Pergamon

TETRAHEDRON

Tetrahedron 56 (2000) 9401–9408

Dissecting the Chain Length Specificity in Bacterial Aromatic Polyketide Synthases using Chimeric Genes

Kim K. Burson† and Chaitan Khosla*

Departments of Chemistry and Chemical Engineering, Stanford University, Stanford, CA 94305, USA

Received 10 February 2000; accepted 13 July 2000

Abstract—Many aromatic polyketides are synthesized by bacterial enzyme complexes called polyketide synthases (PKSs). Previous work has indicated that two subunits in the PKS, called the ketosynthase (KS) and chain length factor (CLF), play a central role in determining the chain length of the polyketide product. The purpose of this study was to construct and analyze chimeric KS and CLF proteins in order to identify the structural determinants of chain length. The genes encoding KS and CLF subunits from different aromatic PKSs share significant (50–70%) sequence identity. This similarity provides a means to genetically engineer the PKSs by directed gene shuffling methods. A set of hybrid KS and CLF genes were designed based on the sequences of the actinorhodin (a 16 carbon polyketide) and tetracenomycin (a 20 carbon polyketide) PKS genes. Analysis of these chimeric genes ruled out a role for most of the KS and reinforced the importance of the CLF in controlling polyketide chain length. Homology modeling of the quaternary structures of the functional hybrid KS–CLF dimers based on recent crystal structures of the ketosynthase homodimer from *E. coli* suggested that the regions of the KS and CLF that are most important for activity and specificity are located at the interface of this dimer. © 2000 Elsevier Science Ltd. All rights reserved.

Introduction

Microbial polyketides are a diverse group of bioactive products whose carbon chain backbones are derived from multiple rounds of decarboxylative condensations between carboxylic acid monomers. They are produced as secondary metabolites in fungi and bacteria, especially those belonging to the actinomyces family of bacteria. Although they are structurally diverse, their carbon chain backbones are synthesized by homologous multifunctional enzymes called polyketide synthases (PKSs), which in turn are related to fatty acid synthases (FASs).[1–4]

Bacterial aromatic PKSs comprise a sub-class within the PKS superfamily, and catalyze the biosynthesis of a variety of polyfunctional aromatic natural products such as actinorhodin, tetracenomycin, oxytetracycline, frenolicin, and doxorubicin.[5] Each bacterial aromatic PKS consists of a core set of proteins called the 'minimal' PKS, which includes the ketosynthase (KS), the chain length factor (CLF) and the acyl carrier protein (ACP). (There is some debate regarding the role of the malonyl transferase (MAT) from the fatty acid synthase in polyketide biosynthesis;[6] however, several laboratories have established that at physiologically relevant concentrations of individual proteins, the MAT-dependent pathway for malonyl-S-ACP formation is dominant.[7–9]) These PKS subunits, which are responsible for synthesizing nascent poly-β-ketone backbones of defined lengths (Fig. 1), are highly conserved (50–70% identical). The KS and CLF in the minimal PKS have been implicated in determining the chain length of the final product.[10,11] Since the diversity of aromatic polyketides is due in part to variability in chain length specificity, there is considerable interest in understanding the structural and mechanistic basis for chain length control by these enzymes.

Early work on bacterial aromatic PKSs was aimed at understanding the properties of individual subunits by deletion and recombination of intact genes. Toward this end, a *Streptomyces* host-vector system was developed, which proved useful for the construction and analysis of recombinant aromatic PKS gene sets.[10] The host strain, *S. coelicolor* CH999, was constructed by deleting the entire actinorhodin (*act*) gene cluster through homologous recombination. A bifunctional shuttle vector, pRM5, was used to express hybrid PKS genes in CH999. A similar system was also developed in *S. glaucescens*.[12] A series of investigations have been conducted using the *act*, granaticin (*gra*), frenolicin (*fren*), griseusin (*gris*), tetracenomycin (*tcm*),

Keywords: polyketides; polyketide synthases; ketosynthase; chain length specificity.

* Corresponding author. Tel./fax: +650-723-6538; e-mail: k@chemeng.stanford.edu

Abbreviations: CoA, coenzyme A; PKS, polyketide synthase; KS, ketosynthase; CLF, chain length factor; ACP, acyl carrier protein; KR, ketoreductase; ARO, aromatase; CYC, cyclase; *act*, actinorhodin; *tcm*, tetracenomycin; KAS II, β-ketoacyl-acyl carrier protein synthase II; FAS, fatty acid synthase.

† This research was supported by a grant CA77248 from the NIH. KB was supported in part by predoctoral fellowships from the NSF, Evelyn Laing McBain Foundation, and Ford Foundation. We thank Dr Ylva Lindqvist for providing us with coordinates for the crystal structure of KAS II.

0040–4020/00/$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved.
PII: S0040-4020(00)00824-3

