



**New insights into the echinocandins and other fungal non-ribosomal peptides and peptaibiotics**

Journal:	<i>Natural Product Reports</i>
Manuscript ID:	NP-REV-04-2014-000046.R1
Article Type:	Review Article
Date Submitted by the Author:	05-Jul-2014
Complete List of Authors:	Bills, Gerald; The University of Texas Health Science Centre, Texas Therapeutics Institute Li, Yan; The University of Texas Health Science Centre, Texas Therapeutics Institute Chen, Li; The University of Texas Health Science Centre, Texas Therapeutics Institute Yue, Qun; The University of Texas Health Science Centre, Texas Therapeutics Institute Niu, Xuemei; Yunnan University, Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of the Ministry of Education An, Zhiqiang; The University of Texas Health Science Centre, Texas Therapeutics Institute

## REVIEW

## New insights into the echinocandins and other fungal non-ribosomal peptides and peptaibiotics

Cite this: DOI: 10.1039/x0xx00000x

Gerald Bills<sup>a,\*</sup>, Yan Li<sup>a</sup>, Li Chen<sup>a</sup>, Qun Yue<sup>a</sup>, Xue-Mei Niu<sup>a,b</sup> and Zhiqiang An<sup>a</sup>

Received 00th January 2012,  
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Non-ribosomal peptide synthetases (NRPSs) are a primary modality for fungal peptidic natural product assembly and are responsible for some of the best known, most useful, and most destructive fungal metabolites. Through genome sequencing and computer-assisted recognition of modular motifs of catalytic domains, one can now confidently identify most NRPS biosynthetic genes of a fungal strain. The biosynthetic gene clusters responsible for two of the most important classes of NRP fungal derived drugs, cyclosporine and the echinocandins, have been recently characterized by genomic sequencing and annotation. Complete biosynthetic gene clusters for the pneumocandins and echinocandins have been mapped at the genetic level and functionally characterized to some extent. Genomic sequencing of representative strains of most of the variants in the echinocandin family, including the wild-type of the three fungal strains employed for industrial-scale production of caspofungin, micafungin and anidulofungin, has enabled characterization of the basic architecture of the echinocandin NRPS pathways. A comparative analysis of how pathway genes cause variations in lipoinitiation, biosynthesis of the non-proteinogenic amino acids, amino acid substitutions, and hydroxylations and sulfonations of the core peptide and contribute to the molecular diversity of the family is presented. We also review new information on the natural functions of NRPs, the differences between fungal and bacterial NRPSs, and functional characterization of selected NRPS gene clusters. Continuing discovery of the new fungal nonribosomal peptides has contributed new structural diversity and potential insights into their biological functions among other natural peptides and peptaibiotics. We therefore provide an update on new peptides, depsipeptides and peptaibols discovered in the Fungi since 2009.

- 1 Introduction and scope of the review
- 2 Natural functions of non-ribosomal peptides in fungi
- 3 Differences between fungal and bacterial non-ribosomal peptide synthesis
- 4 New understanding of the biosynthesis of pneumocandins and echinocandins
  - 4.1 Pathway discovery
  - 4.2 Structure and gene order conservation of the echinocandin gene clusters
  - 4.3 Evolution of the echinocandin gene clusters
- 5 Other recently characterised NRPS gene clusters
  - 5.1 The apicidin gene cluster and its relationship to HC-toxin
  - 5.2 Other NRPS biosynthetic gene clusters
- 6 New peptides and depsipeptides and their bioactivities
  - 6.1 Peptides
  - 6.2 Depsipeptides
  - 6.3 Peptaibols
- 7 Potential for novel peptides in newly sequenced and annotated fungal genomes

- A adenylation
- C condensation
- E Epimerization
- HDAC histone deacetylase
- HPLC high performance liquid chromatography
- HST host-specific toxin
- MW molecular weight
- NRP nonribosomal peptide
- NRPS nonribosomal peptide synthetase
- PCP peptidyl carrier protein
- ROS reactive oxygen species
- TE thioesterase
- T thiolation

### 1 Introduction and scope of the review

Nonribosomal peptide synthetases are a primary modality for fungal peptidic natural product assembly and are responsible for some of the best known, most useful, and most destructive fungal metabolites. Fungal peptides, depsipeptides, and peptaibiotics have been comprehensively reviewed by Degenkolb<sup>1</sup> and by Anke and Antelo in the *The Mycota*.<sup>2</sup> An in-depth general review of recent discoveries and breakthroughs in the

### Abbreviations

aa amino acid(s)

structural elucidation, molecular mechanism, and chemical biology underlying the discrete domains within NRPSs appeared in 2012,<sup>3, 4</sup> while a synthesis focusing on the consequences of: 1) absence of C domains in NRPSs; and 2) predictions of aa specificity has recently appeared.<sup>5, 6</sup> A recent review on fungal cyclooligomer depsipeptides<sup>7</sup> focussed primarily on the biosynthetic mechanisms involved in assembly of beauvericin, enniatins, hirsutellide A, bassianolide, PF1022, verticillide and the diketomorpholines, bassiatin and lateritin. The biosynthesis of complex *Aspergillus* and *Penicillium* peptidyl alkaloids from anthranilate and tryptophan monomers via iterative bimodular and trimodular gene clusters has been recently reviewed<sup>8, 9</sup> and will be further covered in a companion review by Tang (ref). NRPSs are one of the main chemical weapons of entomopathogenic fungi,<sup>10, 11</sup> and new data on NRPS biosynthesis and functions in these fungi appear in the review by Gibson in this issue (ref).

Our understanding about how fungi biosynthesize non-protein peptides is rapidly expanding, and new insights from newly characterised genomes and biosynthetic gene clusters continually modify previous viewpoints. Through comprehensive genome sequencing and computer-assisted recognition of modular motifs of catalytic domains one can now confidently identify most NRPS biosynthetic genes of a fungal strain.<sup>5, 6, 12</sup> So far as known, fungal NRPS genes tend to be clustered<sup>11</sup> or occur as partial NRPSs embedded as single modules in hybrid biosynthetic genes with polyketide or prenyl transferases.<sup>13-17</sup> The established methods of cloning and genome walking for characterisation of each cluster region have largely become obsolete. Those methods have been superseded by genome sequencing technologies that enable facile and comprehensive analysis of an organism's core biosynthetic gene clusters while including identifications of previously unknown genes in the vicinity of these loci.

The biosynthetic gene clusters responsible for two of the most important classes of NRP fungal derived drugs have been recently characterised by genomic sequencing and annotation. The pathway for echinocandin B in *Aspergillus rugulosus* has recently been identified through genome scanning,<sup>18, 19</sup> while our group recently published the results of the genomic sequencing of *Glarea lozoyensis*, the fungus producing pneumocandin B<sub>0</sub>, the starting metabolite for the three-step manufacturing synthesis of the antifungal drug CANCIDAS.<sup>20</sup> Complete biosynthetic gene clusters for the pneumocandins and echinocandins have been mapped at the genetic level, and functionally characterised to some extent, although many questions remain regarding the regulatory and mechanistic processes at the enzymatic level and as well as the evolutionary origins of the echinocandin gene clusters from within the Fungi.

The genomic sequence of *Tolypocladium inflatum* and RNA-Seq transcriptome analysis has recently defined the complete gene cluster for cyclosporine biosynthesis.<sup>21</sup> Unlike the echinocandins, the formidable cyclosporine NRPS simA seems to have evolved via duplication of simpler NRPSs that are apparently shared among other fungi of the Hypocreales. A domains from several other NRPSs responsible for smaller cyclic depsipeptides in Hypocrealean fungi, such as enniatin, beauvericin and bassianolide synthetases, the aureobasidin A NRPS and two modules of destruxin synthetase appear to form a monophyletic lineage within the SimA clade.<sup>22</sup> We will also review the recent characterisations of some NRPS gene clusters, e.g., the pathways responsible for apicidin and trichovirins. Discovery of the new fungal nonribosomal peptides continues to contribute new structure diversity and

potential insights into their biological functions among other natural peptides and peptaibiotics. We provide an update on new peptides and depsipeptides from the Fungi since 2009.

## 2 Natural functions of non-ribosomal peptides in fungi

Ribosome-independent synthesized peptides and depsipeptides likely mediate a multitude of functions in the microbial cell and its environment, but most functions remain obscure. Some of the best studied functions in fungi are those of extracellular and intracellular siderophores involved in iron transport, storage and homeostasis and plant and animal virulence.<sup>23-25</sup> Fungi generally produce hydroxamate-type siderophores. Siderophore-encoding NRPSs do not follow the typical NRPS co-linearity rule, but instead iteratively use of A domains, activating an identical aa more than once in an NRPS catalytic cycle, and employ additional T-C domains as partial modules to extend the NRP products beyond the number of complete A-T-C modules in the NRPSs.<sup>23</sup> Their synthesis starts with the N5-hydroxylation of L-ornithine catalysed by an L-ornithine-N5-monoxygenase and then is followed by an acylation to form the hydroxamate group; an NRPS assembles the final siderophore by linkage of hydroxamate groups, or other aa in the case of ferrichrome-type siderophores. Fungal siderophores are categorized into four structural types, ferrichromes, fusarinines, coprogens and rhodotorulic acid.<sup>23, 26</sup> Fungal siderophores are believed to have arisen before the Ascomycete and Basidiomycete lineages diverged. Aligned A domains of siderophore NRPSs form a well-defined clade that includes both basidiomycete and ascomycete ferrichrome synthetases.<sup>22, 24</sup> Other non-ribosomal cyclic depsipeptides, e.g., beauvericin, enniatins, isariins, pithiomycolide, and PF1022A also have structures with general cation chelating and ionophore activities.<sup>27</sup>

Siderophores can be virulence factors in animal pathogenesis.<sup>25, 28</sup> Silencing of SidA (L-ornithine N5-oxygenase) in *Aspergillus fumigatus* prevents infections in murine models of pulmonary aspergillosis. *Aspergillus fumigatus* produces four siderophores: fusarinine C encoded by NRPS SidD and used for extracellular iron acquisition; triacetylfusarinine C which is derived via fusarinine C via SidG (N5-hydroxyornithine-acetylase) and used for extracellular iron acquisition; ferricrocin which is used for intracellular iron storage, and hydroxyferricrocin which is used for conidial iron acquisition and storage during infection, and both are encoded by SidC. Another *A. fumigatus* peptide synthetase gene, *sidE*, encodes a bimodular NRPS with 2,109 aa that is located near the *sidG* locus. Thus SidE was anticipated to be involved in siderophore production; however, unlike other siderophore encoding genes, its expression depended on LaeA, a positive regulator of virulence and secondary metabolism in *A. fumigatus*. Expression of *sidE* was upregulated during iron depletion, high-iron conditions, exposure to H<sub>2</sub>O<sub>2</sub>, and high temperatures. The *sidE* was up-regulated 16-fold in nascent *Aspergillus* infections in the neutropenic murine lung and therefore, was believed to be involved in virulence. Both *in vitro* and comparative *in vivo* studies under heat-stressed growth determined that fumarylalanine (**1**) was the pathway product. SidE orthologs appear to form a unique lineage of fungal NRPSs that are common among other fungi. The authors suggested that fumarylalanine (**1**) might be involved in local immunosuppression and related to the *A. fumigatus* infective process. Another virulence-associated pathway in *A. fumigatus* is the *has* pathway.<sup>29</sup> The *has* is an eight-gene cluster that encodes putative transcription factors, HasA and HasF, a

transporter HasB, an O-methyltransferase HasC, a NRPS HasD, a 7-dimethylallyltryptophan synthase HasE, a putative FAD-binding protein HasG, and a cytochrome P450 HasH. The pathway's product was identified as a tryptophan-derived iron(III)-complex, hexadecahydro-astechrome (2). The authors speculated that the precursor of this metabolite could function as a siderophore, or compound 2 could act as oxidative virulence factor by releasing astechrome monomers into the host tissue.

Structures 1 to 18 here

Siderophores play a role in fungus-plant interactions. Coprogens encoded by NPS6 in *Cochliobolus heterostrophus* enhance virulence to maize and resistance to oxidative stress.<sup>30</sup> Orthologues of these genes with conserved functions appear to be widely distributed among filamentous Ascomycetes. AaNPS6 of *Alternaria alternata* is responsible for dimethyl coprogen siderophores and was necessary for siderophores production and virulence on citrus.<sup>31</sup> *Magnaporthe grisea* also excretes coprogen-type siderophores for iron acquisition while employing ferricrocin for intracellular iron storage.<sup>32</sup> Disruption of the SSM2 NRPS gene resulted in a loss of the production of coprogens, slower growth rate, fewer conidia and increased sensitivity to oxidative stress. The complete pathway for the siderophore ferrichrome A (3) has been characterised in the maize smut fungus, *Ustilago maydis*. The pathway consists of four genes (*fer3*, *fer4*, *fer5* and *hcs1*) where an enoyl-CoA hydratase Fer4 converts HMG-CoA to methylglutaconyl-CoA, the acylase Fer5 transfers the methylglutaconyl-moiety of methylglutaconyl-CoA to hydroxyornithine to generate methylglutaconyl hydroxyornithine that was then used by NRPS Fer3 as the monomer to generate ferrichrome A (3).<sup>33</sup> The gene *fsn1* was isolated from the sugar-cane pathogen *Fusarium sacchari* and shows the same domain architecture as other SidC proteins and encodes a 4707-aa NRPS consisting of three complete A-T-C modules followed by two additional T and C domain repeats. This structure is similar to that of ferricrocin synthetase, which makes a siderophore that is involved in intracellular iron storage in other filamentous fungi.<sup>34</sup> The complete FSN1 gene was reassembled under the control of the starch- and maltose-inducible *amyB* promoter and was transformed into *Aspergillus oryzae* which resulted in the accumulation ferrirhodin (4). The same siderophore was then detected in both the mycelium and culture filtrates of *F. sacchari*, while ferricrocin was only found in the mycelium indicating that ferricrocin (5) was an intracellular storage siderophore while ferrirhodin was secreted for iron acquisition.<sup>34</sup> This example constituted one of the few successful examples of heterologous expression of a relatively large fungal NRPS gene, thus far.

