A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles

Jörn Piel*

Department of Bioorganic Chemistry, Max Planck Institute for Chemical Ecology, Winzerlaer Strasse 10, Beutenberg Campus, 07745 Jena, Germany

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved September 5, 2002 (received for review August 9, 2002)

Many drug candidates from marine and terrestrial invertebrates are suspected metabolites of uncultured bacterial symbionts. The antitumor polyketides of the pederin family, isolated from beetles and sponges, are an example. Drug development from such sources is commonly hampered by low yields and the difficulty of sustaining invertebrate cultures. To obtain insight into the true producer and find alternative supplies of these rare drug candidates, the putative pederin biosynthesis genes were cloned from total DNA of *Paederus fuscipes* beetles, which use this compound for chemical defense. Sequence analysis of the gene cluster and adjacent regions revealed the presence of ORFs with typical bacterial architecture and homologies. The *ped* cluster, which is present only in beetle specimens with high pederin content, is located on a 54-kb region bordered by transposase pseudogenes and encodes a mixed modular polyketide synthase/nonribosomal peptide synthetase. Notably, none of the modules contains regions with homology to acyltransferase domains, but two copies of isolated monodomain acyltransferase genes were found at the upstream end of the cluster. In line with an involvement in pederin biosynthesis, the upstream cluster region perfectly mirrors pederin structure. The unexpected presence of additional polyketide synthase/nonribosomal peptide synthetase modules reveals surprising insights into the evolutionary relationship between pederin-type pathways in beetles and sponges.

Invertebrates, particularly those from marine environments, are an important source of natural products with high therapeutic potential (1). The low availability of most of these metabolites, however, represents a serious impediment to drug development (2). Because many invertebrates are difficult to cultivate and chemical synthesis in most cases is not economical, alternative and ecologically sound sources of such natural products are needed urgently. The actual producers of many drug candidates isolated from invertebrates may well be symbiotic bacteria (3–5), but thus far no producing symbiont has ever been cultured successfully. Therefore, despite good circumstantial evidence, up to now no direct proof exists for such symbioses. Bacterial genes from secondary metabolism are usually clustered, which simplifies their cloning and transfer into a heterologous host. Heterologous expression in a culturable bacterium therefore could generate a long-term supply of rare symbiont-derived drug candidates isolated from invertebrates.

The strongest evidence for such a bacterial biosynthesis is represented by the highly active antitumor compounds of the pederin group (ref. 6; Fig. 1). Pederin and mycalamides A and B exhibit IC_{50} values in the subnanomolar range for some tumor model systems and prolong the life span of mice bearing a variety of ascitic and solid tumors (6). Mycalamide A has also been shown to induce apoptosis (7). Although almost all the pederin members were isolated from marine sponges, pederin itself is known exclusively from terrestrial *Paederus* and *Paederidus* beetles. These notorious insects use pederin as chemical weapon against predators and cause severe dermatitis when accidentally crushed on the human skin (8). In all *Paederus* species studied thus far, up to 90% of the females contain high levels of pederin, and only these (+)-females transfer this trait to their female offspring (9). Pederin-free (−)-females do not produce (+)-offspring unless they are fed eggs of (+)-females. This non-Mendelian mode of inheritance can be prevented if the eggs are treated previously with antibiotics, which strongly suggests bacterially mediated pederin biosynthesis (10). The presence of a predominant bacterium with a close relationship to *Pseudomonas aeruginosa* only in (+)-females, as revealed by 16S rRNA analysis, further supports such a symbiosis (11). In stark contrast to the large number of suspected symbiont drug candidates from marine invertebrates, pederin is the only terrestrial example known to date. In this work, an architecturally unusual biosynthetic gene cluster is described that resides on the genome of an uncultured bacterial symbiont of *Paederus fuscipes* beetles.

Materials and Methods

Pederin Analysis. Beetle species were identified by R. L. L. Kellner and H. Başpınar. Beetles were stored individually in 100% ethanol immediately after collection to preserve the DNA. For pederin analysis the ethanol was concentrated *in vacuo* to 50 μl, and 10 μl were spotted on a silica gel TLC plate (Merck). The plate then was developed in ethyl acetate and stained with anisaldehyde reagent. A pink spot at $R_F = 0.22$ was specific for pederin (9).

Cloning and Sequencing of the *ped* Genes. A QIAamp DNA mini kit (Qiagen, Hilden, Germany) was used to extract DNA from adult beetles with known pederin content. This DNA was used as PCR template. For egg DNA templates, one egg was ground in PCR buffer at 0°C by using a Wheaton homogenizer (previously treated with concentrated HCl for 15 min and washed with sterile H_2O), transferred into a PCR tube, frozen and thawed three times, and subsequently boiled for 5 min before the remaining PCR components were added. For all initial reactions, the ketosynthase (KS)-specific primers KSDPQQF (5′-MGNGARGCNNWNSMNATGGAYCCNCARCANMG-3′) and KSHGTGR (5′-GGRTCNCCNARNSWNGTNCCNGTNCCRTG-3′) and Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen) were used. Each PCR experiment was performed at least in triplicate for each sex except for the rarer *Paederus litoralis* adults, where only two reactions could be run. PCR products were ligated into the pGEM-T Easy vector (Promega) and digested with *Rsa*I. Plasmids showing a unique restriction pattern were sequenced by using the BigDye Terminator Ready mix (Applied Biosystems) and an ABI 3700 sequencer (Applied Biosystems). From these sequences the following primer pairs specific for single modules were designed: 5′-TGGCATCGT-

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: KS, ketosynthase; PKS, polyketide synthase; AT, acyltransferase; NRPS, nonribosomal peptide synthetase.

