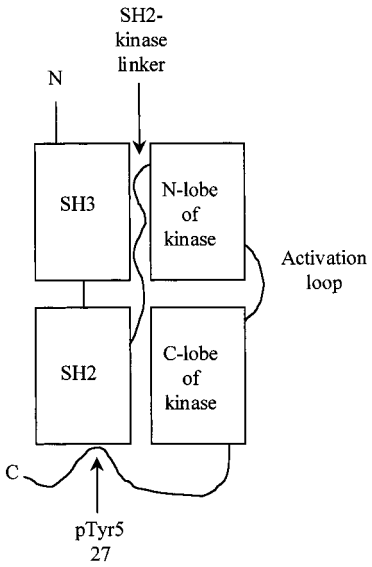

FIGURE 2: Schematic diagram of the modular organization of Src kinase and its regulatory domains. The SH2-kinase linker, the activation loop, and the C-terminal tail participate in specific interactions with the SH3 domain, the active site, and the SH2 domain, respectively. Each interaction helps to control the kinase activity.

II (PPII) helix. In the overall organization of these domains within Src kinase, the N-terminal myristoylation site is followed by the SH3 and SH2 domains and then the bilobal kinase domain whose active site resides between the N- and C-terminal lobes.

The activity of the kinase is regulated through two different elements. The first is the phosphorylation of the activation loop, which can directly block the active site. The second requires many intramolecular interactions far from the active site that involve the SH3 domain, SH2 domain, the non-catalytic face of the N-terminal kinase lobe, the SH2-kinase linker, and phosphorylation of the C-terminal tail (*6*). Specifically, phosphorylation of Tyr527 of the C-terminal tail allows its recognition by the SH2 domain, thereby bringing the C-terminal kinase lobe into close contact with the SH2 domain. To stabilize this conformational change, the SH2-kinase linker is bound by the SH3 domain (adopting a PPII helical structure despite having only one proline) and interacts with the N-terminal kinase lobe.

Although the composition of the SH2-kinase linker is not strictly conserved, the hydrophobic region, where contacts with the SH3 and the kinase domains occur, is conserved (*6*). Furthermore, replacement of individual amino acids has highlighted their relative significance. Replacement of Pro250 presumably eliminates the PPII helix and resulted in release of the SH3 domain, while substitution of Leu255 with Ala or Val led to constitutive activity of the kinase domain (*7*). Clearly, this linker region in Src kinase plays an important role in regulation and serves as more than a mere connection between recognition and catalytic domains.

Methionine Synthase

B_{12}-dependent methionine synthase is a structurally well-defined multidomain metalloprotein that catalyzes a methyl transfer from methyltetrahydrofolate to homocysteine to form methionine. The B_{12} (cobalamin) cofactor primarily cycles between two oxidation states: cob(I)alamin [Co(I)] and cob(III)alamin [Co(III)] (*8*). The cob(I)alamin form of the enzyme accepts a methyl group from methyltetrahydrofolate to generate the methylcob(III)alamin which then donates the methyl group to homocysteine to generate methionine and to regenerate cob(I)alamin. This cycle is occasionally temporarily disrupted by oxidation of cob(I)alamin to inactive cob(II)alamin (*9*). Reactivation of the enzyme to the methylcob(III)alamin form requires a methyl transfer from adenosylmethionine (AdoMet) as well as a single electron transfer (*10*). As indicated by limited proteolysis experiments of the 136 kDa holoprotein from *Escherichia coli* and expression of truncated protein, methionine synthase can be partitioned into four distinct domains connected by flexible linkers (*11*−*15*) (Figure 3). These domains include a cobalamin-binding domain and three substrate-binding domains for the three substrates—homocysteine, methyltetrahydrofolate, and adenosylmethionine.

The coexistence of the primary catalytic cycle and the adenosylmethionine-dependent reactivation cycle require access of both methyltetrahydrofolate and adenosylmethionine to the same coordinate relative to the cobalt atom of cobalamin. This coexistence suggests that the enzyme undergoes a conformational change that is facilitated by the presence of the interdomain linkers and that independently gives the two methyl donor substrates access to the cobalamin (*11*). On the basis of more detailed limited proteolysis, N_2O inactivation, mutagenesis, and EPR and UV/visible spectroscopy experiments, it has been hypothesized that the enzyme exists in two conformational states, each one roughly corresponding to the active and inactive state (*16*). During the primary catalytic cycle, the enzyme is organized such that the methyltetrahydrofolate binding domain is in proximity to the cobalamin cofactor, and the adenosylmethionine-binding domain is distal to the cofactor. However, when the cofactor is oxidized to the inactive cob(II)alamin form, the presence of the interdomain flexible linkers allows the methytetrahydrofolate-binding domain to move away from the cofactor and allows the reactivating adenosylmethionine-binding domain to move in. As in the case of the Src kinase, the role of interdomain flexible linkers in facilitating conformational changes is a theme that is repeated in methionine synthase and many multidomain systems.

