Chemistry & Biology
Article

Genome-Wide High-Throughput Mining of Natural-Product Biosynthetic Gene Clusters by Phage Display

Jun Yin,[1,4,6] Paul D. Straight,[2,6] Siniša Hrvatin,[1] Pieter C. Dorrestein,[3,5] Stefanie B. Bumpus,[3] Cindy Jao,[1] Neil L. Kelleher,[3] Roberto Kolter,[2,*] and Christopher T. Walsh[1,*]
[1] Department of Biological Chemistry and Molecular Pharmacology
[2] Deparment of Microbiology and Molecular Genetics
Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, USA
[3] Department of Chemistry, University of Illinois, Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, USA
[4] Present address: Department of Chemistry, University of Chicago, 929 E. 57th Street, CIS E505A, Chicago, IL 60637, USA.
[5] Present address: Skaggs School of Pharmacy and Pharmaceutical Sciences and Departments of Pharmacology, Chemistry, and Biochemistry, University of California, San Diego, Biomedical Science Building (BSB), 9500 Gilman Drive, La Jolla, CA 92093, USA.
[6] These authors contributed equally to this work.
*Correspondence: rkolter@hms.harvard.edu (R.K.), christopher_walsh@hms.harvard.edu (C.T.W.)
DOI 10.1016/j.chembiol.2007.01.006

SUMMARY

We have developed a phage-display method for high-throughput mining of bacterial gene clusters encoding the natural-product biosynthetic enzymes, polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs). This method uses the phosphopantetheinyl transferase activity of Sfp to specifically biotinylate NRPS and PKS carrier-protein domains expressed from a library of random genome fragments fused to a gene encoding a phage coat protein. Subsequently, the biotinylated phages are enriched through selection on streptavidin-coated plates. Using this method, we isolated phage clones from the multiple NRPS and PKS gene clusters encoded in the genomes of *Bacillus subtilis* and *Myxococcus xanthus*. Due to the rapid and unambiguous identification of carrier domains, this method will provide an efficient tool for high-throughput cloning of NRPS and PKS gene clusters from many individual bacterial genomes and multigenome environmental DNA.

INTRODUCTION

Natural-product biosynthetic gene clusters have been found in the genomes of a wide range of organisms, including bacteria, marine invertebrates, algae, and plants, that inhabit virtually every niche of the biosphere. Among them, polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) gene clusters are responsible for the biosynthesis of two large families of natural products, polyketides and nonribosomal peptides, many of which have important antibiotic, anticancer, and immunosuppressant activities [1–5]. Despite a great interest in producing analogs of natural polyketides and nonribosomal peptides for drug development, the structural and biosynthetic complexity of these natural products poses a daunting challenge to synthetic chemists. To overcome the difficulties in organic synthesis, the alternative strategy of combinatorial biosynthesis holds promise for the production of structurally diversified natural products [6, 7]. Using a combinatorial approach, biosynthetic gene clusters are cloned from their native hosts, and the catalytic modules in the gene clusters are modified or recombined to generate diversified products with designated structural modifications. As an example, a number of structural analogs of the polyketide antibiotic erythromycin and the nonribosomal peptide antibiotic vancomycin have recently been produced [8–11].

Although gene clusters of a number of NRPSs and PKSs have been cloned and many cryptic secondary-metabolite pathways have been uncovered through whole-genome sequencing [12], an alternative means of cloning NRPS and PKS gene clusters would greatly benefit the production of novel compounds. The conventional cloning protocol involves the construction of cosmid libraries of a bacterial genome; screening the libraries with radioactive, degenerate DNA probes or PCR primers, which target conserved regions of PKS or NRPS gene clusters; and chromosome walking from the identified genes to retrieve the sequence of the entire gene cluster [13, 14]. Recently, a genome-scanning method was developed in which a shotgun genomic library was constructed and sequenced randomly to identify genes that might be part of NRPS or PKS clusters based on sequence homology [15]. The identified putative gene fragments were then used as probes to screen the cosmid library of the same genome in order to find the full-length cluster.

Here, we report an alternative method to sequence-based screening that utilizes efficient enzymatic selection of NRPS- and PKS-encoding gene fragments. This method uses phage display for the profiling of NRPS and PKS biosynthetic gene clusters in a bacterial genome

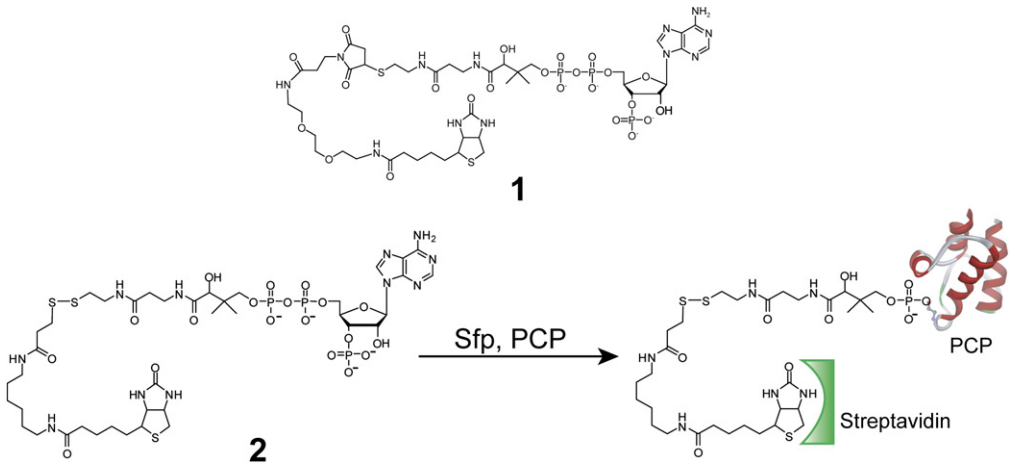

Figure 1. Biotin Labeling of PCP Catalyzed by Sfp Phosphopantetheinyl Transferase with Either Biotin-CoA, 1, or Biotin-SS-CoA, 2, as the Substrate

and identifies NRPS and PKS genes by high-throughput phage selection. Such an approach takes advantage of the capacity for Sfp phosphopantetheinyl transferase from *Bacillus subtilis* to covalently modify carrier-protein domains of NRPS or PKS enzymes with biotin-coenzyme A (CoA) substrate through a phosphopantetheinyl (Ppant) linkage [16, 17]. A shotgun library of the bacterial genome is displayed on the phage surface and undergoes iterative selections for Sfp-modified NRPS or PKS subfragments. The corresponding biosynthetic genes are then identified by sequencing the selected phage clones. The full-length PKS or NRPS clusters can then be reassembled by polymerase chain reactions (PCRs) based on the DNA sequence of the NRPS and PKS modules identified in the phage selection. The phage-display selection offers a simplified alternative to screening cosmid libraries to identify novel NRPS and PKS gene clusters. Here, we demonstrate, using *B. subtilis* and *M. xanthus*, that this selection-based approach can be used for the rapid cloning of NRPS and PKS gene fragments from single bacterial genomes. We propose that our method would be useful for high-throughput cloning of biosynthetic clusters from not only a single bacterial genome, but also environmental samples composed of multiple genomes (metagenome or environmental DNA), providing opportunities for accessing the potentially rich source of natural products from unculturable microbes.