Data deposition: The nucleotide sequence reported in this paper has been deposited in the GenBank database (accession no. AY059471).

*E-mail: piel@ice.mpg.de.

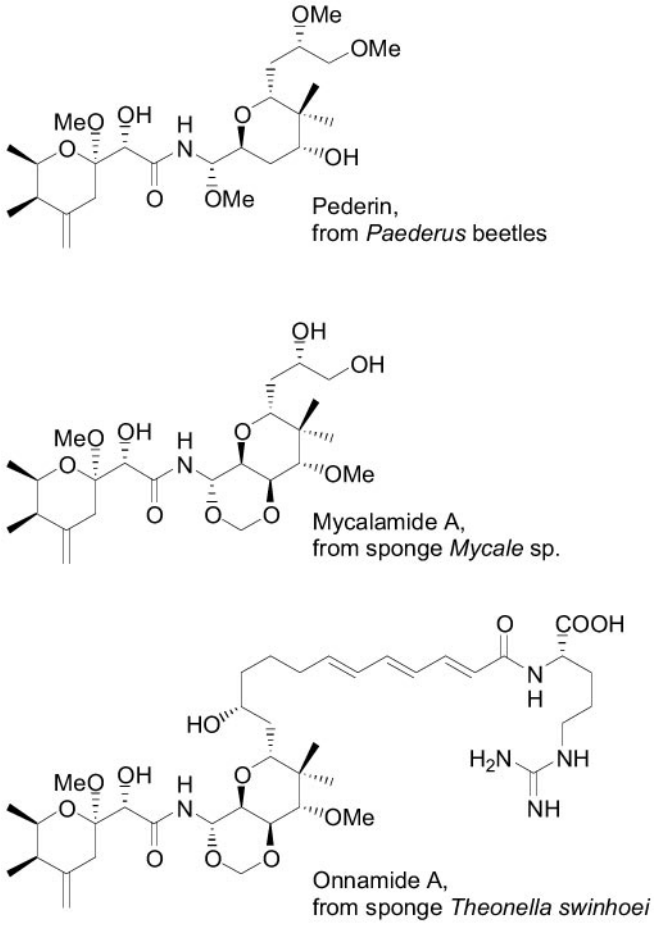

Fig. 1. Some members of the pederin family of antitumor compounds isolated from terrestrial beetles and marine sponges.

GGGGAAAGGCTG-3′ and 5′-GGCGCAGGTGCTGACACGC-3′ (KS1F–KS1R), 5′-TTAGCCATCGAGAGTTACAGCTC-3′ and 5′-AATCGCCGATAGCCATCGCCG-3′ (KS2F–KS2R), 5′-GACGCCATGGATGCACTGCAC-3′ and 5′-TATTGGATGCTCAGCACCGCAC-3′ (KS3F–KS3R), and 5′-GGGCTCAGTTTCCACCCTTATG-3′ and 5′-CCGGCGCTGCAGAGCCAGG-3′ (KS4F–KS4R). A cosmid library was prepared from total DNA of 12 *P. fuscipes* (+)-females collected in Aydin, Turkey (12), by using the pWEB cosmid-cloning kit (Epicentre Technologies, Madison, WI). The library was plated at concentrations to yield ≈300 colonies per plate. The bacteria from each plate were combined, and the complete plasmid DNA isolated from 12 plate pools was screened by diagnostic PCR using the specific primers. Positive pools were plated at numbers of 50 per plate and rescreened. This procedure was repeated once until single positive colonies could be identified. Positive cosmids were sonicated, end-repaired by BAL-31 and Klenow fragment, and size-fractionated by gel electrophoresis to yield fragments of 1- to 2-kb lengths. These fragments were ligated into the *Eco*RV site of pBluescript II SK(−) (Stratagene) and end-sequenced. Remaining gaps were filled by using specifically designed primers and by targeted subcloning. Sequence analysis was performed by using BLASTX, PROSITE, FRAMEPLOT, and the Lasergene DNASTAR software package.

Results

Amplification of Pederin-Specific Polyketide Synthase (PKS) PCR Products. The structure of pederin and early labeling studies (13) indicate that the metabolite is largely synthesized from malonyl-