Pyruvate Dehydrogenase Complex

The pyruvate dehydrogenase complex (PDC) is a multidomain enzyme system that has been best characterized from the Gram-negative bacteria *Azotobacter vinelandii* and *Escherichia coli* (*17*). Composed of three components—E1 (pyruvate dehydrogenase/decarboxylase), E2 (acetyltransferase), and E3 (dehydrolipoamide dehydrogenase)—PDC is an extremely large complex (700 kDa in *A. vinelandii* and 4.5 MDa in *E. coli*) that catalyzes the oxidative decarboxylation of pyruvate and subsequent transfer of the resultant acetyl group to coenzyme A to form acetyl-CoA. E1 initiates the catalytic cycle through the oxidative decarboxylation of pyruvate. The resulting acetyl group is transferred by a lipoyl prosthetic group of E2 to CoASH to form acetyl-CoA. Finally, E3 regenerates the lipoyl disulfide bond in a NAD^+-dependent oxidation. The E2 component forms the catalytic core of the system with 24 polypeptides arranged in octahedral symmetry. The catalytic core is organized with

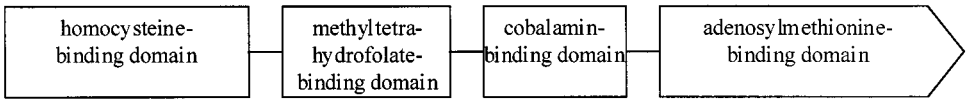

FIGURE 3: Modular organization of B_{12}-dependent methionine synthase. The four domains shown in blocks are separated by short, flexible linkers that are necessary for the catalytic and regenerative cycles of the enzyme.

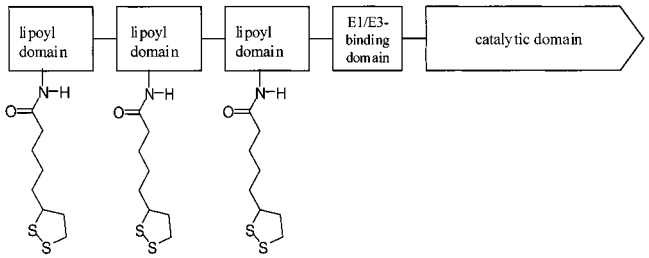

FIGURE 4: Domain organization of the E2 enzyme of the pyruvate dehydrogenase complex from *A. vinelandii*. The five domains shown in blocks are connected by flexible 20−40 amino acid linkers. In addition, each of the three lipoyl domains is posttranslationally modified with lipoyl arms at conserved lysine residues that facilitate channeling of intermediates between active sites of the enzyme complex.

E1 preferentially associated with the 12 edges of the cube and with E3 preferentially associated with the six faces (*18*, *19*). It has been shown through limited proteolysis experiments that the N-terminal 40 amino acids of E1 interact with E2 and are functionally independent of the catalytic decarboxylase activity of E1 (*20*). Similarly, E2 can be partitioned into multiple independent domains—a catalytic domain, an E1/E3 binding domain, and one to three lipoyl domains—each separated by flexible linkers of 20−40 amino acids (Figure 4).

Each lipoyl domain is posttranslationally modified with a lipoyl group which is functionally analogous to the phosphopantetheine arm of the acyl carrier protein (ACP) and peptidyl carrier protein (PCP) domains that, as will be discussed later, are most commonly associated with fatty acid synthases, polyketide synthases, and nonribosomal peptide synthetases (*21*). The lipoyl groups are attached at conserved lysine residues of the lipoyl domain of E2. Together with the interdomain linkers, they are responsible for channeling intermediates between active sites of the complex. According to this active site coupling mechanism, four sets of lipoyl domains (from four different E2 components) can theoretically service the same E1 polypeptide (*22*). However, multidimensional NMR experiments have suggested that the lipoyl-lysine side chain possesses a certain amount of conformational restriction that precludes accessibility to its theoretical range of motion (*23*). Furthermore, while lipoyl moieties must be able to undergo transacylation with E1 and E3 active sites, it is still not clear whether the multiple lipoyl moieties on a single E2 polypeptide are functionally equivalent (*24*−*26*) and, if so, whether these lipoyl arms can transacylate between themselves (*27*, *28*). Notwithstanding these unanswered questions, the role of the lipoyl domains in PDC provides a relevant framework for discussing the mechanism of phosphopantetheine arms in PKSs and NRPSs.

