J Mol Evol (2003) 56:446–457
DOI: 10.1007/s00239-002-2415-0

JOURNAL OF MOLECULAR EVOLUTION



# Evolutionary Affiliations Within the Superfamily of Ketosynthases Reflect Complex Pathway Associations

**Michelle C. Moffitt, Brett A. Neilan**

School of Microbiology and Immunology, The University of New South Wales, Sydney NSW, 2052, Australia

Received: 8 July 2002 / Accepted: 4 November 2002

**Abstract.** Type I polyketide synthases are known to produce a wide range of medically and industrially important polyketides. The ketosynthase (KS) domain is required for the condensation of an extender unit onto the growing polyketide chain during polyketide biosynthesis. KSs represent a superfamily of complex biosynthetic pathway-associated enzymes found in prokaryotes, fungi, and plants. Although themselves functionally conserved, KSs are involved in the production of a structurally diverse range of metabolites. Degenerate oligonucleotide primers, designed for the amplification of KS domains, amplified KS domains from a range of organisms including cyanobacterial and dinoflagellates. KS domains detected in dinoflagellate cultures appear to have been amplified from the less than 3-μm filtrate of the nonaxenic culture. Phylogenetic analysis of sequences obtained during this study enabled the specific identification of KS domains of hybrid or mixed polyketide synthase/peptide synthetase complexes, required for the condensation of an extender unit onto an amino acid starter unit. The primer sets described in this study were also used for the detection of novel KS domains directly from environmental samples. The ability to predict function based on primary molecular structure will be critical for future discovery and rational engineering of polyketides.

*Correspondence to:* B.A. Neilan; *email:* b.neilan@unsw.edu.au

**Key words:** Polyketide synthase — Nonribosomal peptide synthetase — Ketosynthase — Cyanobacteria — Polyketide — Nonribosomal peptide

## Introduction

Polyketide synthases (PKSs) are multienzyme complexes which have been implicated in the biosynthesis of polyketide structures including pharmaceutically and industrially important bioactive compounds and are being engineered for the combinatorial biosynthesis of "unnatural" natural polyketide products. Similar to fatty acid biosynthesis, polyketides are derived from a starter unit such as acetyl-coenzyme A (CoA) or related acyl-CoA's. Extension of the polyketide chain occurs via condensation of extender units, also acyl-CoA's, to the growing polyketide chain.

Biosynthesis of polyketides typically occurs via the action of acyltransferase (AT), ketosynthase (KS), and acyl carrier protein (ACP) domains. The AT domain is responsible for selecting the starter or extender unit, which is then covalently attached to a phosphopantetheine carrier of the adjacent ACP domain. Condensation of the extender unit to the growing polyketide chain occurs via the active-site cysteine of the KS domain. Further additions to the structural novelty of the molecule are the enzymatic functions of a ketoreductase, a dehydratase, or an enoyl reductase. A thioesterase domain is usually required for termination of biosynthesis and removal of the final product from the PKS complex, which

may occur via cyclization or transfer to an amino or aminoacyl acceptor.

PKSs have been divided into three main subclasses. Type I PKSs are large multifunctional enzyme complexes with a single enzyme complex responsible for the biosynthesis of the polyketide chain. Bacterial type I polyketide synthesis is nonreiterative, with a linear order of catalytic sites. The antibiotic erythromycin A and the immunosuppressant rapamycin are examples of polyketides which are biosynthesized by bacterial type I PKSs (Donadio et al. 1991; Schwecke et al. 1995). Fungal type I polyketide synthesis, which occurs via a single module containing all catalytic sites, may be partially reiterative. Recently, fungal type I PKSs have been classified into three subclasses, a nonreduced or "WA" type, including those responsible for the biosynthesis of pigments and aflatoxins; a partially reduced or "MSAS" type, related to the 6-methylsalicylic acid biosynthetic pathway; and a highly reduced type, which includes the anticholesterol polyketide lovastatin (Fig. 1) (Bingle et al. 1999; Nicholson et al. 2001). These PKS types are grouped according to the amount of ketoreductase activity required during biosynthesis. Sequence analyses of KS domains from these three types have shown them to be phylogenetically distinct (Nicholson et al. 2001).

Type II PKSs are made up of several separate monofunctional enzymes and are generally responsible for the formation of cyclic aromatic compounds. Type III PKS systems are the simplest PKSs and are typically found in plants but have also recently been found in bacteria (Funa et al. 1999). These type III PKSs are homodimeric iterative and lack an ACP domain, instead acting on CoA esters of simple carboxylic acids. The biosynthesis of chalcone and 2-pyrone in plants occurs via type III PKSs (Kreuzaler et al. 1979).

Mixed or hybrid multienzyme complexes containing both PKS and nonribosomal peptide synthetase (NRPS) modules have recently been identified. The rapamycin biosynthetic gene cluster was the first gene cluster identified as having both multienzyme complexes within the one system (Schwecke et al. 1995). Mixed multienzyme complexes, with both NRPS and PKS enzymatic abilities, are required for the biosynthesis of bleomycin from *Streptomyces verticillus* and the siderophores yersiniabactin and mycobactin, isolated from *Yersinia pestis* and *Mycobacterium tuberculosis*, respectively (Gehring et al. 1998; Quadri et al. 1998; Shen et al. 1999). The biosynthesis of the polyketide immunosuppressants rapamycin, FK506, and FK520 also requires the activity of an NRPS module, which is responsible for the addition of an amino acid moiety to the end of the polyketide chain (Motamedi et al. 1997; Schwecke et al. 1995; Wu et al. 2000). Recently, a small group of multienzyme complexes has been identified in which a PKS and an NRPS module exist within the same open reading frame (ORF). These hybrid nonribosomal systems occur in the mycosubtilin, myxothiazol, and microcystin synthetases (Duitman. et al. 1999; Silakowski et al. 1999; Tillett et al. 2000).

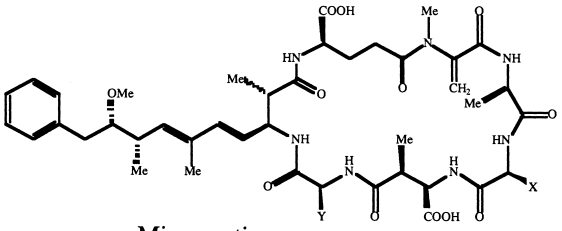
Previously, ketosynthase-specific PCR has been used to screen environmental soil samples and laboratory *Streptomyces* strains for type II PKS genes (Metsa-Ketela et al. 1999; Seow et al. 1997). A similar study analyzed the distribution of iterative type I PKS genes within several fungal species (Nicholson et al. 2001). Widescreen analysis of bacterial laboratory strains and samples from diverse environments for modular type I PKS biosynthetic genes has not been previously published. Type I PKSs are modular and it has recently been shown that replacement of a domain or module within the multienzyme complex can lead to a functional complex which will synthesize a novel polyketide (Gokhale et al. 1999; McDaniel et al. 1999). Therefore recognition of the availability and diversity of KS domains in the environment is important for future drug discovery and combinatorial biosynthesis efforts. During this study, degenerate oligonucleotide primers were designed for the amplification of KS domains from type I PKS modules. In this way, the diversity of modular type I PKS genes was studied in a range of bacterial and dinoflagellate laboratory strains and environmental samples. Using the sequence data obtained, phylogenetic analyses were performed to predict function. In addition, the genes encoding the KS domains involved in the biosynthesis of glycolipids in differentiated cyanobacterial heterocysts were amplified and sequenced. Phylogenetic analysis indicates that this group of PKSs is distinct from other PKSs. These evolutionary analyses are important for achieving a better understanding of the diversity of polyketide biosynthetic systems and the bioactive polyketides produced. These novel PKSs may then be harnessed in combinatorial biosynthesis research to design new drugs.