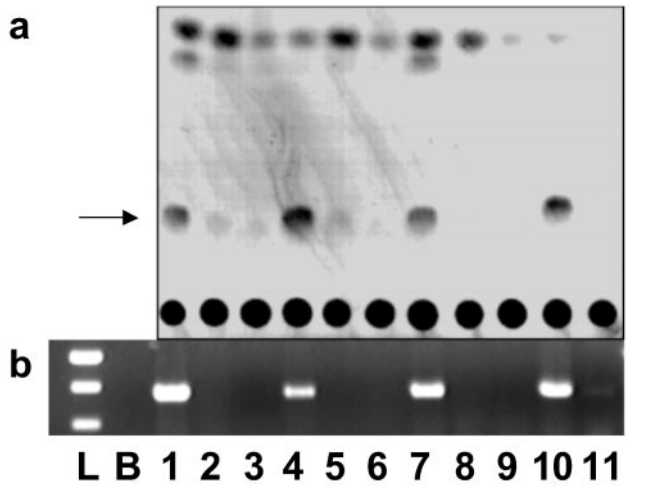

Fig. 2. PCR amplification of PKS gene fragments from beetle total DNA. (*a*) TLC analysis of the ethanolic beetle extracts. The arrow indicates the position of pederin. Small amounts of pederin found in males and (−)-females are due to pederin transfer from (+)-females to the offspring. (*b*) Agarose gel of the PCR products obtained from total DNA of the same beetle specimens that were used for pederin extraction. L, 1-kb DNA ladder; B, negative PCR control without template DNA; 1, *P. fuscipes*, (+)-female, collected at Jena, Germany; 2, *P. fuscipes*, (−)-female, Jena; 3, *P. fuscipes*, male, Jena; 4, *P. fuscipes*, (+)-female, collected at Aydın, Turkey; 5, *P. fuscipes*, (−)-female, Aydın; 6, *P. fuscipes*, male, Aydın; 7, *Paederidus rubrothoracicus*, (+)-female, Aydın; 8, *Pd. rubrothoracicus*, (−)-female, Aydın; 9, *Pd. rubrothoracicus*, male, Aydın; 10, *P. litoralis*, (+)-female, Jena; 11, *P. litoralis*, male, Jena. No (−)-female of *P. litoralis* was available.

and methylmalonyl-CoA units by a type I PKS. Such megasynthases consist of repeated modules, along which the growing polyketide chain is processed in an assembly line-like fashion (14). Each module minimally carries KS, acyltransferase (AT), and acyl carrier protein (ACP) domains to perform one chain-elongation cycle and optional additional domains to catalyze further modifications. To clone the pederin cluster, a PCR strategy was pursued involving degenerate primers based on universally conserved motifs of KS domains (15). Total DNA, isolated from different beetle specimens, was used as a PCR template. Analysis of three species of the genera *Paederus* and *Paederidus* collected at two different locations consistently revealed that only those adult beetles with high pederin content gave the PCR product expected to arise from the presence of PKS genes (Fig. 2). Amplification products were also obtained by using eggs from (+)-females of a fourth species, *Paederus riparius*, collected at Bayreuth, Germany (data not shown). Sequencing of the amplified DNA showed in all cases the exclusive presence of the same group of 3–4 different sequences exhibiting strong homology to KS domains of bacterial type I PKSs. The perfect correlation between pederin content and PKS sequences, independent of species and locality of collection, suggests that the amplified fragments belong to the pederin cluster. These DNA fragments therefore were used to locate the cluster.

Localization of the PKS Gene Cluster on the Symbiont Genome. A metagenomic cosmid library consisting of 80,000 clones was constructed from total DNA of *P. fuscipes* beetles. Initial quantitative PCR experiments had indicated that this library size was sufficient for a multiple coverage of the symbiont genome. By screening the library with specific PCR primers derived from the amplified sequences (see *Materials and Methods*), eight positive cosmids were identified. Sequencing of a 53.8-kb region revealed ORFs homologous to type I PKS genes, designated as