Similarities between Eukaryotic Fatty Acid Synthases and PKS Modules

Vertebrate and yeast fatty acid synthases (FASs) are multidomain enzymes that organize all of their catalytic domains on one and two polypeptides, respectively. The domains required for fatty acid biosynthesis include a ketosynthase, acyltransferase, acyl carrier protein (ACP), ketoreductase, enoylreductase, and dehydratase. The vertebrate FAS in particular serves as an excellent model for a single module of a PKS; indeed, the arrangement of domains in these two homodimeric systems appears identical. The organization and interactions of domains within these FASs and PKS modules have been the subject of extensive studies and will not be the subject of this review (*29*−*32*). We will simply highlight the role of the phosphopantetheine arm of the ACP domain in these interactions. In particular, the active site for condensation of an extender unit onto the growing fatty acid/polyketide chain is derived from a ketosynthase and an ACP. The preferred mode of interaction is between domains from different subunits of the homodimer, although recent studies with the vertebrate FAS have demonstrated that the ketosynthase and ACP of the same subunit can also interact (albeit with lower kinetic efficiency) (*33*). The postcondensational reductions and dehydrations are carried out on a single polypeptide. The active sites of the ketosynthase and ACP from opposing monomers can be cross-linked, suggesting that they are within 20 Å of each other (the phosphopantetheine moiety is only approximately 20 Å in length) (*34*−*37*). In contrast, fluorescent energy transfer experiments have estimated the ACP−thioesterase and ketoreductase−enoylreductase distance between active sites to be 37−48 and >49 Å, respectively, suggesting that a conformational change is likely to be required for catalysis (*38*−*40*). Thus, in conjunction with the "swinging arm", protein−protein interactions are likely to play an important role in intramodular chemistry within a FAS homodimer. The relative contributions of the swinging arm, interdomain interactions, linkers, and active site−substrate interactions to the FAS catalytic cycle, as well as their precise mechanisms, remain to be elucidated and are likely to have direct relevance to our understanding of intramodular chemistry within a PKS. Additionally, modular PKSs have another property that, although absent in FASs, is surrounded by similar issues. Instead of reusing the same module of catalysts for multiple chain extension cycles, after every round of condensation a PKS module transfers its product to the next module. Our current understanding of the role of protein−protein interactions in *inter*modular chemistry, which refers to the directed transfer of intermediates between different PKS modules, is the final subject of this review.

Protein−Protein Interactions in Intermodular Communication in PKSs and NRPSs

Similar to vertebrate FAS, modular PKSs are organized in multidomain polypeptides containing a ketosynthase, acyltransferase, and ACP domain within each module (*41*, *42*). As illustrated in the deoxyerythronolide B synthase (DEBS) model system (Figure 5), these three core domains of a single module catalyze the condensation of a single extender unit onto the polyketide intermediate before passing