## Materials and Methods

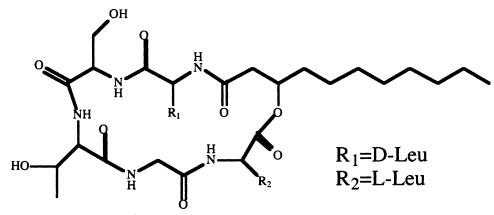
### *Bacterial Strains and Environmental Samples*

Environmental samples were isolated from the Colo River sediment, Australia, and a bird bath, Sydney, Australia. A smooth microbial mat and a pustular microbial mat were sampled from Shark Bay, Australia. Epilithic (granite) Antarctic cyanobacteria were supplied by John Bowman (University of Tasmania, Australia). *Prorocentrum* strains were supplied by Susan Blackburn (Marine Laboratories, CSIRO, Australia). Contaminating microorganisms were separated from dinoflagellate cells by filtering cultures through a 3.0-μm SS Type MF-Millipore filter (Millipore, Bedford, MA). Other bacterial strains used in this study were obtained from the School of Microbiology and Immunology Culture Collection, University of New South Wales, Australia.

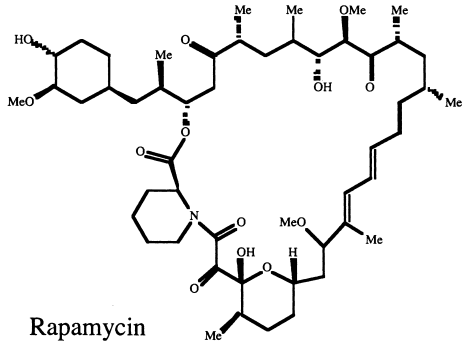
A) Bacterial polyketides with mixed biosynthetic pathways



Microcystin

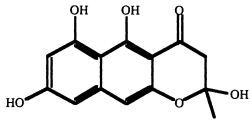



Serrawettin W2

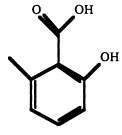


Rapamycin

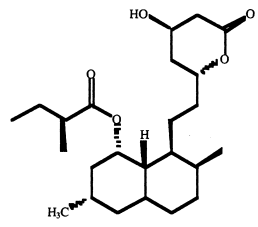
B) Fungal type I polyketides



YwA15

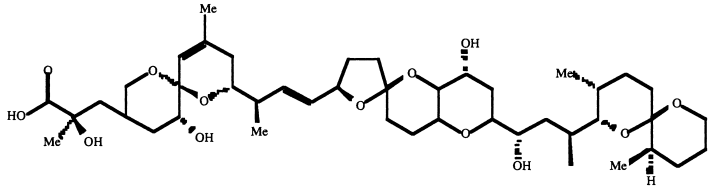


6MSA



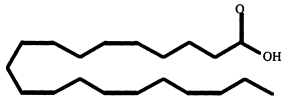
Lovastatin

C) Algal polyketide

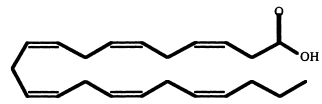


Dinophysistoxin-1

D) Fatty acids



Eicosapentaenoic acid (EPA)



Docosahexaenoic acid (DHA)

**Fig. 1.** Chemical structures of polyketides biosynthesized by PKS systems analyzed in this study. **A** Microcystin, serawettin W2, and rapamycin produced by mixed PKS and NRPS systems. **B** Type I fungal polyketides YwA15, a nonreduced polyketide, 6-MSA, a partially reduced polyketide, and lovastatin, a highly reduced polyketide. **C** Dinophysistoxin, a toxic algal polyketide of unknown biosynthetic pathway. **D** $\omega$-3-Fatty acid eicosapentaenoic acid (EPA) produced by *Shewanella* sp. and docosahexaenoic acid (DHA) isolated from *Moritella marina*.

### *DNA Extraction and Amplification*

Total genomic DNA was extracted from cultures and environmental samples using an SDS/lysozyme-based method as described previously by Neilan et al. (1995) Amplification of KS regions was performed using the degenerate oligonucleotide primers DKF (GTG CCG GTN CCR TGN GYY TC) and DKR (GCG ATG GAY CCN CAR MG). Degenerate oligonucleotide primers HGLR (GGT GAR GGT ATH GGN ATG) and HGLR (TTG GTT CCG CCR AAN CCR AA) were used for the amplification of heterocyst glycolipid biosynthetic KS domains. Thermal cycling was performed in a PCR Sprint Temperature Cycling System Machine (Hybaid Ltd., Middlesex, UK) or a GeneAmp PCR System 2400 Thermocycler (Perkin Elmer, Norwalk, CT). The initial denaturation step at 94°C for 2 minutes was followed by 30 cycles of DNA denaturation at 94°C for 5 s, primer annealing at the corresponding annealing temperature for 10 s, and DNA strand extension at 72°C for 1 min, and a final extension step at 72°C for 7 min. A primer annealing temperature of 55°C was used for the amplification of DNA using the DKF/DKR primer pair and 50°C for the HGLF/HGLR primer pair. PCR products were cloned directly using the TOPO-TA cloning system (Invitrogen).

### *Sequencing, Phylogenetic Analysis and Functional Prediction*

Automated sequencing was performed using the PRISM Big Dye cycle sequencing system and a Model 373 sequencer (Applied Biosystems Inc., Foster City, CA). Sequence data were analyzed using the Applied Biosystems Auto-Assembler computer program, and percentage similarity and identity to other translated sequences determined using BLAST in conjunction with the National Center for Biotechnology Information.

PKS protein sequences used in this study were aligned using the program Pileup from GCG and the multiple-sequence alignment tool from Clustal W (Thompson et al.1994) and the PHYLIP package (version 3.57c) (Felsenstein, 1989). Sequence manipulation and phylogeny programs were accessed through the Australian National Genome Information Service (Sydney). The nucleotide sequences reported in this study have been deposited in the NCBI/GenBank necleotide sequence database under accession numbers AY210783 to AY210796 and AY211087 to AY211090.

## Results and Discussion

### *Distribution of KS Genes*

Detection of modular type I PKS genes was performed using degenerate oligonucleotide primers. These primers were designed from the alignment of KS domains of known PKS gene clusters from bacteria including cyanobacteria and mycobacteria. The primers were designed at conserved regions, the downstream primer designed to encompass the second active-site histidine His340 (Huang et al. 1998) and the upstream primer directed to the conserved region PQQR, approximately 80 amino acids (aa) upstream of the active-site cysteine. The priming sites were distinct from other KS primers designed to screen for type I fungal polyketide domains and type II bacterial polyketide domains (Metsa-Ketela et al. 1999; Nicholson et al. 2001), however, a recent study screening the myxobacterium *Stigmatella aurantiaca* used primers designed within the same conserved regions (Silakowski et al. 2001). This degenerate PCR was designed to detect KS domains from a range of organisms. KS domains approximately 700 bp in size (Fig. 2) were detected by degenerate PCR in members of the cyanobacteria, actinobacteria, and proteobacteria (Table 1). In some cases, the degenerate primers amplified more than one KS domain from the modular PKS gene clusters for which the sequence was not determined (Table 1).