In fungi forming mutualistic mycorrhizal relationships with plants, siderophore production is believed to benefit their host plants by improving solubilisation of insoluble iron oxides and thus positively affecting iron nutrition of the host plant.<sup>35, 36</sup> New evidence comes from newly discovered trishydroxamate siderophores, the basidiochromes (6) that were purified from iron-deficient culture filtrates of *Ceratobasidium* and *Rhizoctonia* species that form mycorrhizae with orchid roots.<sup>36</sup> Epichloënin A (7) is an unusual octapeptide member of the ferrichrome family and was found to be encoded by *sidN*.<sup>37, 38</sup> It comprises a cyclic peptide of four glycines, a glutamine and three N<sup>δ</sup>-*trans*-anhydromethylalanyl-N<sup>δ</sup>-hydroxyornithine (*trans*-AMHO) moieties. This extracellular siderophore is essential for

maintenance of the mutualistic endophytic growth of *Epichloë festucae* within tissues of perennial ryegrass (*Lolium perenne*). In addition to SidN's role in the grass-fungus mutualism, this protein is significant because its third A domain was heterologously expressed and was the first eukaryotic NRPS domain to be purified and characterised at the structural level.<sup>39</sup> Substrate activation of the SidN A domain 3 was modelled with *cis*-AMHO obtained by alkaline hydrolysis of fusigen. X-ray crystallography and ligand-docking experiments determined that a large 17-residue binding pocket is needed to accommodate the bulky aa substrate, *cis*-AMHO. The specific activation of *cis*-AMHO was confirmed biochemically, and an AMHO moiety was unequivocally identified as a constituent of the fungal siderophore. This case clearly illustrated the divergence of A domain signature sequences between prokaryotes and eukaryotes and why substrate prediction methods based on bacterial NRPS A domains generally fail for eukaryotes.<sup>5</sup> Thus, the structure of the *E. festucae* SidN A domain will contribute to improved prediction methods for eukaryotic NRPS enzymes.

Host-specific toxins (HSTs) mediate virulence and pathogenicity among certain plant pathogenic species of the Dothideomycetes. HSTs encompass about 20 different structurally diverse metabolites, including peptides and decapeptides (AM-toxin, HC-toxin (8), victorin, phomalide, dephalairdin, BZR-cotoxins).<sup>40-42</sup> Other non-host specific fungal peptides, e.g., destruxins and roseotoxins, also have been implicated as plant pathogenicity or necrosis factors.<sup>42-44</sup>

Some widespread NRPSs may structurally support fungal cell walls and contribute to their hydrophobicity. FgNRPS4 in *F. graminearum* is a five-module NRPS and is conserved throughout four genome-sequenced *Fusarium* species, *F. graminearum*, *F. oxysporum*, *F. solani*, and *F. verticillioides*. Disruption of *fgnrps4* in *F. graminearum* resulted in loss of mycelial water repellency and led to altered conidial morphology.<sup>45</sup> ChNRPS4 of *Cochliobolus heterostrophus* has a similar five-module structure and appears to be an orthologue of FgNRPS4; its disruption also abolishes mycelial surface hydrophobicity and water repellency.<sup>46</sup> NRPS2 in *Alternaria brassicicola* appears to be another orthologue, and deletion mutants show a similar decreased hydrophobicity, distorted conidial cell wall morphology, poor sporulation and germination, and thus, reduced infection capability.<sup>47</sup> BLAST searches with the protein sequence of FgNRPS4 retrieve a wide range of five-module NRPSs with similarity scores >40%, including GLNRPS6 from *G. lozoyensis* (45% similarity)<sup>20</sup>. Despite persistent attempts at identification,<sup>45</sup> the products of these NRPSs have remained elusive. The eventual identification of these putative pentapeptides will be of great interest to phytopathologists and fungal biologists.

Peptaibols and other peptaibiotics, as defined by Degenkolb & Brückner,<sup>48</sup> are characteristic of mycoparasitic species of the Hypocreales, and are occasionally detected in other Ascomycete fungi. Their biology and chemistry have been reviewed extensively.<sup>49-51</sup> These specialized peptides have structural properties that induce spontaneous formation of voltage-gated channels in bilayer lipid membranes, and have been observed to have antibacterial, antifungal, and a myriad of other biological activities. A comprehensive database with 1043 entries of peptaibiotics listing compound name, aa sequence, sequence length, producing organism, subfamily classification, molecular formula, monoisotopic masses, was recently made available for download.<sup>52</sup> Their producing fungi, particularly *Trichoderma* (= *Hypocrea*) and *Clonostachys* species, have

been intensively exploited as biological control agents for plant pathogenic fungi, with over 950 different peptide sequences known from the *Trichoderma* lineage alone.<sup>51</sup>

Two recent studies have investigated the question of whether peptaibiotics are produced in natural ecosystems. The mycoparasite *Trichoderma phellincola* naturally attacks the polypore, *Phellinus ferruginosus*. LC-ESI-HR-TOF analysis of field-collected *T. phellincola* specimens and naturally infected *P. ferruginosus* basidiomata revealed that a family of 20-residue peptaibols, named the hypophellins (**9**), were biosynthesized in natural mycelia and stromata and were subsequently released into the host fungus.<sup>53</sup> The structures of naturally occurring peptaibols were independently confirmed by analysing agar cultures of *T. phellincola*. *In vitro*, this strain produced 39 new 17-, 18-, and 19-residue hypophellins that displayed the same biosynthetic scheme as the 20-residue hypophellins found in natural specimens. Two of the 19-residue peptaibols were thought to carry tyrosinol, a novel C-terminal residue, as predicted by HR tandem-MS. A parallel investigation of *Hypocrea pulvinata* infecting the birch polypore, *Piptoporus betulinus* and the red-banded polypore, *Fomitopsis pinicola*, revealed novel, mostly 19-residue peptaibiotics, named the hypopulvins (**10**).<sup>54</sup> Extracts of field-collected stromata of *H. pulvinata* were also dominated by the same 19-residue peptaibols, along with some minor 18-residue deletion analogues. The same 19-residue peptaibols also predominated in agar cultures, along with two new 20-residue minor compounds similar in structure to trichosporin B, and 11-residue peptaibols (**11**). These data supported the hypothesis of a mycoparasitic lifestyle for both *T. phellincola* and *H. pulvinata* and demonstrated their potential use as biocontrol agents. This comparative approach of interrogating freshly collected ascomata and cultures in parallel was recently extended to another nine more European specimens, including seven previously uninvestigated mycoparasitic or saprobic *Trichoderma* and *Hypocrea* species.<sup>51</sup> This study detailed the characterization of 78 peptaibiotics of which 56 (72%) were new sequences. Among these new peptaibiotics were new 18-residue peptaibols, named thelephoricolins (e.g., **12**) from *H. thelephoricola* that share some structural similarity with the trichotoxins, 18-residue sequences, named the gelatinosins B (e.g., **13**) from *H. gelatinosa*, new 19-residue peptaibols named the voglmayrins (e.g., **14**) from *H. voglmayrii* that partly resemble the sequences of trichokonin V and trichorzianins B, new 19-residue peptaibols, named the minutisporins (e.g., **15**) from *H. minutispora*, and 19-residue peptaibols named the hypocitrins (e.g., **16**) from *H. citrina*.

The antimicrobial peptide Trichogin GA IV (GA IV) (**17**) from *T. longibrachiatum* alters membrane permeability. Previous data have shown that pore formation is related to GA IV aggregation and insertion in the hydrophobic core of the membrane. This behaviour is similar to that of the peptaibol alamethicin,<sup>55</sup> and is consistent with barrel-stave mechanism where aggregated transmembrane oriented peptides open a channel. Alamethicin has a length (19 aa) similar to the membrane thickness, but GA IV is 10-mer peptide, thus, its helix is about half the normal bilayer thickness. Neutron reflectivity measurements and molecular dynamics simulations demonstrated that GA IV induced membrane thinning.<sup>56</sup> The physical and free energy effects of inserting a peptide into a bilayer with an equilibrium hydrophobic thickness differing from that of the inclusion were modelled. GA IV interacted strongly with the polar phospholipid head groups, pulling them towards its N- and C-termini, inducing membrane thinning. The

free energy cost of aqueous exposure of hydrophobic groups of the protein or lipids, due to their different thickness, was higher than that of distorting the lipids from their equilibrium conformation<sup>56</sup>. As a consequence, the membrane thickness locally adapted to the size of the inclusion, although other effects may be involved. GA IV-induced membrane thinning, thus, allowed the peptide to traverse both sides of the bilayer.

Previous study showed that trichokonins, a type of peptaibol from *T. pseudokoningii*, exhibited antibiotic activities against plant fungal pathogens. A recent study presented the first in-depth investigation of antibiotic mechanism of peptaibols against fungal plant pathogens. Trichokonin VI (TK VI) (**18**) induced extensive apoptotic programmed cell death in plant fungal pathogens.<sup>57</sup> To better understand the apoptotic mechanism involved in the effects of TK VI, cells of the mycoparasitic host, *Fusarium oxysporum*, were observed. When cells of *F. oxysporum* were treated with TK VI, they exhibited apoptotic morphological symptoms, e.g., phosphatidylserine exposure, the appearance of ROS, chromosome condensation, and nuclear DNA fragmentation. Moreover, TK VI-treated cells accumulated cytoplasmic vacuoles and lost mitochondrial transmembrane potential. Cell death was metacaspase independent, and therefore, TK VI induced metacaspase-independent apoptotic cell death in *F. oxysporum*.

### 3 Differences between fungal and bacterial non-ribosomal peptide synthesis

NRPSs synthesize peptides by employing modular sets of core domains that minimally consist of three domains: 1) an A domain that recognizes and activates the substrate via adenylation with ATP, 2) a T- or PCP-domain which binds the activated substrate to a 4'-phosphopantetheine cofactor via a thioester bond and transfers the substrate to 3) a C-domain which catalyses peptide bond formation between tethered substrates on the megasynthase complex. Major features that set fungal NRPSs apart from bacterial NRPSs are their aa compositions. Fungal and bacterial NRP composition was compared based on an analysis of the Norine database<sup>58, 59</sup> that catalogues >1000 peptides representing more than 10,000 monomer occurrences and >500 different monomer types.<sup>60</sup> A total of 34% of the database was comprised of fungal peptides. Correlation coefficients between fungi and bacteria were low for peptide size distribution (0.236) and for monomer composition (0.252) which supported phylogenetic inferences of an independent fungal- and Ascomycete-specific expansion of NRPSs.<sup>22</sup> The database's most frequent monomer was 2-aminoisobutyric acid (Aib) which is a hallmark of fungal peptaibols that can occur multiple times in the same peptaibol (mean = 6) and explained why Aib was the database's most frequent monomer. Other monomers are specific to fungal NRPs, especially those in peptaibols, e.g., isovaline, phenylalaninol, leucinol, and valinol.

Structures 19 and 20 here (*Figs for NPR-part3*)

Bacterial NRPSs employ thioesterase domains to catalyse peptide cyclization, while fungal NRPSs use an alternative enzymatic mechanism.<sup>61</sup> The folds of fungal C<sub>T</sub> domains and bacterial TE domains are completely different and utilize distinct mechanisms to catalyse the same reaction. To date, all fungal NRPSs that assemble macrocyclic peptides terminate with a condensation-like C<sub>T</sub> domain that catalyses the cyclization reaction, and thus, is a near universal macrocyclization strategy used by fungal NRPSs. Fungal C<sub>T</sub>

domains are phylogenetically distinct from chain-extending C domains, and in fact, genome sequencing data indicated that a majority (60–90%) of the NRPSs found in filamentous fungi have a terminal C<sub>T</sub> domain.<sup>61</sup> The biochemical mechanism of the cyclization reaction was recently investigated by comparing the *in vitro* reconstitution of the *Penicillium aethiopicum* trimodular NRPS TqaA, responsible for transforming the linear anthranilate-D-tryptophan-L-alanyl tripeptide into fumiquinazoline F (19) with that of the bimodular NRPS AnaPS that produces 3-indolylmethyl-3,4-dihydrobenzo-1,4-diazepine-2,5-dione precursor R-2 of acetylazonalenin (20) in *Neosartorya fischeri*. The first two modules of TqaA are identical to AnaPS, however the addition of the third module radically alters the structure of the final cyclic alkaloid products. Both gene clusters were expressed in *Saccharomyces cerevisiae*. Extensive biochemical and mutational studies confirmed that differences in the cyclization patterns are likely due to the precise control of the reactive dipeptide D-4 intermediate, and the use of a C-like cyclization domain that regioselectively directs the intramolecular attack of the anthranilic acid amine nucleophile. A catalytic histidine at the C<sub>T</sub>-active site deprotonates the amine nucleophile, thus, facilitating its attack on the thioester carbonyl leading to cyclization and product release. The C<sub>T</sub> domain also differs mechanistically by its requirement for a specific companion T domain in the C<sub>T</sub> domain catalysed cyclization reaction. This study showed that only peptide substrates tethered to the natural T domain portion of the C<sub>T</sub> domain are efficiently recognized; peptidyl-SNAC or peptidyl-CoA substrates were not recognized. Such a mechanism differs from the usual bacterial TE domains where small-molecule mimics of the peptidyl-thioester can be used to probe the cyclization functions.

#### 4 New understanding of the biosynthesis of pneumocandins and echinocandins

##### 4.1 Pathway discovery

The echinocandin family of lipopeptides has been investigated for over three decades and has formed the basis of a successful new class of antifungal drugs. Their chemistry, biology, and distribution among different kinds of Ascomycete fungi have been reviewed extensively.<sup>62–66</sup> Until recently, knowledge of their biosynthesis was largely inferred from the patterns of incorporation of the eight different biosynthetic precursors molecules into the hexapeptide core and acyl side chain of the pneumocandins and subsequent decorations with hydroxyl and methyl groups.<sup>67</sup> During the last two years, genomic sequencing of representative strains of most of the important variants in the family, including the parent wild types of three fungal strains employed for industrial scale production of caspofungin, micafungin and anidulofungin, has enabled characterisation of the basic architecture of the echinocandin NRPS pathways. Thus, an analysis of how pathway genes produce variations in lipoinitiation, biosynthesis of the non-proteinogenic aa, aa substitutions, hydroxylations, and sulfonations of the core peptide contribute to the molecular diversity of the family is now possible.