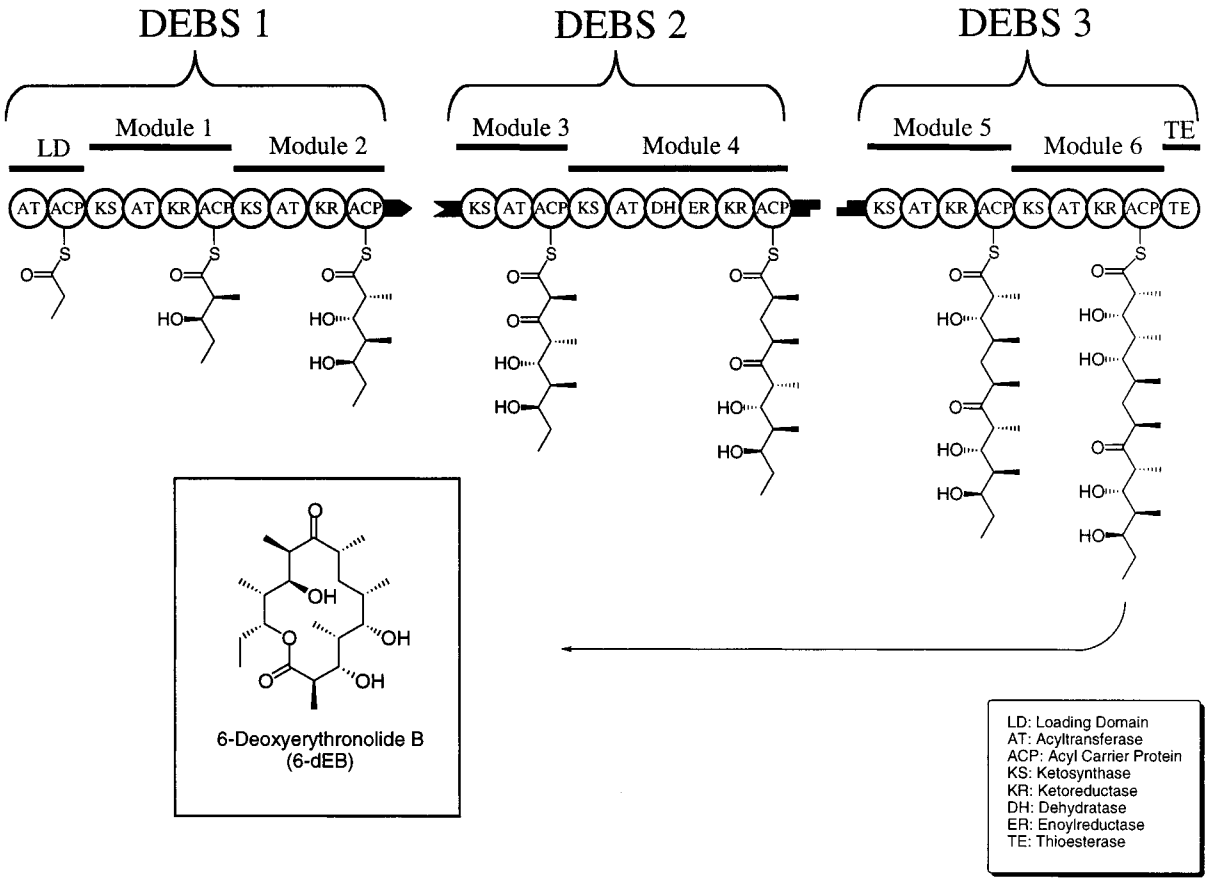

FIGURE 5: Deoxyerythronolide B synthase (DEBS) system with its intermediates and final product. There are six catalytic modules distributed over three large polypeptides. Each module catalyzes one elongation reaction together with a variable number of chain modification reactions. Abbreviations: KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein; KR, ketoreductase; DH, dehydratase; ER, enoylreductase; TE, thioesterase. Two types of intermodular linkers are referred to in the text. The intrapolypeptide linkers, which connect modules 1 and 2, modules 3 and 4, and modules 5 and 6, are not shown explicitly. Interpolypeptide linkers flank DEBS1/DEBS2 and DEBS2/DEBS3 and are shown as matched pairs.

the newly elongated chain to the next module for another round of condensation. (A variable number of postcondensation reactions can be catalyzed within each module, depending on the types of active sites available for such chemistry.) This extension process occurs in an assembly line fashion through the entire sequence of modules. Since PKS modules comprise natural catalytic units, reorganizing entire modules, rather than individual domains, provides an attractive method for exploiting the combinatorial potential of this biochemistry. Before realizing this strategy, the recognition features of modules need to be defined. In wild-type PKS systems, the channeling of intermediates proceeds through a precise sequence of modules such that each module accepts a defined intermediate from its upstream neighbor. Adherence to the strict sequence of chain transfer events can be controlled by two possible variables: substrate recognition by the recipient module of active sites and/or dynamic protein interactions. Although recent experiments indicate that both variables play significant roles in dictating intermodular chain transfer (*43*−*45*, manuscript in preparation), this brief discussion focuses on the progress made toward dissecting the contributions of protein−protein interactions.

In addition to protein interactions across the dimeric interface of the polypeptides (*46*−*48*), protein interactions on a slightly more macroscopic level between individual modules facilitate chain transfer between donor and acceptor modules on the same polypeptide as well as on separate polypeptides. Linker regions between modules on the same polypeptide are referred to as *intra*polypeptide linkers, while linker regions at the N- and C-termini of donor and acceptor modules, respectively, on separate polypeptides are referred to as *inter*polypeptide linkers. The significance of these linker regions was not recognized until recently, perhaps because they lack catalytic activity or obvious sequence conservation. Nonetheless, their crucial contributions to the proper functioning of intermodular transfer are slowly being illuminated.

The role of intrapolypeptide linkers was first illustrated through in vivo studies (*43*). More recently, kinetic analysis of individual modules of DEBS as well as selected hybrid bimodular derivatives has demonstrated not only the importance of conserving intrapolypeptide linkers to enable chain transfer between covalently attached modules but also the robustness of the strategy. For example, as seen in Figure 1, modules 2, 5, and 6 of DEBS catalyze the same set of reactions with identical stereocontrol, albeit on very different substrates. Although module 6 of DEBS can accept and elongate an *N*-acetylcysteamine thioester of the natural substrate of module 2 with comparable steady-state kinetic parameters as module 2 itself, module 5 is a considerably poorer catalyst of this elongation reaction (Figure 6A−C) (*44*). In contrast, the k_{cat} values for formation of the same polyketide product by the bimodular fusion proteins module 1 + module 5 + TE (M1 + M5 + TE) and module 1 + module 6 + TE (M1 + M6 + TE) were comparable to that

