Nonribosomal peptides in fungal cell factories: from genome mining to optimized heterologous production

Antoine Vassaux, Loïc Meunier, Micheline Vandenbol, Denis Baurain, Patrick Fickers, Philippe Jacques, Valérie Leclère

A TERRA Teaching and Research Centre, Microbial Processes and Interactions, Gembloux Agro-Bio Tech, University of Liege, Avenue de la Faculté d’Agronomie, B5030 Gembloux, Belgium
b Univ. Lille, INRA, ISA, Univ. Artois, Univ. Littoral Côte d’Opale, EA 7394-ICV-Institut Charles Vielmetter, F-59000 Lille, France
c InBioS-PhytoSYSTEMS, Eukaryotic Phylogenomics, University of Liege, Boulevard du Rectorat 27, B-4000 Liège, Belgium
d TERRA Teaching and Research Centre, Microbiologie et Génomique, Gembloux Agro-Bio Tech, University of Liege, Avenue de la Faculté d’Agronomie, B5030 Gembloux, Belgium

ARTICLE INFO

Keywords:
Bioinformatics tools
Fungal cells
Genome mining
Heterologous production
Large gene cluster
Multi-modular enzyme
Nonribosomal peptide
Nonribosomal peptide synthetase
Secondary metabolite

ABSTRACT

Fungi are notoriously prolific producers of secondary metabolites including nonribosomal peptides (NRPs). The structural complexity of NRPs grants them interesting activities such as antibiotic, anti-cancer, and anti-inflammatory properties. The discovery of these compounds with attractive activities can be achieved by using two approaches: either by screening samples originating from various environments for their biological activities, or by identifying the related clusters in genomic sequences thanks to bioinformatics tools. This genome mining approach has grown tremendously due to recent advances in genome sequencing, which have provided an incredible amount of genomic data from hundreds of microbial species. Regarding fungal organisms, the genomic data have revealed the presence of an unexpected number of putative NRP-related gene clusters. This highlights fungi as a goldmine for the discovery of putative novel bioactive compounds. Recent development of NRP dedicated bioinformatics tools have increased the capacity to identify these gene clusters and to deduce NRPs structures, speeding-up the screening process for novel metabolites discovery. Unfortunately, the newly identified compound is frequently not or poorly produced by native producers due to a lack of expression of the related genes cluster. A frequently employed strategy to increase production rates consists in transferring the related biosynthetic pathway in heterologous hosts. This review aims to provide a comprehensive overview about the topic of NRPs discovery, from gene cluster identification by genome mining to the heterologous production in fungal hosts. The main computational tools and methods for genome mining are herein presented with an emphasis on the particularities of the fungal systems. The different steps of the reconstitution of NRP biosynthetic pathway in heterologous fungal cell factories will be discussed, as well as the key factors to consider for maximizing productivity. Several examples will be developed to illustrate the potential of heterologous production to both discover uncharacterized novel compounds predicted in silico by genome mining, and to enhance the productivity of interesting bio-active natural products.

1. Introduction

The era of nonribosomal peptides (NRPs) started with the brilliance of Fleming to analyze the unexpected ability of a filamentous Penicillium strain to impair bacterial growth. These observations led to the discovery of the penicillin antibiotic, a NRP derived compound which was not identified as such at that time (Fleming, 1929). Indeed, the biosynthesis mechanism of this type of peptides, named thiotemplate mechanism, was elucidated only 42 years later with the description of the Non Ribosomal Peptide Synthetase (NRPS) of the gramicidin

Abbreviations: A-domain, Adenylation domain; ACV, d-(L-α-aminoadipyl)-L-cysteinyl-D-valine; BCGs, Biosynthetic gene clusters; C_term-domain, Terminal condensation domain; NRP, Nonribosomal peptide; NRPS, Nonribosomal peptides synthetase; ORF, Open reading frame; pHMMs, Profile hidden markov models; PKS, Polyketide synthase; PPTase, Phosphopantetheine transferase; SGS, Second generation sequencing; SMs, Secondary metabolites; TAR cloning, Transformation-associated recombination cloning; TGS, Third generation sequencing

⁎ Corresponding author.
E-mail address: valerie.leclere@univ-lille.fr (V. Leclère).