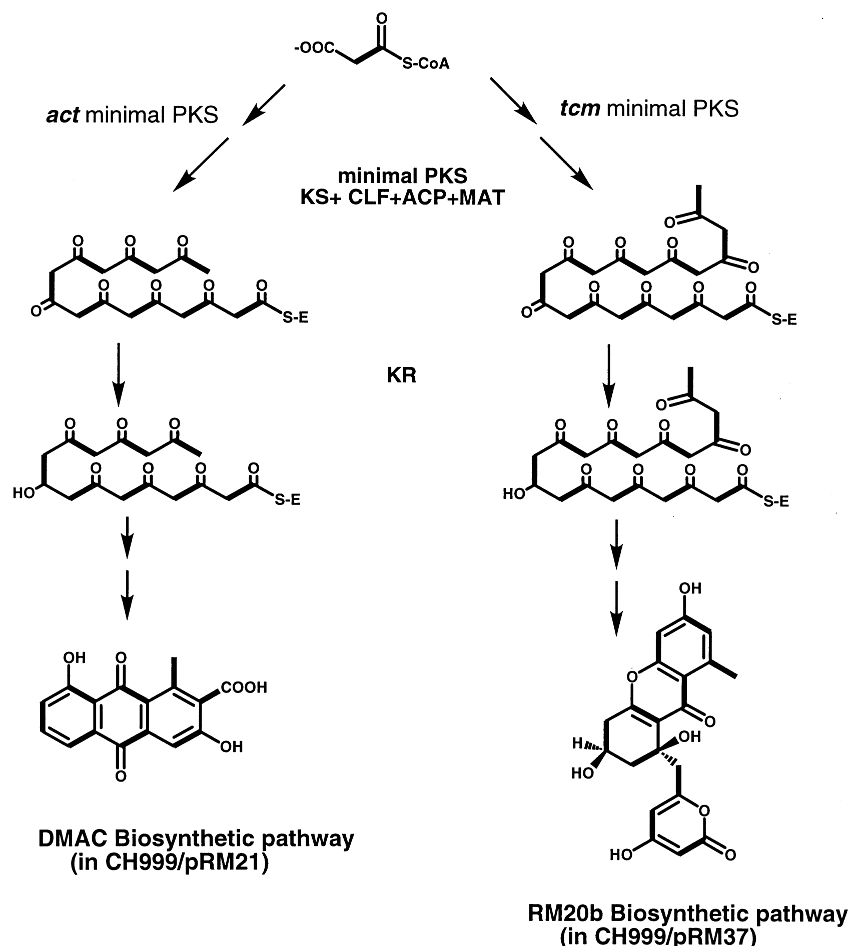

Figure 1. Proposed biosynthetic pathways of catalyzed by the *act* and the *tcm* minimal PKSs in *S. coelicolor* CH999/pRM21 and CH999/pRM37, respectively.

oxytetracycline (*otc*) and the *whiE* spore pigment PKS genes.[10–23] Among other properties of aromatic PKSs, these results revealed that (i) in each case where a product was isolated from a recombinant PKS, the chain length of the polyketide corresponded to the source of the CLF, (ii) the ACP did not affect the chain length of the product, and (iii) since not all hybrid KS–CLF were functional, it was not possible to rule out any influence of the KS on chain length control.

Table 1. Analysis of chimeric KS proteins

Chimera #	First segment[a]	Second segment[a]	Third segment[a]	PRM21[b] (*act* CLF)	pRM37[b] (*tcm*CLF)
60	*tcm* A	–	*act* B–F	16C	N/P
61	*act* A	*tcm* B	*act* C–F	N/P	N/P
62	*act* A–B	*tcm* C	*act* D–F	16C	N/P
63	*act* A–C	*tcm* D	*act* E–F	N/P	20C
64	*act* A–D	*tcm* E	*act* F	16C	N/P
65	*act* A–E	–	*tcm* F	16C	N/P
66	*act* A	–	*tcm* B–F	N/P	N/P
67	*tcm* A	*act* B	*tcm* C–F	16C	20C
68	*tcm* A–B	*act* C	*tcm* D–F	ND	ND
69	*tcm* A–C	*act* D	*tcm* E–F	16C	N/P
70	*tcm* A–D	*act* E	*tcm* F	16C	N/P
71	*tcm* A–E	–	*act* F	N/P	N/P

[a] Each segment is defined by the cassettes it includes, as illustrated in Fig. 3. For information regarding the precise sequence of fusion junctions, see Fig. 2 and the Experimental Procedures.

[b] The KS library was inserted into either pRM21 or pRM37 shuttle plasmid and transformed into CH999. The hybrid produced either a 16 carbon (16C) or a 20 carbon (20C) polyketide or was a non-producer (N/P). ND=not determined.