RESULTS

Phage-Selection Strategy for the High-Throughput Profiling of NRPS and PKS Clusters in Bacterial Genomes

Phage display has been widely used for selecting peptide or protein ligands with high binding affinities [18] and for evolving enzymes or catalytic antibodies with higher catalytic activity or altered substrate specificity [19–22]. The displayed peptides or proteins are anchored to the surface of the phage by fusion with the phage capsid proteins. The genes encoding the displayed entities are encapsulated inside the phage particle as part of the phagemid or phage genome [23]. In this way, there is a direct physical linkage of the displayed peptide or protein with the encoding gene. The identity of the gene can be easily determined by sequencing the DNA from the phagemid.

The phage selection for NRPS or PKS genes takes advantage of the posttranslational modification of peptidyl carrier protein (PCP) domains and acyl carrier protein (ACP) domains embedded within the NRPS or PKS clusters catalyzed by phosphopantetheinyl transferases (PPTases) [24]. Type I PKSs and NRPSs are large, multimodular proteins that function with assembly line logic. The enzymes are composed of repeating catalytic modules, each of which is usually responsible for one cycle of polyketide or polypeptide chain elongation [3, 4]. Each module contains a thiolation (T) domain specifically referred to as PCP or ACP for NRPSs or PKSs, respectively. In order for the enzymes to be functional, Ppant prosthetic groups derived from coenzyme A (CoA-SH) need to be attached to all ACP or PCP domains by a Ppant transferase at a conserved Ser residue. The Ppant appendages act as swinging arms to transport the elongating polyketide or peptide chain along the assembly line [24].

Recently, we showed that Sfp PPTase from *B. subtilis* can attach a biotin-modified Ppant group to apo PCPs excised from the NRPS for enterobactin and gramicidin biosynthesis by using biotin-CoA (**1**) as the substrate [17] (Figure 1). We also showed that a folded PCP domain can be displayed on the surface of M13 phage and labeled with biotin by Sfp-catalyzed site-specific transfer of the biotin group from biotin-CoA (**1**) or biotin-SS-CoA (**2**) to the PCP domain anchored on the phage surface [25]. Subsequently, the phage particles displaying biotinylated PCP can be enriched more than 2000-fold through a single round of selection with immobilized streptavidin [25]. We reasoned that streptavidin-based selection of phage particles displaying biotinylated PCP or ACP domains could

 ©2007 Elsevier Ltd All rights reserved

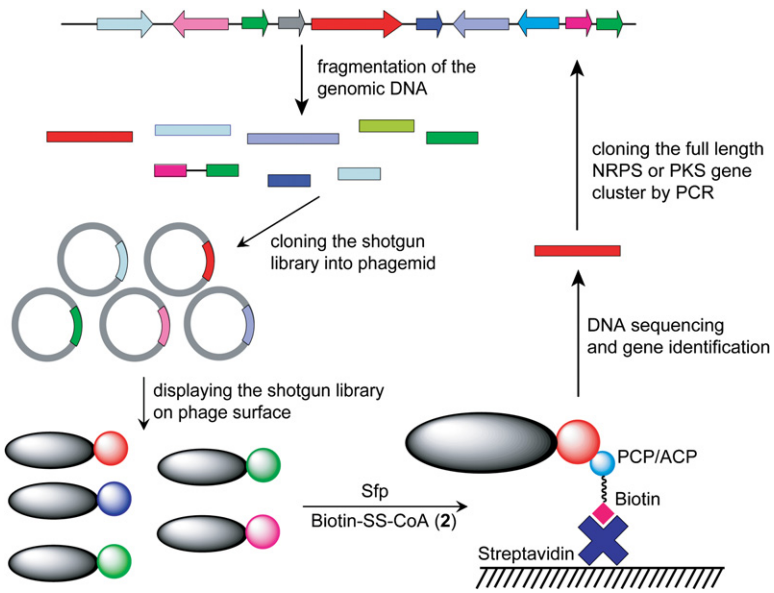

Figure 2. Strategy for High-Throughput Profiling of NRPS and PKS Gene Clusters by Phage Display

be a potential method for genome-wide functional profiling of biosynthetic gene clusters without prior knowledge of the DNA sequence. In this approach, we first construct a shotgun genomic library for the bacterial genome containing PKS or NRPS clusters (Figure 2). Proteins encoded by the genomic DNA fragments are displayed on the phage particles, and Sfp and biotin-SS-CoA (**2**) are added to covalently label the phage-displayed apo PCP or apo ACP domains with biotin. Phage particles displaying biotin-labeled PCPs or ACPs are then selectively bound to streptavidin-coated plates. The bound phage particles are then recovered and the DNA sequenced to determine corresponding biosynthetic genes.

Profiling PKS and NRPS Clusters in *B. subtilis* by Phage Display

To validate the phage-display strategy for profiling PKS and NRPS biosynthetic gene clusters, we first used *B. subtilis* as a model system since its genome (4.2×10^6 bp) has been sequenced [26] and Sfp is the native PPTase of this Gram-positive bacterium [27]. The genome encodes two NRPS gene clusters, *srfA* and *pps*, and one hybrid NRPS/PKS gene cluster, *pks*, that are responsible for the biosynthesis of surfactin, plipastatin, and bacillaene, respectively [28–30]. Also, the *dhb* cluster in the *B. subtilis* genome encodes a small NRPS required for the synthesis of the siderophore bacillibactin [31]. The genomic library of *B. subtilis* was constructed by fragmenting the genomic DNA by sonication into the range of 500–1500 bp and cloning the fragments into the phagemid vector pComb3H for M13 phage display [23]. The size of the phage library was $\sim 10^8$ clones, sufficient to cover the *B. subtilis* genome, assuming an average DNA insert size of 1000 bp: 1000 bp $\times 10^8 = 1 \times 10^{11}$ bp, 2.4×10^4 the size of the *B. subtilis* genome (4.2×10^6 bp).

To select for carrier proteins fused to the phage, the library was packaged into M13 phage particles. The Sfp enzyme and the disulfide-containing substrate, biotin-SS-CoA (**2**) (Figure 1), were added to a suspension of phage particles to covalently attach biotin-Ppant groups specifically to phage-displayed carrier proteins. After the biotin-labeling reaction, the modified phage particles were bound to a streptavidin-coated 96-well plate, and unbound phages were washed away. The phage particles bound to the streptavidin-coated plate via biotin-SS-CoA were eluted by disulfide reduction with 20 mM dithiothreitol (DTT). This process was repeated multiple times to selectively enrich Sfp-modified phage particles. Control reactions were carried out in parallel with the exclusion of Sfp or biotin-SS-CoA (**2**) during the labeling process. The titer of phage particles recovered from the streptavidin-coated plate was used to determine the efficiency of biotin labeling.