https://doi.org/10.1016/j.biotechadv.2019.107449
Received 7 May 2019; Received in revised form 6 September 2019; Accepted 9 September 2019
Available online 10 September 2019
0734-9750/ © 2019 Elsevier Inc. All rights reserved.
antibiotic from \textit{B. subtilis} (Lipmann et al., 1971). Since then, there has been growing interest in the study of this kind of compounds. As a result, NORINE, the first database entirely dedicated to NRPs was established in 2007, starting with 700 NRPs entries (Caboche et al., 2008). At present, over a thousand of NRPs have been characterized and classified (Pupin et al., 2016). These NRPs were mostly found in prokaryotes and fungal eukaryotes (Demain, 2014), and are notably not present in metazoa and plantae eukaryotic clades. Few exceptions were reported such as the production of astins in the plant \textit{Aster tataricus}, later refuted and attributed to a fungal endophyte \textit{Cyanodermella asteris} (Vassaux et al., 2019). Likewise, polytheonamide initially isolated from the marine sponge \textit{Theonella swinhoei} (Hamada et al., 2005), was finally shown to be produced by a symbiotic bacteria (Freeman et al., 2016; Wilson et al., 2014).

Nonribosomal peptides, as suggested by their name, are not assembled on the ribosome but by large multi-modular enzymatic systems. Several domains are presented: A: adenylation; T: thiolation; C: condensation; C\textsubscript{term}/Te: thioesterase or terminal condensation domain (only found in fungal NRPSs); and an example of a potential additional epimerization (E) domain (converting an \textit{l}-form amino acid to its \textit{d}-form). Epimerization reaction can occur either before or after condensation reaction.

\textbf{Fig. 1.} Multiple steps involved in NRP biosynthesis pathway. A- Transcription of the NRPS encoding gene. B- Translation of the NRPS mRNA. C- Folding of the NRPS in multi-enzymatic domains modular organization and addition of phosphopantetheine arm (holo-NRPS). D- Operating way of NRPS assembly line to generate NRP (Mm: monomer). Epimerization reaction can occur either before or after condensation reaction.
complexes that are encoded by gene clusters that often span over tens of kilobases in the genome (Fig. 1). However, NRP synthetases are synthesized on the ribosome, before being activated by a specific phosphopantetheinyl transferase (PPTase). This PPTase is responsible for the conversion of the apo-NRPS into active holo-NRPSs through the addition of a so-called phosphopantetheine arm onto thiolation domains (Beld et al., 2014). NRPSs can be divided into multiple modules, each of them responsible for the incorporation of one peptide monomer. These modules comprise multiple enzymatic domains exhibiting specific functions in the NRPs assembly line (Caboche et al., 2008). The adenylation (A) domain is able to recognize and activate a specific monomer, which is then tethered onto a thiolation (T) / peptidyl carrier protein domain through the phosphopantetheine arm. The flexibility of the multi-enzymatic complex enables the presentation of monomers to the condensation (C) domain which catalyzes the formation of a peptide bond between monomers from two subsequent modules (Winn et al., 2016). In NRPSs involved in lipopeptide biosynthesis, a major NRPs class, a starter C domain priming the synthetase enables to acylate the first peptide monomer with the fatty acid part (Rausch et al., 2007). Some NRPSs also exhibit additional domains responsible for monomer modifications, such as methylation, oxidation and epimerization.

NRPSs predominantly operate according to a linear synthesis, with an initiation module (A-T) able to recognize the first monomer, followed by as many modules (C-A-T) as monomers required to complete the peptide. Nonetheless, some NRPSs rather operate with an iterative or nonlinear organization, and then may not follow the colinearity rule between the number of modules and the size of synthesized peptides (Richter et al., 2014; Scheweke et al., 2006). The aminobactin synthetase found in Aeromonas sp. is a prime example of an atypical assembly line which displays an iterative, alternative and optional mode of synthesis. In this NRPS, the AmoG-A-domain activates alternatively a Trp or a Phe, on which are bonded two dipeptides iteratively generated by the AmoE and AmoF domains, and optionally in two out of four cases, an AmoH domain incorporates a glycine between the 2,3-dihydroxy benzoic acid and the lysine of one of the dipeptides (Eismael et al., 2017). NRPS assembly lines usually end with a termination module comprising a thioesterase domain (TE) responsible for the release, and the cyclization, in most cases, of the final compound (Du and Lou, 2010; Kopp and Marahiel, 2007). In fungi, thioesterase domains can be replaced by condensation terminal domain (Cterm) ending the NRPS and leading to cyclic NRPs (Gao et al., 2012). Alternatively, in NRPSs involved in peptibol biosynthesis (i.e., NRPs harboring a C-terminal alcoholic group instead of carboxyl group), TE is substituted by a reductase domain releasing the peptide with a C-terminal alcohol (Manavalan et al., 2010).