The synergistic application of gene fusion and structural approaches has often proven useful in dissecting divergent properties within families of homologous enzymes. The feasibility of constructing functional hybrids from naturally occurring aromatic PKS subunits has been demonstrated in the cases of the ACP,[24] the KS[25] and the bifunctional aromatase/cyclase[26] proteins. Here we sought to extend this analysis by constructing a systematic series of chimeric KS and CLF subunits. When these studies were initiated, the absence of any structural data for this family of enzymes necessitated exclusive reliance on primary sequence information for the design of these chimeric proteins. Recently, however, two X-ray crystallographic structures of the KAS II ketosynthase dimer encoded by the *fabF* gene of *E. coli* have been reported (both as unbound and cerulenin-bound forms).[27,28] The significant sequence identity between KAS II and the KS (46–47% for *act* and *tcm*) as well as the CLF (38–42% for *act* and *tcm*) proteins provided us with an opportunity to assess the structural implications of our experimental findings. The results of these modeling studies are also presented here.

```
actclf   ~~~~~~~MSV LITGVGVVAP NGLGLAPYWS AVLDGRHGLG
tcmclf   ~~~MSAPAPV VVTGLGIVAP NGTGTEEYWA ATLAGKSGID
actks    ~~~~~LKRRV VITGVGVRAP GGNGTRQFWE LLTSGRTATR
tcmks    ~MTRHAEKRV VITGIGVRAP GGAGTAAFWD LLTAGRTATR
fabF     ~~~~~SKRRV VVTGLGMLSP VGNTVESTWK ALLAGQSGIS

actclf   PVTRFDVSRY PATLAGQIDD FHAPDH.IPG RLLPQTDPST RLALTAADWA
tcmclf   VIQRFDPHGY PVRVGGEVLA FDAAAH.LPG RLLPQTDRMT QHALVAAEWA
actks    RISFFDPSPY RSQVAAEA.D FDPVAEGFGP RELDRMDRAS QFAVACAREA
tcmks    TISLFDAAPY RSRIAGEI.D FDPIGEGLSP RQASTYDRAT QLAVVCAREA
fabF     LIDHFDTSAY ATKFAGLVKD FN.CEDIISR KEQRKMDAFI QYGIVAGVQA

actclf   LQDAKADPES .LTD..YDMG VVTANACGGF DFTHREFRKL ......WSEG
tcmclf   LADAGLEPEK .QDE..YGLG VLTAAGAGGF EFGQREMQKL ......WGTG
actks    FAASGLDPDT L...DPARVG VSLGSAVAAA TSLEREYLLL SDSGRDWEVD
tcmks    LKDSGLDPAA V...NPERIG VSIGTAVGCT TGLDREYARV SEGGSRWLVD
fabF     MQDSGLEITE E...NATRIG AAIGSGIGGL GLIEENHTSL MNGG......

actclf   PKSVSVYESF AWFYAVNTGQ ISIRHGMRGP SSALVAEQAG GLDALGHARR
tcmclf   PERVSAYQSF AWFYAVNTGQ ISIRHGMRGH SSVFVTEQAG GLDAAAHAAR
actks    AAWLSRHMFD YLVPSVMPAE VAWAVGAEGP VTMVSTGCTS GLDSVGNAVR
tcmks    HTLAVEQLFD YFVPTSICRE VAWEAGAEGP VTVVSTGCTS GLDAVGYGTE
fabF     PRKISPFFVP STIVNMVAGH LTIMYGLRGP SISIATACTS GVHNIGHAAR

actclf   TIRRGTP.LV VSGGVDSALD PWGWVS.QIA SGRISTATDP DRAYLPFDER
tcmclf   LLRKGTLNTA LTGGCEASLC PWGLVA.QIP SGFLSEATDP HDAYLPFDAR
actks    AIEEGSADVM FAGAADTPIT PIVVACFDAI RATTARNDDP EHASRPFDGT
tcmks    LIRDGRADVV VCGATDAPIS PITVACFDAI KATSANNDDP AHASRPFDRN
fabF     IIAYGDADVM VAGGAEKAST PLGVGGFGAA RALSTRNDNP QAASRPWDKE

actclf   AAGYVPGEG. GAILVLEDSA AAEARGRHDA YGELAGCAST FDP..APG.S
tcmclf   AAGYVPGEG. GAMLVAERAD SARERDAATV YGRIAGHAST FDA..RPG.T
actks    RDGFVLAEG. AAMFVLEDYD SALARGAR.I HAEISGYATR CNAYHMTGLK
tcmks    RDGFVLGEG. SAVFVLEELS AARRRGAH.A YAEVRGFATR SNAFHMTGLK
fabF     RDGFVLGDG. AGMLVLEEYE HAKKRGAK.I YAELVGFGMS SDAYHMTSPP

actclf   GRPAGLERAI RLALNDAGTG PEDVDVVFAD GAGVPELDAA EARAIGRVFG
tcmclf   GRPTGPARAI RLALEEARVA PEDVDVVYAD AAGVPALDRA EAEALAEVFG
actks    ADGREMAETI RVALDESRTD ATDIDYINAH GSGTRQNDRH ETAAYKRALG
tcmks    PDGREMAEAI TAALDQARRT GDDLHYINAH GSGTRQNDRH ETAAFKRSLG
fabF     ENGAGAALAM ANALRDAGIE ASQIGYVNAH GTSTPAGDKA EAQAVKTIFG

actclf   REG..VPVTV PKTTTGRLYS GGGPLDVVTA LMSLREGVIA PTAGVTSVPR
tcmclf   PGA..VPVTA PKTMTGRLYA GGAALDVATA LLSIRDCVVP PTVGHRCARA
actks    EHARRTPVSS IKSMVGHSLG AIGSLEIAAV VLALEHGVVP PTANLRTSDP
tcmks    QRAYDVPVSS IKSMIGHSLG AIGSLELAAC ALAIEHGVIP PTANYEEPDP
fabF     EAASRVLVSS TKSMTGHLLG AAGAVESIYS ILALRDQAVP PTINLDNPDE

actclf   EYGIDLVLGE PRSTA.PRTA LVLARGRWGF NSAAVLR.RF APTP~~
tcmclf   GLGIDLVLHQ PRELR.VDTA LVVARGMGGF NSALVVRRHG ~~~~~~
actks    ECDLDYVPLE ARE.RKLRSV LTVGSGFGGF QSAMVLRDAE TAGAAA
tcmks    ECDLDYVPNV ARE.QRVDTV LSVGSGFGGF QSAAVLARPK ETRS~~
fabF     GCDLDFVPHE ARQVSGMEYT LCNSFGFGGT NGSLIFKKI~ ~~~~~~
```

Figure 2. Sequence alignments between KS, CLF and KAS II (fabF). The fusion junctions in the KS are highlighted in turquoise, whereas the fusion junctions in the CLF are highlighted in red. The active site cysteine in the KS is indicated in dark blue. The region of the KS shown in purple was found in an earlier study to be interchangeable without affecting chain length.[25]

Results

Design and analysis of chimeric KS genes

As described in Table 1, eleven chimeric KS genes were constructed, each containing substitution of a single cassette from the *act* or *tcm* gene with homologous DNA from its counterpart (Fig. 3). (The twelfth chimera could not be constructed for technical reasons.) Since fusion junctions were engineered at highly conserved stretches of residues, they can be expected to lie at overlapping positions in the superimposed 3-dimensional structures (as yet unknown) of the two ketosynthases. By doing so, we hoped to maximize the likelihood of obtaining chimeric polypeptides that were capable of folding into native KS-like folds.

Table 1 lists the phenotypes of eleven chimeric KS genes when co-expressed with either the wild-type *act* CLF (in