Figure 1 here

Fig. 1. Representation of the pneumocandin gene cluster from *Glarea lozoyensis* ATCC 20868 and the echinocandin B gene cluster, including the disjunct L-homotyrosine gene cluster, from *Aspergillus rugulosus* NRRL 111440.

The gene cluster responsible for the synthesis of echinocandin B (21) of *Aspergillus rugulosus* (*ecd*) (Fig. 1, 2) was characterised via genome sequencing and bioinformatic analysis and was confirmed by functional knock-out experiments and heterologous expression of selected pathway

genes.<sup>18, 19</sup> The *ecd* gene cluster spans 12 genes (*ecdA* to *ecdL*). The core NRPS gene *ecdA* encodes for a six-module megasynthetase with each module consisting of a C, A, and T domain. An extra T<sub>0</sub> domain initiates the first module at the N-terminus, while the terminal module ends with a C<sub>T</sub> domain. Gene *ecdI* encodes a fatty-acyl-AMP ligase responsible for activating lineolate and loads it onto the first thiolation domain of *ecdA*. EcdA loads L-Orn as the first module. Iron oxygenases EcdG, EcdH and EcdK mediate hydroxylations of the core hexapeptide.<sup>18</sup> Other genes flanking *ecdA* include a transcription factor gene *edcB*, and three transporter proteins, *ecdC*, *ecdD* and *ecdL*.<sup>19</sup> Furthermore, the *hty* gene cluster for L-homotyrosine biosynthesis was discovered on another scaffold, and its disruption silenced echinocandin production.<sup>19</sup> The *hty* gene cluster consists of an isopropyl malate synthase gene (*htyA*), a transaminase (*htyB*), an isopropyl malate dehydrogenase (*htyC*), and a aconitase gene (*htyD*), along with two oxygenase genes (*htyF*, *htyE*) for further hydroxylation of homotyrosine (Fig. 1).

Structures 21 to 29 here (Figs for NPR- part4)

Pneumocandin A<sub>0</sub> (22) is the major antifungal product of wild-type *Glarea lozoyensis* and of certain *Pezizula* (= *Cryptosporiopsis*) species. Like echinocandin B (21), five of the same six aa in the cyclic hexapeptides are incorporated in the same order with the same hydroxylation pattern, the only difference being hydroxy-L-glutamine in position 5 in *G. lozoyensis* instead of L-threonine. Hydroxylation of the two proline residues in pneumocandin A<sub>0</sub> are catalysed by a proline-3-hydroxylase and a proline-4-hydroxylase.<sup>68</sup> The enzyme responsible for hydroxylation of 4-methyl-proline derived from leucine in pneumocandin A<sub>0</sub> may also be a proline-3-hydroxylase because 4-methyl-proline is an analogue of L-proline.<sup>68</sup> Shortly after the publication of the echinocandin B gene cluster, our group published the genomic sequence of the Ascomycete *G. lozoyensis* and annotated and functionally confirmed the biosynthetic gene cluster (Fig. 1) responsible for the antifungal lipohexapeptides pneumocandins A<sub>0</sub> (22) and B<sub>0</sub> (23).<sup>20</sup>

Figure 2 here

Fig. 2. Comparative organization of the principal echinocandin gene clusters. Genes in black are of unknown function or believed not to participate in the biosynthesis.

The commonalities of pneumocandin and echinocandin B pathways are striking (Fig. 1) because most of clusters' genes appear to be orthologs despite some organizational differences. Pneumocandin's core NRPS, GLNRPS4, and EcdA (T<sub>0</sub>CATCATCATCATCATC<sub>T</sub>) are equivalent with same direction of transcription and sharing a high nucleic acid identity (60.8% identity over 22.7 kb, 55.2% identity over 7218 aa). However, unlike *ecdA*, *glnrps4* sits downstream of, and

adjacent to, a highly reducing polyketide synthase gene, *glpks4*. These two core genes are centrally located in the pneumocandin gene cluster and are independently transcribed and translated. Instead of the pneumocandin acyl side chain originating from cytosolic fatty acids as proposed in *ecd*, GLPKS4 synthesizes 10,12-dimethylmyristate. The 10,12-dimethylmyristoyl side chain is released from GLPKS4 as a carboxylic acid and subsequently converted to a CoA thioester by an acyl-CoA ligase (GLAREA10043), and then presumably loaded onto the acyltransferase (GLAREA10021). However, we have preliminary data indicating the acyltransferase step may be dispensable (Chen, unpublished). The polyketide intermediate is then shuttled to the first thiolation domain of GLNRPS4, followed by its acylation to 4,5-dihydroxy-ornithine to trigger elongation of the cyclic hexapeptide. Threonine, 4-hydroxy-proline, 4,5-dihydroxy-homotyrosine, 3-hydroxy-glutamine and 3-hydroxy-proline or 3-hydroxy-4-methyl-proline are sequentially added to the growing chain in the same sequence as in *EcdA*. Like *EcdA*, the last C domain of GLNRPS4 is a C<sub>t</sub> and is responsible for cyclization by condensation to form the peptide bond between 4,5-dihydroxy-ornithine and 3-hydroxy-4-methyl-proline.

Other gene downstream of the *glnrps4* (Fig. 1) involved in the biosynthesis is GLAREA10043 encoding an acyl-CoA ligase which shares 43% identity with *EasD* which converts a polyketide carboxylic acid to a CoA thioester during emericellamide A (**24**) and B (**25**) biosynthesis in *A. nidulans*.<sup>20, 69</sup> This gene's existence suggested that the polyketide intermediate was first synthesized by GLPKS4, and then transferred to the T<sub>0</sub> domain of GLNRPS4, which would be mechanistically equivalent to the emericellamide biosynthetic pathway.<sup>69</sup> Further contributing to the autonomy of pneumocandin gene cluster, and unlike the echinocandin B pathway (Fig. 1), a set of genes orthogonal to the echinocandin *hty* pathway for the homotyrosine residue of the pneumocandin peptide core sits downstream of the NRPS.<sup>19</sup> Both pathways also contain a number of oxygenases that tailor the multiple hydroxyl or diol groups of the peptide core, but their physical linkage is rearranged.

DNA sequences for the gene clusters of additional echinocandin-type pathways recently have been made available at the NCBI GenBank and at the US Department of Energy's Joint Genome Institute. These sequences include the gene cluster for aculeacin A (**26**)<sup>70, 71</sup> from *A. aculeatus* (ATCC 16872 = JCM 22898) and the sulfonated homotyrosine echinocandin-pneumocandin variants: FR190293 (**27**) from a Leotiomyces fungus related to *Phialophora hyalina*<sup>72</sup> (referred to from here on as *P. cf. hyalina*, it previously was misidentified as *Tolyopocladium parasiticum*<sup>73</sup>), which is essentially a sulfonated pneumocandin A<sub>0</sub>; FR901379 (WF11899A) (**28**) from *Coleophoma empetri*<sup>74</sup>, the starting molecule for micafungin; and FR209602 (**29**) from *C. crateriformis*.<sup>75</sup> The availability of these sequences of the gene clusters responsible for almost the entire range of echinocandin complexity now enables development of hypotheses about the functions of specific pathway genes where their presence or absence in certain pathways can be correlated with specific differences in the pathway products (Fig. 2). The sequences also present the opportunity to trace the evolutionary history of the echinocandin-pneumocandin gene clusters, as well as investigate the significance of these potent cell wall-modifying metabolites to the fungi that produce them.

These newly characterized gene clusters are distributed among fungi from two major evolutionary lineages,

Eurotiomycetes (*A. rugulosa*, *A. nidulans* var. *echinulatus*, *A. aculeatus*) and Leotiomyces (*G. lozoyensis*, *C. empetri*, *C. crateriformis*, *P. cf. hyalina*) that diverged around 290 to 390 million of years ago.<sup>76</sup> Explanation of this pattern based on independent parallel evolution seems improbable. Echinocandin-producing fungi vary in their ecology<sup>62, 66</sup> and range from soil and litter saprobes (*Aspergillus* species, possibly *G. lozoyensis*, *P. cf. hyalina*) and endophytes and weak pathogens of woody plants (*Pezizula* and *Coleophoma* species). All these strains were discovered because of their fermentations were potently active in antifungal screens; therefore, the pathways are operative. In our experience, echinocandin biosynthesis is not tightly regulated in the laboratory.<sup>66, 77, 78</sup> The natural function of echinocandins remains uninvestigated; however, the ability to produce null pathway mutants<sup>19, 20</sup> opens up interesting possibilities for experimentation in fungal microcosms.

#### 4.2 Structure and gene order conservation of the echinocandin gene clusters

The functions determined to date for enzymes in the echinocandin pathway<sup>18-20</sup> are consistent with the presence or absence of specific genes each strains and their specific pathway end products. Genes in those pathways lacking experimentally determined roles can be linked with hypothesized steps on the basis of the functions predicted from their sequences and their presence or absence in clusters among the various strains compared to those of *G. lozoyensis* and *A. rugulosus* where some pathway genes have been functionally characterised (Fig. 1). For example, clusters that incorporate a highly reducing PKS (*G. lozoyensis*, *P. cf. hyalina*, Fig. 2) have a branched acyl side chain. Other genes responsible for structural differences can be inferred, such as in FR209602 (**29**) from *C. crateriformis*, where the single hydroxylation of the homotyrosine is likely due to the absence of one of the oxygenases of the *hty* pathway, i.e. the orthologue of *htyF* and GLAREA10030. Likewise, the absence of orthologs to oxygenase GLAREA10042 in the *Aspergillus* echinocandin pathways points to their function in the hydroxylation of hydroxyl-glutamine at position 4. Finally, orthologs of GLAREA10045 and *ecdJ* are in all echinocandin pathways (Fig. 2). Its consistent presence in all pathways suggests an important function that remains unknown. In summary, by comparing common pathway elements, the minimal elements needed to biosynthesize a functional echinocandin would include a six-module NRPS, an acyl-AMP ligase, the four orthologs of *HtyA-D* of the *hty* pathway to synthesize L-homotyrosine, and the three orthologs of oxygenases *EcdG*, *K*, and *H* which hydroxylate the L-ornithine, L-homotyrosine, and leucine precursor of the fifth 4-methyl-L-proline residue.

The most obvious dichotomies between pathways are those that incorporate a dedicated highly reducing PKS for biosynthesis of a methylated, branched acyl side (*G. lozoyensis*, *P. cf. hyalina*) and pathways that incorporate linear fatty acid-derived acyl side chains from the cytosol fatty acids, and those pathways where the C4 of the homotyrosine residue is sulfonated (*Coleophoma* spp. and *P. cf. hyalina*). Thus, the most compact pathway architectures would correspond to that of the aculeacin pathway where the *hty* genes sit immediately downstream of, and separated by, two unknown genes and a cytochrome P450, from the core NRPS gene (Fig. 2). In contrast, the most elaborate cluster is assumed to be that of FR190293 (**27**) which not only recruits a highly reducing PKS for the side chain synthesis, but also has a sulfonated

homotryosine residue. To date, the enzyme(s) responsible for the sulfonation reaction remains unknown. An obvious candidate sulfotransferase gene responsible for the homotryosine sulfonation is not evident in the clusters. Therefore, sulfonation is assumed to occur post-synthesis of the core hexapeptide and that the responsible gene(s) is located at some distance from the cluster. The vicinal dihydroxy groups in hydroxyhomotryosine, like the hydroxyl group of phenol, would have a strong activating effect on the aromatic ring because of its high electron density within the ring. Chemically this directing influence is strong enough that ortho sulfonation of phenols in the presence of sulphuric acid occurs at room temperature without a catalyst.<sup>79</sup>

Predicted protein sequence similarities of the core NRPS range from near 100% between the two most similar clusters, the two *Coleophoma* species, and the two echinocandin B NRPSs from the different species of the *A. nidulans* complex, while the most dissimilar NRPSs are those of *A. aculeatus* and *A. rugulosus* and *A. nidulans* var. *echinulatus* versus *G. lozoyensis* (71%). Furthermore, a dichotomy between the Leotiomycete and Eurotiomycete gene clusters is consistently evident across all pairwise comparisons between the protein sequences of *G. lozoyensis* and those of the other fungi (Chen, Bills, unpublished).

#### 4.3 Evolution of the echinocandin gene clusters

BLAST and HMMER searches with the GLNRPS4 and EcdA across the NCBI and the Joint Genome Institute databases<sup>80</sup> retrieved only homologues of echinocandin gene clusters. BLAST searches with individual A domains of Ecd4 and GLNRPS4 also indicated that uncharacterised dipeptides NRPS AN3496 (InpB) from *A. nidulans*, and its orthologues As12236 in *A. sydowii* and Av13511 in *A. versicolor* were related (Fig. 3). The top 50 BLAST hits to EcdA and GLNRPS4 from the NCBI database failed to retrieve any bacterial sequences. More targeted BLAST searches with individual catalytic domains (A, T, C) of the NRPSs only retrieved other Ascomycete NRPS domains. These searches suggested that the echinocandins likely evolved by internal duplication of modules of an ancestral module originating within the Ascomycete fungi rather than through HGT from bacteria.

Figure 3 here

Fig. 3. Phylogeny of selected fungal NRPS A domains, including A domains of the principal echinocandin-type NRPSs (see Fig. 2). Data were partially resampled from Bushley and Turgeon<sup>22</sup>. Protein sequences were annotated by using the HMMER analytical tool. The A domain aa sequences were aligned with M-Coffee. A maximum likelihood tree was constructed by applying WAG plus gamma model using RAxML analytical software with 100 replicates. Bootstrap values greater than 50% are shown above branches. The A domains corresponding compounds that indicated by different symbols are listed on the left side. A domains of selected fungal metabolites (left) were color-coded and mapped on the tree (right).