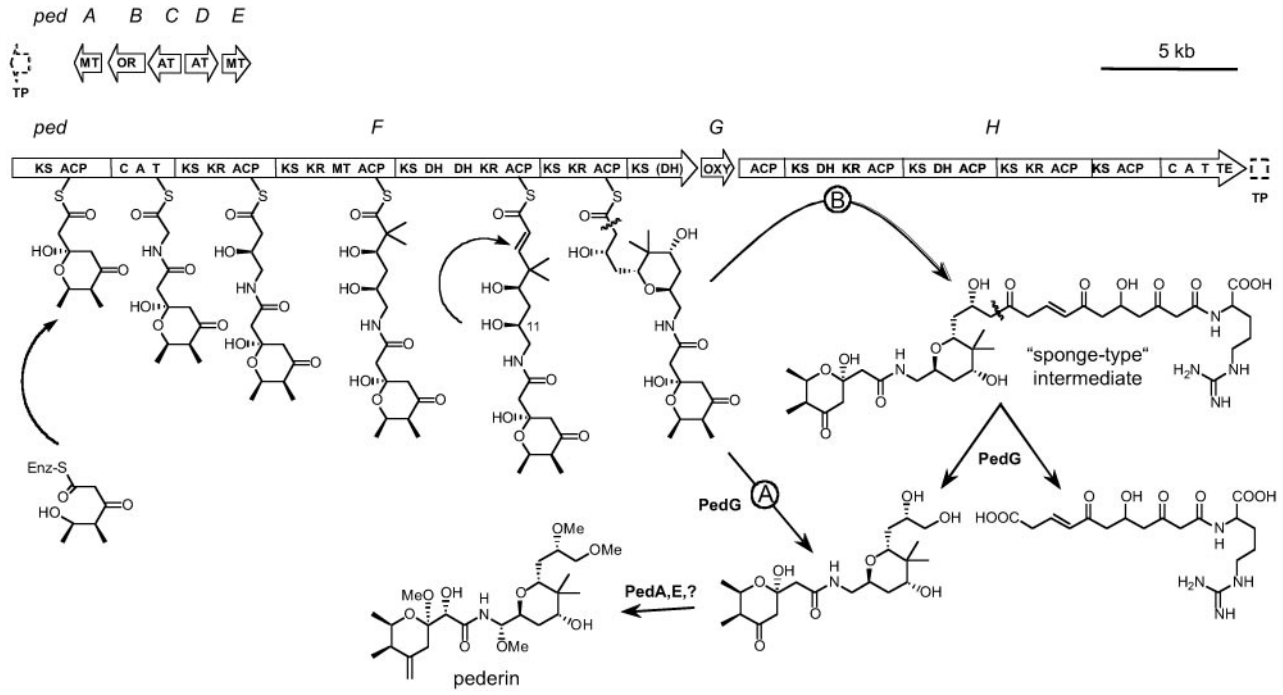

Fig. 3. Map of the sequenced *ped* genes and proposed pederin biosynthesis pathway. The product of PedF could be directly cleaved off the enzyme (route A) or elongated further by PedH to yield an onnamide-type intermediate and subsequently cleaved (route B). TP, transposase pseudogene; MT, methyltransferase; OR, oxidoreductase; OXY, oxygenase; C, NRPS condensation domain; A, NRPS adenylation domain; T, NRPS thiolation domain; DH, putative nonfunctional dehydratase domain.

ped genes (Fig. 3 and Table 1). All the amplified KS sequences could be found on these ORFs. Additional regions covering ≈65 kb outside of the cluster were obtained on two cosmids isolated by chromosome walking and subjected to extensive spot sequencing. All putative genes present on this >110-kb region exhibit typical bacterial features: they are tightly packed, free of introns and polyadenylation sites, and preceded by Shine–Dalgarno patterns in appropriate distances to the start codons. Furthermore, when subjected to database homology searches, each of the translated ORFs exhibited the highest similarity to bacterial proteins. Of the 68 deduced proteins analyzed, 27 are known exclusively from prokaryotes including enzymes used in vitamin B_6 and B_{12} biosynthesis and regulatory proteins of the LuxR and LysR families. Interestingly, most ORFs located outside of the *ped* region, including almost all housekeeping genes, exhibit an extraordinary similarity to genes from *P. aeruginosa*, with an identity of >80% at the nucleotide level. This result is in perfect agreement with the 16S rRNA data (11) and provides clear evidence that the *ped* cluster belongs to a bacterial symbiont of the genus *Pseudomonas*.

Domain Analysis of the *ped* Cluster. Fig. 3 summarizes the results of the completely sequenced 53.8-kb *ped* region. The predicted gene products of the ORFs *pedF* and *pedH* are giant 8,601- and 6,266-aa proteins resembling mixed modular PKSs/nonribosomal peptide synthetases (NRPSs) (16, 17). In striking contrast to nearly all known PKSs, AT homologies are completely absent on these proteins. Their alignment with other known type I PKSs revealed that a continuous ≈300-aa region of each AT domain,

Table 1. Deduced functions of the *ped* ORFs shown in Fig. 3

Protein	Amino acids	Proposed function	Sequence similarity (protein, origin)	Identity/ similarity, %	Accession no.
PedA	312 or 323 from an alternative start codon	*O*-methyltransferase	AdpE, *Anabaena* sp. 90	44/62	CAC01607
PedB	464	FMN-dependent oxidoreductase	MmpIII (OR domain), *Pseudomonas fluorescens*	54/73	AAM12912
PedC	334	Acyltransferase	MmpIII (AT domain), *P. fluorescens*	38/55	AAM12912
PedD	364	Acyltransferase	PksC, *Bacillus subtilis*	52/68	CAB13582
PedE	273	*O*-methyltransferase	SnogM, *Streptomyces nogalater*	38/55	AAG42853
PedF	8,601	PKS/NRPS	PksK, *B. subtilis*	≈40/50 module homology	AAA85144
PedG	285	FAD-dependent monooxygenase	SCO0122, *Streptomyces coelicolor*	33/48	CAB52349
PedH	6,266	PKS/NRPS	PksP, *B. subtilis*	≈40/50 module homology	CAB13605

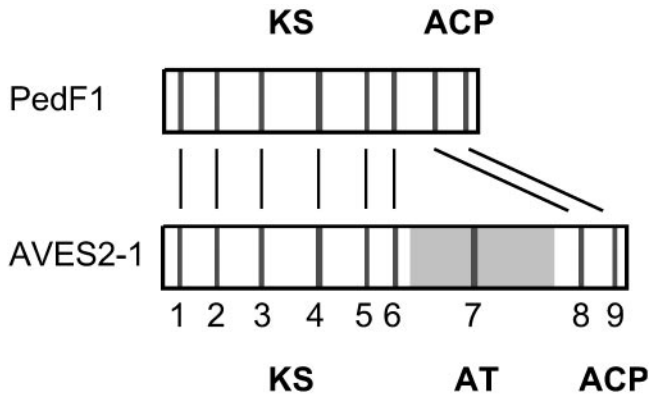

Fig. 4. Comparison of the first modules of PedF and the avermectin AVES2. The shaded area in AVES2-1 indicates the AT region that is deleted in PedF1. Conserved motifs are shown as vertical bars. Motifs are: 1, EPIAIV; 2, DPQQRL; 3, CSSS; 4, HGTGTxLGD; 5, GxGGxNAHVILEE; 6, AYTLQxGR; 7, GHSxG; 8, YPF; 9, GxDS. ACP, acyl carrier protein.