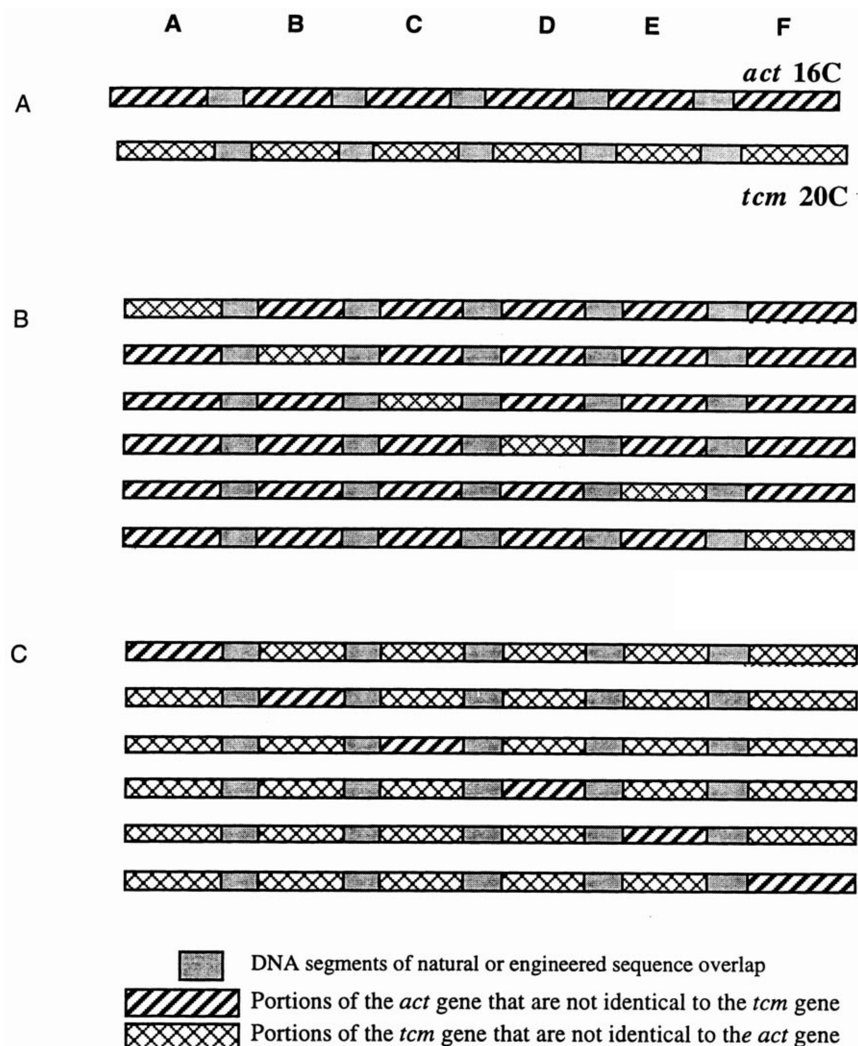

Figure 3. Design and construction of a library of chimeric KS and CLF genes. As shown in (A), each *act* and the *tcm* gene was divided into six 'cassettes', designated A–F. The 'junctions' between adjacent cassettes correspond to highly conserved segments highlighted in Fig. 2. In cases of sequence differences between corresponding junctions, the *tcm* sequence was converted into the *act* sequence. Each cassette was individually amplified via PCR together with its flanking junctions. In order to construct the chimeric genes described in (B) and (C), an appropriate set of cassettes was then mixed together and amplified with suitable oligonucleotides corresponding to the 5′ and 3′ ends of the full-length chimeric gene. As controls, the intact *act* and *tcm* genes shown in (A) were also constructed and their activity was confirmed.

pRM21) or the *tcm* CLF (in pRM37) genes. In the presence of the *act* CLF, 7 out of 11 chimeric genes yielded polyketide products. In contrast only 2 chimeras produced polyketides when co-expressed with the *tcm* CLF gene. In every case where a 16 carbon or 20 carbon product was detected (Table 1), the chain length corresponded to that dictated by the CLF. These findings are consistent with earlier studies involving intact gene replacements where polyketide chain length invariably tracked with the source of the CLF gene.[10,11] A similar conclusion is also derived from the analysis of the KS chimeras #63 and 69. Chimera #63 is identical to the *act* KS except for cassette D, which includes the 50 amino acid residues immediately upstream of the active site cysteine. When co-expressed with the *tcm* CLF, this chimeric KS yields a 20-carbon backbone. Reciprocally, in chimera #69, cassette D from the *act* KS is used to replace its counterpart in the *tcm* KS. When co-expressed with the *act* CLF, this chimeric KS yields a 16-carbon backbone. Taken together, these results demonstrate that the influence, if any, of the KS on chain length specificity is limited to cassette D. This cassette corresponds to residues 112–174 in the *act* KS.

Design and analysis of chimeric CLF genes

Analogous to the KS gene shuffling studies described above, twelve chimeric CLF genes were also designed and constructed (Figs. 2 and 3). Each chimera was independently co-expressed with either the *act* KS (in pRM21) or the *tcm* KS (in pRM37) gene. Out of a total of 24 recombinant PKSs, only five PKSs were active, as judged by polyketide production in vivo (Table 2). Four of these (#77–79 and 82) included the wild-type *tcm* KS. Two conclusions can be drawn from the product profiles of these clones. First, chimera #77 produces a 16-carbon backbone when co-expressed with either the *act* or the *tcm* KS. Since cassette F of this hybrid protein corresponds to the *tcm* CLF, one can conclude that this portion of the CLF does not influence chain length. Second, the phenotypes of chimeras #78, 79, and 82, when co-expressed with the *tcm* KS, can be

Table 2. Analysis of chimeric CLF proteins

Chimera #	First segment[a]	Second segment[a]	Third segment[a]	pRM21[b] (*act* KS)	PRM37[b] (*tcm* KS)
72	*tcm* A	–	*act* B–F	N/P	N/P
73	*act* A	*tcm* B	*act* C–F	N/P	N/P
74	*act* A–B	*tcm* C	*act* D–F	N/P	N/P
75	*act* A–C	*tcm* D	*act* E–F	N/P	N/P
76	*act* A–D	*tcm* E	*act* F	N/P	N/P
77	*act* A–E	–	*tcm* F	16C	16C
78	*act* A	–	*tcm* B–F	N/P	20C
79	*tcm* A	*act* B	*tcm* C–F	N/P	20C
80	*tcm* A–B	*act* C	*tcm* D–F	N/P	N/P
81	*tcm* A–C	*act* D	*tcm* E–F	N/P	N/P
82	*tcm* A–D	*act* E	*tcm* F	N/P	20C
83	*tcm* A–E	–	*act* F	N/P	N/P