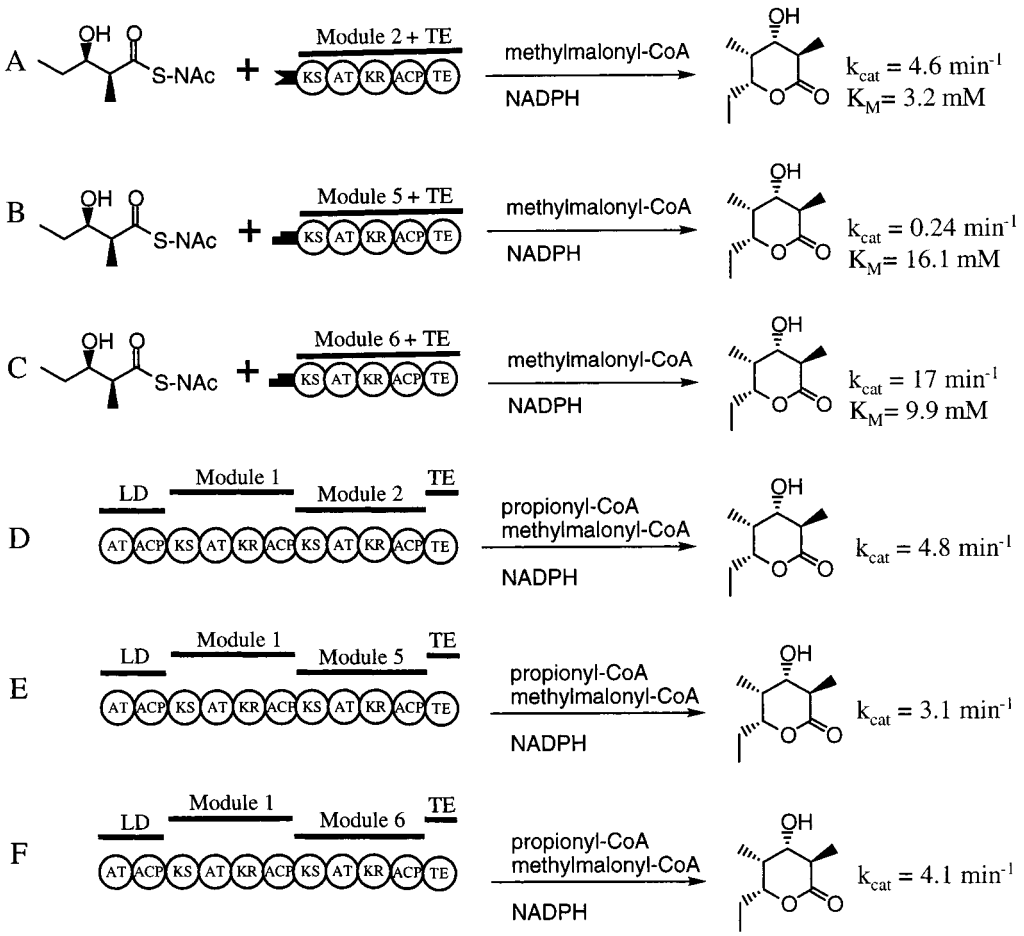

FIGURE 6: Reactions A−C: Comparison of the k_{cat} values of individual DEBS modules with *N*-acetylcysteamine-activated (2*R*,3*S*)-diketide. (A) Module 2 + TE. (B) Module 5 + TE. (C) Module 6 + TE. Reactions D−F: Comparison of the k_{cat} values of bimodular DEBS constructs. (D) DEBS1 + TE. (E) Module 1 + module 5 + TE. (F) Module 1 + module 6 + TE.

for wild-type DEBS1 + TE (Figure 6D−F) (manuscript in preparation). Thus, the ability of module 5 to accept and elongate the diketide product of module 1 is significantly superior in the context of M1 + M5 + TE as compared to M5 + TE alone. This intrapolypeptide effect could result from the existence of fewer nonproductive substrate binding modes during chain transfer when the diketide is presented on a pantetheine arm as compared to an *N*-acetylcysteamine group, or it might result from a specific contribution by the intrapolypeptide linker, or even an interaction between ACP1 and KS5. In fact, alteration of two amino acid residues at the C-terminal end of the intrapolypeptide linker in M1 + M6 + TE is adequate to abolish chain transfer without disrupting the activity of the downstream module (manuscript in preparation). Thus, similar to the SH2-kinase linker of Src, mutating specific residues in a linking region can completely alter interactions between the flanking modules.

The role of interpolypeptide linkers was also first intimated through in vivo studies in which module 5 from the rifamycin synthase system was used to replace module 2 of DEBS (*43*). By conserving the module 2−module 3 interpolypeptide linker pair, intermodular communication between rifamycin module 5 and DEBS module 3 was maintained to yield the expected 14-membered macrolactone. Further studies conducted in vitro by monitoring interpolypeptide transfer between individually expressed DEBS module 2 and module 3 verified the importance of these linkers for chain transfer (*45*). A DEBS module 2−module 3 system in which the wild-type interpolypeptide linker pair was replaced with the matched linker pair from module 4 and module 5 manifests similar kinetic properties, indicating that the linkers themselves are modular (Figure 7A,B). In contrast, experiments conducted with the same pair of modules, but with mismatched linker pairs, led to a dramatic drop in turnover rates (Figure 7C,D). No significant differences were observed in the chain elongation activities of any of the individual modules when assayed separately. These results highlighted the selectivity of the interpolypeptide linkers, since the linker of module 2 mediated efficient transfer when paired with the linker of module 3 (its natural partner), but not with the linker of module 5. Similarly, the linker of module 3 facilitated faster transfer when paired with the linker of module 2 (natural partner) than when paired with the linker of module 4.