Over five iterations of selection, we observed a stepwise increase in the ratio of phages recovered from the reaction with Sfp and biotin-SS-CoA (**2**) versus the control reactions (Figure 3). Recovery of Sfp-modified phages indicated that proteins displayed on the phage surface are folded appropriately to be recognized and specifically biotinylated by Sfp. After the fifth round of selection, the recovery of biotin-modified phage was more than 10^4-fold that seen in the control reactions, suggesting that *B. subtilis* proteins posttranslationally modified by Sfp were substantially enriched.

To identify representative cloned fragments of *B. subtilis* DNA in the recovered phage particles, 100 clones were sequenced, and the sequences were mapped to the *B. subtilis* genome (Figure 4). After five rounds of selection, 85% of the selected clones encoded PCPs or ACPs, and the remaining clones contained in-frame stop codons within the genomic DNA fragments. These "nonsense" clones were presumably due to nonspecific binding of the phages to the streptavidin surface. Multiple clones with the same PCP or ACP gene were also found in the

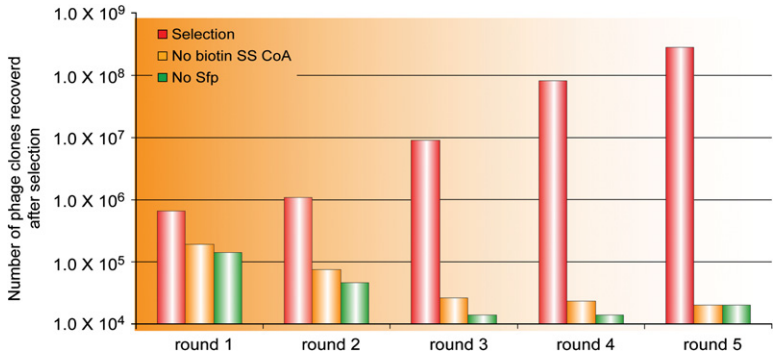

Figure 3. Steady Enrichment of the Phage Clones from the Phage-Displayed Bacterial Genomic Library during Iterative Rounds of Selection with Sfp and Biotin-SS-CoA

Red columns designate the number of phage particles recovered after each round of selection in the phage-labeling reactions with both Sfp and biotin-SS-CoA added. Yellow and green columns designate the corresponding phage recovery for the control reactions without the addition of Sfp or biotin-SS-CoA, respectively.

pool of recovered phage, indicating that coverage of cloned PCPs and ACPs was near saturation after the fifth round of selection. In particular, the first PCP domain in the *pksL* open reading frame (ORF) arose 35 times among the 100 clones sequenced (Figure 4), suggesting that this clone started to dominate the library, and thus no additional phage selection was carried out after the fifth round.

PCPs and ACPs were recovered from all four gene clusters in the *B. subtilis* genome that are responsible for the biosynthesis of the secondary metabolites surfactin [29], a hybrid peptide-polyketide bacillaene [28], plipastatin [26], and bacillibactin [31]. Although not all 39 of the PCP and ACP domains were found among the 100 clones being sequenced, 4 out of 7 PCPs from the surfactin gene cluster, 10 out of 20 PCPs and ACPs from the bacillaene gene cluster, 2 out of 10 PCPs from the plipastatin gene cluster, and 1 out of 2 PCPs from the bacillibactin gene cluster were captured (Figure 4). Overall 17 out of 39 PCPs/ACPs in annotated NRPS- and PKS-encoding genes in the *B. subtilis* genome were identified among the 100 clones sequenced. We found that the PCP and ACP domains were distributed over all three PCP-containing NRPS proteins for surfactin biosynthesis (SrfAA, SrfAB, and SrfAC), all ACP-containing PKSs for polyketide biosynthesis (PksJ, PksL, PksM, PksN, and PksR), two out of five PCP-containing NRPSs for plipastatin biosynthesis (PpsB and PpsD, but not PpsA, PpsC, and PpsE), and the only PCP-containing NRPS for bacillibactin biosynthesis (DhbF) (Figure 4). Thus, after five rounds of Sfp-dependent phage selection, nearly 50% of ACP and PCP domains in the *B. subtilis* genome were cloned from a single genomic library containing shotgun DNA fragments. Interestingly, the ACP domain of the fatty acid synthase (FAS) of *B. subtilis* was not recovered. This may be due to the bias of Sfp recognition for PCPs and ACPs involved in secondary metabolism. However, the fact that Sfp can complement a loss of function in AcpS, the FAS-specific Ppant transferase in *B. subtilis*, seems to suggest that Sfp should be able to modify phage-displayed FAS ACP [32]. We thus speculate that *E. coli* AcpS may preload *B. subtilis* FAS ACP with CoA in situ during the translation and secretion of ACP-pIII fusion protein for the assembly of the phage capsid. In this manner, phage particles displaying *B. subtilis* FAS ACP would be blocked from Sfp modification in vitro and would not be enriched through selection.

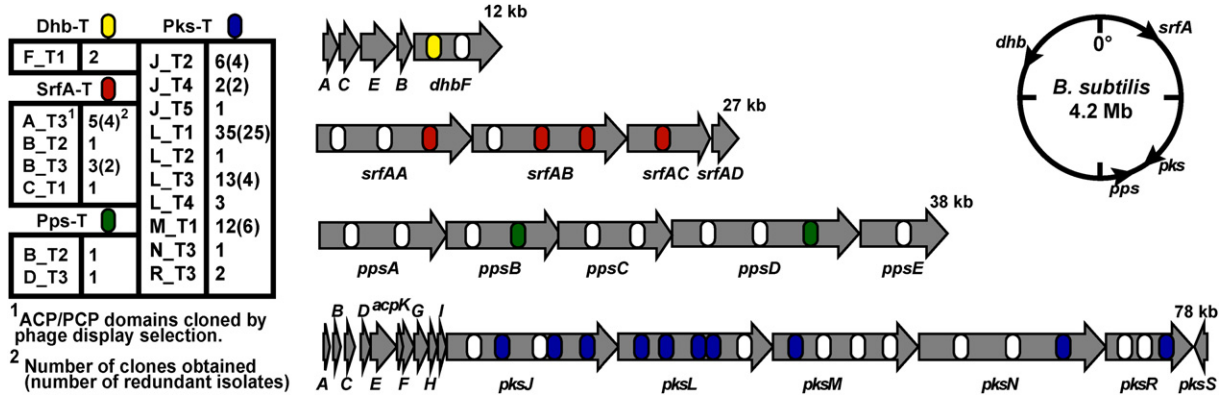

Figure 4. ACP and PCP Fragments Recovered after Phage Display and Selection of Modified *B. subtilis* Genomic Library Clones

Representatives of all four NRPS- and PKS-encoding gene clusters were identified in the selected pools. A schematic of the *B. subtilis* genome, with location of each gene cluster, is shown. Clones from the four gene clusters are: *dhbF*, yellow (bacillibactin); *srfAA*, *srfAB*, and *srfAC*, red (surfactin); *ppsB* and *ppsD*, green (plipastatin); and *pksJ*, *pksL*, *pksM*, *pksN*, and *pksR*, blue (bacillaene). ACP and PCP domains that were not recovered are represented as white ovals. The table lists the relative position of the cloned ACPs and PCPs in the protein primary sequence and the number of clones for each protein fragment.