Cyanobacteria are known to produce a wide range of bioactive secondary metabolites, however, to date, the presence of PKS genes in cyanobacteria has been published only in the microcystin-producing cyanobacterium *Microcystis aeruginosa* and the nodularin-producing *Nodularia spumigena* (Moffitt and Neilan 2001; Tillett et al. 2000). The gene cluster responsible for nostopeptolide biosynthesis also appears to contain both PKS and NRPS modules (accession No. AF204805). A PKS has also been implicated in the biosynthesis of a glycolipid which occurs in the cell wall of differentiated heterocysts during nitrogen starvation (Campbell et al. 1997), however, this PKS is unrelated to the type I PKS genes found in *Microcystis* and *Nodularia*. This study found, by KS domain PCR, that PKS gene clusters were also found in the cyanobacteria *Cylindrospermopsis raciborskii* AWT205 and *Umezakia natans* sp., both of which produce the toxin cylindrospermopsin. The structure of cylindrospermopsin indicated that a type I PKS is involved in biosynthesis (Burgoyne et al. 2000). The sequence of the KS region was also obtained from *Anabaena flos-aquae* strain NRC44-1. In addition, KS domains were amplified from *A. flos-aquae* strain NRC525-17, *Aphanothece* sp., *Nostoc* sp. AWT203, the microviridin A- and aeruginosin 102-A-producing strain *Microcystis viridis* NIES102 (Ishituka et al. 1990; Matsuda et al. 1996), microviridin- and aeruginosin 298-A-producing *M. aeruginosa* NIES298 (Murakami et al. 1994; Okino et al. 1995), and *Oscillatoria agardhii* CCAP1459/21, which produces a novel microcystin isoform (Sano and Kaya 1995).

The degenerate PCR described above was also used to screen a wide range of laboratory bacterial strains and environmental samples. Type I KS domains were amplified and sequenced from *Agrobacterium tumefacians, Brevibacterium flavum, Serratia liquefacians*, and *Rhodococcus* sp. 33. KS domains were also detected but not sequenced from *Proteus mirabilis* and *H.pylori*, *P. mirabilis* is known to produce a peptide with a polyketide side chain, which is involved in swarming (Gaisser and Hughes 1997). The detection of PKS genes in *Helicobacter pylori* has not been reported previously and the function of the putative KS domain amplified by PCR is not known. Analysis of the complete genome sequences

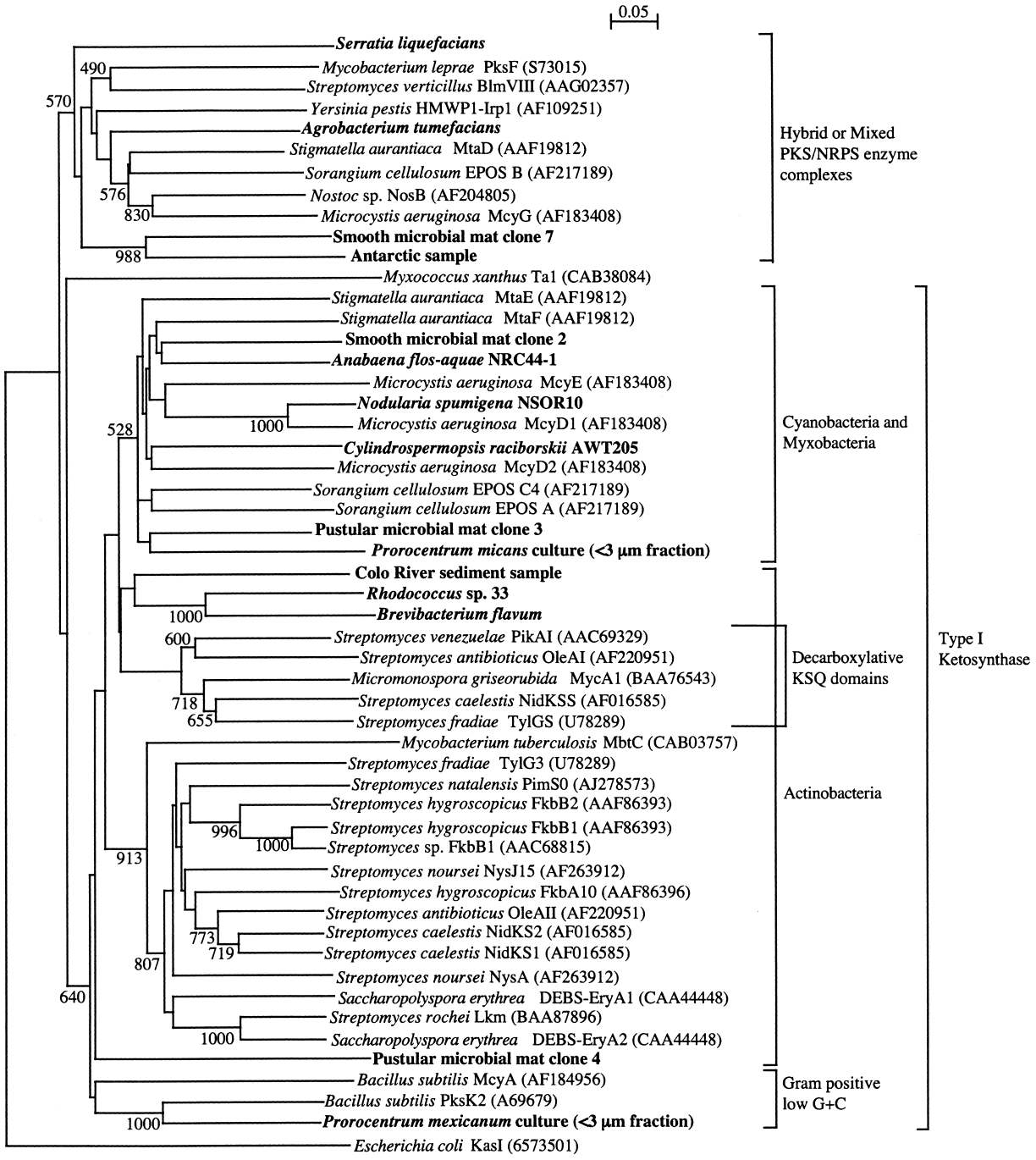



**Fig. 2.** Phylogenetic analysis of type I ketosynthase domains. Sequences obtained during this study using degenerate primers are given in *boldface*. Other sequences were obtained from GenBank; the accession numbers are given in *parentheses*. Sequences were aligned using the Pileup and Clustal X programs. Divergence between amino acid sequences was calculated using a PAM-Dayhoff matrix and tree constructed using the UPGMA method supplied by the PHYLIP package (version 3.57c) (Felsenstein 1989). The scale bar represents 5 substitutions per 100 amino acids. Significant bootstrap values (Felsenstein 1985) calculated from 1000 bootstrap trees are indicated at the nodes.

by homology searches was unable to detect the presence of PKS genes. These results indicate a much broader distribution of KS domains in both Gram-positive and Gram-negative bacteria than previously shown.