To assess the relative importance of linker pairs and substrate−acceptor KS interactions, individual ACP domains have been expressed with natural C-terminal linkers (manuscript in preparation). These ACPs can be chemoenzymatically modified so as to enable them to act as donors for virtually any ketide substrate. Using these modified ACP proteins as substrate donors, kinetic analyses have revealed that mismatched linker pairs have significant effects on the efficiency of chain transfer. The importance these protein−protein interactions was highlighted by noting that matched linkers could facilitate channeling of otherwise poor substrates between modules. For example, although turnover of

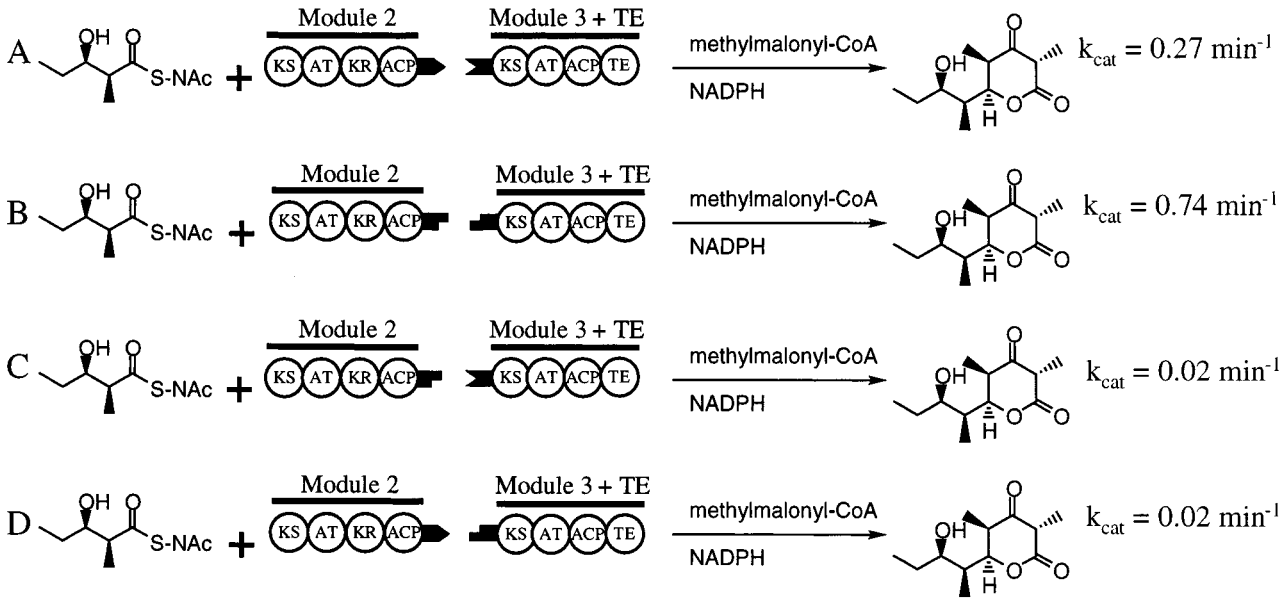

FIGURE 7: Comparison of the effects of swapping interpolypeptide linkers on the k_{cat} values of tetraketide formation by DEBS modules 2 and 3. (A) Reaction with natural, matched interpolypeptide linkers. (B) Reaction with matched linkers from modules 4 and 5. (C) Reaction of module 2 with module 4 linker + module 3 with module 3 linker, such that the interpolypeptide linkers are mismatched. (D) Reaction of module 2 with module 2 linker + module 3 with module 5 linker, such that the interpolypeptide linkers are mismatched.

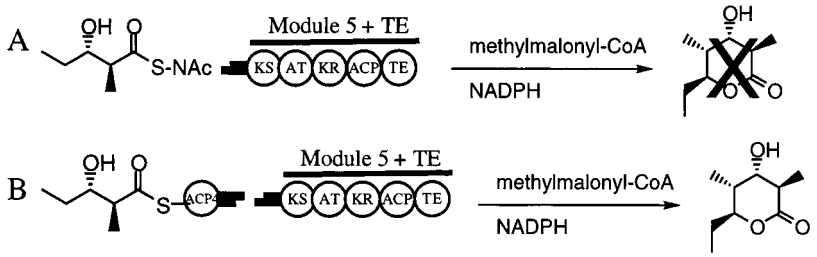

FIGURE 8: Reactions A and B: Comparison of exogenous loading versus channeled loading of KS5 in DEBS. (A) Reaction of *N*-acetylcysteamine-activated (2*S*,3*S*)-diketide with module 5 with module 5 linker gives no triketide product. (B) Reaction of ACP4-loaded (2*S*,3*S*)-diketide with module 5 with its natural linker gives the expected triketide product.

the (2*S*,3*S*)-anti-diketide by module 5 cannot be detected when the diketide is presented as an *N*-acetylcysteamine thioester (Figure 8A), the same substrate can be accepted and extended by module 5 when presented on an ACP with a matched linker (Figure 8B).

Definitive information regarding the structure of intrapolypeptide and interpolypeptide linkers is unavailable, but sequence analysis and preliminary biophysical results suggest that the interpolypeptide linkers may form α-helical coiled coils (*43*, *45*). Given the example of Fos-Jun heterodimerization, the coiled-coil motif offers a framework for mediating specific pairwise interactions of the N- and C-terminal interpolypeptide linkers.