Profiling PKS and NRPS Clusters in *M. xanthus* by Phage Display

Since Sfp is the native PPTase in *B. subtilis* responsible for the posttranslational activation of the *B. subtilis* NRPS and PKS, we sought to test whether Sfp can also be used to profile NRPS and PKS genes in bacteria other than its native host. We chose *M. xanthus* as a test organism, because its genome has recently been sequenced (*M. xanthus* sequence data are deposited in the database of The Institute for Genomic Research [TIGR] at http://www.tigr.org/). *M. xanthus* is a Gram-negative myxobacterium, phylogenetically distant from *B. subtilis*. The size of the *M. xanthus* genome (9.1×10^6 bp) is more than twice that of *B. subtilis* and is among one of the largest microbial genomes that has been sequenced. Although a little more than half of the genes have been annotated, 5 NRPS and 9 hybrid NRPS/PKS gene clusters, containing a total of 81 PCP domains and 38 ACP domains, have been identified in the *M. xanthus* genome based on sequence analysis. Some of those gene clusters have been identified as the NRPS and PKS enzymes for the biosynthesis of antibiotic TA [33] and saframycin Mx1, which has antitumor activity [34]. No gene clusters exclusively encoding PKSs have been identified in the *M. xanthus* genome.

We constructed two genomic DNA libraries of the *M. xanthus* genome in the pComb3H phagemid vector. The construction of libraries differed by the method of DNA fragmentation used, either HaeIII partial digestion or sonication. Each library contained $\sim 10^7$ clones with DNA inserts of 1000 bp on average, a size large enough to cover the entire *M. xanthus* genome (9.1×10^6 bp). We carried out three rounds of phage selection for each library as described previously. Similar to the results with the *B. subtilis* library phage selection, we observed a steady increase in the ratio of phage recovered from the labeling reaction with both Sfp and biotin-SS-CoA (**2**) over that of the control reactions lacking either enzyme or substrate. At the end of the third round of selection, both libraries yielded a greater than 10^3-fold enrichment of modified phage particles, suggesting that carrier proteins from *M. xanthus* were displayed on the phage surface and enriched through iterative selection.

After the selection procedure, we sequenced 100 clones recovered from each library and identified the *M. xanthus* gene fragments cloned (Table 1). Out of the 200 clones sequenced, 171 clones give readable sequences, and all contained fragments of *M. xanthus* genomic DNA. Among them, 131 clones encoded in-frame fusions to pIII. The remaining 40 clones were either out of frame or ligated in the non-sense orientation and would not be expected to express pIII fusions. Among the 131 clones expressing in-frame pIII fusions to *M. xanthus* ORFs, 15 PCP- and 4 ACP-encoding inserts were recovered, and one of them is an ACP associated with FAS (Table 1). Interestingly, we recovered 15 clones from 3 different *M. xanthus* ORFs that share similarity to known PCP sequences but do not correspond to domains annotated as PCP or ACP in the NCBI BLAST conserved domain database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Table 1). Furthermore, among the 131 in-frame clones recovered, we also found 22 clones encoding 6 truncated fragments of *M. xanthus* proteins that share no sequence similarity with known PCP or ACP domains (Table 1). Whether the full-length proteins that correspond to these fragments are modified by Sfp or the endogenous *M. xanthus* PPTases remains to be investigated.

Consistent with the *B. subtilis* results, after multiple rounds of selection, we identified *M. xanthus* clones that predominated the pool of recovered phage. The clones SH290 and SH629 were isolated 12 and 26 times, respectively. SH290 was only recovered from library 1, and SH629 was recovered from library 2. Among the 22 carrier proteins identified, only 4 are ACP domains, out of a total of 38 ACPs in the genome. Thus, Sfp appears to have a bias for modification of PCPs from the *M. xanthus* genomic library. There are three ORFs in the *M. xanthus* genome annotated as putative PPTases by TIGR: one (MXAN_4350 in the TIGR database) is an AcpS based on its sequence homology with AcpS from *E. coli* [35], and the others (MXAN_3485 and MXAN_4192 in the TIGR database) are homologous to Sfp from *B. subtilis* [24]. The existence of two Sfp-like PPTases in the *M. xanthus* genome raises the possibility that these are differentially selective for posttranslational priming of PCPs versus ACPs. Future work could address this hypothesis by repeating phage selection with the three purified *M. xanthus* PPTases in parallel.

Cloned ORFs that Expand the Repertoire of Sfp beyond the NRPS and PKS

To our surprise, we recovered from the *B. subtilis* and *M. xanthus* libraries phage clones expressing unconventional proteins, not annotated as domains of NRPSs, PKSs, or FAS. We followed this up in the case of *B. subtilis*, in which truncated proteins encoded by fragments of the *gcvT* and *ybbR* ORFs were recovered repeatedly during phage selection. The *gcvT* gene encodes a subunit of the glycine-cleavage system that is predicted to function in conjunction with *gcvH*, *gcvPA*, and *gcvPB*. The *ybbR* gene encodes a protein of unknown function and is a component of a putative three ORF operon that appears to be highly conserved in Gram-positive bacteria. All of the cloned fragments of *gcvT* and *ybbR* contained overlapping residues, suggesting that Sfp modified a specific residue within these overlapping fragments. Because Sfp is known to show strong selectivity for folded peptide sequences containing the consensus serine for modification [36], we sought to determine the site of Ppant modification of these proteins.

One of the cloned *gcvT* fragments (aa 29–204) was expressed as a soluble protein in *E. coli*, and posttranslational phosphopantetheinylation of this fragment was confirmed by mass spectrometry and immunoblot. High-resolution mass spectrometry of the GcvT truncate treated with Sfp and CoA gave a mass of 22,650.09 Da, well in agreement with the calculated mass of 22,649.98