[a] Each segment is defined by the cassettes it includes, as illustrated in Fig. 3. For information regarding the precise sequence of fusion junctions, see Fig. 2 and the Experimental Procedures.
[b] The CLF library was inserted into either pRM21 or pRM37 shuttle plasmid and transformed into CH999. The hybrid produced either a 16 carbon (16C) or a 20 carbon (20C) polyketide or was a non-producer (N/P)

compared with a heterodimer comprised of the *tcm* KS and the *act* CLF. Earlier studies[10] showed that the *tcm* KS/*act* CLF heterodimer produces a 16-carbon polyketide. In contrast, all three chimeric CLF proteins produce 20-carbon polyketides when co-expressed with the *tcm* KS. Therefore replacement of either cassettes B–F (#78), or B and D–F (#79), or A–D and F (#82) from the *tcm* CLF into the *act* CLF is adequate to reverse the substrate specificity of the *tcm* KS/*act* CLF heterodimer. Taken together, the results summarized in Table 2 suggest that cassettes C and D in the CLF are pivotal for chain length control. This corresponds to residues 132–294 of the *act* CLF.

Analysis of the above results within the framework of the KAS II crystal structure

The crystal structure of a ketosynthase (KAS II) has been reported.[27] This homodimeric enzyme catalyzes chain growth in fatty acid biosynthesis in *E. coli*. Each subunit consists of two mixed five-stranded β-sheets packed against α-helices and folded into five layers, α–β–α–β–α. The conserved active site, C163, is located at the bottom of a hydrophobic pocket near the dimer interface, and is 25 Å away from the active site cysteine in the other subunit (Fig. 4).

More recently, the same protein was co-crystallized with the inhibitor cerulenin bound to its active site.[28] Analysis of the co-crystal structure facilitated better definition of the hydrophobic substrate-binding pocket in the KAS II homodimer. Cerulenin was observed to fit snugly in this pocket, and was surrounded by residues located at the homodimer interface. Notably, the size of the chain-binding pocket was constrained by a residue (F133) from the second subunit. The high degree of sequence similarity between the KS–CLF heterodimer and the KAS II homodimer (38% identity between the *act* KS and KAS II, and 30% identity between the *act* CLF and KAS II) suggests that the α–β–α–β–α fold is conserved. Moreover, it is likely that the binding pocket of the KS–CLF heterodimer is also located at the dimer interface. If so, then our experimental results are in excellent agreement with the KAS II structure, since homology modeling suggests that the PKS binding pocket is primarily composed of residues drawn from cassette D of the KS (residues 112–174) and cassettes C–D of the CLF (residues 132–294) (Fig. 4).

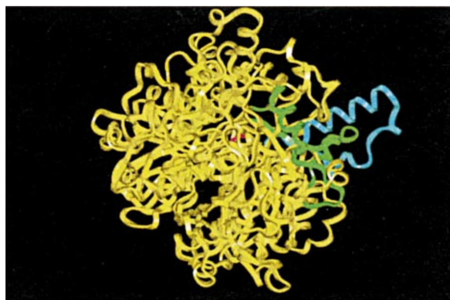

Figure 4. Structure of the KAS II homodimer. The active site cysteine, which is located in a deep cleft within the dimer, is indicated in red. The dimer interface, where the chain length specificity is presumably controlled, is shown in blue and green for the two subunits.

Discussion

The construction and analysis of hybrid enzymes has been an important component of the protein engineering toolbox for dissecting enzyme mechanism and specificity. For example, chimeras of trypsin and chymotrypsin,[31] and of trypsin and elastase[32] have been useful in understanding the structural basis for substrate specificity in these proteases. Similar studies have also been performed on sesquiterpene and triterpene cyclases involved in natural product biosynthesis.[33] Indeed, the host specificity and chain length specificity of proteins involved in Nod factor biosynthetic pathways in *Rhizobium* species has also been dissected via this approach.[34,35] Although the analysis of such data is ultimately limited by the availability of high-resolution structural information regarding the target proteins, the structures of homologous proteins can often ameliorate this limitation. In this study we have sought to combine the complementary capabilities of gene shuffling with those of structure-based homology modeling to gain further insights into the structural basis for chain length control by bacterial aromatic PKSs.

A series of chimeric KS and CLF genes were constructed

and analyzed, as reported in Tables 1 and 2. In these experiments the absence of polyketide biosynthetic activity is uninformative, since it may result from a variety of translational or post-translational factors. Therefore, we have chosen to limit our data analysis only to those chimeric PKSs that produce polyketides in vivo. Several conclusions have emerged from this analysis. First, a surprisingly large number of chimeric proteins are active. This suggests that the folding patterns and packing geometries of homologous PKS subunits are indeed very similar. Second, a significantly greater fraction of PKSs containing chimeric KS subunits were active as compared to PKSs containing chimeric CLF subunits (41 v/s 21%). Perhaps this reflects greater divergence in CLF structure than KS structure among this family of PKSs. This hypothesis is consistent with the notion that the most divergent property within the PKS family (chain length specificity) is primarily dictated by the CLF. Third, from Tables 1 and 2 as well as earlier data, one can conclude that the differential chain length specificity of the *act* and the *tcm* PKSs is programmed within residues 112–174 of the KS and 132–294 of the CLF (*act* numbering). Finally, by comparing these findings with the crystal structure of the KAS II ketosynthase from *E. coli*, we suggest that chain length control in bacterial aromatic PKSs is also dictated by a substrate-binding pocket located at the KS–CLF dimer interface.

This study provides a foundation for two further directions of investigation. On one hand the data presented here should be valuable in the analysis of high-resolution KS–CLF crystal structures that are likely to emerge in the future. At the same time, the above delineation of residues involved in chain length control can facilitate the design of more sophisticated protein engineering strategies aimed at expanding the chain length specificity of naturally occurring aromatic PKSs.