Analysis of hybrid nonribosomal peptide synthetase (NRPS) constructs also suggests that protein interactions may play an important role in selecting the proper peptide intermediate for extension (*49*). However, the junctions between the peptidyl carrier protein (PCP; functional equivalent of the ACP) and condensation domains in NRPSs seem to be more tolerant of heterologous fusion than in PKSs, whereas the condensation and adenylation domains of NRPSs work most efficiently as a pair due to the "editing" function of the C domain (*50*). Thus, instead of matching ACP and KS domains in PKSs, NRPSs appear to have matched the condensation and adenylation domains over evolutionary time. Since all of the modular engineering of NRPSs has occurred as fusions within a polypeptide, the features associated with interpolypeptide transfer have not been investigated in NRPS systems. Such an investigation might demonstrate protein selectivity similar to that seen in PKSs, given that successful interactions between PCP and condensation domains of consecutive modules are dictated by protein−protein interactions.

Outstanding Questions

In contrast to some of the structurally well-defined systems described above, high-resolution structures of neither PKS nor NRPS modules are available at present. Such insights would allow evaluation of the interactions between domains within and between modules. By understanding how these domains communicate with each other, the protein engineer would be better equipped to construct chimeric modules with activity comparable to wild-type systems. Similarly, structures of multimodular proteins would allow analysis of how modules interact with each other.

Several specific types of macromolecular interactions in PKSs and NRPSs are worthy of investigation at an atomic level. First, although the functional relevance of maintaining linker features in both intra- and interpolypeptide transfer has been established, structural evidence would allow insight into the mechanism of recognition and transfer as facilitated by these linkers. Second, as in the case of the phosphopan-

tetheine arm of the vertebrate FAS and the lipoyl arm of PDC, the phosphopantetheine arm of PKSs and NRPSs bridges the multiple intermodule and intramodule active sites. Although no structural information is available, it is expected that the distance between relevant active sites will be significantly greater than the 20 Å reach of the phosphopantetheine arm, suggesting the necessity of interdomain flexible linkers (such as those found in methionine synthase and between lipoyl domains in PDC) to reduce the relative distances between active sites. In addition, interaction of multiple domains requires coordination of domain movements similar to that suggested in methionine synthase. The control of such coordination is still an open question. Third, specificity of transfer also reflects the vectorial motion of the growing chain, meaning that the overall movement flows from the KS to the ACP of the same module and then to the KS of the next module. However, reverse (upstream) movement from an acceptor module KS to a donor module ACP has been shown to be possible in the absence of downstream traffic (manuscript in preparation). More interestingly, evidence suggesting the occurrence of "lateral" movement via transacylation across ACP domains within the same homodimer (such as that suggested between lipoyl domains in PDC) has also been obtained (*51*). Thus, whether the apparent unidirectional movement in an ordinarily operating PKS system results from an ordered motion of the pantetheinyl arm or whether the pantetheinyl arm can simply swing freely is yet to be determined. In either case, the controls for such motions are still to be determined. Finally, a growing body of evidence argues that chain transfer between PKS modules is mediated by a combination of several features—substrate recognition, linker interaction, and perhaps even ACP−KS interaction. Of these individual factors, only the linker interactions are obviously modular, since they can be engineered independent of the catalytic activities of flanking modules. However, although linker-mediated interactions by themselves may be adequate to facilitate chain transfer of some otherwise poor substrates, not all unfavorable substrate recognition features can be overcome in this manner. Accordingly, heterologous transfer of inherently poor substrates is likely to require alteration of the acceptor KS domain to modify its recognition features. Here, too, structural insights should be invaluable to facilitate such engineering.

NOTE ADDED IN PROOF

The importance of interdomain linkers in the human fatty acid synthase has been vividly highlighted by Wakil and coworkers in a recent report (*52*).