Dinoflagellates are known to produce a wide range of complex polyketides including toxins, however, PKS genes have not been reported in these organisms (Rein and Borrone 1999). A KS region was amplified from nonaxenic cultures of the dinoflagellates *Prorocentrum mexicanum* CS-292/2 and *Prorocentrum micans* CS-293, both of which produce the diahoretic shellfish poison, dinophysistoxin (Fig. 1). These nonaxenic cultures were filtered by size exclu-

**Table 1.** Bacterial, algal, and environmental strains analyzed in this study by degenerate ketosynthase PCR (DPK-PCR) and degenerate heterocyst glycolipid PCR (HGL-PCR)

| Strain | Origin | DPK-PCR | HGL-PCR |
|---|---|---|---|
| *Agrobacterium tumefacians* SU585 | Australia | +;S | ND |
| *Anabaena circinalis* AWQC131C | Australia | – | + |
| *Anabaena circinalis* AWQC148C | Australia | – | + |
| *Anabaena cylindrica* NIES19 | England | – | +;S |
| *Anabaena flos-aquae* NRC44-1 | Canada | +;S | + |
| *Anabaena flos-aquae* NRC525-17 | Canada | + | + |
| *Anabaenopsis circularis* NIES21 | Japan | – | ND |
| *Aphanocapsa* sp. | | – | ND |
| *Aphanothece* sp. | | + | ND |
| *Aphanizomenon flos-aquae* NIES81 | Japan | – | + |
| *Bacillus stearothermophilus* 7914 | Australia | – | ND |
| *Brevibacterium flavum* BF4 | Australia | +;S | ND |
| *Corynebacterium michisanense* DAR25721 | Australia | – | ND |
| *Cylindrospermopsis raciborskii* AWT205 | Australia | +;S | + |
| *Klebsiella pneumoniae* A3 | Scotland | – | ND |
| *Lactobacillus acidophilus* ATCC4356 | Human | – | ND |
| *Lyngbya* sp. AWT211 | Australia | – | ND |
| *Micrococcus luteus* ATCC10240 | Australia | – | ND |
| *Microcystis aeruginosa* NIES298 | Japan | + | ND |
| *Microcystis aeruginosa* PCC7005 | United States | – | – |
| *Microcystis viridis* NIES102 | Japan | + | ND |
| *Nodularia spumigena* NSOR10 | Australia | +;S | + |
| *Nostoc* sp. AWT203 | Australia | + | + ;S |
| *Oscillatoria agardhii* CCAP1459/22 | Norway | + | ND |
| *Prorocentrum mexicanum* CS-292/2 | Australia | +;S | ND |
| *Prorocentrum micans* CS-293 | Australia | +;S | ND |
| *Proteus mirabilis* J | Rat | + | ND |
| *Pseudoalteromonas tunicata* D2T | Australia | – | ND |
| *Pseudoanabaena* sp. | Australia | – | ND |
| *Pseudomonas putida* ATCC17453 | Australia | – | ND |
| *Rhizobium meliloti* SU101 | Root nodule | – | ND |
| *Rhodococcus* sp. 33 | Australia | +;S | ND |
| *Serratia liquefacians* MG1 | Australia | +;S | ND |
| *Sphingomonas alaskensis* RB2256 | United States | – | ND |
| *Staphylococcus aureus* ATCC12598 | Australia | – | ND |
| *Streptococcus pyogenes* UNSW026701 | Australia | – | ND |
| *Synechocystis* sp. PCC6803 | United States | – | – |
| *Umezakia natans* | Australia | + | – |
| *Vibrio angustum* S14 | Australia | – | ND |
| Bird bath sample | Australia | + | – |
| Colo River sediment sample | Australia | +;S | – |
| Epilithic (granite) cyanobacteria | Antarctic | +;S | – |
| Shark Bay smooth microbial mat | Australia | +;S | +;S |
| Shark Bay pustular microbial mat | Australia | +;S | +;S |

Note. + indicates a positive PCR result. – indicates a negative PCR result. S indicates that sequence data were obtained for that strain. ND indicates that PCR was not performed on that strain.

sion to remove all algal cells and to analyzse for the presence of contaminating microorganisms. Analysis of both the filtered algal cells and the < 3-μm filtrate by the degenerate PCR indicated that the KS domain was amplified from other microorganisms coexisting with the *Prorocentrum* cultures. Based on a negative result for the universal amplification of the eukaryotic 18S rRNA gene from the filtrate (data not shown), the KS domain was determined to be of bacterial or archaeal origin. However, a survey of archaeal genomes has not revealed the presence of KS domains in their genomes (Neil Saunders, personal communication). 16S rDNA PCR on the filtrate indicated that more than one species of bacteria may be present within the nonaxenic dinoflagellate culture, and hence the species containing the amplified KS domain has not been determined.

### *KS Domains Involved in Amino Acid Condensation Are Structurally Distinct*

Type I KS domain protein sequences appeared to cluster phylogenetically into two functional groups. While the first group represents KS domains which use acyl-CoA's as their starter or extender unit, a

distinct cluster of the type I KS domains is comprised of those sequences isolated from hybrid or mixed PKS/NRPS systems which condense amino acids onto a polyketide extender unit (Fig. 2).

The phylogenetic relationships between the amino acid condensing KS domains indicates functional similarity and substrate specificity within the group. The KS domain of *mcyG*, which contains both NRPS and PKS modules within the same ORF, is thought to be responsible for the decarboxylative condensation of a phenylacetate thioester to the malonyl-CoA extender unit (Tillett et al. 2000). Here the phenylacetate is activated by an NRPS module, which requires ATP. Similarly, during epothilone synthesis, the KS domain of EPOS C is thought to catalyze the addition of the 2-methyl-4-carbonyl thiazole from the phosphopantetheinyl carrier protein (PCP) domain of the preceding NRPS module to the extender unit, methylmalonyl-CoA (Molnar et al. 2000). The related KS domains isolated from *Y. pestis* and *Streptomyces* have also been shown to possess a role in condensing amino acids to the polyketide extender (Gehring et al. 1998; Shen et al. 1999). Complete analysis of the nostopeptolide gene cluster indicates that the KS domain of NosB is also responsible for the condensation of a polyketide extender unit onto an activated amino acid moiety (accession No. AF204805). Recent studies have also shown that a phylogenetic trend is shown throughout KS domains of this type (Du et al. 2001).

KS domain sequences isolated in this study from *S. liquefacians* and *A. tumefacians* were also members of this functionally distinct group of KS domains, suggesting that they may be a part of a mixed or hybrid PKS/NRPS multienzyme system (Fig. 2). Analysis of the genome of *A. tumefacians* indicates the presence of a gene cluster containing both PKS and NRPS modules, suggesting that the sequence is a part of this functional group (Goodner et al. 2001; Wood et al. 2001). *S. liquefacians* contains an NRPS system which has been shown to be involved in the biosynthesis of a swarming agent, serrawettin W2 (Fig. 1) (Lindum et al. 1998). The chemical structure prediction of serrawettin W2 suggests that it may be biosynthesized via a mixed PKS/NRPS biosynthetic pathway, suggesting that the KS domain isolated from *Serratia* in this study belongs within this group.

Samples isolated from the Antarctic and the smooth microbial mat also contained sequences related to the KS domains of hybrid or mixed PKS/NRPS biosynthetic systems. This suggests that these mixed biosynthetic pathways are ubiquitous to a range of environments and bacterial lineages.

The KS domains within the FK506, FK520, and rapamycin synthases, which also contain both NRPS and PKS modules, did not show significant amino acid sequence similarities to the distinct clade of KS domains from mixed and hybrid PKS/NRPS biosynthetic systems. The condensation domain of the NRPS modules required for the biosynthesis of these immunosuppressants is responsible for the condensation of the amino acid moiety to the end of the polyketide chain (Motamedi et al. 1997; Schwecke et al. 1995; Wu et al. 2000).

Structurally distinct KS domain, therefore, appear to occur in coupled PKS/NRPS systems only when the KS domain is responsible for the condensation of a polyketide extender unit onto an amino acid moiety, which often acts as a starter unit for the polyketide chain.