Experimental Procedures

Materials, strains, and recombinant DNA procedures

Restriction enzymes and other enzymes were obtained from New England Biolabs (Beverly, MA), BRL Gibco (Gaithersburg, MD) or Boehringer Mannheim (Indianapolis, IN) and used according to the instructions provided by the suppliers. All recombinant DNA manipulations were according to standard protocols.[29] All *Streptomyces*-related procedures were also according to standard protocols.[30] The principal plasmids used in this study were pRM21 and pRM37, whose construction has been described earlier.[13] Plasmid pRM21 contains the *act* minimal PKS gene set, with the KS gene flanked by unique *PacI* and *NsiI* restriction sites at the 5′ and 3′ ends, respectively. Immediately downstream of the KS gene is the CLF gene flanked by unique *NsiI* and *XbaI* sites. Similarly, pRM37 carries the *tcm* minimal PKS genes flanked by the same set of restriction sites. Therefore, substitution of either the KS or the CLF gene or both is a straightforward matter in both plasmids. The principal metabolite of CH999/pRM21 is the octaketide 3,8-dihydroxy-1-methylanthraquinone carboxylic acid (DMAC),[10] whereas the principal metabolite of CH999/pRM37 is the decaketide RM21b.[17]

Design of chimeric *act* and *tcm* KS and CLF genes

Earlier studies had revealed that the C-terminal ends (163 residues, starting from the conserved sequence EISGY) of the *act* and *tcm* KS genes could be interchanged without affecting chain length specificity of the resulting proteins.[25] Therefore, our studies here focused on generating chimeric KS genes with fusion junctions localized upstream of this C-terminal segment. To design a set of *act*/*tcm* chimeras, the amino acid sequences of the KS and CLF pairs were aligned (Fig. 2). The strategy for chimeric gene design is illustrated in Fig. 3. In each polypeptide five approximately evenly spaced segments of 6–7 amino acid residues were identified, where protein sequence was identical or nearly identical. These short segments were designated as 'junctions' for the design of chimeric KS and CLF genes. In those junctions where amino acid changes were required, the *tcm* residue was replaced with the corresponding *act* residue. Likewise, where silent nucleotide changes were required, the *tcm* nucleotide was replaced with the corresponding *act* nucleotide. In only three cases were relatively radical amino acid substitutions required (G→S; A→S and A→P). As described below, all these changes were phenotypically silent, since the resulting *act* and *tcm* KS and CLF genes were indistinguishable from wild-type genes based on in vivo product analysis.

Each chimeric KS and CLF gene was constructed through the assembly of a defined set of six interchangeable cassettes. Individual cassettes were generated via PCR amplification using primers based on the conserved junction sequences described above. Thus, twelve primers were designed for each gene; their sequences are listed below. Two primers, one forward (odd numbers, e.g. actks 1) and one reverse (even numbers, e.g. actks 2), were generated based on the sequence of each junction. Because the *tcm* KS and CLF sequences were changed to the corresponding *act* sequences in these junctions, it was necessary to extend the *tcm* primers at their 3′ ends. Some primers are common to both genes, and are indicated as such. Primers flanking the ends of the genes (actks 1, acttcmks 12, tcmks 1, acttcmclf 1, actclf 12, and tcmclf 12) included suitable restriction sites to facilitate cloning of chimeric genes in pRM21 or pRM37. The following primers were used (all sequences are 5′ to 3′; base changes are indicated in bold):

actks 1 primer, TTTAAGCTTAATTAAGGAGGACCATCATGAAGCGCAGA;
actks 2 primer, GAGGTGAGCAGTTCCCAGAA;
actks 3 primer, TTCTGGGAACTGCTCACCTC;
actks 4 primer, CTCGGCGACCGGGTCGAAGT;
actks 5 primer, ACTTCGACCCGGTCGCCGAG;
actks 6 primer, TACTCGCGCTCCAGGCTGGT;
actks 7 primer, ACCAGCCTGGAGCGCGAGTA;
actks 8 primer, AGGCCCGAGGTGCAGCCGGT;
actks 9 primer, ACCGGCTGCACCTCGGGCCT;
actks 10 primer, ATCGCGTCGAAGCAGGCGAC;
actks 11 primer, GTCGCCTGCTTCGACGCGAT;
acttcmks 12 primer, GCGTCGCGTACCCCGAGATCT (common);
tcmks 1 primer, TTTAAGCTTAATTAAGGAGGACCATCATGA;

tcmks 2 primer, G**A**GGTGAGCAG**T**TCCCAGAAGG-CCGCGGTTCC;
tcmks 3 primer, TTCTGGGA**A**CTGCTCACC**T**CCGG-ACGCACCGCCACCAGG;
tcmks 4 primer, CTCG**G**CGA**C**CGGGTCGAAGTCGA-TCTCTC;
tcmks 5 primer, ACTTCGACCCG**G**TCG**C**CGAGGGC-CTGTCGCCCCGGC;
tcmks 6 primer, TACTCGCG**C**TCCAGGC**T**GGTGG-TGCAGCCGACGGCGG;
tcmks 7 primer, ACC**A**GCCTGGA**G**CGCGAGTACGC-GCGGGTCAG;
tcmks 8 primer, AGGCCCGAGGTGCA**G**CCGGTCGA-CACCACCGTCAC;
tcmks 9 primer, ACCGG**C**TGCACCTCGGGCCTCG-ACGC;
tcmks 10 primer, AT**C**GCGTCGAAGCAGGC**G**ACGG-TGATGGGGGAGATCGG;
tcmks 11 primer, GT**C**GCCTGCTTCGACGC**G**ATCA-AGGCGACGTCGGCCAA;
acttcmks 12 primer, GCGTCGCGTACCCCGAGATCT (common);
acttcmclf 1 primer, TTTAAGCTTATGCATGGAGGA-GCCATC (common);
actclf 2 primer, GGTCCGTCTGCGGCAGCAAC;
actclf 3 primer, GTTGCTGCCGCAGACGGACC;
acttcmclf 4 primer, CGGATGGAGATCTGGCCGGT (common);
acttcmclf 5 primer, ACCGGCCAGATCTCCATCCG (common);
actclf 6 primer, GAACGTAGCCGGCCGCCCGC;
actclf 7 primer, GCGGGCGGCCGGCTACGTTCCG;
actclf 8 primer, ACGACGTCGACGTCCTCGGGACCG;
actclf 9 primer, CCCGAGGACGTCGACGTCGT;
actclf 10 primer, ACGTCGAGCGGGCCGCCGCC;
actclf 11 primer, GGCGGCGGCCCGCTCGACGT;
actclf 12 primer, TTTGAATTCTAGATTACGGGGTC-GGTGCGAAA;
acttcmclf 1 primer, TTTAAGCTTATGCATGGAGGA-GCCATC (common);
tcmclf 2 primer, GGTC**C**GTCTGCGGCAGCA**A**CCGC-CCGGGGAGATGGGCGGC;
tcmclf 3 primer, **GT**TGCTGCCGCAGAC**G**GACCGCA-TGACCCAGCACG;
tcmclf 6 primer, G**A**ACGTA**G**CCGGCCGCCCGCG-CGTCGAA;
tcmclf 7 primer, GCGGGCGGCCGG**C**TACGT**T**CCC-GGCGAGGGCGGAGCCAT;
tcmclf 8 primer, AC**G**ACGTCGACGTCCTCGGGCGC;
tcmclf 9 primer, CCCGAGGACGTCGACGT**C**GTC-TACGCCGACGCCGCTGG;
tcmclf 10 primer, ACGTCGAG**C**G**GG**CC**G**CC**G**CCG-GCGTAGAGCCGTCCGGT;
tcmclf 11 primer, GG**C**GG**C**G**GC**CC**G**CTCGACGTC-GCGACCGCGCT;
tcmclf 12 primer, TTTGAATTCTAGATTAGCCGTG-GCGCCGGACGAC.