including the active-site GHS motif, are deleted in PedF and PedG, with no other homologies replacing them (Fig. 4). This architecture results in a considerable shortening of each module. The missing AT domains are possibly complemented by the two isolated genes *pedC* and *pedD* upstream of *pedF*, which exhibit high similarity to monodomain ATs. Apart from the lacking ATs, however, PedF is perfectly colinear with pederin structure. A particularly characteristic feature is the second module (PedF2) with homology to NRPSs, catalytic entities involved in the attachment of a wide range of proteinogenic and nonproteinogenic amino acids. Previous studies have identified sequence patterns of NRPS adenylation (A) domains that allow a prediction of the incorporated amino acid (18, 19). The structure of pederin suggests that glycine is selected by PedF2. In accordance, a pattern analysis of its A domain revealed 100% identity to the known consensus nonribosomal code for this amino acid (ref. 20; Table 2). A second distinctive property of PedF is the rare methyltransferase (MT) domain of the PKS module PedF4. MTs integrated into PKS modules are known from yersiniabactin (21) and epothilone (22) biosynthesis, where they generate geminal dimethyl groups. Such a moiety is present also at the corresponding position of the pederin structure. A third notable feature is an unprecedented repeated dehydratase (DH) domain in the next module, PedF5. This module should be responsible for the closure of the dimethylated tetrahydropyrane ring, most likely via dehydration and subsequent Michael addition of the C-11 hydroxyl group, which is the reverse reaction. The catalysis of the identical transformation sequence by two DH copies is known also from nonactin biosynthesis, although the DH performing the addition step is encoded on a separate gene, and the product is a five-membered ring (23). Taken together, the architecture of PedF clearly indicates that this protein assembles the largest part of the pederin molecule except for the six-membered ring bearing the exomethylene group. A corresponding gene candidate responsible for the biosynthesis of this moiety could not be identified on the cluster.

Downstream of *pedF* lies *pedH*, encoding four additional PKS modules and one NRPS module. The presence of this gene was surprising at first, because it does not correspond to any part of the pederin structure. However, a detailed module analysis of *pedH* in fact revealed further substantial evidence for an involvement of the cluster in pederin biosynthesis. The terminal NRPS module PedH6 features a nonribosomal code most similar to the patterns of the arginine-specific syringomycin (24) and microcystin adenylation domains (ref. 25; Table 2) and contains a thioesterase domain invariably involved in release of assembled polyketide and peptide chains from the enzymatic machinery. PedH therefore should theoretically further elongate the pederin-type intermediate generated by PedF and release a compound having the hypothetical extended structure shown in Fig. 3 (route B). The remarkable similarity to most members of the pederin group isolated from sponges is obvious. These compounds include the 13 known onnamides and icadamide A, which all feature a variable polyketide side chain and a terminal arginine residue.

Tailoring and Accessory Genes. The cluster contains a number of genes in addition to the PKS/NRPS ORFs. The *pedA* and *pedE* products are similar to methyltransferases involved in the methylation of polyketide and peptide heteroatoms. Because four hydroxyl groups of pederin are methylated, the presence of such genes was expected. PedB exhibits good homology to oxidoreductases that use flavin mononucleotide (FMN) as cofactor. Interestingly, all related proteins are either parts of type I PKS systems lacking AT domains (see *Discussion*) or suspected enoyl-CoA reductases of the recently discovered bacterial polyunsaturated fatty acid biosynthesis clusters (26, 27). The closest homologue is MmpIII, a bifunctional oxidoreductase/AT from the mupirocin cluster. The ubiquitous presence of these redox enzymes in such clusters suggests that PedB is involved in important processes at the early biosynthesis stage such as the supply of small molecule precursors. Another putative redox enzyme is the deduced product of *pedG*, which resembles FAD-dependent monooxygenases. Its location between the two module fragments of *pedF* and *pedH* is unusual and raises the possibility that the module halves resulted from an insertion event of *pedG* into a formerly intact module.

Discussion

Although the existence of bacterial symbionts as producers of invertebrate drug candidates has been proposed for many years, the general inability to grow such bacteria outside the host has prevented a direct proof of this hypothesis thus far. Our current view on such symbioses has been shaped largely by cell-separation studies conducted by Bewley and Faulkner (3), who showed that various sponge natural products in fact are localized in associated bacterial cells. However, these results do not rule out the possibility that bacteria are just importing metabolites

Table 2. Prediction of the amino acid specificity of the PedF2 and PedH6 NRPS modules based on known nonribosomal codes