### *The Phylogeny of KS Domains is Related to the Organism of Origin*

Of the type I bacterial ketosynthases, which condense acyl-CoA subunits, subclusters represented the bacterial family from which the polyketide synthase gene cluster was isolated (Fig. 2). Cyanobacterial KS domain sequences clustered with KS domains of microcystin synthetase and the *Nodularia* KS domain. Interestingly, myxobacterial KS domains appeared to be phylogenetically similar to cyanobacteria, indicating an evolutionary relationship between the enzymes of these two bacterial families. Clone 2 from the Shark Bay smooth microbial mat was also closely related to this group, suggesting that the KS domain may have been amplified from a cyanobacterium or myxobacterium associated with the microbial mat. Shark Bay pustular microbial mat clone number 3, the *Prorocentrum* contaminant, and the stromatolite 2 clone were also associated with this group, suggesting that this KS domain is of a similar origin. Of the actinobacteria, *Streptomyces* KS domains analyzed from gene databases formed a single phylogenetic cluster. Coryneform bacterial sequences formed a small cluster within the actinobacteria group, as did those from mycobacteria.

A small cluster of sequences within the actinobacteria group are found to have a glutamine residue in place of the active-site cysteine of a type I KS domain (KSQ). These KSQ domains act like a chain initiation factor found in aromatic PKSs (KS β) (Bisang et al. 1999). The glutamine in the active site provides a decarboxylative activity not seen in other type I KS domains. Decarboxylative activity has been measured in the KSQ domains of niddamycin synthase (nidKSS) and tylosin synthase (Tyll) (Bisang et al. 1999). The functional difference between the KS domains and the KSQ domain is apparent in this phylogenetic analysis (Fig. 2), with the KSQ domains of niddamycin and tylosin synthases in a separate lineage from other KS domains in the same biosynthetic cluster. The KSQ domains isolated from mycinamicin synthase and oleandomycin synthase clustered phylogenetically with *nidKSS* and *tylI*,

however, the function of these domains is yet to be biochemically determined (accession No. BAA76543 [Shah et al. 2000]). Active-site tyrosine (KSY) or serine (KSS) KS domains have been described in the first PKS modules of epothilone synthase and pimaricin and nystatin synthases, respectively (Aparicio et al. 2000; Brautaset et al. 2000; Molnar et al. 2000). Phylogenetic analysis did not, however, indicate any structural similarities to the KSQ domains. The results of a study by Witkowski et al. (1999) indicate that when Cys161 is replaced by Ser at the active site, the β-KAS becomes a weak decarboxylase which retains some condensation activity, however, the functional role of the KSS and KSY domains of epothilone synthase and pimaricin and nystatin synthases remains to be biochemically determined.

The KS domain is the most highly conserved domain of the PKS. In the present study, the active-site sequence motif was analyzed in relation to the functionality of the KS domain (Fig. 3). Generally, the conserved sequence motif DTACSSSLVA was present in all KS domains analyzed. Of the decarboxylative KSQ domains, the active-site cysteine is replaced by a glutamine and the preceding alanine is less highly conserved, replaced by a glycine in the KS domains of NidSS, OleA1, and PikAI. In addition, a glutamine, 12 aa C-terminal to the active site, is conserved throughout the KSQ domains (Fig. 3). This glutamic acid residue may also be involved in the decarboxylative activity of the KSQ domains. The KSY domain of EpoA and the KSS domains of PimSO and NysA do not contain this conserved glutamic acid moiety, further supporting that these domains do not possess decarboxylative activity. Analysis of the active-site sequence motif of the KS domains that use an amino acid moiety as a starter unit found that aspartate, 3 aa N-terminal to the active-site cysteine, is replaced by glutamine. This functionally distinct group of KS domains also contains a conserved sequence motif, NDKD, 22 aa N-terminal to the active site. The role of these conserved motifs is also unknown but may be responsible for KS enzymatic function or substrate recognition.

*Analysis of Structurally Diverse KS Domains*

Phylogeny was performed on KS domains involved in the biosynthesis of heterocyst glycolipids in cyanobacteria to determine their structural relatedness to other well-described groups of KS domains. Heterocyst glycolipids are present in the cell wall of heterocysts of cyanobacteria. Degenerate oligonucleotide primers were designed within *hglE* (Campbell et al. 1997) to screen bacteria for a similar biosynthetic system (Table 1). This primer set identified *hglE* sequences in other *Nostoc* strains and also amplified a similar region from the Shark Bay microbial mats used in this study. 16S rDNA sequencing and microscopy did not, however, suggest the presence of heterocyst-forming cyanobacteria (Burns and Neilan, manuscript in preparation). This suggests the presence of a similar biosynthetic pathway in these environments. The *hglE* amino acid sequences also showed significant homology to regions of ω-3-fatty acid eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) biosynthetic gene clusters isolated from *Shewanella* sp. and *Moritella marina*, respectively (Tanaka et al. 1999; Yazawa 1996).

Polyketide and fatty acid KS domain sequences from bacteria, fungi, and plants were compared to the "HglE-like" sequences (Fig. 4). As expected, three distinct clades represent the type I, type II, and type III KS domains. The type I, II, and III clusters also separate into distinctly defined functional groups. Subclasses of the type I KS domains include the bacterial, bacterial "hybrid," fungal "partially reduced, MSAS type," and fungal "nonreduced, WA type," which do not contain a ketoreductase domain (Bingle et al. 1999; Nicholson et al. 2001). The type II clade (Fig. 4) is made up of three subclasses: bacterial; bacterial KS β domains, required for chain initiation and decarboxylase activity; and bacterial fatty acid KAS domains. Chain initiation factors of aromatic PKSs are required for chain initiation and decarboxylase activity. Chain initiation factors are evident by the replacement of the active-site cysteine with a glutamine (Bisang et al. 1999). The type III clade is grouped into those isolated from bacteria and those isolated from plants. It appears that the HglE-like sequences represent a distinct group of KS domains, not belonging to the previously described three main types, and may represent a type IV group of KS domains (Fig. 4). This distinct group is also defined by two subgroups. The first subgroup contains the HglE sequence from *N. punctiforme* ATCC29133 and the HglE sequences obtained during this study. This group also contained a region of ORF8 from the DHA gene cluster and a region of ORF5 from the EPA gene cluster (Tanaka. et al. 1999; Yazawa 1996). The second subgroup contained HglD, also required for heterocyst glycolipid biosynthesis, and a region of ORF7 and ORF10 from the EPA and DHA gene clusters. This group of sequences was also closely related to a 3-oxoacyl [acyl-carrier-protein] synthase domain of a multidomain KS isolated from *Streptomyces coelicolor*, the role of which is unknown (accession No. T37056). The active sites of the HglE-like domains were aligned with the active-site sequence of *E. coli* KASI. This group of KS domains contains a cysteine active site with the exception of HglE, which has a tryptophan residue at this position. An alanine residue 2 aa N-terminal to the active-site cysteine is conserved throughout the HglE-like group, along with a lysine residue, 8 aa C-