PCR procedures

Six cassettes for each of the four genes were amplified via PCR using pRM21 and pRM37 as templates. Each 50 μL PCR reaction contained 0.1 μg template, 100 pmoles of each primer, 1 mM each dNTP, and 2.5 units Pfu polymerase (Stratagene). A hot start PCR program was used on a Perkin Elmer Gene Amp 2400 with the following parameters: 95°C for 2 min; 25× (95°C for 1 min, 65 or 66°C for 1 min, 75°C for 1 min); 75°C for 7 min. (For KS cassettes an annealing temperature of 65°C was used, whereas for CLF cassettes an annealing temperature of 66°C was used.) Cassettes were individually gel-purified by using Qiaquick gel extraction columns (Qiagen).

Purified KS cassettes were assembled into full length DNA fragments (encoding only the N-terminal two-thirds of the KS gene) as shown in Table 1. For each chimera, a cocktail containing 2 μL of each cassette, the supplier's buffer for the Pfu polymerase (Stratagene; containing 2 mM Mg^{+2}), 100 μM each dNTP, and 1.2 units of Klenow polymerase (Stratagene) was prepared in a final volume of 10 μL. After a 30 min incubation at 37°C in the absence of any primers, this reaction mixture was supplemented with 100 pmol each of the appropriate end-primers, 1 mM final concentration of each dNTP (a 1:1 mixture of dGTP:7-deaza-dGTP was used), 2.5 units of Pfu polymerase, and additional supplier's buffer in a final volume of 100 μL. The following PCR protocol was used: 95°C for 2 min, 25 X (94°C for 1 min, 64°C for 1 min, and 72°C 1 min), and 72°C for 10 min. In every case except for chimera #68 (Table 1) gel electophoresis revealed that an expected DNA fragment of 828 bp was generated. These DNA fragments were gel-purified and cloned into pRM21 or pRM37. In all cases, as described above, the C-terminal end of the KS genes were identical.

A similar protocol was used to assemble purified CLF cassettes into full-length chimeric genes as shown in Table 2. The principal differences were as follows. In the case of CLF genes, a single cycle of Klenow-catalyzed assembly was inadequate. Instead, two cycles of Pfu-catalyzed assembly without primers were required (94°C for 2 min, 2× (94°C for 30 s and 70°C for 5 min), and 72°C for 7 min). Second, these PCR mixtures contained 1.5-fold greater concentrations of each cassette. Third, 50 pmol of each end primer was used in the amplification step. Finally, the amplification protocol involved 25× (94°C for 1 min, 68°C for 1 min, 75°C for 1 min); and 75°C for 7 min.

Precautions to minimize the implications of unexpected PCR errors

Although unexpected mutations could have been generated via the PCR protocols described above, it was impractical to sequence the large number of chimeric genes described in this study. To minimize the likelihood that such PCR errors might lead to inaccurate conclusions, several precautions were taken. First, every PCR protocol was assessed to ascertain that the number of PCR cycles used were kept to a minimum. Second, a high fidelity thermostable DNA polymerase was used throughout the study. Third, as described below, two independent clones were analyzed for every chimera. Fourth, after cloning, each chimeric gene was validated for the incorporation of the expected heterologous cassette by PCR analysis using sequence

specific oligonucleotides. Finally, intact *act* and *tcm* KS and CLF genes were generated as controls using the above protocols and the appropriate set of cassettes. When introduced into *S. coelicolor* CH999, 95% of the colonies derived from these plasmids were phenotypically unaltered.

Analysis of polyketide products

For each construct carrying a chimeric KS or CLF gene, two independent transformants of *S. coelicolor* CH999 were analyzed as described earlier (13). Briefly, each clone was streaked onto two fresh R2YE plates (ca. 35 mL agar media per plate). After 7–10 days of growth, the agar was finely chopped and extracted two times with ethyl acetate and 1% acetic acid. Following concentration by evaporation under vacuum at room temperature, the extract was flashed through a silica gel column using ethyl acetate +1% acetic acid as the solvent. Colored fractions were collected, pooled and analyzed via reverse phase HPLC (Beckman ultrasphere C-18 ion pair column; 0–100% acetonitrile/water/1% acetic acid gradient over 45 min). Absorbances were monitored at 280 and 410 nm. CH999/pRM21 extracts and purified DMAC were used as controls to identify strains producing 16 carbon polyketides, whereas CH999/pRM37 extracts and purified RM20b were used to identify 20 carbon producers. Non-producers were identified by comparison with CH999 lacking any plasmid.