Module	235	236	239	278	299	301	322	330	331	517	Similarity, %
Gly consensus (20)	**D**	**I**	**L**	**Q**	**L**	**G**	**L**	**I**	**W**	**K**	
PedF2	D	I	L	Q	L	G	L	I	W	K	100
Arg consensus	**D**	**V**	**X**	**D/T**	**I/V**	**G**	**A**	**I/V**	**D**	**K**	
PedH6	D	A	E	D	I	G	A	I	T	K	70
SyrE5	D	V	A	D	V	G	A	I	D	K	90
McyC	D	V	W	T	I	G	A	V	D	K	90

synthesized by their hosts. To unequivocally identify an as-yet-uncultured bacterium as producer, the biosynthetic capabilities have to be examined at the molecular level. Gene cloning and heterologous expression studies, besides providing a glimpse into the metabolism of an otherwise unaccessible microorganism, should also unveil mechanisms of symbiosis and provide long-term supplies of rare invertebrate drug candidates (28). With this intention pederin biosynthesis in *Paederus* beetles was selected as a model for chemically mediated symbiosis. This singular terrestrial system has several advantages over the many suspected marine associations: an unparalleled amount of evidence that symbionts indeed exist (10), the likely presence of an almost pure bacterial culture as revealed by 16S rRNA analysis (11), and the unique opportunity to subsequently access highly complex and otherwise hard-to-study sponge systems by direct comparison of terrestrial and marine pederin-type gene clusters.

These cloning efforts led to the isolation of a contiguous 110-kb genomic region, which also was sequenced outside of the *ped* cluster to gain insight into the biology of the pederin producer. Because most ORFs, significantly those from primary metabolism, remarkably resemble genes of *P. aeruginosa*, the symbiont should be a very close relative of this species. This classification is supported also by recently conducted 16S rRNA gene studies (11) that revealed the presence of a bacterium with closest similarity to *P. aeruginosa* only in the ''pederin-producing'' (+)-females. *P. aeruginosa* is an opportunistic pathogen that causes serious and often fatal infections in immunocompromised and cystic fibrosis patients. It is virulent in a remarkably large variety of organisms including insects, which are used therefore as model infection hosts (29). An ancient infection event followed by genetic adaptation of the partners also may have resulted in the establishment of pederin symbiosis. Regardless of the evolutionary events bringing forth this association, the close relationship between the two bacteria now allows a direct comparison of symbiont genes with the completely sequenced *P. aeruginosa* genome. This makes the pederin symbiont a unique model organism not only to identify candidate genes governing symbiosis but also to help pinpointing pathogenic determinants in *P. aeruginosa*. Horizontal gene transfer allows bacteria to evolve in quantum leaps and is therefore the major driving force of bacterial adaptation. The importance of horizontally transferred genomic islands for the development of virulent or symbiotic traits in bacteria is revealed by an increasing number of studies (30). One cluster of genes that has no counterpart in *P. aeruginosa* and seems to have been horizontally acquired is the *ped* region. It is bordered on both sides by transposase pseudogenes, indicating that it has been inserted into the genome during a transposition event. This region indeed bears all the architectural and functional hallmarks of mobile symbiosis islands, which could explain the widespread occurrence of pederin-type metabolites in diverse organisms.

An unexpected finding was the complete absence of AT domains in the *ped* cluster. In modular PKSs, these domains are crucial for the selection of the correct acyl-CoA unit in each extension cycle and their attachment to the acyl carrier protein domain (31), raising the question of how this process is catalyzed during the assembly of pederin. PedC and PedD, encoded upstream of *pedF* and resembling monodomain ATs, could perform this function in *trans*. The *ped* cluster then would encode a type I protein complemented by repetitive type II enzymes. Recent studies indicate that such unusual mixed type I/type II PKS systems are more widespread in nature. A similar PKS containing such putative ''super ATs'' is the only partly described *pksX* cluster (32) on the genome of *Bacillus subtilis*. Its gene products consist of a large number of PKS modules without AT domains and, encoded at the upstream end of the cluster, three isolated ATs. The secondary metabolite generated by these proteins is not known. Another related system is the mupirocin PKS from *Pseudomonas fluorescens* recently submitted to GenBank (accession no. AF318063; A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, and C. M. Thomas, unpublished data). This cluster contains a single gene with two AT domains. Apart from these two systems, a small number of further PKS modules lacking AT domains are known from the albicidin (33) and TA biosynthesis gene clusters (34). Both systems have only been analyzed in part thus far, but it is very likely that they also contain discrete AT genes. Sequence comparison with other known PKSs (data not shown) reveals that all enzymes lacking AT domains are related more closely to each other than to any other known modular PKS, independent of bacterial phylogeny. They therefore could belong to a phylogenetically distinct subgroup of functionally novel PKS systems.