To infer phylogenies of the NRPSs components, the deduced aa sequences for all the echinocandin A domains were aligned with homologues throughout the ascomycetes by resampling and updating a previously published data set of fungal NRPS A domains.<sup>22</sup> A maximum likelihood tree (Fig. 3) constructed from the A domain protein alignments indicated all

the echinocandin A domains and the A domains of the predicted dipeptides NRPS AN3496 (InpB) and its orthologues As 12236 and Av 13511 formed a distinctive clade (59% bootstrap value). The echinocandin A domain clade clearly nested within an Ascomycete-exclusive subfamily of NRPSs previously designated as the euascomycete only subclade<sup>22</sup> (Fig. 3). The same analysis<sup>22</sup> detected the singularity of the AN3496 NRPS where it appeared as a weakly supported branch and a sister group to the A domains of the emericellamide (EAS, AN2545) NRPS<sup>69</sup>. Thus, the echinocandin NRPSs along with the AN3496 NRPS appear to form a novel lineage within the Ascomycete-exclusive subclade of fungal NRPSs.

In addition to the echinocandin A domain genes forming a monophyletic clade, the intra-clade relationships among the six individual A domains were consistent. Remarkably, the internal topology of the cluster clade was resolved into six subclades each consisting of the corresponding A domain modules for each aa positions in the echinocandin nucleus (Fig. 3). This topology was supported by strong internal bootstrap values and indicated that the A domains for each aa position from each fungus were more similar to each other than the individual A domains from a single given echinocandin NRPS from any one species. This one-to-one correspondence between modules and aa specificity of different species is consistent with models of tandem duplication and subfunctionalization prior to divergence from a hypothetical ancestor, as has been hypothesized to occur during the evolution of cyclosporine and peptaibol NRPSs.<sup>21,22</sup>

We have constructed gene phylogenies for rest of the dedicated pathway genes, including the cytochrome p450s, oxygenases, enzymes of the *hty* pathway, the acyl-ligase, and the ABC transporter. When examined in the context of their respective fungal orthologs, each group of the echinocandin genes form well-supported monophyletic groups with tree topologies similar to that of the core NRPS (unpublished). In a few cases, e.g., some NRPS A domains, acyl-ligases from close neighbours of the genus *Aspergillus* (*A. nidulans* and *A. sydowii*) were sister groups to the clade formed by the clustered echinocandin genes. The presence of these gene relatives in related Aspergilli suggests an origin of the pathway in this genus, possibly in common ancestors of sections *Nidulantes* and *Nigri*.<sup>81</sup> This high degree of functional similarity conflicts with established phylogenies of the Eurotiomycetes, and Leotiomycetes.<sup>82</sup> The echinocandin pathways among these divergent fungal species from distinct evolutionary lineages appear to have shared a common ancestor, and a minimal ancestral echinocandin pathway likely existed before radiation into or between the two fungal lineages.

#### 5 Other recently characterised NRPS gene clusters

##### 5.1 The apicidin gene cluster and its relationship to HC-toxin.

The cyclic tetrapeptide apicidin A (**30**), produced by various *Fusarium* species, was identified as a broad-spectrum *in vitro* and *in vivo* agent effective against a range of apicomplexan parasites, including *Plasmodium berghei* malaria in mice<sup>83, 84</sup> and been used as molecular template for synthesis of HDAC inhibitors.<sup>85</sup> Apicidin and related fungal tetrapeptides, e.g., trapoxin, HC-toxin (**8**), and chlamydocin are also potent inhibitors of other protozoans and affect the HDACs of many eukaryotic organisms. Unlike apicidin, similar cyclic peptides including HC-toxin (**8**), trapoxins and others contain an epoxyketone moiety on the 2-amino-decanoic acid residue. These compounds irreversibly inhibit HDAC at nanomolar



concentrations by covalently binding to the enzyme through the epoxide group. The apicidins produced by plant-inhabiting *Fusaria* may function as a phytotoxin in a manner similar to the HST, HC-toxin (**8**) produced by *Cochliobolus carbonum*.<sup>86</sup> The apicidin gene cluster (ASP1) has been characterised in *F. semitectum* by the sequencing of three overlapping cosmid clones covering a 63 kb genomic region with 19 ORFs of which 10 genes (ASP1-ASP11) appear to be involved in synthesis and regulation.<sup>87</sup> The APS1 gene consists of four peptide synthase modules with an E domain in module 1. The modules sequentially incorporate L-pipecolic acid, L-isoleucine, *S*-*N*-methoxy-tryptophan, and *S*-2-amino-8-oxodecanoic acid (AODA). ASP2 has significant homology to HC-toxin synthase HTS1 (43%) (Fig. 3).

Structures 30 to 51 here (*Figs for NPR-part5*)

Mutation of APS3 gene resulted in apicidin B (**31**), i.e. substitution of D-pipecolic acid by proline at position 1, thus leading to the conclusion that the APS3 may be involved in reducing  $\Delta^1$ -pyrroline-6-carboxylate into L-pipecolic acid, while the E domain of APS1 converts the latter to D-pipecolic acid. The first A domain of ASP1 was sufficiently flexible to accept proline as a substrate. The ASP2 gene was predicted to be a transcription factor that was essential for pathway expression. ASP5 is an  $\alpha$ -subunit of a fatty acid synthase dedicated to the production of the lipid chain of AODA. It was thought that  $\beta$ -subunit from cytosolic fatty synthesis complemented the biosynthesis. Mutants of the cytochrome P450 ASP7 accumulated apicidin E (**32**) indicating that this gene is responsible for the conversion of *S*-2-amino-decanoic acid to 2*S*,8*S*-2-amino-hydroxydecanoic acid. Further side-chain oxidations are mediated by the oxygenase ASP9 resulting in AODA; loss of ASP9 resulted in production of apicidin D<sub>2</sub> (**33**). ASP8, another cytochrome P450, was also thought to be involved in the sequential oxidative formation of AODA. ASP4 was found to be an aminotransferase, and orthologue to TOXF. ASP6, an O-methyltransferase, was believed to methylate tryptophan. Viable deletion mutants of the predicted efflux pump ASP11 could not be obtained, indicating that efflux of apicidin out the cells was essential to the fungus.

The genomic sequence of the causal agent of bakanae disease of rice,<sup>88</sup> *F. fujikuroi*, revealed an apicidin-like gene cluster (NRPS31) largely syntenic to the organization of the APS cluster in *F. semitectum*. Transcriptome analysis revealed expression of the *F. fujikuroi* APS cluster homologue during growth in acidic high-nitrogen conditions and possible gene silencing under the low-nitrogen growth conditions. Overexpression of the pathway-specific transcription factor-encoding gene led to increased accumulation of an unknown metabolite with the molecular formula (C<sub>35</sub>H<sub>43</sub>N<sub>5</sub>O<sub>7</sub>) similar to that of **30** (C<sub>34</sub>H<sub>49</sub>N<sub>5</sub>O<sub>6</sub>). NMR, MS and acid hydrolysis revealed a new apicidin, apicidin F (**34**), consisting L-*N*-methoxytryptophan, D-pipecolic acid, and L-phenylalanine and L-2-aminooctanedioic acid<sup>89</sup>. It was about three-fold less potent than apicidin A (**30**) in a *Plasmodium falciparum* inhibition assay.

HC-toxin (**8**) was unexpectedly detected in extracts of *Alternaria jesenskae*<sup>90</sup> while screening for new HDAC inhibitors.<sup>91</sup> Genomic sequencing of *A. jesenskae* revealed that the major genes (designated AjTOX2) involved in the biosynthesis of *A. jesenskae* HC-toxin were orthologous to those of TOX2, and their proteins shared 75-85% aa identity (Fig. 3). Similarly in both fungi, the genes for HC-toxin

biosynthesis were duplicated. The genomic organization of TOX2 clusters in the two fungi showed a similar but not identical partial clustering arrangement.

## 5.2 Other NRPS biosynthetic gene clusters

Monomodular NRPSs with an A-T-TE domain structure have been found to synthesize 1,4-benzoquinone compounds from symmetric condensation of two aromatic 2-oxo acids; they include metabolites such as terrequinone A (**35**) encoded by the TdiA gene cluster.<sup>15, 92</sup> Recently an efficient expression of TdiA in *E. coli* was reported with the objective of making enzymatic preparations of bis-indolylquinone metabolites.<sup>93</sup> The pigment atromentin (**36**) from various mushrooms in the Boletales is a terphenylquinone arising from the adenylation and dimerization of 4-hydroxyphenylpyruvic acid. The biosynthesis is mediated by atromentin synthetase AtrA, encoding a quinone synthetase, and an aminotransferase, AtrD, that adenylates and dimerizes 4-hydroxyphenylpyruvic acid.<sup>94</sup> One of these enzymes, GreA, has been identified in *Suillus grevillei* and was heterologously expressed and biochemically characterised.<sup>95</sup> Atromentin biosynthesis was reconstituted *in vitro*. Another A-T-TE NRPS-like enzyme, AN3396.4 from *A. nidulans* has also been heterologously expressed in *A. niger* to confirm its function and was shown to produce a furanone compound, designated as microperfurane (**37**).<sup>96</sup> Cosmid libraries of the mushroom *Armillaria mellea* were genetically screened for genes encoding natural product biosynthetic pathways, thus leading to the recognition of a *tdiA*-like gene (*arma*) encoding a A-T-TE domain structure.<sup>97</sup> This locus lacked a typical neighbouring aminotransferase indicating affinity for substrates other than 2-oxo acids. Two genes coding for flavin-dependent halogenases, *armH1* and *armH2* were located nearby to *arma*. The cDNA of *arma* was heterologously expressed to determine the substrate specificity of its A domain employing an ATP-pyrophosphate exchange assay. Surprisingly, considering that ArMA lacks a C domain, the preferred substrates were leucine, threonine, and other hydrophobic aa, leading to the conclusion that these novel gene cluster may represent a new alternative route to diketopiperazine formation.

WYK-1 (**38**) from *Aspergillus oryzae* is an isoquinoline derived from a tripeptide consisting L-tryptophan, L-phenylalanine or L-tyrosine, and L-leucine.<sup>98</sup> The *A. oryzae* genome contained four NRPS genes that were approximately 9 to 16 kb long. Comparative analysis of expression levels of these NRPS genes by RT-PCR under WYK-1 producing or non-producing conditions detected that only expression of locus AOR-1-1360164 was correlated with WYK-1 production. This locus encoded a trimodular peptide synthase WykN, and transcription analysis identified 10 more genes that were coexpressed with wykN. The proposed biosynthesis commenced with the tripeptide synthesis, followed by methylation by WykI. Cyclization of the phenylalanine or tyrosine residue by the oxidase, WykG, would form an isoquinoline ring, followed by methylation by *N*-methyltransferase, WykH, and hydroxylation by phenol hydroxylase WykA to yield WYK-1 (**38**).

Genome sequencing projects of three representative *Trichoderma* strains from the genus' major evolutionary lineages indicated the existence of as many as three kinds of NRPSs with 7, 14, or 18–20 modules. The major peptaibols of biocontrol strain of *T. virens* were determined to be trichovirins of two classes, 11- and 14-residues<sup>99</sup>. Of the 35 11-residue peptaibols sequenced, 18 were new, and all of the 53 14-residue sequences were new. A 14-module NRPS, trichovirin A

synthetase, was identified by genome scanning and its disruption caused the loss of both 11- and 14-residue peptaibol families. The mechanism of dual peptide formation was further investigated in *T. reesei* which also produced both 11- and 14-residue peptaibols.<sup>100</sup> The *nrps2* with a predicted size of 16,847 aa was identified as a SF4-type peptaibol synthetase based on identical substrate binding codes in domain positions 2, 5, 6, 9, and 10 and homology to the *T. virens* trichovirin A synthetase.<sup>99</sup> Bioinformatic analysis of the three-dimensional structure of the NRPSs resulted in a “direct-transfer” model, different from the typical “pass-along” model. This model assumed that the C-domain-attached tripeptide intermediate could be transferred from module 3 to either the module 4 or module 7 C domains, and that the module 7 C domain could accept both module 3 and module 6 carrier domains with their respective tri- and hexapeptide intermediates. In either case, C domain 7 could not discriminate tri- and hexapeptidyl intermediates.

Gene knock-out experiments have identified a new iterative tetra-modular NRPS, designated *HcpA*, in *Penicillium chrysogenum* and *A. niger*.<sup>101</sup> The *HcpA* gene consisted of 6064 aa and a domain architecture of A<sub>1</sub>-PCP<sub>1</sub>-E-C<sub>2</sub>-A<sub>4</sub>-A<sub>2</sub>-PCP<sub>2</sub>-C<sub>3</sub>-A<sub>3</sub>-PCP<sub>3</sub>-E-C<sub>4</sub>-PCP<sub>4</sub>-C-PCP that was unusual because the adjacent A<sub>2</sub> and A<sub>4</sub> domains flanked by a single C and PCP domain. Furthermore, the N-termini of these NRPSs appeared to be incomplete, lacking an adjacent A domain, leading to an incomplete module C<sub>4</sub>-PCP<sub>4</sub>. *HcpA* was confirmed to be responsible for encoding the biosynthesis of the metabolite fungisporin (39) along with a series of new cyclic and linear peptides (40-51). Microheterogeneity of each adenylation domain towards different aa monomers resulted in a range of cyclic tetrapeptides produced by this single NRP synthetase. *HcpA* was 54% identical with an orthologous protein from *A. niger* that had the same module organization. Furthermore, all cyclic products (39-48) present in *P. chrysogenum* could be detected in the culture broth of an *A. niger* strain. Parallel deletion of *HcpA* in this strain abolished their production leading to the conclusion that *HcpA* encodes for the production the same series of metabolites in *A. niger*.