Filamentous fungi are considered as prolific producers of NRPs. A large number of clusters involved in NRPS synthesis has been detected for example in Penicillium thymicola (17 clusters), Aspergillus fumigatus (14 clusters), Talaromyces islandicus, Beauveria bassiana (13 clusters), and Pestalotiopsis fici (12 clusters) (Chiang et al., 2014; Gibson et al., 2014; Schafhauser et al., 2016; Tang et al., 2018; Wang et al., 2015). The high potential of fungal organisms for NRPs production can be exploited for the discovery of novel compounds of interest. The search for novel compounds in fungi is typically achieved following two different approaches: one starting from the biological activities and the other, more recent, from genome sequences. The principle of the former consists in screening samples originating from various environments for their biological activities. Once the targeted bioactivity is detected, the active natural product is purified and characterized (Kildgaard et al., 2017; Kunakom and Eustáquio, 2019). Usually bioactivities are first screened at small-scale and then up-scaled to reach higher production rates. However, this strategy is very restrictive as only compounds displaying the screened activity and being produced in sufficient amount to induce an observable activity, will be detected (Luo et al., 2014). In contrast, the second approach is based on the use of bioinformatics tools in order to identify the SM gene clusters of interest (Luo et al., 2014; Mustaq et al., 2018). Frequently, the associated compound is not produced because the related cluster is cryptic and the cultivation conditions were probably not optimal to succeed in its awakening (Anyaogu and Mortensen, 2015; Lazarus et al., 2014). Therefore, the second approach often requires genetic engineering either to activate the expression in the native host (homologous production) or to transfer the biosynthesis pathway in a heterologous host (heterologous production) (Luo et al., 2016).

In the past few years, microbial genome sequencing and annotation methods have made an incredible leap forward, providing the scientific community a wealth of publicly available fungal genomes (Grigoriev et al., 2014). This technological progress has opened new perspectives for the discovery of novel compounds through the bioinformatic approach. The development of novel bioinformatics tools enables, at present, the analysis of multi-modular complexes such as NRPSs, but also polyketide synthases (PKSs) or hybrid complexes PKS-NRPS (Weber and Kim, 2016). The case of polyketides, assembled from acyl- and malonyl- coenzyme A by PKSs, is already well documented and is beyond the scope of this review (Chooi and Tang, 2012; Schümann and Hertweck, 2006; Zhang et al., 2011). Nonetheless, this study presents the multiple bioinformatics tools developed to predict the modular organization of NRPSs, and the substrate specificity of their adenylation domains, with an emphasis on the specificities of fungal systems. Bearing in mind that these NRPs identified by genome mining are frequently poorly synthesized under laboratory culture conditions, and taking into account for most of these fungi that adapted molecular tools are lacking, the heterologous expression of their biosynthetic pathways is often a necessary step (Nah et al., 2017; Sung et al., 2017; Tsumematsu et al., 2013a). The heterologous production of NRPs in bacterial organisms such as Escherichia coli and Streptomyces spp. have been widely reviewed (Beites and Mendes, 2015; Nah et al., 2017; Weber et al., 2015b; Zhang et al., 2011), but only few studies have investigated the potential of fungal organisms to produce such foreign compounds. Therefore, this review discusses, through multiple successful examples, the possibility to use fungal cell factories to heterologously express NRPS-encoding genes identified by genome mining. Finally, the multiple key factors to consider obtaining an efficient NRPs heterologous producer are described and then illustrated onto the example of the NRP-derivative penicillin, a multi-decades case of study.