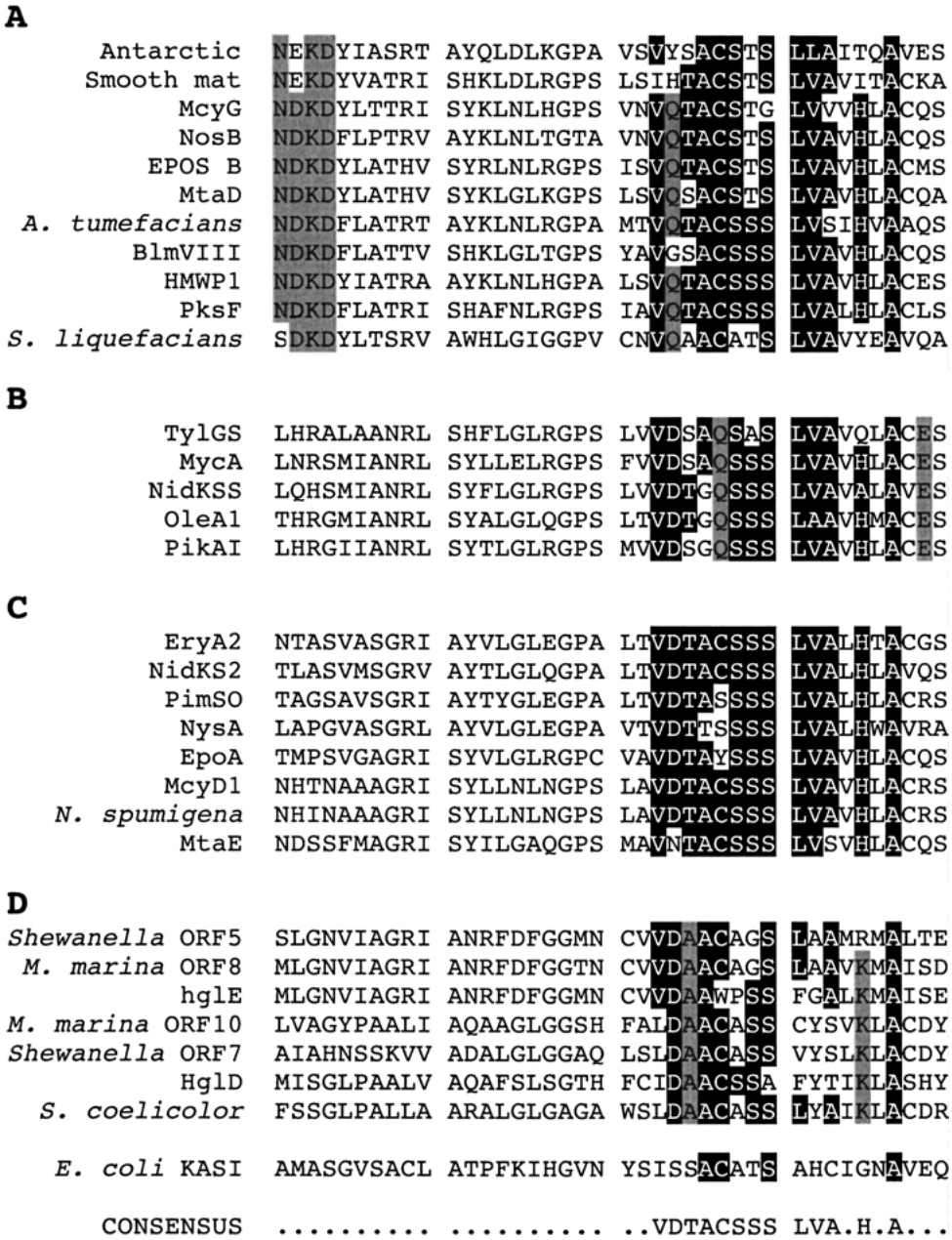



**Fig. 3.** Alignment of the active sites of type I ketosynthase (KS) domains with distinct functions. Conserved residues are highlighted in *black*. Residues conserved within each functional group are highlighted in *gray*. **A** Representatives of KS domains specific for amino acid starter units in mixed or hybrid systems. Sequences analyzed here include those from the Antarctic sample, the Shark Bay smooth microbial mat, *Serratia liquefacians*, and *Agrobacterium tumefacians* isolated during this study. Also analyzed are McyG from *Microcystis aeruginosa* (accession No. AF183408), NosB from *Nostoc* sp. (accession No. AF204805), EPOS B from *Sorangium cellulosum* (accession No. AF217189), MtaD from *Stigmatella aurantiaca* (accession No. AAF19812), BlmVIII from *Streptomyces verticillus* (accession No. AAG02357), HMWP1 from *Yersinia pestis* (accession No. AF109251), and PksF from *Mycobacterium leprae* (accession No. S73015). **B** Representatives of KSQ domains that have decarboxylative activity. Sequences analyzed here include TylGS from *Streptomyces fradiae* (accession No. U78289), MycA from *Micromonospora griseorubida* (accession No. BAA76543), NidKSS from *Streptomyces caelestis* (accession No. AF016585), OleA1 from *Streptomyces antibioticus* (accession No. AF220951), land PikA1 from *Streptomyces venezuelae* (accession No. AAC69329). **C** Representatives of other type I bacterial KS domains. Sequences aligned are EryA2 from *Saccharopolysyora erythrea* (accession No. CAA44448), NidKS2 from *Streptomyces caelestis* (accession No. AF016585), the KSS domains PimSO from *Streptomyces natalensis* (accession No. AJ278573) and NysA from *Streptomyces noursei* (accession No. AF263912), the KSY domain EpoA from *Sorangium cellulosum* (accession No. AF217189), McyD1 from *Microcystis aeruginosa* (accession No. AF183408), *Nodularia spumigena* NSOR10 isolated during this study, and MbtC from *Mycobacterium tuberculosis* (accession No. CAB03757). **D** Representatices of the HglE-like group of KS domains. Sequences aligned are ORF5 and ORF8 from *Shewanella* sp. (accession Nos. T30183 and T30185), ORF8 and ORF10 from *Moritella marine* (accession Nos. BAA89384 and BAA89384), HglE from *Nostoc punctiforme* (accession No. T37056), HgID from *Anabaena* sp. (accession No. AAA93154), and *Streptomyces coelicolor* (accession No. AAB82059). The Type I fatty acid sythase KS domain sequence from *Escherichia coli* (accession No. 6573501) is also given as a comparison and the consensus sequence is shown.