## 6 New peptides and depsipeptides and their bioactivities

New peptides and depsipeptides continue to be discovered either by empirical spectral probing of fermentation extracts or because of detection of interesting *in vitro* biological characteristics, such as antiplasmodial, antiviral, insecticidal, cytotoxic, and antiproliferative properties. In a few cases, in-depth investigation of their biology has led to the discovery of unique modes of action hypotheses.

### 6.1 Peptides

A fungal cell wall inhibition assay measuring differential response in a two-plate assay with a wild-type and a wall-less mutant of *Neurospora crassa* led to the isolation of a linear hexapeptide (52) (L-threonine, D-alanine, L-alanine, L-alanine, D-tyrosine, L-valine) with a  $\delta$ -hydroxymyristic acid amide substituted N-terminus on the threonine residue from a strain of *Pochonia bulbinosa*.<sup>102</sup> Although the peptide strongly inhibited fungal growth and caused protoplast formation in some fungi, it failed to inhibit chitin synthase or  $\beta$ -1,3-glucan synthetase, suggesting that its mechanism may be via a regulatory unit of cell wall polymerase enzyme complex. A new cyclic heptapeptide, unguisin E (53), comprised of cyclo alanyl-tryptophyl-4-aminobutanoyl-alanyl-valyl- $\beta$ -methyl-phenylalanyl-valyl, was isolated from the fermentation an

unidentified *Aspergillus* strain.<sup>103</sup> Fermentation extracts of *A. sclerotiorum* yielded 11 new aspochracin-type cyclic tripeptides, sclerotiotides A-K.<sup>104</sup> Sclerotiotide A (54) consisted cyclo N-Me-L-alanine, L-valine, NR-octa-2,4,6-trienoyl-L-ornithine. Sclerotiotides A, B, F, and I selectively inhibited growth of *Candida albicans*. A strain of *Penicillium canescens* yielded the linear tetrapeptide D-phenylalanine, L-valine, D-valine, L-tryosine (55).<sup>105</sup> The peptide inhibited growth of *Fusarium virguliforme* and *Bacillus subtilis*.

An unidentified vegetative *Xylaria* strain produced two new cyclopentapeptides (56, 57).<sup>106</sup> These pentapeptides had an aa sequence of cyclo N-methyl-L-phenylalanine, L-valine or L-isoleucine, D-isoleucine, L-leucine, L-proline. Although not specifically antifungal, the peptide enhanced the activity of ketoconazole. A novel cyclic octapeptide, epichlicin (58) was found from *Epichloe typhina*, a biotrophic symbiont of timothy (*Phelum pratense*).<sup>107</sup> The cyclic peptide sequence consisted of 3-amino-tetradecanoic acid, L-asparagine, L-tyrosine, L-asparagine, L-glutamine, L-serine, L-asparagine, and L-proline. Although its role in the symbiosis remains unknown, the peptide inhibited conidial germination of *Cladosporium pheii*, a pathogen of timothy.

The family of fellutamide lipopeptide-aldehyde proteasome inhibitors has continued to expand. Fellutamides C (59) and D (60) were discovered by screening with the genome-wide *Candida albicans* fitness test and analysis of the haploinsufficient heterozygote test set indicated that their mode of action was mediated through inhibition of the eukaryote 20S proteasome complex.<sup>108</sup> The new fellutamides were produced from an undescribed species of *Metulocladosporiella*. The specific active extract was found by screening microfermentations and standard flask fermentations in which only one of 12 conditions produced fellutamides. Another new fellutamide, designated fellutamide F (61) was isolated by cytotoxicity-guided fractionation of an extract from *A. versicolor*.<sup>109</sup>

*Pseudallescheria boydii*, a common environmental fungus and invasive fungal pathogen, produced a family of cyclic peptides designated pseudacyclins A-E.<sup>110</sup> The peptide sequence of pseudacyclin A (62) was comprised of three isoleucine residues (one N-acetylated) and one residue each of phenylalanine, ornithine, and proline. The cyclopeptide could also be detected in the fungal conidia by semiquantitative MS. Besides the pseudacyclin family, cyclic peptide of phenylalanyl-proline was found to be a major metabolite in one of investigated strains of *P. boydii*. Pseudacyclins A was weakly cytotoxic to activated lymphocytes, but not resting lymphocytes. *Onychocola sclerotica* is a rare keratinophilic fungus. Arrays of microfermentation extracts from the type strain were examined to survey its metabolic capacity. In one of the eight media, LC-UV-MS analysis detected three major compounds unrecognized by an in-house microbial natural products dereplication database. The most abundant was determined to be three cyclotetrapeptides (63-65) with an aa sequence of L-N-methyl-phenylalanine, L-valine or L-isoleucine, L-N-methyl-phenylalanine, and L-valine or L-isoleucine.<sup>111</sup> The compounds were moderately active in cardiac calcium channel assay, but were not cytotoxic or antibiotic.

Structures 52 to 73 here (Figs for NPR-part6.1)

An investigation of a strain of *Asteromyces cruciatusis* is a rare example of a new metabolite from an obligate marine

fungus. The fungus was grown in a battery of different media conditions intended to diversify the production of low abundance metabolites; UV-spectral analysis led to the isolation and purification of a novel pentapeptide molecule, lajollamide A (**66**), consisting of valine, N-methylleucine, and three leucine residues.<sup>112</sup> The compound was weakly antibacterial towards Gram-positive bacteria. During isolation of cytotoxic cytochalasins from an unidentified *Penicillium* sp., penilumamide (**67**), a novel lumazine peptide was also isolated.<sup>113</sup> Penilumamide consisted of a unique combination of a 1,3-dimethylumazine-6-carboxylic acid, a methionine sulfoxide group and an anthranilic acid methyl ester. No biological activity was identified for the molecule.

Secondary metabolite phenotypes of nine species of the *Hamigera* were analysed and correlated with phylogeny.<sup>114</sup> HPLC-UV-based chemical analysis revealed distinctive intrageneric patterns of secondary metabolite production. Fifteen of 19 strains among 9 species produced one or more of the peptides, avellanins A (**68**) and B (**69**). Avellanins A and B consist of five aa: alanine, phenylalanine, proline and anthranilic acid as common monomers and isoleucine for avellanin A and valine for avellanin B. *Hamigera paravellanea* and *H. ingelheimensis* produced a new pentapeptide avellanin C (**70**) in which the isoleucine residue of avellanin A was replaced by phenylalanine. A biosynthetically related series of peptides, PF1171B (**71**), the leucine-version of the pentapeptide avellanin A, and PF1171A (**72**) and PF1171C (**73**) were also variably present among several species.

## 6.2 Depsipeptides

Verticilide A1 (**74**) is a cyclodepsipeptide related to bassianolide and enniatin.<sup>115, 116</sup> Acyl-CoA:cholesterol acyltransferase (ACAT) has been proposed as a target for a new type of anti-atherosclerotic agent. Screening for ACAT2-selective inhibitors with a CHO cell-based assay discovered a series of new verticilide cyclodepsipeptides, A2 (**75**), A3 (**76**) and B1 (**77**).<sup>117</sup> Extracts obtained from mixed fermentation of *Fusarium tricinctum* and *Fusarium begonia* on white-bean medium were analysed by HPLC, and compared to those from fermentation of monocultures on the same medium. HPLC-UV-spectra indicated extracts of the mixed culture fungi differed significantly.<sup>118</sup> These induced metabolites were determined to be new non-cyclic forms of enniatins and were designated, subenniatis A (**78**) and B (**79**). Unlike, the cyclic enniatins, the linear forms were determined to be inactive in antimicrobial and cytotoxicity assays.

Structures 74 to 91 here (*Figs for NPR-part6.2-1*)

Pullularin A (**80**) was isolated as the most abundant cyclodepsipeptide from an ethyl acetate extract of an *Aureobasidium* sp.<sup>119</sup> In addition to pullularins A, C (**81**) and D (**82**), new pullularins, E (**83**) and F (**84**), were isolated from a strain of the ubiquitous *Bionectria ochroleuca* (previously named *Gliocladium roseum*), from the mangrove *Sonneratia caseolaris*.<sup>120</sup> Pullularin E (**83**) had an aa sequence of D-3-phenyl-lactate, L-proline, L-alanine, N-methyl-L-isoleucine, O-prenyl-L-tyrosine, N-methyl-L-alanine. Subsequent testing of the pullularins demonstrated they were cytotoxic towards a murine lymphoma cell line. F04W2166A, a proteasome inhibitor from a *Verticillium* strain is apparently equivalent to pullularin C (**81**), and therefore proteasome inhibition may be responsible for the cytotoxicity observed in previous studies.<sup>121</sup>

Fusaristatins A (**85**) and B (**86**) are cyclic lipotetradepsipeptides consisting of glutamine or 2-amino adipic acid, dehydroalanine and  $\beta$ -aminoisobutyric acid and a  $\beta$ -hydroxy C24 fatty acid isolated from an unidentified *Fusarium* strain.<sup>122</sup> Although the fusaristatins were not antibiotic to common human pathogenic yeasts and bacteria, their similarity to the cyclic lipopeptide topostatin<sup>123</sup> prompted testing of their activity towards human DNA topoisomerases I and II by employing relaxation assays with supercoiled plasmid DNA. Fusaristatin B was found to be weakly active against calf thymus topoisomerase I and human topoisomerase II.

The *Candida albicans* genome-wide fitness test revealed two mechanistically related antifungal cyclic lipodepsipeptides.<sup>124, 125</sup> The first of these, phomafungin (**87**), was observed in multiple strains of a widespread *Phoma*-like fungus.<sup>124</sup> Phomafungin is an octapeptide linked to a  $\beta$ -hydroxy- $\gamma$ -methyl-hexadecanoic acid. Phomafungin-containing extracts and purified phomafungin produced a complex and unique response in the population of *C. albicans* haploinsufficient heterozygotes resulting in four hypersensitive strains that corresponded to genes involved in sphingolipid biosynthesis (LCB2, RTA2/RSB1, MIT1/SUR1/CSH1, AUR1) along with strains involved in the regulation of  $Ca^{2+}$ , e.g. calmodulin (CMD1),  $Ca^{2+}$ -binding protein (FRQ1) and a vacuolar  $Ca^{2+}$ -ATPase involved in ion regulation (PMCI).<sup>124</sup> Later, during investigations of the mechanism of action of phaeofungin (**88**), phomafungin (**87**) was observed to potentiate the effects of cyclosporin A, indicating it may exert its antifungal activity through the calcineurin pathway. Phaeofungin (**88**), from a *Phaeosphaeria*-like fungus, consists of seven aa and a  $\beta,\gamma$ -dihydroxy- $\gamma$ -methylhexadecanoic acid. Response of heterozygote strains to phaeofungin in the *Candida albicans* fitness test partially overlapped with that of phomafungin, in particular, affecting heterozygotes deficient in genes involved in sphingolipid biosynthesis. It also induced hypersensitivity of two GSK1+/- strains and a group of double heterozygous deletion strains implicated in the action of inhibitors of fungal glucan synthase. Consistent with these effects on heterozygotes for sphingolipid and glucan synthase genes, phaeofungin synergized the activity of aureobasidin A, a sphingolipid biosynthesis inhibitor, and caspofungin, a glucan synthase inhibitor. The authors speculated that phomafungin's and phaeofungin's antifungal effects might be differentially exerted through the *C. albicans* cell membrane resulting in changes in sphingolipid content,  $Ca^{2+}$  concentration, and cell wall integrity.

Organic extracts of a rice-cultivation of *Eupenicillium javanicum* were highly and specifically inhibitory to growth of *A. fumigatus*. Bioassay-guided purification yielded a cyclic depsipeptide consisting of nine aa linked by a lactic acid residue that was named eujavanicin A (**89**).<sup>126</sup> A fungal isolate (MSX 51320), tentatively identified as a *Trichothecium* sp. by sequence homology of LSU rDNA gene and by production of some known *Trichothecium*-associated metabolites, produced cyclodepsipeptides designated as trichodepsipeptides A (**90**) and B (**91**).<sup>127</sup> Despite the acute cytotoxicity of other fermentation extract components, trichodepsipeptides A and B were inactive in a human tumor cell-line panel. A cyclic pentadepsipeptide, alternaramide (**92**) consisting of  $\alpha$ -L-hydroxyisovaleric acid, L-proline, D-phenylalanine, L-proline, D-phenylalanine was isolated from an unidentified *Alternaria* strain.<sup>128</sup> The peptide, designated alternaramide, was weakly inhibitory to Gram-positive bacteria.

Structures 92-95 here (*Figs for NPR-part6.2-2*)

New cyclic depsipeptides, JBIR-113 (**93**), JBIR-114 (**94**) and JBIR-115 (**95**) were isolated from a nutrient-supplemented rice culture of *Penicillium* sp. based on spectral analysis.<sup>129</sup> An attempt to associate the compounds with a biological function, found that the compounds were not cytotoxic to a human cell lines, nor inhibited growth of a Gram-positive or a Gram-negative bacterium.

### 6.3 Peptaibols

Although the chemical diversity of peptaibols and peptaibiotics are among the best-studied of all fungal NRPs, new peptide sequences continue to be reported. Three strains of *Trichoderma atroviride* were found to produce typical 19-residue peptaibols, however, three of them exhibited unusual peptidic sodium-adduct  $[M + 2Na]^{2+}$  ions.<sup>130</sup> Peptide sequencing revealed two series of unprecedented 17-residues with an atypical C-terminus (e.g., **96**) and, thus, they were recognized as a new peptaibiotic family. Comparison of sequences for 17- and 19-residue peptides showed that positions 1 to 16 were similar, but differences occurred at position 17 where the typical glutamine was replaced with an unknown residue of a MW of 129 Da. The authors hypothesized that this new unexpected residue, possibly dihydro-glutamine or N<sup>6</sup>-hydroxyornithine, may have prematurely terminated peptide elongation. These 17-residue peptaibiotics were weakly toxic towards human oral epidermoid carcinoma cells. *Trichoderma citrinoviride*, associated with *Quercus suber* (cork oak), has been observed to antagonize various fungal stem pathogens involved in oak decline. Fermentation extracts were screened for the presence of antagonistic secondary metabolites. From liquid culture, a mixture of peptaibols of the paracelsin family was characterised.<sup>131</sup> Twenty-eight 20-mer peptides with C-terminal phenylalaninol and N-terminal acetylation were identified, and aa substitutions occurred at positions 6, 9, 12 and 17, including seven new peptide sequences (e.g., **97**). As expected, the peptide mixture was strongly inhibitory to growth of the oak bark pathogen *Biscogniauxia mediterranea*.