2. Computational tools and methods for genome mining

Bioinformatic gene clusters (BGCs) involved in the synthesis of secondary metabolites (SMs) usually contain core synthase genes, regulatory and transport-related elements, and sometimes genes conferring a resistance to the related SM. Most of the developed bioinformatics tools can be applied to the analysis of BGCs related to multiple type of SMs (e.g. NRPS, PKS) coming out from either bacterial or fungal genomic data. Regarding to NRPS analysis, bioinformatics methods have been developed based on the high conservation of the NRPS thiotemplate mechanism and decades of experimental characterization (Weber, 2014; Ziemert et al., 2016). Applications of these algorithms can be used to identify NRPS cluster genes, their subsequent annotation, the prediction of the backbone of the putative NRPs produced, and the search for related NRPs. The in silico annotation of BGCs relies on the algorithm efficiency, but also on the genome assembly quality, as the quality and the completeness of the annotation depend directly on the input sequences (Yandell and Ence, 2012). These two aspects will be further discussed in the following sections.

2.1. Recent advances in genome sequencing and annotation

From the mid-2000s, Second Generation Sequencing (SGS) technologies, also named Next Generation Sequencing (including Illumina and Ion Torrent technologies), revolutionized genomics and many other biological fields, by offering low-cost genome sequencing (Loman et al.,
These developments generated an exponential growth in the genome sequences available in public databases (more than 190,000 prokaryotic genome assemblies available on NCBI databases as of March 2019, https://www.ncbi.nlm.nih.gov/genome/browse/) (Tatusova et al., 2014). SGS technologies are based on the principle of a high-throughput method, with a short read length up to a few hundred bases (e.g., 250 bp for Illumina HiSeq), and usually a low error rate. The drawbacks of short read genome assemblies are related to low contiguity and poor gene completeness (Heather and Chan, 2016; Koren and Phillippy, 2015). As NRPS biosynthetic gene clusters can span several tens of thousands bases - even more when considering “accessory genes” -, BGCs full sequence might be spread into several contigs (Conway and Boddy, 2013; Vesth et al., 2016). This causes complex issues for reconstructing BGCs from draft genomes, as well as determining if several unknown gene clusters in an organism are related to only one or several NRPSs.

The recent emergence of Third Generation Sequencing (TGS), with PacBio and Oxford Nanopore technologies, brings new interesting features for BGC annotation. These two methods yield long read length (from a few tens of kb for PacBio, up to 1 Mb for Oxford Nanopore, depending on the quality of the DNA extraction (Jain et al., 2018)), hence generating high contiguity assemblies (Lu et al., 2016). This long read length may present a real advantage in the case of BGC studies, allowing for obtaining complete sequences, and therefore the entire set of involved genes. Yet, in contrast to SGS, TGS displays a lower throughput and a higher error rate (Rhoads and Au, 2015). Several software packages, such as Flye (Kolmogorov et al., 2019) and Canu (koren et al., 2017), provide efficient methods for TGS data assembly. When the read depth coverage is not sufficient, hybrid strategies taking advantage of the low error rate of SGS can be used for polishing TGS reads or contigs, allowing for obtaining a high-quality genome assembly (Bashir et al., 2012). For instance, SPAdes (Antipov et al., 2016) offers hybrid assemblies, while Pilon (Walker et al., 2014) allows the user to polish TGS assemblies.

Regarding the NRPS identification and annotation, these steps are driven by fast and reasonably accurate bioinformatics methods, which have now become one of the most predominant approaches to support the discovery of novel SMs. This emerging field is promising, as a result of constant effort in software development (Chavali and Rhee, 2018; Weber and Kim, 2016; Ziemert et al., 2016) and the availability of genomic data. Most methods for annotating NRPSs rely on similarity searches, which is a robust approach as these multi-enzymatic systems are composed of repeated, well conserved domains. The detection of core biosynthetic components is commonly conducted by protein signature searches (Chavali and Rhee, 2018). These signatures, also called PHMMs (profile Hidden Markov Models) are probabilistic models capturing the versatile information contained in a multiple sequence alignment (i.e., amino acid composition and insertions/deletions per site) to perform specific and sensible annotations (Eddy, 2011; Eddy, 1998). Hence, from the prediction of synthetase genes, it is then possible to deduce the NRPS architecture (i.e., domain composition), and search for additional genes surrounding this core (e.g., regulation and transport related genes, tailoring enzymes).