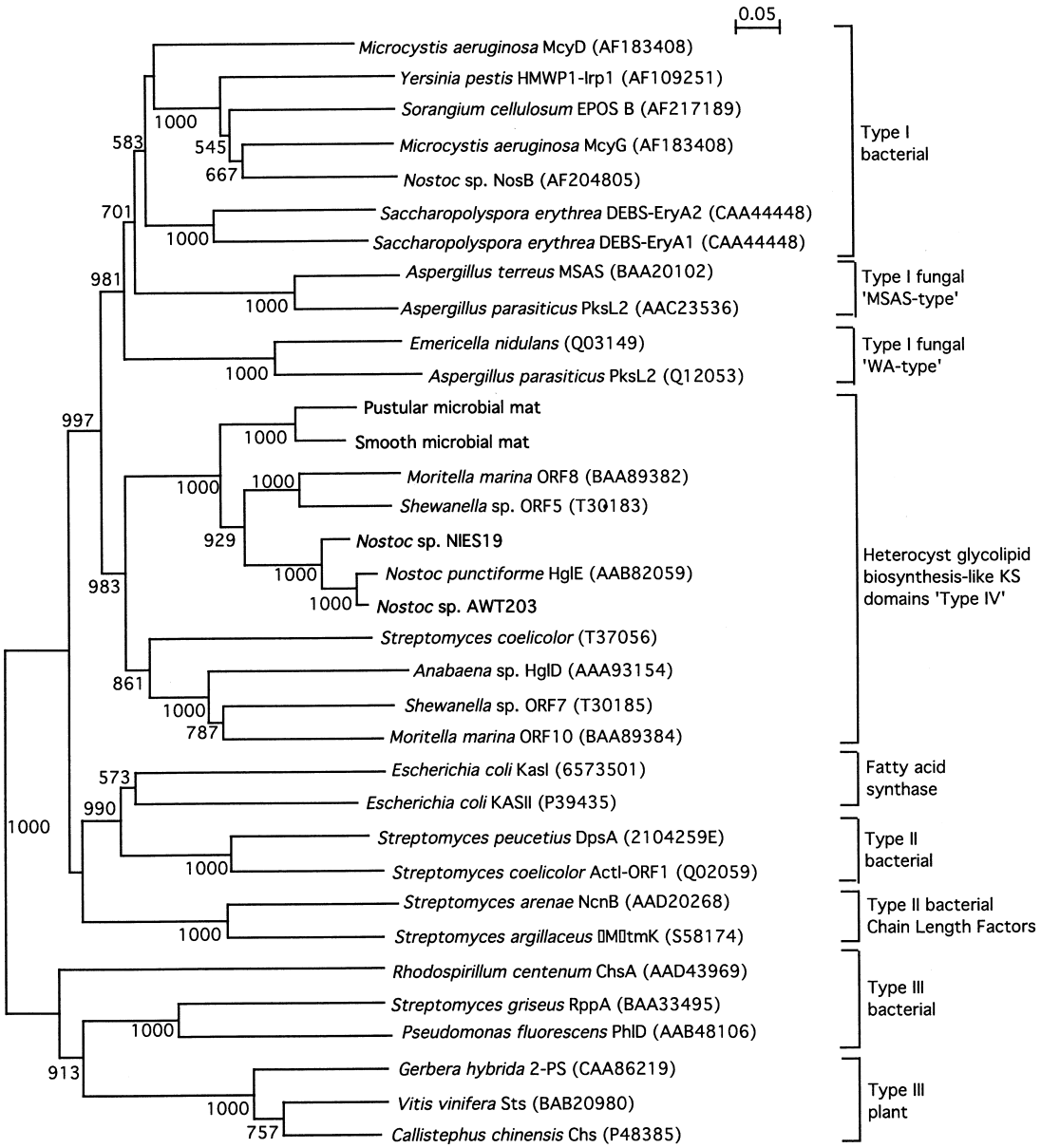



**Fig. 4.** Phylogenetic analysis of heterocyst glycolipid biosynthetic ketosynthase regions with respect to a diverse range of ketosynthase domains, including types I, II, and III. Sequences obtained during this study using degenerate primers are given in *boldface*. Other sequences were obtained from GenBank; the accession numbers are given in *parentheses*. Sequences were aligned using the Pileup and Clustal X programs. Divergence between amino acid sequences was calculated using a PAM-Dayhoff matrix and tree constructed using the UPGMA method supplied by the PHYLIP package (version 3.57c) (Felsenstein 1989). The scale bar represents 5 substitutions per 100 amino acids. Significant bootstrap values (Felsenstein 1985) calculated from 1000 bootstrap trees are indicated at the nodes.

terminal to the active-site cysteine (Fig. 3). It seems that, although little is known of the function of this HglE-like group of KS domains, this phylogenetically distinct clade may represent a new subclass of PKSs. Biochemical analysis of the role of these domains and the conserved residues is required to define further the type of KS domains to which they belong and their function.

Of particular significance for future studies, it has been shown that the degenerate primers designed during this study are suitable for the detection of type I KS domains of PKSs in both laboratory strains and the environment. This included the amplification of KS domains from a wide range of bacteria, including cyanobacteria and dinoflagellate cultures. While the biosynthesis of polyketide toxins from the dinoflagellates *Prorocentrum* sp. remains to be determined, KS domains were amplified from two species. Further analysis indicated that these domains were amplified from another organism in the culture.

Sequencing, analysis of the active site, and phylogenetic reconstruction gives us an insight into the

activity of the KS domain within the multienzyme complex. Decarboxylative KSQ domains represented a distinct phylogenetic cluster and type II KS β domains also represented a distinct cluster. Phylogeny and sequence analyses were unable to distinguish between the specificities for substrate of KS domains. Phylogenetic analysis was, however, able to distinguish KS domains that accepted an activated amino acid from an NRPS domain, as opposed to those systems where an amino acid is fused to the extended polyketide. Conserved regions were identified within this group of KS domains, which may be responsible for their specific activity. Such KS domains occurred in both hybrid and mixed PKS/NRPS systems. In addition, this study reports that KS domains required for the biosynthesis of $\omega$-3-fatty acids and glycolipids represent a distinct and as yet undefined group of KS domains. These novel domains were also detected in the environment.

This study reports that the use of degenerate primers accompanied by phylogenetic analysis is useful for screening genomes and environmental samples for the presence of bacterial type I KS domains and therefore the presence of novel polyketides. The indication that KS domains which link a polyketide extender unit with an amino acid are structurally distinct may be utilized for the engineering of novel polyketides onto an amino acid moiety or short peptide, further expanding the realm of unnatural product biosynthesis. The molecular predictions presented here based on phylogenetic analyses afford an efficient means of determining the components of rational biochemical engineering.

*Acknowledgments.* The Australian Research Council (ARC) financially supported this work. M.C.M. is jointly funded by the ARC and the CRC for Water Quality and Treatment. B.A.N. is a fellow of the ARC. We gratefully acknowledge Graham Garnsey and Amanda Bass for excellent technical assistance.

## References

Aparicio JF, Fouces R, Mendes MV, Olivera N, Martin JF (2000) A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimaricin in *Streptomyces natalensis*. Chem Biol 7:895–905

Bingle LE, Simpson TJ, Lazarus CM (1999) Ketosynthase domain probes identify two subclasses of fungal polyketide synthase genes. Fungal Genet Biol 26:209–223

Bisang C, Long PF, Cortes J, Westcott J, Crosby J, Matharu AL, Cox RJ, Simpson TJ, Staunton J, Leadlay PF (1999) A chain initiation factor common to both modular and aromatic polyketide synthases. Nature 401:502–505

Brautaset T, Sekurova ON, Sletta H, Ellingsen TE, Strom AR, Valla S, Zotchev SB (2000) Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: Analysis of the gene cluster and deduction of the biosynthetic pathway. Chem Biol 7:395–403

Burgoyne DL, Hemscheidt TK, Moore RE, Runnegar MT (2000) Biosynthesis of cylindrospermopsin. J Org Chem 65:152–156

Campbell EL, Cohen MF, Meeks JC (1997) A polyketide-synthase-like gene is involved in the synthesis of heterocyst glycolipids in *Nostoc punctiforme* strain ATCC 29133. Arch Microbiol 167:251–258

Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. Science 252:675–679

Du L, Sanchez C, Shen B (2001) Hybrid peptide-polyketide natural products: Biosynthesis and prospects toward engineering natural molecules. Metab Eng 3:78–95

Duitman EH, Hamoen LW, Rembold M, Venema G, Seitz H, Saenger W, Bernhard F, Reinhardt R, Schmidt M, Ullrich C, Stein T, Leenders F, Vater J (1999) The mycosubtilin synthetase of *Bacillus subtilis* ATCC 6633: A multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. Proc Natl Acad Sci USA 96:13294–13299

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:166–170

Felsenstein J (1989) PHYLIP. Phylogeny inference package. Cladistics 5:164–166

Funa N, Ohnishi Y, Fujii I, Shibuya M, Ebizuka Y, Horinouchi S (1999) A new pathway for polyketide synthesis in microorganisms. Nature 400:798–799

Gaisser S, Hughes C (1997) A locus coding for putative non-ribosomal peptide/polyketide synthase functions is mutated in a swarming-defective *Proteus mirabilis* strain. Mol Gen Genet 253:415–427