Table 1. Summary of Phage Selection of the *M. xanthus* Genomic Library

Clone ID	Gene ID	PCP/ACP Selected	Gene Cluster	Times Selected	Library (1 or 2)
SH219	MXAN_1607	MXAN_1607.5 PCP	NRPS 2	13+2	1 and 2
SH227	MXAN_4402	MXAN_4402.1 PCP	NRPS 4	9	1
SH235	MXAN_4402	MXAN_4402.3 PCP	NRPS 4	1	1
SH571	MXAN_4403	MXAN_4403.1 PCP	NRPS 4	4	2
SH262	MXAN_4403	MXAN_4403.4 PCP	NRPS 4	1	1
SH616	MXAN_4601	MXAN_4601.1 PCP	NRPS 5	2	2
SH586	MXAN_3618	MXAN_3618.1 PCP	NRPS/PKS 2	2	2
SH582	MXAN_3634	MXAN_3634.5 PCP	NRPS/PKS 3	4	2
SH595	MXAN_3636	MXAN_3636.2 PCP	NRPS/PKS 3	9	2
SH466	MXAN_3636	MXAN_3636.4 PCP	NRPS/PKS 3	1	2
SH296	MXAN_3636	MXAN_3636.6 PCP	NRPS/PKS 3	3	1
SH261	MXAN_3636	MXAN_3636.8 PCP	NRPS/PKS 3	5	1
SH193	MXAN_3635	MXAN_3635.1 ACP	NRPS/PKS 3	1	1
SH320	MXAN_3935	MXAN_3935.1 PCP	NRPS/PKS 4	3+2	1 and 2
SH629	MXAN_3935	MXAN_3935.3 ACP	NRPS/PKS 4	26	2
SH518	MXAN_3935	MXAN_3935.5 ACP	NRPS/PKS 4	3	2
SH607	MXAN_4299	MXAN_4299.1 PCP	NRPS/PKS 7	1	2
SH573	MXAN_4525	MXAN_4525.1 PCP	NRPS/PKS 9	1	2
SH588	MXAN_6392	MXAN_6392 ACP	FAS	1	2
Clones Not Annotated by NCBI					
SH58S	MXAN_4601	MXAN_4601.0 PCP	NRPS/PKS 5	2	1
SH290	MXAN_4414	MXAN_4414.0 PCP	NRPS/PKS 8	12	1
JY1247	MXAN_4532	MXAN_4532.0 PCP	NRPS/PKS 9	1	1
Total	14	22	9	109	
Other Clones		Regional Homology			
SH238	MXAN_1740	cytochrome c-type protein		1	1
SH251	MXAN_1001	putative transporter		1	1
JY1236	MXAN_2231	putative lipoprotein		1	1
SH552	MXAN_1082	NADH dehydrogenase I		8	2
SH221	MXAN_0536	Zn-dependent hydrolases		1	1
SH233	MXAN_4123	no homology		10	1
Total	6	6		22	

Selected phage clones were listed with the clone identification number sent for DNA sequencing ("Clone ID"), the corresponding gene in the TIGR sequence database ("Gene ID"), the number of times the clones had been identified within the 200 clones sent for sequencing ("Times Selected"), and the library from which the clones were selected ("Library [1 or 2]"). Library 1, *M. xanthus* genomic DNA with HaeIII-digested fragments; library 2, sonicated fragments. The domains were annotated as PCP if they were located after an adenylation (A) domain and as ACP if they followed a ketosynthase (KS) or acyltransferase (AT) domain. Each gene cluster was defined as the set of genes adjacent to or in the immediate vicinity of each other. "Other Clones" denotes the clones that cannot be aligned with known PCP or ACP domains; their homology in peptide sequences with other proteins is also listed ("Regional Homology").

Da (Figure S1; see the Supplemental Data available with this article online). We observed a 340.23 Da difference between the modified mass and the apo form of the protein (observed mass is 22,309.86 Da), agreeing with the addition of a phosphopantetheine. The GcvT residue modified by Sfp was identified by using FTMS methodology (see Supplemental Data and Figure S2). By comparing masses of peptide fragment ions by MS, we localized the phosphopantetheinylated residue to the peptide GNDSL. This fragment contains the Ser62 residue (corresponding

to the full-length GcvT protein) embedded in the consensus motif for Sfp Ppant recognition, DSL. To confirm that GcvT is a target for Sfp, the full-length protein and an engineered Ser62Ala mutant were expressed and analyzed by anti-biotin immunoblot. Sfp-dependent biotinylation of the GcvT protein requires the Ser62 residue, positively identifying this Ser as the site of Ppant modification (Figure S3).

Using a similar strategy, truncated YbbR protein (aa 229–278) was expressed and confirmed to undergo posttranslational phosphopantetheinylation, and the site of Ppant modification was mapped to Ser274 by FTMS [37]. Full-length YbbR protein was also cloned and expressed. In vitro Ppant modification of the full-length protein was confirmed by FTMS, showing an observed mass shift of 344.28 Da for the Ppant-modified form from the apo form (the expected mass shift for phosphopantetheinylation is 340.1 Da) (Figure S4). The native function of ybbR in *B. subtilis* is not yet understood, but we have utilized YbbR modification by Sfp as a starting point to determine that a minimal YbbR-derived peptide of 11 residues suffices as a peptide tag that can be inserted into other proteins and modified by Sfp [37]. This finding highlights an unexpected benefit of identification from novel posttranslational protein modifications uncovered by our phage-display method.

Phage selection with Sfp also identified a number of ORFs from the *M. xanthus* genomic libraries that can be modified by Sfp by using biotin-SS-CoA (**2**) but that do not share significant sequence homology with ACP or PCP domains as the conventional substrates of Sfp-catalyzed posttranslational modification (Table 1). The protein sequences of ORFs recovered from these phage clones are listed in Figure S5. We determined, by using an ELISA assay, that these fragments displayed on the phage surface are efficiently labeled with biotin by Sfp, confirming that the phage clones were enriched from the *M. xanthus* genomic library and not simply background contaminants (Figure S6). It would be interesting in the future to verify if the corresponding full-length proteins are phosphopantetheinylated by *M. xanthus* PPTases as well as the effect of Ppant modification on the function of those proteins.

DISCUSSION

The current study presents a high-throughput method for the cloning of NRPS and PKS natural-product biosynthetic gene clusters from bacterial genomes. This method is based on Sfp PPTase-catalyzed biotin-Ppant modification of PCP and ACP domains displayed on the phage surface as part of the library encoded by bacterial genomic DNA fragments. Subsequent selection of biotin-labeled phages for streptavidin binding enriches phage particles displaying fragments of NRPS and PKS; sequencing of the enriched phage clones provides the DNA sequence of carrier-protein domain-containing fragments of NRPS and PKS genes.

Sfp is the native PPTase in *B. subtilis* [24] and has been shown to posttranslationally modify ACP and PCP domains from a variety of bacteria such as *Amycolatopsis mediterranei* [38], *Bacillus brevis* [39], *Escherichia coli* [27], *Mycobacterium tuberculosis* [40], *Pseudomonas syringae* [41], *Sorangium cellulosum* [42], *Streptomyces spheroides* [43], *Vibro cholerae* [44], and *Yersinia pestis* [45]. Thus, Sfp is likely to be useful for the mining of NRPS and PKS clusters from a wide range of bacterial species by phage display. Yet, the Sfp enzyme is only one of a family of PPTases that have wide substrate specificity with ACPs and PCPs. Other PPTases such as Gsp from *Bacillus brevis* [46], AcpS from *Escherichia coli* [47, 48], and PcpS from *Pseudomonas aeruginosa* [49] have been cloned and used for posttranslational modification of ACPs or PCPs from bacteria other than their native hosts.

PPTases have been classified into two superfamilies based on their sequence homologies and substrate preferences [24]. The first family is represented by AcpS of *Escherichia coli*, which has a restricted function for FAS and FAS-related PKSs [35]. Sfp from *B. subtilis* belongs to the second family of PPTases, which exhibits a broad substrate specificity, modifying both ACPs from PKSs and PCPs from NRPSs [24]. A recent phylogenetic analysis of the Sfp-type PPTases also identified several subclasses of the enzymes originating from a broad spectrum of bacterial genera, including *Bacillus*, *Escherichia*, *Streptomyces*, and representatives of the cyanobacteria and myxobacteria [50]. Although PPTases in the Sfp family have shown cross-species or cross-phyla activities for the modification of ACPs and PCPs, it is conceivable that one type of PPTase would show bias toward carrier proteins from its original species or closely related species. Supporting this notion, Sfp profiling identified 17 out of 39 carrier proteins in *B. subtilis*, almost half of the carrier proteins encoded in the genome. In contrast, profiling with Sfp on the *M. xanthus* genome identified 22 carrier-protein clones from 119 carrier proteins annotated in the genome. These results suggest that preference for endogenous transferases may carry over into phage-display modification of carrier proteins. The use of multiple Ppant transferases would likely broaden the spectrum of modified carrier proteins to be cloned by this approach.