Structures 96 to 103 here (*Figs for NPR-part6.3.*)

Suzukacillins are one of oldest known peptaibol families.<sup>132</sup> The classic suzukacillin-producing strain *Trichoderma viride* strain 63 C-1 produced two microheterogeneous groups of suzukacillin peptaibols in fermentations.<sup>133</sup> Both groups were easily distinguishable by TLC, and the major group was designated suzukacillin-A and the minor group suzukacillin-B. Fourteen peptides comprised of the C-terminal sequence Pro<sup>6</sup>-Lxx-Lxx-Aib-Pro-Vxxol/Lxxol were identified among the latter group,<sup>11</sup> typical for 11-mer peptaibols and one peptide tentatively identified as a ten-residue sequence, where the C-terminal 1,2-amino alcohol was missing, thus terminating in free proline. Trichoderins A (**98**), A1 (**99**) and B (**100**) are new aminolipopeptaibiotics from a *Trichoderma* sp. and are related to leucinostatins, trichopolyns, helioferins, and roseoferins.<sup>134</sup> They were discovered with a screening system against dormant mycobacteria and demonstrated potent anti-mycobacterial activity against *Mycobacterium smegmatis*, *M. bovis* and *M. tuberculosis* under both aerobic conditions and latency-inducing hypoxic conditions. Rice-culture extracts of two strains of fungi of the Hypocreales were investigated for the cause of cytotoxicity to a human non-small cell lung

carcinoma cell line. Extract fractionation yielded series of 12 peptaibols, eight of which represented new sequences (e.g., atroviridin D **101**).<sup>136</sup> The set of peptaibols exhibited a range of bioactivities against tumour cell lines, Gram-positive bacteria, and *Haemonchus contortus* larval motility. Screening for molecules inhibiting *Trypanosoma brucei* led to the discovery of new trichosporins B-VIIa (**102**) and B-VIIb (**103**) from a *Trichoderma polysporum*.<sup>137</sup>

### 7 Potential for novel peptides in newly sequenced and annotated fungal genomes

A global overview of NRPS evolution and diversification in the Fungi is now emerging, and significant progress has been made in developing a predictive framework for classifying fungal NRPSs.<sup>5</sup> Few NRPSs are fully conserved across ascomycete fungi.<sup>22, 138</sup> As a result, the high diversity of domain structures, and the patchy distribution of equivalent A domains complicates ortholog recognition among different fungal species. As illustrated above, NRP biosyntheses in model fungi, major human, animal, and plant pathogens has received considerable attention because of its involvement in pathogenicity and the origin of drug-like peptides. However, genome sequencing projects covering the whole spectrum of environmental fungi are providing fertile ground for directed peptide discovery through pathway prediction, activation, heterologous expression, and pathway engineering.

Figure 4 here

Fig. 4. Frequency of putative NRPS gene clusters in selected genome sequenced species of the Basidiomycota. NRPS gene clusters with >5 modules are absent.

Recently one of the first evaluations of NRPS complexity in a lower filamentous ascomycete, *Pyronema confluens*, has confirmed that typical Ascomycete secondary metabolites are underrepresented in the Pezizales.<sup>139</sup> This fungus possessed seven putative NRPS genes, and one PKS gene, substantially fewer than in the genomes of the kinds of Ascomycetes described above. PCON\_02859 was predicted to encode an NRPS which had the typical domain structure of siderophore NRPSs and was part of a cluster of gene homologues which are involved in siderophore biosynthesis in other fungi. A second putative NRPS gene (PCON\_07777) was not part of cluster and lacked homology to NRPSs with known functions. The remaining five NRPS genes all had a domain structure that is typical for alpha-amino adipate reductase (AAR)-type NRPSs, and (with the exception of PCON\_04030) all have high sequence similarity to amino adipate semialdehyde dehydrogenase, an enzyme for lysine biosynthesis that is conserved in fungi. Curiously, most fungi usually have only a single AAR-type NRPS, therefore the multiplicity of AAR-type NRPS genes in *P. confluens* presently appears to be an anomaly. Possible reason for this duplicity may be selective amplification of the specific gene family or the loss of other NRPS genes leaving only AAR-type NRPS genes. One of these AAR-type NRPS genes, PCON\_04030 was located adjacent to the single PKS gene (PCON\_04029) in a gene cluster that also contained other genes encoding enzymes that might participate in the biosynthesis of a hybrid PK-NRP, rather than in lysine biosynthesis. The existence of such structured gene clusters encoding separate PKS and NRPS proteins that act in a common biosynthetic pathway is reminiscent of the cluster responsible for the production of emericellamide and the

echinocandins. Such examples are fertile targets for new metabolite discovery through genome mining techniques.

Figure 5 here

Fig. 5. Frequency of putative NRPS gene clusters in selected genome sequenced species of the Ascomycota. NRPS gene clusters with >5 modules are indicated in maroon.

Many single modular NRPS-like proteins do not consist of the typical A-T-C module architecture, but often consist of an A-T bidomain followed by a variety of C-terminal domains.<sup>22, 138</sup> These monomodular NRPS-like proteins may or may not be involved in secondary metabolite biosynthesis; for example, the amino adipate reductase LYS2 in lysine biosynthesis pathway which terminates with a thioester reductase domain. Macrocyclic peptides and peptaibiotics have a high affinity for biological targets<sup>48, 140</sup> and would be desirable objectives for genome mining and gene cluster activation. The Norine and The Comprehensive Peptaibiotics Database projects have plotted size distributions for published fungal peptides.<sup>52, 60</sup> As a simple illustration, we analysed the frequency and species distribution of large peptide synthases (>5 A domains) and in a selection of recently sequenced fungal genomes as an example for future exploration, and as a retrospective analysis of whether natural products discovery paradigms have been effective in unearthing these generally highly bioactive molecules (Figs. 4, 5).

Because the correlation of NRPS genes with their products has become straightforward, the focus now will shift to the question of how best to obtain the products of these pathways in sufficient quantities and channel them into screening libraries where their chemical biology can be explored. Although vast numbers of predicted NRPS genes have been annotated in fungal genomes, still only a few dozen pathways have been identified and characterised experimentally, including those for cryptic or silent gene loci. Several strategies have been developed for discovering the products of these biosynthetic gene clusters. As illustrated in the above examples, gene knockouts are probably the most direct way. For example, Chiang et al.<sup>69</sup> randomly deleted six NRPS genes in *A. nidulans*, and compared the metabolic profiles of the wild-type strain with the six deletion mutants. They found five compounds disappeared in the AN2545.3 deletion mutant. However, four new compounds appeared. After isolation and structural characterisation; they were found to be emericellamides C-F. In some cases, biosynthetic gene clusters will be silent or their expression levels will be very low, making metabolites difficult to detect in typical fermentation conditions, and particularly if many other metabolites are present. For this situation, other strategies would be needed, for example induced expression of transcriptional activator and heterologous gene expression. A PKS-NRPS hybrid gene cluster (*apdA*, *apdB*, *apdC*, *apdD*, *apdE*, *apdG* and *apdR*) in *A. nidulans* was activated by induced expression of transcriptional activator ApdR, and two new compounds were found and named aspyridones A-B.<sup>141</sup> Xu et al.<sup>142</sup> expressed the ApdA (PKS-NRPS hybrid) and ApdC (ER) in *Saccharomyces cerevisiae*, reconstituted their functions *in vitro* and in *S. cerevisiae*, and isolated a new compound named preaspyridone 7. Wasil et al.<sup>143</sup> heterologously expressed genes from the aspyridone biosynthetic cluster in *A. oryzae*, and detected eight different compounds in addition to aspyridone A. Moreover, Halo et al.<sup>144</sup> expressed *tens*, a PKS-NRPS hybrid gene

responsible for the production of tenellin, with *orf3* (*trans*-acting ER) in *A. oryzae* and obtained three new compounds. These successful experiments have demonstrated that heterologous gene expression is an efficient method to attain the target gene's product. However, many fungal NRPSs contain more than five A domains which means their genes exceed 15 kb, and the largest ones can consist of 19 A domains encoded by a 66 kb gene (Fig. 5). Such large genes are very difficult to amplify and heterologously express with conventional molecular biology methods. Gressler et al.<sup>145</sup> outlined a strategy that might work around this limitation. They searched for the compound biosynthesized by PKS-NRPS hybrid ATEG00325 in *A. terreus*. They constructed a *lacZ* reporter strain of *A. terreus* to reveal conditions for expression, which finally led to the isolation of three compounds, isoflavipucine, dihydroisoflavipucine, and a phytotoxin, under specific physiological conditions.

Finally, as in any other natural products discovery process, after overcoming the technical challenges of activation, expression, detection, and production of the products, one needs to seek biological functions. Therefore, a major challenge for the field as a whole remains overcoming the logistical, infrastructural and administrative barriers that prevent flow of compounds from the laboratories of individual research groups into the appropriate screening centres where they can be efficiently accessed by biologists. Taking full advantage of these new sources of molecules will require the building of cohesive scientific teams with creative know-how in exploiting sources of natural chemical novelty and biological functions, along with those seeking new molecular entities.

## Acknowledgements

This work was sponsored by the National Basic Research Program of China (2013CB127505), National Natural Science Foundation of China (U1036602, 31328001 to ZA), the National High Technology Research and Development Program of China (2011AA10A205), a scholarship awarded to XN (201207035005) from China Scholarship Council, the University of Texas Health Science Centre at Houston new faculty start-up funds (to GFB), the Texas Emerging Technology Fund and the Welch Foundation (AU00024 to ZA).

## Notes and references

<sup>a</sup> Texas Therapeutics Institute, The Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Centre at Houston, Houston, Texas 77054, USA.

<sup>b</sup> Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming, 650091, P. R. China.

1. T. Degenkolb, W. Gams and H. Brückner, *Chem. Biodivers.*, 2008, **5**, 693-706.
2. H. Anke and L. Antelo, in *Physiology and Genetics*, eds. T. Anke and D. Weber, Springer Berlin Heidelberg 2009, vol. 15, ch. 13, pp. 273-296.
3. G. H. Hur, C. R. Vickery and M. D. Burkart, *Nat. Prod. Rep.*, 2012, **29**, 1074-1098.
4. H. L. Conurso and S. D. Bruner, *Nat. Prod. Rep.*, 2012, **29**, 1099-1110.
5. D. Kalb, G. Lackner and D. Hoffmeister, *Fungal Biol. Rev.*, 2013, **27**, 43-50.