To ultimately prove the involvement of the *ped* genes in pederin biosynthesis, loss-of-function studies after gene inactivation are needed. Naturally, the inability to culture the *Paederus* symbionts forbids such an approach. However, several lines of evidence provide compelling evidence that the cluster indeed encodes the pederin pathway. First, there is a strict correlation between the ability to synthesize pederin and the presence of identical PKS PCR products. These products are present exclusively in (+)-females and their eggs and are amplified from a wide range of species regardless of the locality of collection. Second, the fact that all PCR products belong to a single cluster argues against the possibility that pederin biosynthesis is encoded on another, undetected PKS cluster. The degenerate PCR primers were not designed originally for the amplification of sequences from aberrant PKS clusters lacking AT domains. Nevertheless, 4 of a possible 10 PCR products from the *ped* system were obtained, documenting the broad specificity of these primers for modular PKSs. It seems unlikely that a second large cluster comprised of at least eight PKS modules would not produce any amplification products. Third and most important, the architecture of the *ped* cluster perfectly matches the hypothetical biosynthetic pathway of pederin-type compounds (Fig. 3). Every single domain of PedF agrees with the corresponding pederin moiety, which includes diagnostically valuable and rare features such as the NRPS module with glycine nonribosomal code, the methyltransferase domain presumably attaching the geminal dimethyl group, and the repeated dehydratase domain most likely responsible for the pyran ring formation. Also in agreement with pederin biosynthesis are the post-PKS tailoring genes. Pederin contains four *O*-methyl groups, of which at least two could be attached by the methyltransferases PedA and PedE. One of the most distinctive features of pederin, however, is the oxidized single-carbon terminus that is extremely rare among nonrelated polyketides. This moiety should result from an oxidative cleavage of a two-carbon extension unit. Strikingly, *pedG* encoding an FAD-dependent monooxygenase is inserted precisely behind the last complete module corresponding to this unit. Members of this enzyme family are known to catalyze oxidative cleavages in a Baeyer–Villiger fashion and include MtmOIV, which cleaves a precursor of the aromatic polyketide mithramycin (35), and cyclohexanone monooxygenase (36). PedG therefore seems to be a good candidate to perform the proposed pederin cleavage.

An interesting issue is the function of the additional PKS-NRPS gene *pedH*. Although the lack of a structural counterpart to this gene at first glance seems puzzling, a more detailed analysis of its architecture strongly suggests its involvement in the synthesis of a pederin-type metabolite. The striking and unexpected similarity of the deduced full-length biosynthetic product of PedF and PedH to the large series of marine pederin analogues with terminal arginine residues raises fascinating new questions about the evolutionary relationship between marine and terrestrial pederin members. Fig. 3 shows two possible mechanistic hypotheses for pederin assembly. It could be envi-

sioned that the *ped* cluster originally generated a sponge-type compound but that in the course of evolution the oxygenase gene *pedG* inserted into the cluster, thereby destroying one module (route A). Consequently, this would lead to a premature termination of the PKS elongation process at the module PedF6, which is exactly the module corresponding to the terminal pederin extension unit. Because a thioesterase domain is lacking on this gene, the chain would remain attached to this module until it is cleaved off oxidatively by PedG. This scenario would not only explain the module halves on both sides of *pedG* but also why *pedG* is located at exactly this position of the cluster. Alternatively, the *ped* system could first synthesize an extended sponge-type intermediate, which then is cleaved by PedG in a Baeyer–Villiger oxidation to give an appropriately shortened pederin precursor and a second fragment (route B). If this mechanism is correct, then the fragmented module should be functional. Such subdivided but functional PKS modules are unusual but indeed exist, as a single known example from the myxalamid cluster documents (37).

A number of genes required for pederin biosynthesis were not found on the cluster. Missing are a small PKS comprising the first three modules, a protein necessary for the attachment of the exomethylene group, and the enzyme catalyzing the formation of the aminal moiety. Although in bacteria genes from secondary metabolism are typically clustered together, there are a few known exceptions to this rule. Examples of incomplete clusters include the ansamitocin PKS genes, which are located on two regions separated by 30 kb of nonrelated genes (38), and the enterocin cluster, which lacks a dehydrogenase participating in the formation of the polyketide starter unit (39). The presence of transposase pseudogenes at both sides of the *ped* region strongly suggests that the cluster has been fragmented during genomic rearrangement processes.

This work provides insight into the biosynthesis of a structurally unique insect defense compound and the evolutionary relationship to the related sponge metabolites. It also shows for the first time that genes responsible for the biosynthesis of rare invertebrate drug candidates can be cloned from bacterial symbionts without the need for culturing. Because whole sets of type I PKS genes have been functionally expressed even in *Escherichia coli* (40), a similar production of "invertebrate" natural products in a suitable host now has become a realistic scenario.

I thank R. L. L. Kellner for providing eggs of *P. riparius* and valuable advice; H. Başpınar and E. Şavk for helping with the collections of Turkish beetle specimens; I. Höfer for technical assistance; U. Hentschel for suggestions on the manuscript; and W. Boland and B. S. Moore for support and helpful discussions. This work was supported by Deutsche Forschungsgemeinschaft Research Grant PI 430/1-1.