2.2. Bioinformatics tools dedicated to NRPS study

Publicly available bioinformatics tools for annotating NRPS started to be devised a decade ago (Blin et al., 2017a), since then, a large number of software packages and methodologies for genome mining analyses has been released (Fig. 2). The Secondary Metabolite Bioinformatics Portal represents a very useful online resource (http://www.secondarymetabolites.org/) facilitating the exploration of numerous software packages. It gathers information about the majority of genome mining and annotation tools, as well as databases applied to secondary metabolism investigation (Weber and Kim, 2016). The portion of these tools implying specific functionalities for NRPS and closely related PKS is mostly constituted with generic algorithms applicable for both bacteria and fungi, even if the training of their algorithm was mainly based on bacterial data. For instance, NRPS annotation: NP.searcher (Li et al., 2009), ClustScan (Starcevic et al., 2008), and antiSMASH (Blin et al., 2017a) are generic tools for the overall annotation of NRPS. The latter is currently the most comprehensive and up-to-date pipeline for BGC analysis, providing a great deal of available functionalities for refining NRPS annotations (Blin et al., 2017a; Weber and Kim, 2016). Among the small portion of tools dedicated to Fungi, SMURF (Khalid et al., 2010) is also a well-known software performing the basic identification of putative synthase and auxiliary enzyme genes. It is also noteworthy that antiSMASH possesses a version for fungal data (i.e., fungiSMASH - http://fungismash.secondarymetabolites.org) (Blin et al., 2017c) which adapts the internal gene predictor for an eukaryotic one (GlimmerHMM). In addition, fungiSMASH also offers to apply CASSIS (Wolf et al., 2016), an algorithm aiming the improvement of NRPS/PKS BGC boundaries delimitation by searching the presence of common regulation pattern in BGC promoters.

In addition to these general NRPS annotation tools, some are entirely devoted to the annotation of NRPS biosynthetic pathways, focusing on specific features of NRPSs and their products. Indeed, these tools aim to predict the peptidyl backbone of the putative NPR produced by analyzing two elements: substrate specificity of adenylation domains and condensation domain subtypes classes.

Substrate selection is based on the specificity of adenylation domains. Determining this specificity is a key element in the structural characterization of the putative synthesized compound. The specificity-conferring code, known as “Stachelhaus or NRPS code”, was the first prediction method based on the identity of ten amino acid residues located in the phenylalanine-binding pocket of PheA (adenylation domain of the first module of the gramicidin synthetase) (Stachelhaus et al., 1999). These critical residues were demonstrated to be involved in the substrate specificity as they define the binding-pocket recruiting specifically the various monomers. Therefore, with multiple alignments of the already known NRPS A-domains, it was possible to establish a code that correlate these residues to specifically recognized amino acid substrates (Challis et al., 2000). Therefrom, several substrate specificity prediction methods were developed such as the pHMM-based approach analyzing the specific protein signatures in the active site (Mimowa et al., 2007). This method was integrated and further developed in NRPSp (http://www.nrpssp.com/) and NRPS-PKS-substrate-predictor (http://www.cmbi.ru.nl/NRPS-PKS-substrate-predictor/) (Khayatt et al., 2013; Prieto et al., 2012). Another method, based on a Support Vector Machine approach, is exploited in NRPSpred2. This software package provides predictions according to the physicochemical properties of the 34 residues in a radius of 8 angstrom around the PheA binding pocket (Rausch et al., 2005; Röttig et al., 2011). Substrate prediction obtained through the above-mentioned methods are usually accurate when dealing with bacterial data but are still not entirely reliable with fungal data. Three reasons can justify these limitations regarding substrate prediction: (1) low sequence conservation between A-domains (10-40% identity) (Agüero-Chapin et al., 2016), (2) the existence of different A-domain patterns for a same recognized amino acid (Khayatt et al., 2013), or on the contrary, (3) A-domains displaying the relaxed selection of more than one substrate (Belshaw et al., 1999; Esmael et al., 2016; Richter et al., 2014). Therefore, new strategies have been developed to address the still unresolved challenge of substrate specificity prediction. One recently developed approach, SEQL-NRPS (http://services.birc.au.dk/seql-nrps/), enables to predict substrate specificities based on an alignment-free method detecting occurrences of motifs in domains primary structure (Knudsen et al., 2016). Recently, predicAT, a phylogenetic-based algorithm enabling to estimate the degree of predictability of A-domain, has been developed and integrated in SANDPUMA (Chevrette et al., 2017). SANDPUMA is a prediction framework combining three methods: (1) a retrained version
of the Support Vector Machine models applied in NRPSpredictor2, (2) a search for exact matches from active site motifs as in the Stachelhaus et al. (1999) approach and (3) a pHMMs analysis like the Khayatt et al. (2013) approach. The last antiSMASH version (4.0) modified its previous substrate predictors (Stachelhaus code, Minowa’s pHMMs and NRPSpredictor2) to implement SANDPUMA (Blin et al., 2017c).