Phage selection with Sfp also showed a preference for the enrichment of PCP clones versus ACP clones. For example, the *M. xanthus* genome annotation reveals a 2.1:1 ratio of PCPs over ACPs in the genome (81 PCPs and 38 ACPs annotated). However, we detected a 4.5:1 bias in the Sfp selection reported here (18 PCPs and 4 ACPs identified by phage selection). Such a bias is likely due to the substrate preference of Sfp for modifying PCPs. To address this asymmetry that is almost certain to be associated with any phage selection with one particular PPTase, phage-display selection could be carried out in parallel or with a mixture of two or three PPTases with broad carrier-protein specificity to increase the range of carrier proteins that are modified.

The use of multiple PPTases during phage selection would also contribute to a more even distribution of carrier-protein domains in the cloned pool than was

obtained by using Sfp alone. During phage-display selection with both the *B. subtilis* and *M. xanthus* libraries, individual clones arose disproportionately within the sequenced pool. For example, the first PCP of *pksL* predominated the clones after the fifth round of selection with the *B. subtilis* genome, and clones SH290 and SH629 were highly represented in the *M. xanthus* library after the third round of selection. We offer three possible explanations for the predominance of individual clones. Sfp could have a preference for the folded fragments of some carrier proteins over others. In support of this hypothesis is the finding that, among the overrepresented clones, multiple nonidentical fragments were recovered. Alternatively, particular fragments may be more favorable for expression on the phage coat, rendering them overrepresented in the unselected library. A third possibility, supported by nonoverlapping sets of clones arising from *M. xanthus* libraries 1 and 2, is a stochastic bias arising from enzymatic selection of phage-displayed protein. We have aligned the peptide sequences of the carrier-protein clones identified by Sfp profiling in the *B. subtilis* and *M. xanthus* genomes with the sequences of carrier proteins not identified in the selection, and we found no obvious difference between the two pools of carrier proteins based on sequence phylogeny. Phage-display selection with multiple PPTases will likely reduce bias in the cloned fragments despite the cause, and it has the added benefit of enhancing the effectiveness of selection from a broader range of DNA sources.

We propose that this phage-display method will be useful for mining and cloning NRPS or PKS clusters from bacterial genomes isolated from environmental samples (metagenome) [51]. It has been estimated that soil contains more than 4000 species of bacteria per gram, and that nutrient-rich fresh water contains about 160 bacterial species per 10 liters [52]. However, as many as 99% of the microorganisms from the environmental samples cannot be cultured by standard techniques, and the uncultured fraction contains diverse organisms that are only distantly related to the cultured ones [53]. Thus, in order to access the natural-product repertoire produced by the uncultured bacteria, NRPS and PKS clusters from these microorganisms need to be cloned from DNA directly isolated from environmental samples and transferred to a culturable heterologous host for natural-product production. The library size required to obtain sufficient coverage of the simplest metagenome represents a daunting challenge for hybridization screening or shotgun sequencing used in the conventional methods. As an alternative to conventional methods, phage display provides the advantage that a shotgun library of the metagenome can cover 10^{12} bp of sequence space, approximating the collective genome size of 10,000 different bacterial species. Moreover, since the selection of NRPS or PKS genes by phage display can be carried out in one pool, the phage-display method is truly high throughput and offers a unique opportunity for rapid cloning of NRPS and PKS cluster fragments from environmental DNA.

SIGNIFICANCE

Polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) are two important types of natural-product biosynthetic enzymes in bacteria, having great potential for producing diverse and medically useful compounds. We have developed a phage-display method for high-throughput cloning of carrier proteins from bacterial PKS and NRPS gene clusters. Clones from shotgun libraries of two bacterial genomes were displayed on the surface of phage. Phage clones displaying protein domains that are recognized as substrates for the Sfp PPTase are enriched by iterative selection for site-specific attachment of biotin by Sfp. In this way, PKS and NRPS gene fragments in the genomic library are significantly enriched in the selected pool and do not require extensive sequencing of the bacterial genome. The large sequence space covered by phage libraries and the high-throughput phage-selection protocol will allow this method to be applied to a collection of bacterial genomes for mining PKS and NRPS enzymes from environmental sources of DNA.

EXPERIMENTAL PROCEDURES

Construction of the *B. subtilis* Library

Genomic DNA was extracted from 50 ml of a log-phase culture of *B. subtilis* strain 3610 [54]. A total of 100 μg RNaseA-treated genomic DNA was randomly sheared with two 5 s pulses with a sonicator (Misonix, Inc., Farmingdale, NY) set at 35% power. Sheared DNA was size selected on an agarose prep gel for fragments ranging from 500 bp to 1.5 kb. A total of 10 μg sheared DNA fragments were treated with 15 U T4 polymerase to create blunt-ended fragments. The pComb phagemid vector, which was modified to contain EcoRI and HindIII restriction sites (pComb-HE), was digested with SacI and SpeI and treated with T4 polymerase to create blunt ends, and with calf intestinal phosphatase to prevent vector self-ligation. The blunt fragments were ligated into the pComb-HE phagemid vector by using 5 μg *B. subtilis* genomic DNA fragments and 17 μg pComb-HE vector. Ligated plasmids were cloned into electrocompetent XL1-Blue *E. coli* (Stratagene, La Jolla, CA). *E. coli* clones containing the vector were screened by colony PCR for insertion of *B. subtilis* genomic DNA fragments. The efficiency of insertion was ~90% of the total clones recovered.

Construction of the *M. xanthus* Library

The *M. xanthus* DNA was isolated by incubating the cells in 8 ml TE (pH 7.5), containing 0.5% SDS, 0.1 mg/ml Proteinase K, and 50 μg/ml RNase A. After 1 hr of incubation at 37°C and two phenol:chloroform: isoamyl alcohol (25:24:1) extractions in 50 ml phase-lock tubes (Eppendorf, Hamburg, Germany), the DNA was precipitated by adding 0.7 volumes of isopropanol, spooled from the solution, washed with 70% ethanol, and resuspended in TE buffer. For library 1, the DNA was partly digested with NEB HaeIII to an approximate size of 500–1500 bp, followed by DNA electrophoresis size selection in 1× TAE with 1% agarose. The pComb3H vector was digested by SacI and SpeI, blunt ended by T4 DNA polymerase, and treated with calf intestinal alkaline phosphatase. For library 2, *M. xanthus* DNA was sheared by sonication to ~1–3 kb and was treated with the T4 DNA polymerase for 15 min at 12°C; 500–1500 bp fragments were selected as described. The pComb3H vector was digested by EcoRV. For both libraries, the ligation of the vectors with *M. xanthus* fragments was performed by using the T4 DNA ligase at 16°C for 16 hr, followed by electroporation into XL1-Blue electrocompetent *E. coli* cells (Stratagene).