1. Faulkner, D. J. (2000) *Nat. Prod. Rep.* **17,** 1–6.
2. Munro, M. H. G., Blunt, J. W., Dumdei, E. J., Hickford, S. J. H., Lill, R. E., Li, S. X., Battershill, C. N. & Duckworth, A. R. (1999) *J. Biotechnol.* **70,** 15–25.
3. Bewley, C. A. & Faulkner, D. J. (1998) *Angew. Chem. Int. Ed. Engl.* **37,** 2163–2178.
4. Jensen, P. R. & Fenical, W. (1994) *Annu. Rev. Microbiol.* **48,** 559–584.
5. Kobayashi, J. & Ishibashi, M. (1993) *Chem. Rev. (Washington, DC)* **93,** 1753–1769.
6. Narquizian, R. & Kocienski, P. J. (2000) in *The Role of Natural Products in Drug Discovery* (Springer, Heidelberg, Germany), Vol. 32, pp. 25–56.
7. Hood, K. A., West, L. M., Northcote, P. T., Berridge, M. V. & Miller, J. H. (2001) *Apoptosis* **6,** 207–219.
8. Borroni, G., Brazzelli, V., Rosso, R. & Pavan, M. (1991) *Am. J. Dermatopathol.* **13,** 467–474.
9. Kellner, R. L. L. & Dettner, K. (1995) *J. Chem. Ecol.* **21,** 1719–1733.
10. Kellner, R. L. L. (2001) *J. Insect Physiol.* **47,** 475–483.
11. Kellner, R. L. L. (2002) *Insect Biochem. Mol. Biol.* **32,** 389–395.
12. Şendur, N., Şavk, E. & Karaman, G. (1999) *Dermatology* **199,** 353–355.
13. Cardani, C., Fuganti, C., Ghiringh, D., Grassell, P., Pavan, M. & Valcuron, M. (1973) *Tetrahedron Lett.* **30,** 2815–2818.
14. Staunton, J. & Weissman, K. J. (2001) *Nat. Prod. Rep.* **18,** 380–416.
15. Beyer, S., Kunze, B., Silakowski, B. & Müller, R. (1999) *Biochim. Biophys. Acta* **1445,** 185–195.
16. Doekel, S. & Marahiel, M. A. (2001) *Metabol. Eng.* **3,** 64–77.
17. Cane, D. E. & Walsh, C. T. (1999) *Chem. Biol.* **6,** R319–R325.
18. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. (1999) *Chem. Biol.* **6,** 493–505.
19. Challis, G. L., Ravel, J. & Townsend, C. A. (2000) *Chem. Biol.* **7,** 211–224.
20. Du, L. C., Sanchez, C., Chen, M., Edwards, D. J. & Shen, B. (2000) *Chem. Biol.* **7,** 623–642.
21. Miller, D. A., Walsh, C. T. & Luo, L. S. (2001) *J. Am. Chem. Soc.* **123,** 8434–8435.
22. Molnar, I., Schupp, T., Ono, M., Zirkle, R. E., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D. D., *et al.* (2000) *Chem. Biol.* **7,** 97–109.
23. Woo, A. J., Strohl, W. R. & Priestley, N. D. (1999) *Antimicrob. Agents Chemother.* **43,** 1662–1668.
24. Zhang, J. H., Quigley, N. B. & Gross, D. C. (1995) *J. Bacteriol.* **177,** 4009–4020.
25. Tillett, D., Dittmann, E., Erhard, M., von Dohren, H., Borner, T. & Neilan, B. A. (2000) *Chem. Biol.* **7,** 753–764.
26. Metz, J. G., Roessler, P., Facciotti, D., Levering, C., Dittrich, F., Lassner, M., Valentine, R., Lardizabal, K., Domergue, F., Yamada, A., *et al.* (2001) *Science* **293,** 290–293.
27. Kaulmann, U. & Hertweck, C. (2002) *Angew. Chem. Int. Ed. Engl.* **41,** 1866–1869.
28. Davidson, S. K., Allen, S. W., Lim, G. E., Anderson, C. M. & Haygood, M. G. (2001) *Appl. Environ. Microbiol.* **67,** 4531–4537.
29. Jander, G., Rahme, L. G. & Ausubel, F. M. (2000) *J. Bacteriol.* **182,** 3843–3845.
30. Hentschel, U., Steinert, M. & Hacker, J. (2000) *Trends Microbiol.* **8,** 226–231.
31. Haydock, S. F., Aparicio, J. F., Molnar, I., Schwecke, T., Khaw, L. E., Konig, A., Marsden, A. F. A., Galloway, I. S., Staunton, J. & Leadlay, P. F. (1995) *FEBS Lett.* **374,** 246–248.
32. Albertini, A. M., Caramori, T., Scoffone, F., Scotti, C. & Galizzi, A. (1995) *Microbiology* **141,** 299–309.
33. Huang, G. Z., Zhang, L. H. & Birch, R. G. (2001) *Microbiology* **147,** 631–642.
34. Paitan, Y., Alon, G., Orr, E., Ron, E. Z. & Rosenberg, E. (1999) *J. Mol. Biol.* **286,** 465–474.
35. Prado, L., Fernandez, E., Weissbach, U., Blanco, G., Quiros, L. M., Brana, A. F., Mendez, C., Rohr, J. & Salas, J. A. (1999) *Chem. Biol.* **6,** 19–30.
36. Stewart, J. D. (1998) *Curr. Org. Chem.* **2,** 195–216.
37. Silakowski, B., Nordsiek, G., Kunze, B., Blocker, H. & Müller, R. (2001) *Chem. Biol.* **8,** 59–69.
38. Yu, T. W., Bai, L. Q., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K. Q., Xu, J., Moss, S. J., Leistner, E. & Floss, H. G. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 7968–7973.
39. Piel, J., Hertweck, C., Shipley, P. R., Hunt, D. M., Newman, M. S. & Moore, B. S. (2000) *Chem. Biol.* **7,** 943–955.
40. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. (2001) *Science* **291,** 1790–1792.