6. C. Boddy, *J. Ind. Microbiol. Biotechnol.*, 2013, 1-8.
7. R. Süßmuth, J. Müller, H. Von Döhren and I. Molnár, *Nat. Prod. Rep.*, 2011, **28**, 99-124.
8. C. T. Walsh, S. W. Haynes, B. D. Ames, X. Gao and Y. Tang, *ACS Chem. Biol.*, 2013, **8**, 1366-1382.
9. J. M. Finefield, J. C. Frisvad, D. H. Sherman and R. M. Williams, *J. Nat. Prod.*, 2012, **75**, 812-833.
10. Q. Gao, K. Jin, S. H. Ying, Y. Zhang, G. Xiao, Y. Shang, Z. Duan, X. Hu, X. Q. Xie, G. Zhou, G. Peng, Z. Luo, W. Huang, B. Wang, W. Fang, S. Wang, Y. Zhong, L. J. Ma, R. J. St. Leger, G. P. Zhao, Y. Pei, M. G. Feng, Y. Xia and C. Wang, *PLoS Genet.*, 2011, **7**.
11. B. Wang, Q. Kang, Y. Lu, L. Bai and C. Wang, *Proc. Natl. Acad. Sci. USA*, 2012, **109**, 1287-1292.
12. K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano and T. Weber, *Nucl. Acids Res.*, 2013, **41**, W204-W212.
13. D. Boettger and C. Hertweck, *ChemBioChem*, 2013, **14**, 28-42.
14. C. L. Schardl, C. A. Young, U. Hesse, S. G. Amyotte, K. Andreeva, P. J. Calie, D. J. Fleetwood, D. C. Haws, N. Moore, B. Oeser, D. G. Panaccione, K. K. Schweri, C. R. Voisey, M. L. Farman, J. W. Jaromczyk, B. A. Roe, D. M. O'Sullivan, B. Scott, P. Tudzynski, Z. An, E. G. Arnaudova, C. T. Bullock, N. D. Charlton, L. Chen, M. Cox, R. D. Dinkins, S. Florea, A. E. Glenn, A. Gordon, U. Güldener, D. R. Harris, W. Hollin, J. Jaromczyk, R. D. Johnson, A. K. Khan, E. Leistner, A. Leuchtmann, C. Li, J. G. Liu, J. Liu, M. Liu, W. Mace, C. Machado, P. Nagabhyru, J. Pan, J. Schmid, K. Sugawara, U. Steiner, J. E. Takach, E. Tanaka, J. S. Webb, E. V. Wilson, J. L. Wiseman, R. Yoshida and Z. Zeng, *PLoS Gen.*, 2013, **9**.
15. C. J. Balibar, A. R. Howard-Jones and C. T. Walsh, *Nat. Chem. Biol.*, 2007, **3**, 584-592.
16. K. Qiao, H. Zhou, W. Xu, W. Zhang, N. Garg and Y. Tang, *Org. Lett.*, 2011, **13**, 1758-1761.
17. W. B. Yin, A. Grundmann, J. Cheng and S. M. Li, *J. Biol. Chem.*, 2009, **284**, 100-109.
18. W. Jiang, R. A. Cacho, G. Chiou, N. K. Garg, Y. Tang and C. T. Walsh, *J Am Chem Soc*, 2013, **135**, 4457-4466.
19. R. A. Cacho, W. Jiang, Y. H. Chooi, C. T. Walsh and Y. Tang, *J Am Chem Soc*, 2012, **134**, 16781-16790.
20. L. Chen, Q. Yue, X. Zhang, M. Xiang, C. Wang, S. Li, Y. Che, F. J. Ortiz-Lopez, G. F. Bills, X. Liu and Z. An, *BMC Genom.*, 2013, **14**, 339.
21. K. E. Bushley, R. Raja, P. Jaiswal, J. S. Cumbie, M. Nonogaki, A. E. Boyd, C. A. Owensby, B. J. Knaus, J. Elser, D. Miller, Y. Di, K. L. McPhail and J. W. Spatafora, *PLoS Genet.*, 2013, **9**, e1003496.
22. K. E. Bushley and B. G. Turgeon, *BMC Evolutionary Biology*, 2010, **10**, 26.
23. L. Johnson, *Mycol. Res.*, 2008, **112**, 170-183.
24. H. Haas, M. Eisendle and B. G. Turgeon, *Ann. Rev. Phytopath.*, 2008, **46**, 149-187.
25. H. Haas, *Front. Microbiol.*, 2012, **3**, 28.
26. J. C. Renshaw, G. D. Robson, A. P. J. Trinci, M. G. Wiebe, F. R. Livens, D. Collison and R. J. Taylor, *Mycol. Res.*, 2002, **106**, 1123-1142.
27. R. Lemmens-Gruber, M. R. Kamyar and R. Dornetshuber, *Curr. Med. Chem.*, 2009, **16**, 1122-1137.
28. M. Blatzer, M. Schrettl, B. Sarg, H. H. Lindner, K. Pfaller and H. Haas, *Appl. Environ. Microbiol.*, 2011, **77**, 4959-4966.
29. W. B. Yin, J. A. Baccile, J. W. Bok, Y. Chen, N. P. Keller and F. C. Schroeder, *J. Amer. Chem. Soc.*, 2013, **135**, 2064-2067.
30. B. N. Lee, S. Kroken, D. Y. T. Chou, B. Robbertse, O. C. Yoder and B. G. Turgeon, *Euk. Cell.*, 2005, **4**, 545-555.
31. L.-H. Chen, C.-H. Lin and K.-R. Chung, *Molec. Plant Path.*, 2013, **14**, 497-505.
32. C. Hof, K. Eisfeld, L. Antelo, A. J. Foster and H. Anke, *Fungal Gen. Biol.*, 2009, **46**, 321-332.
33. B. Winterberg, S. Uhlmann, U. Linne, F. Lessing, M. A. Marahiel, H. Eichhorn, R. Kahmann and J. Schirawski, *Molec. Microbiol.*, 2010, **75**, 1260-1271.
34. A. Munawar, J. W. Marshall, R. J. Cox, A. M. Bailey and C. M. Lazarus, *ChemBioChem*, 2013, **14**, 388-394.
35. K. Haselwandter, *Mineral. Mag.*, 2008, **72**, 61-64.
36. K. Haselwandter, V. Passler, S. Reiter, D. G. Schmid, G. Nicholson, P. Hentschel, K. Albert and G. Winkelmann, *Biometals*, 2006, **19**, 335-343.
37. L. J. Johnson, A. Koulman, M. Christensen, G. A. Lane, K. Fraser, N. Forester, R. D. Johnson, G. T. Bryan and S. Rasmussen, *PLoS Path.*, 2013, **9**.
38. A. Koulman, T. V. Lee, K. Fraser, L. Johnson, V. Arcus, J. S. Lott, S. Rasmussen and G. Lane, *Phytochem.*, 2012, **75**, 128-139.
39. T. V. Lee, L. J. Johnson, R. D. Johnson, A. Koulman, G. A. Lane, J. S. Lott and V. L. Arcus, *J. Biol. Chem.*, 2010, **285**, 2415-2427.
40. I. Stergiopoulos, J. Collemare, R. Mehrabi and P. J. De Wit, *FEMS Microbiol. Rev.*, 2013, **37**, 67-93.
41. T. Tsuge, Y. Harimoto, K. Akimitsu, K. Ohtani, M. Kodama, Y. Akagi, M. Egusa, M. Yamamoto and H. Otani, *FEMS Microbiol. Rev.*, 2013, **37**, 44-66.
42. M. S. Pedras, C. J. Biesenthal and I. L. Zaharia, *Plant Sci.*, 2000, **156**, 185-192.
43. R. Y. Parada, K. Oka, D. Yamagishi, M. Kodama and H. Otani, *Physiol. Mol. Plant Path.*, 2007, **71**, 48-54.
44. M. Zabka, K. Drastichova, A. Jegorov, J. Soukupova and L. Nedbal, *Mycopathol.*, 2006, **162**, 65-68.
45. F. T. Hansen, A. Droce, J. L. Sørensen, P. Fojan, H. Giese and T. E. Sondergaard, *Fungal Biol.*, 2012, **116**, 855-862.
46. B. Gillian Turgeon, S. Oide and K. Bushley, *Mycol. Res.*, 2008, **112**, 200-206.
47. K. H. Kim, Y. Cho, M. La Rota, R. A. Cramer Jr and C. B. Lawrence, *Molec. Plant Path.*, 2007, **8**, 23-39.
48. T. Degenkolb and H. Brückner, *Chem. Biodivers.*, 2008, **5**, 1817-1843.
49. H. Ducholier, *Chemistry and Biodiversity*, 2007, **4**, 1023-1026.
50. B. Leitgeb, A. Szekeres, L. Manczinger, C. Vágvolgyi and L. Kredics, *Chem. Biodivers.*, 2007, **4**, 1027-1051.
51. C. Röhrich, W. Jaklitsch, H. Voglmayr, A. Iversen, A. Vilcinskas, K. Nielsen, U. Thrane, H. von Döhren, H. Brückner and T. Degenkolb, *Fungal Diversity*, 2014, In press.
52. N. Stoppacher, N. K. N. Neumann, L. Burgstaller, S. Zeilinger, T. Degenkolb, H. Brückner and R. Schuhmacher, *Chem. Biodivers.*, 2013, **10**, 734-743.

53. C. R. Röhrich, A. Iversen, W. M. Jaklitsch, H. Voglmayr, A. Vilcinskas, K. F. Nielsen, U. Thrane, H. Von Döhren, H. Brückner and T. Degenkolb, *Chem. Biodivers.*, 2013, **10**, 787-812.
54. C. R. Röhrich, A. Iversen, W. M. Jaklitsch, H. Voglmayr, A. Berg, H. Dörfelt, U. Thrane, A. Vilcinskas, K. F. Nielsen, H. von Döhren, H. Brückner and T. Degenkolb, *Fungal Biol.*, 2012, **116**, 1219-1231.
55. J. Pan, S. Tristram-Nagle and J. F. Nagle, *J. Membrane Biol.*, 2009, **231**, 11-27.
56. S. Bobone, Y. Gerelli, M. De Zotti, G. Bocchinfuso, A. Farrotti, B. Orioni, F. Sebastiani, E. Latter, J. Penfold, R. Senesi, F. Formaggio, A. Palleschi, C. Toniolo, G. Fragneto and L. Stella, *Biochim. Biophys. Act.*, 2013, **1828**, 1013-1024.
57. M. Shi, L. Chen, X.-W. Wang, T. Zhang, P.-B. Zhao, X.-Y. Song, C.-Y. Sun, X.-L. Chen, B.-C. Zhou and Y.-Z. Zhang, *Microbiol.*, 2012, **158**, 166-175.
58. S. Caboche, M. Pupin, V. Leclère, A. Fontaine, P. Jacques and G. Kucherov, *Nucl. Acids Res.*, 2008, **36**, D326-D331.
59. S. Caboche, M. Pupin, V. Leclere, P. Jacques and G. Kucherov, *BMC Struct. Biol.*, 2009, **9**, 15.
60. S. Caboche, V. Leclère, M. Pupin, G. Kucherov and P. Jacques, *J. Bacteriol.*, 2010, **192**, 5143-5150.
61. X. Gao, S. W. Haynes, B. D. Ames, P. Wang, L. P. Vien, C. T. Walsh and Y. Tang, *Nat. Chem. Biol.*, 2012, **8**, 823-830.
62. M. Hino, A. Fujie, T. Iwamoto, Y. Hori, M. Hashimoto, Y. Tsurumi, K. Sakamoto, S. Takase and S. Hashimoto, *J. Industr. Microbiol. Biotechnol.*, 2001, **27**, 157-162.
63. T. Emri, L. Majoros, V. Tóth and I. Pócsi, *Appl. Microbiol. Biotechnol.*, 2013, **97**, 3267-3284.
64. J. M. Balkovec, D. L. Hughes, P. Masurekar, C. A. Sable, R. A. Schwartz and S. B. Singh, *Nat. Prod. Rep.*, 2013, **31**, 15-34.
65. S. Hashimoto, *J. Antibiotics*, 2009, **62**, 27-35.
66. F. Peláez, J. Collado, G. Platas, D. P. Overy, J. Martín, F. Vicente, A. González del Val, A. Basilio, M. De la Cruz, J. R. Tormo, A. Fillola, F. Arenal, M. Villareal, V. Rubio, H. O. Baral, R. Galán and G. F. Bills, *Mycology*, 2011, **2**, 1-17.
67. N. Connors and D. Pollard, in *Handbook of Industrial Mycology*, ed. Z. An, Marcel Dekker, New York 2005, pp. 515-538.
68. L. Petersen, R. Olewinski, P. Salmon and N. Connors, *Appl. Microbiol. Biotechnol.*, 2003, **62**, 263-267.
69. Y. M. Chiang, E. Szewczyk, T. Nayak, A. D. Davidson, J. F. Sanchez, H. C. Lo, W. Y. Ho, H. Simityan, E. Kuo, A. Praseuth, K. Watanabe, B. R. Oakley and C. C. C. Wang, *Chem. Biol.*, 2008, **15**, 527-532.
70. K. Mizuno, A. Yagi, S. Sato, M. Takada and M. Hayashi, *J. Antibiotics*, 1977, **30**, 297-302.
71. S. Sato, A. Yagi, K. Asano, K. Mizuno and T. Watanabe, *J. Antibiotics*, 1977, **30**, 303-307.
72. M. Bogale, M. J. Orr, M. J. O'Hara and W. A. Untereiner, *Fungal Biol.*, 2010, **114**, 396-409.
73. R. Kanasaki, M. Kobayashi, K. Fujine, I. Sato, M. Hashimoto, S. Takase, Y. Tsurumi, A. Fujie, M. Hino, S. Hashimoto and Y. Hori, *J. Antibiotics*, 2006, **59**, 158-167.
74. R. Kanasaki, F. Abe, M. Kobayashi, M. Katsuoka, M. Hashimoto, S. Takase, Y. Tsurumi, A. Fujie, M. Hino, S. Hashimoto and Y. Hori, *J. Antibiotics*, 2006, **59**, 149-157.
75. R. Kanasaki, K. Sakamoto, M. Hashimoto, S. Takase, Y. Tsurumi, A. Fujie, M. Hino, S. Hashimoto and Y. Hori, *J. Antibiotics*, 2006, **59**, 137-144.
76. C. Gueidan, C. Ruibal, G. S. De Hoog and H. Schneider, *Fungal Biol.*, 2011, **115**, 987-996.
77. M. De La Cruz, J. Martín, V. González-Menéndez, I. Pérez-Victoria, C. Moreno, J. R. Tormo, N. El Aouad, J. Guarro, F. Vicente, F. Reyes and G. F. Bills, *Chem. Biodivers.*, 2012, **9**, 1095-1113.
78. G. Bills, G. Platas, A. Fillola, M. R. Jiménez, J. Collado, F. Vicente, J. Martín, A. González de Val, J. Bur-Zimmermann, J. R. Tormo and F. Peláez, *J. Appl. Microbiol.*, 2008, **104**, 1644-1658.
79. R. J. W. Cremllyn, *Chlorosulfonic Acid, A Versatile Reagent*, Royal Society of Chemistry, Cambridge, UK, 2002.
80. I. V. Grigoriev, D. Cullen, S. B. Goodwin, D. Hibbett, T. W. Jeffries, C. P. Kubicek, C. Kuske, J. K. Magnuson, F. Martin, J. W. Spatafora, A. Tsang and S. E. Baker, *Mycology*, 2011, **2**, 192-209.
81. S. W. Peterson, *Mycologia*, 2008, **100**, 205-226.
82. C. L. Schoch, G. H. Sung, F. Lopez-Giraldez, J. P. Townsend, J. Miadlikowska, V. Hofstetter, B. Robbertse, P. B. Matheny, F. Kauff, Z. Wang, C. Gueidan, R. M. Andrie, K. Trippe, L. M. Ciufetti, A. Wynns, E. Fraker, B. P. Hodkinson, G. Bonito, J. Z. Groenewald, M. Arzanlou, G. S. de Hoog, P. W. Crous, D. Hewitt, D. H. Pfister, K. Peterson, M. Gryzenhout, M. J. Wingfield, A. Aptroot, S. O. Suh, M. Blackwell, D. M. Hillis, G. W. Griffith, L. A. Castlebury, A. Y. Rossman, H. T. Lumbsch, R. Lucking, B. Budel, A. Rauhut, P. Diederich, D. Ertz, D. M. Geiser, K. Hosaka, P. Inderbitzin, J. Kohlmeyer, B. Volkmann-Kohlmeyer, L. Mostert, K. O'Donnell, H. Sipman, J. D. Rogers, R. A. Shoemaker, J. Sugiyama, R. C. Summerbell, W. Untereiner, P. R. Johnston, S. Stenroos, A. Zuccaro, P. S. Dyer, P. D. Crittenden, M. S. Cole, K. Hansen, J. M. Trappe, R. Yahr, F. Lutzoni and J. W. Spatafora, *Syst. Biol.*, 2009, **58**, 224-239.
83. S. B. Singh, D. L. Zink, J. M. Liesch, R. T. Mosley, A. W. Dombrowski, G. F. Bills, S. J. Darkin-Rattray, D. M. Schmatz and M. A. Goetz, *J. Org. Chem.*, 2002, **67**, 815-825.
84. S. J. Darkin-Rattray, A. M. Gurnett, R. M. Myers, P. M. Dulski, T. M. Crumley, J. J. Alloco, C. Cannova, P. T. Meinke, S. L. Colletti, M. A. Bednarek, S. B. Singh, M. A. Goetz, A. W. Dombrowski, J. D. Polishook and D. M. Schmatz, *Proc. Nat. Acad. Sci. USA*, 1996, **93**, 13143-13147.
85. F. Berst, M. Ladlow and A. B. Holmes, *Chemical Communications*, 2002, 508-509.
86. J. D. Walton, *Phytochem.*, 2006, **67**, 1406-1413.
87. J. M. Jin, S. Lee, J. Lee, S. R. Baek, J. C. Kim, S. H. Yun, S. Y. Park, S. Kang and Y. W. Lee, *Molec. Microbiol.*, 2010, **76**, 456-466.
88. P. Wiemann, C. M. K. Sieber, K. W. von Bargaen, L. Studt, E. M. Niehaus, J. J. Espino, K. Huß, C. B. Michiels, S. Albermann, D. Wagner, S. V. Bergner, L. R. Connolly, A. Fischer, G. Reuter, K. Kleigrew, T. Bald, B. D. Wingfield, R. Ophir, S. Freeman, M. Hippler, K. M. Smith, D. W. Brown, R. H. Proctor, M. Münsterkötter, M. Freitag, H. U. Humpf, U. Güldener and B. Tudzynski, *PLoS Path.*, 2013, **9**.
89. K. W. von Bargaen, E.-M. Niehaus, K. Bergander, R. Brun, B. Tudzynski and H.-U. Humpf, *J. Nat. Prod.*, 2013.
90. R. Labuda, P. Eliáš Jr, H. Sert and K. Sterflinger, *Microbiol. Res.*, 2008, **163**, 208-214.