 ©2007 Elsevier Ltd All rights reserved

Library 1 was plated on LB agar supplemented with 2% glucose and 100 μg/ml ampicillin, while library 2 was plated in a liquid gel culture containing LB, 0.5% SeaPrep agarose supplemented with 2% glucose and 100 μg/ml ampicillin [55].

Phage Selection

Phage-displayed *B. subtilis* or *M. xanthus* genomic libraries were first labeled with biotin by Sfp by using biotin-SS-CoA (**2**) as the substrate. For the first round of selection, the labeling reactions were carried out with 10^{12} phage particles in 1 ml 10 mM $MgCl_2$ and 50 mM HEPES (pH 7.5) with 5 μM biotin-SS-CoA (**2**), 1 μM Sfp and were incubated at 30°C for 30 min. For the subsequent rounds of selections, the number of input phage particles, the concentration of enzymes and biotin-SS-CoA (**2**), and the reaction time were decreased step by step, and, eventually, for the fifth round of selection, only 10^{10} phage particles were incubated with 0.08 μM enzyme and 1 μM biotin-SS-CoA (**2**) for 7 min at 30°C. Control reactions were also run in parallel without the addition of enzymes or biotin-SS-CoA (**2**).

After the labeling reaction, the reaction mixtures were added to 250 μl 20% (w/v) polyethylene glycol 8000 with 2.5 M NaCl, followed by a 10 min incubation on ice. The phage particles in the reaction mixture were then precipitated by centrifugation at 4°C at a speed of 13,000 revolutions per minute (rpm). The phage pellet was then resuspended in 1 ml TBS supplemented with 1% (w/v) BSA and distributed in 100 μl aliquots to the wells of streptavidin-coated 96-well plates (Pierce). The plates were allowed to incubate at room temperature for 1 hr before the supernatant was discarded, and each well was washed 30 times with 0.05% (v/v) Tween 20, 0.05% (v/v) Triton X-100 in TBS and 30 times with TBS; each time the wash was performed with 200 μl solution. After washing, phages bound to the streptavidin surface were eluted by adding 100 μl 20 mM DTT in TBS to each well to induce the cleavage of the disulfide bond that links the biotin group with Ppant. Eluted phage particles were combined, added to 10 ml log-phase XL1-Blue cells, and shaken at 37°C for 1 hr to infect the cells. The cells were then plated on LB agarose plates supplemented with 2% (w/v) glucose and 100 μg/ml ampicillin. After incubation at 37°C overnight, colonies on the plates were scratched, and the phagemid DNA was extracted by a QIAGEN Plasmid Maxi kit. The phagemid DNA was then used for the next round of phage production and selection. Also, phage particles eluted from the wells loaded with either the selection or the control reactions were titered in order to count the number of phage particles selected by each round. After the fifth round of selection, phage clones were sequenced by using the primer Jun13 (5′-ACTTTATGCTTCCGGCTCGTATGT-3′).

Supplemental Data

Supplemental Data include mass spectrometry results, additional Experimental Procedures, and supplemental figures and are available at http://www.chembiol.com/cgi/content/full/14/3/303/DC1/.

ACKNOWLEDGMENTS

P.D.S. was supported by a National Science Foundation Postdoctoral Fellowship in Microbial Biology (DBI-0200307). This work was supported by grants GM20011 (to C.T.W.), GM67725 (to N.L.K.), and GM58213 (to R.K.) from the National Institutes of Health.

Received: October 14, 2006
Revised: December 26, 2006
Accepted: January 9, 2007
Published: March 23, 2007

REFERENCES

1. Clardy, J., and Walsh, C. (2004). Lessons from natural molecules. Nature *432*, 829–837.
2. Khosla, C. (1997). Harnessing the biosynthetic potential of modular polyketide synthases. Chem. Rev. *97*, 2577–2590.
3. Sieber, S.A., and Marahiel, M.A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. Chem. Rev. *105*, 715–738.
4. Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. *18*, 380–416.
5. Walsh, C.T. (2004). Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science *303*, 1805–1810.
6. Cane, D.E., Walsh, C.T., and Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. Science *282*, 63–68.
7. Walsh, C.T. (2002). Combinatorial biosynthesis of antibiotics: challenges and opportunities. ChemBioChem *3*, 125–134.
8. Fu, X., Albermann, C., Jiang, J., Liao, J., Zhang, C., and Thorson, J.S. (2003). Antibiotic optimization via in vitro glycorandomization. Nat. Biotechnol. *21*, 1467–1469.
9. Losey, H.C., Jiang, J., Biggins, J.B., Oberthur, M., Ye, X.Y., Dong, S.D., Kahne, D., Thorson, J.S., and Walsh, C.T. (2002). Incorporation of glucose analogs by GtfE and GtfD from the vancomycin biosynthetic pathway to generate variant glycopeptides. Chem. Biol. *9*, 1305–1314.
10. McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., and 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.
11. Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., Hopwood, D.A., and Santi, D.V. (2005). Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat. Biotechnol. *23*, 1171–1176.
12. Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., and Omura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat. Biotechnol. *21*, 526–531.
13. Liu, W., Christenson, S.D., Standage, S., and Shen, B. (2002). Biosynthesis of the enediyne antitumor antibiotic C-1027. Science *297*, 1170–1173.
14. Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C., and Julien, B. (2000). Cloning and heterologous expression of the epothilone gene cluster. Science *287*, 640–642.
15. Zazopoulos, E., Huang, K., Staffa, A., Liu, W., Bachmann, B.O., Nonaka, K., Ahlert, J., Thorson, J.S., Shen, B., and Farnet, C.M. (2003). A genomics-guided approach for discovering and expressing cryptic metabolic pathways. Nat. Biotechnol. *21*, 187–190.
16. La Clair, J.J., Foley, T.L., Schegg, T.R., Regan, C.M., and Burkart, M.D. (2004). Manipulation of carrier proteins in antibiotic biosynthesis. Chem. Biol. *11*, 195–201.
17. Yin, J., Liu, F., Li, X., and Walsh, C.T. (2004). Labeling proteins with small molecules by site-specific posttranslational modification. J. Am. Chem. Soc. *126*, 7754–7755.
18. Smith, G.P., and Petrenko, V.A. (1997). Phage display. Chem. Rev. *97*, 391–410.
19. Cesaro-Tadic, S., Lagos, D., Honegger, A., Rickard, J.H., Partridge, L.J., Blackburn, G.M., and Pluckthun, A. (2003). Turnover-based in vitro selection and evolution of biocatalysts from a fully synthetic antibody library. Nat. Biotechnol. *21*, 679–685.
20. Janda, K.D., Lo, L.C., Lo, C.H., Sim, M.M., Wang, R., Wong, C.H., and Lerner, R.A. (1997). Chemical selection for catalysis in combinatorial antibody libraries. Science *275*, 945–948.
21. Pedersen, H., Holder, S., Sutherlin, D.P., Schwitter, U., King, D.S., and Schultz, P.G. (1998). A method for directed evolution and functional cloning of enzymes. Proc. Natl. Acad. Sci. USA *95*, 10523–10528.