Condensation domains also exhibit substrate specificity during the elongation reaction, as well as some possible additional activities such as cyclization or epimerization. These domains can be divided in different subtypes (Table 1). The identification of the C domain-subtype is a crucial element for inferring the modifications introduced into the amino acids constituting the final peptide (e.g., identifying D-amino acids or peptide cyclization) (Caradec et al., 2014). Analysis of condensation domains can be achieved with the NaPDoS web application (http://napdos.ucsd.edu), which uses a phylogenetic approach to classify the subtypes of condensation domains and ketosynthase domains from PKSs (Ziemert et al., 2012). Each subtype of condensation domain exhibits specific conserved motifs enabling to efficiently classify them in phylogenetic trees (Caradec et al., 2014). According to its position into the phylogenetic tree, it is possible by inference to determine the subtype of the analyzed C-domain.

When this additional information regarding core domains specificities has been collected, a quite precise determination of a putative NRP compounds architecture is possible by combining the predicted substrates of adenylation domains, with the modifications related to the predicted condensation classes. Florine, a workflow dedicated to the annotation of NRP compounds, presents a detailed decision diagram to carry out this annotation process (Caradec et al., 2014). Nonetheless, the mechanism for substrate selection in NRPS is not fully elucidated and bioinformatics tools are being continuously improved as a result of new discoveries. For instance, the short amino acid regions connecting adjacent modules, called linkers, have been very recently highlighted to also play a role in the determination of substrate specificity. Subsequently to this discovery, a web application just has been released to enable the identification of these inter-modular NRPS linkers (https://nrps-linker.unc.edu) (Farag et al., 2019).

![Fig. 2. Simplified workflow for novel NRP discovery by genome mining with an overview of the main bioinformatics tools available (accessible through hyperlinks).](image-url)

Table 1
List of condensation domain subtypes and their associated reactions.

<table>
<thead>
<tr>
<th>Condensation domain subtype</th>
<th>Reaction description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>C\text{L} _L</td>
<td>Bind L-amino acids to L-amino acids</td>
<td>(Rausch et al., 2007; Ziemert et al., 2012)</td>
</tr>
<tr>
<td>C\text{L} _D</td>
<td>Bind D-amino acids to L-amino acids</td>
<td></td>
</tr>
<tr>
<td>Starter C</td>
<td>Load a fatty acid as first substrate of the synthesis</td>
<td></td>
</tr>
<tr>
<td>Dual C/E</td>
<td>Combine condensation and epimerization activities</td>
<td></td>
</tr>
<tr>
<td>Epimerization</td>
<td>Invert an L-form amino acid to its D-form enantiomer</td>
<td></td>
</tr>
<tr>
<td>Heterocyclization</td>
<td>Catalyse peptide bond formation, followed by cyclization</td>
<td></td>
</tr>
<tr>
<td>C\text{term}</td>
<td>Replace thioesterase activity for the final step of peptide release</td>
<td>(Gao et al., 2012)</td>
</tr>
<tr>
<td>Hybrid C</td>
<td>Bind an amino acid to an aminated polyketide in hybrid PKS-NRPS systems</td>
<td>(Ziemert et al., 2012)</td>
</tr>
</tbody>
</table>