REFERENCES

1. Angel, P., and Karin, M. (1991) *Biochim. Biophys. Acta 1072*, 129−157.
2. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989) *Science 243*, 538−542.
3. O'Shea, E. K., Rutkowski, R., Stafford, W. F. D., and Kim, P. S. (1989) *Science 245*, 646−648.
4. Cohen, G. B., Ren, R., and Baltimore, D. (1995) *Cell 80*, 237−248.
5. Kuriyan, J., and Cowburn, D. (1997) *Annu. Rev. Biophys. Biomol. Struct. 26*, 259−288.
6. Williams, J. C., Wierenga, R. K., and Saraste, M. (1998) *Trends Biochem. Sci. 23*, 179−184.
7. Gonfloni, S., Frischknecht, F., Way, M., and Superti-Furga, G. (1999) *Nat. Struct. Biol. 6*, 760−764.
8. Banerjee, R. V., Frasca, V., Ballou, D. P., and Matthews, R. G. (1990) *Biochemistry 29*, 11101−11109.
9. Frasca, V., Banerjee, R. V., Dunham, W. R., Sands, R. H., and Matthews, R. G. (1988) *Biochemistry 27*, 8458−8465.
10. Banerjee, R. V., Harder, S. R., Ragsdale, S. W., and Matthews, R. G. (1990) *Biochemistry 29*, 1129−1135.
11. Drummond, J. T., Huang, S., Blumenthal, R. M., and Matthews, R. G. (1993) *Biochemistry 32*, 9290−9295.
12. Banerjee, R. V., Johnston, N. L., Sobeski, J. K., Datta, P., and Matthews, R. G. (1989) *J. Biol. Chem. 264*, 13888−13895.
13. Drummond, J. T., Loo, R. R. O., and Matthews, R. G. (1993) *Biochemistry 32*, 9282−9289.
14. Goulding, C. W., and Matthews, R. G. (1996) *FASEB J. 10*, C34−C34.
15. Roberts, D. L., Zhao, S., Doukov, T., and Ragsdale, S. W. (1994) *J. Bacteriol. 176*, 6127−6130.
16. Jarrett, J. T., Huang, S., and Matthews, R. G. (1998) *Biochemistry 37*, 5372−5382.
17. de Kok, A., Hengeveld, A. F., Martin, A., and Westphal, A. H. (1998) *Biochim. Biophys. Acta 1385*, 353−366.
18. Perham, R. N. (1991) *Biochemistry 30*, 8501−8512.
19. Reed, L. J., and Hackert, M. L. (1990) *J. Biol. Chem. 265*, 8971−8974.
20. Hengeveld, A. F., Westphal, A. H., and de Kok, A. (1997) *Eur. J. Biochem. 250*, 260−268.
21. Reed, L. J. (1974) *Acc. Chem. Res. 7*, 40−46.
22. Hackert, M. L., Oliver, R. M., and Reed, L. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2907−2911.
23. Dafydd Jones, D., Stott, K. M., Howard, M. J., and Perham, R. N. (2000) *Biochemistry 39*, 8448−8459.
24. Dave, E., Guest, J. R., and Attwood, M. M. (1995) *Microbiology 141*, 1839−1849.
25. Guest, J. R., Attwood, M. M., Machado, R. S., Matqi, K. Y., Shaw, J. E., and Turner, S. L. (1997) *Microbiology 143*, 457−466.
26. Allen, A. G., Perham, R. N., Allison, N., Miles, J. S., and Guest, J. R. (1989) *J. Mol. Biol. 208*, 623−633.
27. Bates, D. L., Danson, M. J., Hale, G., Hooper, E. A., and Perham, R. N. (1977) *Nature 268*, 313−316.
28. Collins, J. H., and Reed, L. J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4223−4227.
29. Smith, S. (1994) *FASEB J. 8*, 1248−1259.
30. Carreras, C. W., Pieper, R., and Khosla, C. (1997) *Top. Curr. Chem. 188*, 85−126.
31. Staunton, J., and Wilkinson, B. (1997) *Chem. Rev. 97*, 2611−2630.
32. Cane, D. E., and Walsh, C. T. (1999) *Chem. Biol. 6*, R319−R325.
33. Joshi, A. K., Witkowski, A., and Smith, S. (1998) *Biochemistry 37*, 2515−2523.
34. Stoops, J. K., and Wakil, S. J. (1982) *J. Biol. Chem. 257*, 3230.
35. Stoops, J. K., Henry, S. J., and Wakil, S. J. (1983) *J. Biol. Chem. 258*, 2482−2486.
36. Stoops, J. K., and Wakil, S. J. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4544−4548.
37. Stoops, J. K., and Wakil, S. J. (1981) *J. Biol. Chem. 256*, 5128−5133.
38. Foster, R. J., Poulose, A. J., Bonsall, R. F., and Kolattukudy, P. E. (1985) *J. Biol. Chem. 260*, 2826−2831.
39. Yuan, Z. Y., and Hammes, G. G. (1986) *J. Biol. Chem. 261*, 13643.
40. Chang, S. I., and Hammes, G. G. (1989) *Biochemistry 28*, 3781−3788.
41. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J., and Leadlay, P. F. (1990) *Nature 348*, 176−178.
42. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., and Katz, L. *Science 252*, 675−679.
43. Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999) *Science 284*, 482−485.
44. Wu, N., Kudo, F., Cane, D. E., and Khosla, C. (2000) *J. Am. Chem. Soc. 122*, 4847−4852.

45. Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001) *Biochemistry 40*, 2326−2331.
46. Staunton, J., Caffrey, P., Aparicio, J. F., Roberts, G. A., Bethell, S. S., and Leadlay, P. F. (1996) *Nat. Struct. Biol. 3*, 188−192.
47. Kao, C. M., Pieper, R., Cane, D. E., and Khosla, C. (1996) *Biochemistry 35*, 12363−12368.
48. Gokhale, R. S., Lau, J., Cane, D. E., and Khosla, C. (1998) *Biochemistry 37*, 2524−2528.
49. Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 5848−5853.
50. Belshaw, P. J., Walsh, C. T., and Stachelhaus, T. (1999) *Science 284*, 486−489.
51. Wilkinson, B., Foster, G., Rudd, B. A. M., Taylor, N. L., Blackaby, A. P., Sidebottom, P. J., Cooper, D. J., Dawson, M. J., Buss, A. D., Gaisser, S., Bohm, I., Rowe, C. J., Cortes, J., Leadlay, P. F., and Staunton, J. (2000) *Chem. Biol. 7*, 111−117.
52. Chirala, S. S., Jayakumar, A., Gu, Z-W., and Wakil, S. J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* (in press).

BI002462V