Homology modeling

Insight II, Release 95.0 (Biosym/MSI, San Diego, CA) was used to analyze the KAS II structure, and to compare the results from the hybrid KS–CLF heterodimers described here to the KAS II homodimer.

References

1. Hopwood, D. A.; Sherman, D. H. *Annu. Rev. Genet.* **1987**, *24*, 37–66.
2. Katz, L.; Donadio, S. *Ann. Rev. Microbiol.* **1993**, *47*, 875–912.
3. O'Hagan, D. *The Polyketide Metabolites*; Ellis Horwood: Chichester, UK, 1991.
4. Hutchinson, C. R.; Fujii, I. *Ann. Rev. Microbiol.* **1995**, *49*, 201–238.
5. Richardson, M. T.; Khosla, C. In *Biosynthesis and Biodegradation of Polyketides and Other Secondary Metabolites Including Fatty Acids and their Derivatives*; Sankawa, U., Ed.; Elsevier: Amsterdam, 1999; Vol. 1, pp 473–494.
6. (a) Hitchman, T. S.; Crosby, J.: Byrom, K. J.; Cox, R. J.; Simpson, T. *J. Chem. Biol.* **1998**, *5*, 35–47. (b) Matharu, A. L.; Cox, R.; Crosby, J.; Byrom, K. J.; Simpson, T. *J. Chem. Biol.* **1998**, *5*, 699–711
7. Revill, W. P.; Bibb, M. J.; Hopwood, D. A. *J. Bacteriol.* **1995**, *177*, 3946–3952.
8. (a) Carreras, C. W.; Khosla, C. *Biochemistry* **1998**,*37*, 2084–2088. (b) Dreier, J.; Shah, A. N.; Khosla, C. *J. Biol. Chem.* **1999**, *274*, 25108–25112.
9. Bao, W.; Wendt-Pienkowski, E.; Hutchinson, C. R. *Biochemistry* **1998**, *37*, 8132–8138.
10. McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *Science* **1993**, *262*, 1546–1550.
11. McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1993**, *115*, 11671–11675.
12. Shen, B.; Summers, R. G.; Wendt-Pienkowski, E.; Hutchinson, C. R. *J. Am. Chem. Soc.* **1995**, *117*, 6811–6821.
13. McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11542–11546.
14. McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1994**, *116*, 10855–10859.
15. McDaniel, R.; Hutchinson, C. R.; Khosla, C. *J. Am. Chem. Soc.* **1995**, *117*, 6805–6810.
16. McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *Nature* **1995**, *375*, 549–554.
17. Fu, H.; McDaniel, R.; Hopwood, D. A.; Khosla, C. *Biochemistry* **1994**, *33*, 9321–9326.
18. Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1994**, *116*, 4166–4170.
19. Fu, H.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1994**, *116*, 6443–6444.
20. Fu, H.; Hopwood, D. A.; Khosla, C. *Chem. Biol.* **1994**, *1*, 205–210.
21. Alvarez, M. A.; Fu, H.; Khosla, C.; Hopwood, D. A.; Bailey, J. E. *Nat. Biotechnol.* **1996**, *14*, 335–338.
22. Kramer, P. J.; Zawada, R. J. X.; McDaniel, R.; Hutchinson, C. R.; Hopwood, D. A.; Khosla, C. *J. Am. Chem. Soc.* **1997**, *119*, 635–639.
23. Yu, T. W.; Shen, Y.; McDaniel, R.; Floss, H. G.; Khosla, C.; Hopwood, D. A.; Moore, B. S. *J. Am. Chem. Soc.* **1998**, *120*, 7749–7759.
24. Khosla, C.; McDaniel, R.; Ebert-Khosla, S.; Torres, R.; Sherman, D. H.; Bibb, M. J.; Hopwood, D. A. *J. Bacteriol.* **1993**, *175*, 2197–2204.
25. Kramer, P. J. Ph.D. thesis, Stanford University, 1996.
26. Zawada, R. J. X.; Khosla, C. *J. Biol. Chem.* **1997**, *272*, 16184–16188.
27. Huang, W.; Jia, J.; Edwards, P.; Dehesh, K.; Schneider, G.; Lindqvist, Y. *EMBO J.* **1998**, *17*, 1183–1191.
28. Moche, M.; Schneider, G.; Edwards, P.; Dehesh, K.; Lindqvist, Y. *J. Biol. Chem.* **1999**, *274*, 6031–6037.
29. Maniatis, T.; Fritsch, E. F.; Sambrook, J. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory: NY, USA, 1982.
30. Hopwood, D. A.; Bibb, M. J.; Chater, K. F.; Kieser, T.; Bruton, C. J.; Kieser, H. M.; Lydiate, D. J.; Smith, C. P.; Ward, J. M.; Schrempf, H. *Genetic Anipulation of Streptomyces: A Laboratory Manual*; The John Innes Foundation: Norwich, UK, 1985.
31. Hedstrom, L.; Szilagyi, L.; Rutter, W. J. *Science* **1992**, *255*, 1249–1253.
32. Hung, S. H.; Hedstrom, L. *Protein Eng.* **1998**, *11*, 669–673.
33. Kushiro, T.; Shibuya, M.; Ebizuka, Y. *J. Am. Chem. Soc.* **1999**, *121*, 1208–1216.
34. Spaink, H. P.; Weinman, J.; Djordjevic, M. A.; Okker, R. J.; Lugtenberg, B. J. *EMBO J.* **1989**, *8*, 2811–2818.
35. Ritsema, T.; Gehring, A. M.; Stuitje, A. R.; van der Drift, K. M.; Dandal, I.; Lambalot, R. H.; Walsh, C. T.; Thomas-Oates, J. E.; Lugtenberg, B. J.; Spaink, H. P. *Mol. Gen. Genet.* **1998**, *257*, 641–648.