22. Yin, J., Mills, J.H., and Schultz, P.G. (2004). A catalysis-based selection for peroxidase antibodies with increased activity. J. Am. Chem. Soc. *126*, 3006–3007.

23. Barbas, C.F., 3rd, Kang, A.S., Lerner, R.A., and Benkovic, S.J. (1991). Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. Proc. Natl. Acad. Sci. USA *88*, 7978–7982.

24. Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996). A new enzyme superfamily - the phosphopantetheinyl transferases. Chem. Biol. *3*, 923–936.

25. Yin, J., Liu, F., Schinke, M., Daly, C., and Walsh, C.T. (2004). Phagemid encoded small molecules for high throughput screening of chemical libraries. J. Am. Chem. Soc. *126*, 13570–13571.

26. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., et al. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature *390*, 249–256.

27. Quadri, L.E., Weinreb, P.H., Lei, M., Nakano, M.M., Zuber, P., and Walsh, C.T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. Biochemistry *37*, 1585–1595.

28. Chen, X.H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., Koumoutsi, A., Hitzeroth, G., Grammel, N., Strittmatter, A.W., et al. (2006). Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. J. Bacteriol. *188*, 4024–4036.

29. Peypoux, F., Bonmatin, J.M., and Wallach, J. (1999). Recent trends in the biochemistry of surfactin. Appl. Microbiol. Biotechnol. *51*, 553–563.

30. Steller, S., Vollenbroich, D., Leenders, F., Stein, T., Conrad, B., Hofemeister, J., Jacques, P., Thonart, P., and Vater, J. (1999). Structural and functional organization of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213 and A1/3. Chem. Biol. *6*, 31–41.

31. May, J.J., Wendrich, T.M., and Marahiel, M.A. (2001). The dhb operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. J. Biol. Chem. *276*, 7209–7217.

32. Mootz, H.D., Finking, R., and Marahiel, M.A. (2001). 4′-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. J. Biol. Chem. *276*, 37289–37298.

33. Rosenberg, E., Fytlovitch, S., Carmeli, S., and Kashman, Y. (1982). Chemical properties of *Myxococcus xanthus* antibiotic TA. J. Antibiot. (Tokyo) *35*, 788–793.

34. Pospiech, A., Cluzel, B., Bietenhader, J., and Schupp, T. (1995). A new *Myxococcus xanthus* gene cluster for the biosynthesis of the antibiotic saframycin Mx1 encoding a peptide synthetase. Microbiology *141*, 1793–1803.

35. Flugel, R.S., Hwangbo, Y., Lambalot, R.H., Cronan, J.E., Jr., and Walsh, C.T. (2000). Holo-(acyl carrier protein) synthase and phosphopantetheinyl transfer in *Escherichia coli*. J. Biol. Chem. *275*, 959–968.

36. Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. *97*, 2651–2674.

37. Yin, J., Straight, P.D., McLoughlin, S.M., Zhou, Z., Lin, A.J., Golan, D.E., Kelleher, N.L., Kolter, R., and Walsh, C.T. (2005). Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. Proc. Natl. Acad. Sci. USA *102*, 15815–15820.

38. Admiraal, S.J., Khosla, C., and Walsh, C.T. (2002). The loading and initial elongation modules of rifamycin synthetase collaborate to produce mixed aryl ketide products. Biochemistry *41*, 5313–5324.

39. Belshaw, P.J., Walsh, C.T., and Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. Science *284*, 486–489.

40. Quadri, L.E., Sello, J., Keating, T.A., Weinreb, P.H., and Walsh, C.T. (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.

41. Vaillancourt, F.H., Yin, J., and Walsh, C.T. (2005). SyrB2 in syringomycin E biosynthesis is a nonheme FeII α-ketoglutarate- and O2-dependent halogenase. Proc. Natl. Acad. Sci. USA *102*, 10111–10116.

42. Chen, H., O'Connor, S., Cane, D.E., and Walsh, C.T. (2001). Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. Chem. Biol. *8*, 899–912.

43. Chen, H., and Walsh, C.T. (2001). Coumarin formation in novobiocin biosynthesis: β-hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI. Chem. Biol. *8*, 301–312.

44. Keating, T.A., Marshall, C.G., and Walsh, C.T. (2000). Vibriobactin biosynthesis in *Vibrio cholerae*: VibH is an amide synthase homologous to nonribosomal peptide synthetase condensation domains. Biochemistry *39*, 15513–15521.

45. Suo, Z., Tseng, C.C., and Walsh, C.T. (2001). Purification, priming, and catalytic acylation of carrier protein domains in the polyketide synthase and nonribosomal peptidyl synthetase modules of the HMWP1 subunit of yersiniabactin synthetase. Proc. Natl. Acad. Sci. USA *98*, 99–104.

46. Mootz, H.D., Schorgendorfer, K., and Marahiel, M.A. (2002). Functional characterization of 4′-phosphopantetheinyl transferase genes of bacterial and fungal origin by complementation of *Saccharomyces cerevisiae* lys5. FEMS Microbiol. Lett. *213*, 51–57.

47. Carreras, C.W., Gehring, A.M., Walsh, C.T., and Khosla, C. (1997). Utilization of enzymatically phosphopantetheinylated acyl carrier proteins and acetyl-acyl carrier proteins by the actinorhodin polyketide synthase. Biochemistry *36*, 11757–11761.

48. Gehring, A.M., Lambalot, R.H., Vogel, K.W., Drueckhammer, D.G., and Walsh, C.T. (1997). Ability of *Streptomyces spp*. acyl carrier proteins and coenzyme A analogs to serve as substrates in vitro for *E. coli* holo-ACP synthase. Chem. Biol. *4*, 17–24.

49. Finking, R., Solsbacher, J., Konz, D., Schobert, M., Schafer, A., Jahn, D., and Marahiel, M.A. (2002). Characterization of a new type of phosphopantetheinyl transferase for fatty acid and siderophore synthesis in *Pseudomonas aeruginosa*. J. Biol. Chem. *277*, 50293–50302.

50. Copp, J.N., and Neilan, B.A. (2006). The phosphopantetheinyl transferase superfamily: phylogenetic analysis and functional implications in cyanobacteria. Appl. Environ. Microbiol. *72*, 2298–2305.

51. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. *5*, R245–R249.

52. Pettit, R.K. (2004). Soil DNA libraries for anticancer drug discovery. Cancer Chemother. Pharmacol. *54*, 1–6.

53. Torsvik, V., and Ovreas, L. (2002). Microbial diversity and function in soil: from genes to ecosystems. Curr. Opin. Microbiol. *5*, 240–245.

54. Harwood C.R. and Cutting S.M., eds. (1990). Molecular Biological Methods for *Bacillus* (Chichester, UK: John Wiley & Sons).

55. Hrvatin, S., and Piel, J. (2006). Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. J. Microbiol. Methods *68*, 434–436.

 ©2007 Elsevier Ltd All rights reserved