91. W. D. Wight, R. Labuda and J. D. Walton, *BMC Microbiol.*, 2013, **13**.
92. P. Schneider, M. Weber and D. Hoffmeister, *Fungal Gen. Biol.*, 2008, **45**, 302-309.
93. F. Zhang, Y. Shen, H. Li, F. Du, A. Wang and W. Gao, *Afr. J. Microbiol. Res.*, 2012, **6**, 7103-7113.
94. P. Schneider, S. Bouhired and D. Hoffmeister, *Fungal Gen. Biol.*, 2008, **45**, 1487-1496.
95. B. Wackler, G. Lackner, Y. H. Chooi and D. Hoffmeister, *ChemBiochem*, 2012, **13**, 1798-1804.
96. H.-H. Yeh, Y.-M. Chiang, R. Entwistle, M. Ahuja, K.-H. Lee, K. Bruno, T.-K. Wu, B. Oakley and C. C. Wang, *Appl. Microbiol. Biotechnol.*, 2012, **96**, 739-748.
97. M. Misiak, J. Braesel and D. Hoffmeister, *Fungal Biol.*, 2011, **115**, 775-781.
98. K. Imamura, Y. Tsuyama, T. Hirata, S. Shiraishi, K. Sakamoto, O. Yamada, O. Akita and H. Shimoi, *Appl. Environ. Microbiol.*, 2012, **78**, 6996-7002.
99. P. K. Mukherjee, A. Wiest, N. Ruiz, A. Keightley, M. E. Moran-Diez, K. McCluskey, Y. F. Pouchus and C. M. Kenerley, *J. Biol. Chem.*, 2011, **286**, 4544-4554.
100. T. Degenkolb, R. Karimi Aghcheh, R. Dieckmann, T. Neuhofer, S. E. Baker, I. S. Druzhinina, C. P. Kubicek, H. Brückner and H. Von Döhren, *Chem. Biodivers.*, 2012, **9**, 499-535.
101. H. Ali, M. I. Ries, P. P. Lankhorst, R. A. M. van der Hoeven, O. L. Schouten, M. Noga, T. Hankemeier, N. N. M. E. van Peij, R. A. L. Bovenberg, R. J. Vreeken and A. J. M. Driessen, *PLoS One*, 2014, **9**, e98212.
102. F. E. Koehn, D. R. Kirsch, X. Feng, J. Janso and M. Young, *J. Nat. Prod.*, 2008, **71**, 2045-2048.
103. S. Liu and Y. Shen, *Chem Nat Compd*, 2011, **47**, 786-788.
104. J. Zheng, Z. Xu, Y. Wang, K. Hong, P. Liu and W. Zhu, *J. Nat. Prod.*, 2010, **73**, 1133-1137.
105. B. V. Bertinetti, N. I. Pena and G. M. Cabrera, *Chem. Biodivers.*, 2009, **6**, 1178-1184.
106. W. Wu, H. Dai, L. Bao, B. Ren, J. Lu, Y. Luo, L. Guo, L. Zhang and H. Liu, *J. Nat. Prod.*, 2011, **74**, 1303-1308.
107. Y. Seto, K. Takahashi, H. Matsuura, Y. Kogami, H. Yada, T. Yoshihara and K. Nabeta, *Biosci. Biotechnol. Biochem.*, 2007, **71**, 1470-1475.
108. D. Xu, J. Ondeyka, G. H. Harris, D. Zink, J. N. Kahn, H. Wang, G. Bills, G. Platas, W. Wang, A. A. Szewczak, P. Liberator, T. Roemer and S. B. Singh, *J Nat Prod*, 2011, **74**, 1721-1730.
109. Y. M. Lee, H. T. Dang, J. Li, P. Zhang, J. Hong, C. O. Lee, J. H. Jung and J. H. Jung, *Bull. Kor. Chem. Soc.*, **32**, 3817-3820.
110. K. Pavlaskova, J. Nedved, M. Kuzma, M. Zabka, M. Sulc, J. Sklenar, P. Novak, O. Benada, O. Kofronova, M. Hajduch, P. J. Derrick, K. Lemr, A. Jegorov and V. Havlicek, *J. Nat. Prod.*, 2010, **73**, 1027-1032.
111. I. Perez-Victoria, J. Martin, V. Gonzalez-Menendez, N. de Pedro, N. El Aouad, F. J. Ortiz-Lopez, J. R. Tormo, G. Platas, F. Vicente, G. F. Bills, O. Genilloud, M. A. Goetz and F. Reyes, *J Nat Prod*, 2012, **75**, 1210-1214.
112. T. Gulder, H. Hong, J. Correa, E. Egereva, J. Wiese, J. Imhoff and H. Gross, *Marine Drugs*, 2012, **10**, 2912-2935.
113. S. W. Meyer, T. F. Mordhorst, C. Lee, P. R. Jensen, W. Fenical and M. Kock, *Org. Biomol. Chem.*, 2010, **8**, 2158-2163.
114. Y. Igarashia, T. Hanafusa, F. Gohda, S. Peterson and G. Bills, *Mycology*, 2014, in press.
115. K. Shiomi, R. Matsui, A. Kakei, Y. Yamaguchi, R. Masuma, H. Hatano, N. Arai, M. Isozaki, H. Tanaka, S. Kobayashi, A. Turberg and S. Omura, *J. Antibiotics*, 2010, **63**, 77-82.
116. S. Monma, T. Sunazuka, K. Nagai, T. Arai, K. Shiomi, R. Matsui and S. Omura, *Org. Lett.*, 2006, **8**, 5601-5604.
117. T. Ohshiro, D. Matsuda, T. Kazuhiro, R. Uchida, K. Nonaka, R. Masuma and H. Tomoda, *J. Antibiotics*, 2012, **65**, 255-262.
118. J.-p. Wang, W. Lin, V. Wray, D. Lai and P. Proksch, *Tetrahedron Lett.*, 2013, **54**, 2492-2496.
119. M. Isaka, P. Berkaew, K. Intereya, S. Komwijit and T. Sathitkunanon, *Tetrahedron*, 2007, **63**, 6855-6860.
120. W. Ebrahim, J. Kjer, M. El Amrani, V. Wray, W. Lin, R. Ebel, D. Lai and P. Proksch, *Marine Drugs*, 2012, **10**, 1081-1091.
121. Y. T. Wang, A. B. Ke, F. S. Wang, X. L. Zhang, Y. L. Fan, X. H. Lu, Z. H. Zheng, Q. Jiang, H. Zhang and B. H. Zhao, *Chinese J. Antibiotics*, 2011, **36**, 662-666.
122. Y. Shiono, M. Tsuchinari, K. Shimanuki, T. Miyajima, T. Murayama, T. Koseki, H. Laatsch, T. Funakoshi, K. Takanami and K. Suzuki, *J. Antibiotics*, 2007, **60**, 309-316.
123. K. Suzuki, M. Yamaizumi, S. Tateishi, Y. Monnai and M. Uyeda, *J. Antibiotics*, 1999, **52**, 460-465.
124. K. Herath, G. Harris, H. Jayasuriya, D. Zink, S. Smith, F. Vicente, G. Bills, J. Collado, A. González, B. Jiang, J. N. Kahn, S. Galuska, R. Giacobbe, G. Abruzzo, E. Hickey, P. Liberator, D. Xu, T. Roemer and S. B. Singh, *Biorg. Med. Chem.*, 2009, **17**, 1361-1369.
125. S. B. Singh, J. Ondeyka, G. Harris, K. Herath, D. Zink, F. Vicente, G. Bills, J. Collado, G. Platas, A. González Del Val, J. Martin, F. Reyes, H. Wang, J. N. Kahn, S. Galuska, R. Giacobbe, G. Abruzzo, T. Roemer and D. Xu, *J. Nat. Prod.*, 2013, **76**, 334-345.
126. S. Nakadate, K. Nozawa, H. Sato, H. Horie, Y. Fujii, M. Nagai, T. Hosoe, K.-i. Kawai and T. Yaguchi, *J. Nat. Prod.*, 2008, **71**, 1640-1642.
127. A. A. Sy-Cordero, T. N. Graf, A. F. Adcock, D. J. Kroll, Q. Shen, S. M. Swanson, M. C. Wani, C. J. Pearce and N. H. Oberlies, *J. Nat. Prod.*, 2011, **74**, 2137-2142.
128. M. Y. Kim, J. H. Sohn, J. S. Ahn and H. Oh, *J. Nat. Prod.*, 2009, **72**, 2065-2068.
129. T. Kawahara, M. Takagi and K. Shin-ya, *J. Antibiotics*, 2012, **65**, 147-150.
130. A. Carroux, A. I. Van Bohemen, C. Roullier, T. Robiou Du Pont, M. Vansteelandt, A. Bondon, A. Zalouk-Vergnoux, Y. F. Pouchus and N. Ruiz, *Chem. Biodivers.*, 2013, **10**, 772-786.
131. L. Maddau, A. Cabras, A. Franceschini, B. T. Linaldeddu, S. Crobu, T. Roggio and D. Pagnozzi, *Microbiol.*, 2009, **155**, 3371-3381.
132. G. Jung, W. A. König, D. Leibfritz, T. Ooka, K. Janko and G. Boheim, *Biochim. Biophys. Act.*, 1976, **433**, 164-181.
133. C. R. Röhrich, A. Vilcinskas, H. Brückner and T. Degenkolb, *Chem. Biodivers.*, 2013, **10**, 827-837.
134. P. Pruksakorn, M. Arai, L. Liu, P. Moodley, W. R. Jacobs, Jr. and M. Kobayashi, *Biol. Pharm. Bull.*, 2011, **34**, 1287-1290.



135. P. Pruksakorn, M. Arai, N. Kotoku, C. Vilchze, A. D. Baughn, P. Moodley, W. R. Jacobs Jr and M. Kobayashi, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 3658-3663.
136. S. Ayers, B. M. Ehrmann, A. F. Adcock, D. J. Kroll, E. J. Carcache de Blanco, Q. Shen, S. M. Swanson, J. O. Falkinham, M. C. Wani, S. M. Mitchell, C. J. Pearce and N. H. Oberlies, *J. Pept. Sci.*, 2012, **18**, 500-510.
137. M. Iwatsuki, Y. Kinoshita, M. Niitsuma, J. Hashida, M. Mori, A. Ishiyama, M. Namatame, A. Nishihara-Tsukashima, K. Nonaka, R. Masuma, K. Ootoguro, H. Yamada, K. Shiomi and S. Omura, *J. Antibiotics*, 2010, **63**, 331-333.
138. K. Bushley, D. Ripoll and B. G. Turgeon, *BMC Evol. Biol.*, 2008, **8**, 328.
139. S. Traeger, F. Altegoer, M. Freitag, T. Gabaldon, F. Kempken, A. Kumar, M. Marcet-Houben, S. Pöggeler, J. E. Stajich and M. Nowrousian, *PLoS Gen.*, 2013, **9**, e1003820.
140. G. Pirri, A. Giuliani, S. Nicoletto, L. Pizzuto and A. Rinaldi, *Cent. Eur. J. Biol.*, 2009, **4**, 258-273.
141. S. Bergmann, J. Schumann, K. Scherlach, C. Lange, A. A. Brakhage and C. Hertweck, *Nat. Chem. Biol.*, 2007, **3**, 213-217.
142. W. Xu, X. Cai, M. E. Jung and Y. Tang, *J. Amer. Chem. Soc.*, 2010, **132**, 13604-13607.
143. Z. Wasil, K. A. K. Pahirulzaman, C. Butts, T. J. Simpson, C. M. Lazarus and R. J. Cox, *Chem. Sci.*, 2013, **4**, 3845-3856.
144. L. M. Halo, J. W. Marshall, A. A. Yakasai, Z. Song, C. P. Butts, M. P. Crump, M. Heneghan, A. M. Bailey, T. J. Simpson, C. M. Lazarus and R. J. Cox, *ChembioChem*, 2008, **9**, 585-594.
145. M. Gressler, C. Zaehle, K. Scherlach, C. Hertweck and M. Brock, *Chem. Biol.*, 2011, **18**, 198-209.