Gehring AM, DeMoll E, Fetherston JD, Mori I, Mayhew GF, Blattner FR, Walsh CT, Perry RD (1998) Iron acquisition in plague: Modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. Chem Biol 5:573–586

Gokhale RS, Tsuji SY, Cane DE, Khosla C (1999) Dissecting and exploiting intermodular communication in polyketide synthases. Science 284:482–485

Goodner B, Hinkle G, Gattung S, Miller N, Blanchard M, Qurollo B, Goldman B, Cao Y, Askenazi M, Halling C, Mullin L, Houmiel K, Gordon J, Vaudin M, Iartchouk O, Epp A, Liu F, Wollam C, Allinger M, Doughty D, Scott C, Lappas C, Markelz B, Flanagan C, Crowell C, Gurson J, Lomo C, Sear C, Strub G, Cielo C, Slater S (2001) Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science 294:2323–2328

Haydock SF, Aparicio JF, Molnar I, Schwecke T, Khaw LE, Konig A, Haydock SF, Aparicio JF, Molnar I, Schwecke T, Khaw LE, Konig A, Marsden AFA, Galloway IS, Staunton J, Leadlay PF (1995) Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthases. FEBS Lett 374:246–248

Huang W, Jia J, Edwards P, Dehesh K, Schneider G, Lindqvist Y (1998) Crystal structure of beta-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. EMBO J 17:1183–1191

Ishituka M, Kusumi T, Kakisawa H, Kaya K, Watanabe M (1990) Microviridin: A novel tricyclic depsipeptide from the toxic cyanobacterium *Microcystis viridis*. J Am Chem Soc 112:8180–8182

Kreuzaler F, Ragg H, Heller W, Tesch R, Witt I, Hammer D, Hahlbrock K (1979) Flavanone synthase from *Petroselinum hortense*. Molecular weight, subunit composition, size of messenger RNA, and absence of pantetheinyl residue. Eur J Biochem 99:89–96

Lindum PW, Anthoni U, Christophersen C, Eberl L, Molin S, Givskov M (1998) N-Acyl-L-homoserine lactone autoinducers control production of an extracellular lipopeptide biosurfactant

required for swarming motility of *Serratia liquefaciens* MG1. J Bacteriol 180:6384–6388

Matsuda H, Okino T, Murakami M, Yamaguchi K (1996) Aeruginosins 102-A and B, new thrombin inhibitors from the cyanobacterium *Microcystis viridis* (NIES-102). Tetrahedron 52:14501–14506

McDaniel R, Thamchaipenet A, Gustafsson C, Fu H, Betlach M, Ashley G (1999) Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. Proc Natl Acad Sci USA 96:1846–1851

Metsa-Ketela M, Salo V, Halo L, Hautala A, Hakala J, Mantsala P, Ylihonko K (1999) An efficient approach for screening minimal PKS genes from Streptomyces. FEMS Microbiol Lett 180: 1–6

Moffitt MC, Neilan BA (2001) On the presence of peptide synthetase and polyketide synthase genes in the cyanobacterial genus *Nodularia*. FEMS Microbiol Lett 196:207–214

Molnar I, Schupp T, Ono M, Zirkle R, Milnamow M, Nowak-Thompson B, Engel N, Toupet C, Stratmann A, Cyr DD, Gorlach J, Mayo JM, Hu A, Goff S, Schmid J, Ligon JM (2000) The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. Chem Biol 7:97–109

Motamedi H, Cai SJ, Shafiee A, Elliston KO (1997) Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK506. Eur J Biochem 244:74–80

Murakami M, Okita Y, Matsuda H, Okino T, Yamaguchi K (1994) Aeruginosin 298-A, a thrombin and trypsin inhibitor from the blue-green alga *Microcystis aerusinosa* (NIES-298). Tetrahedron Lett 35:3129–3132

Neilan BA, Jacobs D, Goodman AE (1995) Genetic diversity and phylogeny of toxic cyanobacteria determined by DNA polymorphisms within the phycocyanin locus. Appl Environ Microbiol 61:3875–3883

Nicholson TP, Rudd BA, Dawson M, Lazarus CM, Simpson TJ, Cox RJ (2001) Design and utility of oligonucleotide gene probes for fungal polyketide synthases. Chem Biol 8:157–178

Okino T, Matsuda H, Murakami M, Yamaguchi K (1995) New microviridins, elatase inhibitors from the blue-green alga *Microcystis aeruginosa*. Tetrahedron Lett 51:10679–10686

Quadri LE, Sello J, Keating TA, Weinreb PH, Walsh CT (1998) Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem Biol 5:631–645

Rein KS, Borrone J (1999) Polyketides from dinoflagellates: Origins, pharmacology and biosynthesis. Comp Biochem Physiol B Biochem Mol Biol 124:117–131

Sano T, Kaya K (1995) A 2-amino-2-butenoic acid (Dhb)-containing microcystin isolated from *Oscillatoria agardhii*. Tetrahedron Lett 36:8603–8606

Schwecke T, Aparicio JF, Molnar I, Konig A, Khaw LE, Haydock SF, Oliynyk M, Caffrey P, Cortes J, Lester JB, Bohm GA, Staunton J, Leadlay PF (1995) The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. Proc Natl Acad Sci USA 92:7839–7843

Seow KT, Meurer G, Gerlitz M, Wendt-Pienkowski E, Hutchinson CR, Davies J (1997) A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: A means to access and use genes from uncultured microorganisms. J Bacteriol 179:7360–7368

Shah S, Xue Q, Tang L, Carney JR, Betlach M, McDaniel R (2000) Cloning, characterization and heterologous expression of a polyketide synthase and P-450 oxidase involved in the biosynthesis of the antibiotic oleandomycin. J Antibiot (Tokyo) 53:502–508

Shen B, Du LC, Sanchez C, Chen M, Edwards DJ (1999) Bleomycin biosynthesis in *Streptomyces verticillus* ATCC 15003: A model of hybrid peptide and polyketide biosynthesis. Bioorganic Chem 27:155–171

Silakowski B, Schairer HU, Ehret H, Kunze B, Weinig S, Nordsiek G, Brandt P, Blocker H, Hofle G, Beyer S, Muller R (1999) New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. J Biol Chem 274:37391–37399

Silakowski B, Kunze B, Muller R (2001) Multiple hybrid polyketide synthase/non-ribosomal peptide synthetase gene clusters in the myxobacterium *Stigmatella aurantiaca*. Gene 275:233–240

Tanaka M, Ueno A, Kawasaki K, Yumoto I, Ohgiya S, Hoshino T, Ishizaki K, Okuyama H, Morita N (1999) Isolation of clustered genes that are notably homologous to the eicosapentaenoic acid biosynthesis gene cluster from the docosahexaenoic acid-producing bacterium *Vibrio marinus* strain MP-1. Biotechnol Lett 21:939–945

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tillett D, Dittmann E, Erhard M, von Dohren H, Borner T, Neilan BA (2000) Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC 7806: An integrated peptide-polyketide synthetase system. Chem Biol 7:753–764

Witkowski A, Joshi AK, Lindqvist Y, Smith S (1999) Conversion of a β-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. Biochemistry 38:11643–11650

Wood DW, Setubal JC, Kaul R, et al. (2001) The genome of the natural genetic engineer *Agrobacterium tumefacians*. Science 294:2317–2323

Wu K, Chung L, Revill WP, Katz L, Reeves CD (2000) The FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. Gene 251:81–90

Yazawa K (1996) Production of eicosapentaenoic acid from marine bacteria. Lipids 31:S297–S300