A lot of useful tools are available to perform this dereplication steps, but only few of them will be described thereafter. The standard approach for analyzing BGC is to use the Basic Local Alignment Search Tool (BLAST) search program hosted on the NCBI website (http://www.ncbi.nlm.nih.gov/BLAST). This application enables to perform a sequence similarity search relative for instance to sequences listed in the “nucleotide” (BLASTn) or the “non-redundant protein sequences” (BLASTp) collections of the NCBI (McGinnis and Madden, 2004; NCBI Resource Coordinators, 2016). Alternatively, it is possible to query dedicated databases for BGCs, NRPSs and their products. Regarding BGC pathways, ClusterMine360 is a PKS/NRPS database mainly updated by crowdsourcing, which offers a two-way approach where nearly 1000 gene clusters can be browsed directly from a cluster list or following a search from more than 200 compound families (Conway and Boddy, 2013). The antiSMASH database is another relevant resource, encompassing currently a collection of antiSMASH results for almost 25,000 genomes - but mainly of bacterial origin - and offering the possibility to build precise queries (Blin et al., 2017b). MultiGeneBlast tool enables identifying homologs of multigene modules, such as NRPSs, on the basis of similarity and synteny criteria (Medema et al., 2013). MultiGeneBlast is also integrated in antiSMASH, where it suggests some potentially related clusters (or the identical cluster itself) taken from different BGC databases of varying levels of curation. For example, the “KnownClusterBlast” option refers to a curated database containing experimentally characterized clusters, while the “ClusterBlast” option uses an extended database composed mostly of only hypothetical clusters (Weber et al., 2015a).

The only database fully dedicated to NRP compounds, Norine, currently comprises 1190 peptides, including 346 NRPs of fungal origin distributed in 46 families. Norine offers different search modes: (1) an “annotation search” engine, which includes queries with all annotations (e.g., biological activity, molecular weight, monomer count, or producing organism), and (2) a “structure mode” allowing searches based on monomer composition fingerprints, sequence similarity or whole structures. The analysis of predicted peptides permits searching for related analogous compounds already characterized, or to check if some amino acid patterns involved in some biological activities can be spotted (Pupin et al., 2016). Only a few main databases were mentioned in this section, but a more exhaustive list is provided on the Secondary Metabolite Bioinformatics Portal (Weber and Kim, 2016).

2.4. Fungal NRPS analysis challenges

Most of the methods and bioinformatic tools available so far were designed both for the analysis of NRPS of fungal and bacterial origins. However, these organisms are taxonomically distant and some differences at the genomic level can be readily identified. Indeed, in comparison with prokaryotic genomes, fungal genomes are characterized by a larger size (i.e., several tens of millions bases), a higher number of chromosomes, a lower gene density, and the presence of introns (Elliott and Gregory, 2015; Setubal et al., 2018). BGC analysis pipelines, such as antiSMASH, offer an internal gene prediction before applying protein signature searches. Structural annotations (i.e., the prediction of whole gene structure) performed by automated internal gene predictors (e.g., Glimmer3 in antiSMASH (Medema et al., 2011)) are sufficiently accurate for bacteria, whereas annotation of genes from eukaryotic genomes is usually more complex. Dedicated gene prediction pipelines for eukaryotes often need some external data (gene expression and/or homologous protein data) to support ab initio predictions and provide higher quality annotation (Yandell and Ence, 2012). As it is essential to have coding sequences correctly annotated (i.e. with intron-exon and intergenic structures well resolved) to accurately predict NRPS gene clusters, fungal genomes are likely to benefit from a preliminary gene prediction, before being processed by BGC analysis tools. The available tools, dedicated to eukaryotic genome analysis, to perform this preliminary gene prediction are depicted in a relevant review (Yandell and Ence, 2012).

When pioneering predictive models for the substrate specificity of adenylation domains were developed a decade ago (Minowa et al., 2007; Rausch et al., 2005), very few fungal data was available. Therefore, these models provide predictions based mainly on monomers identified as a substrate for the few well-characterized bacterial NRPSs at that time. Consequently, these bioinformatics tools cannot predict with high accuracy the substrate specificity of most of the Ado-domains of eukaryotic NRPSs. Furthermore, based on the Norine database, it appears that NRPSs of fungal and bacterial origin display only a moderate range of overlapping substrate and incorporate a different set of monomers into the NRPs (Fig. 3). For example, Aib (2-
aminoisobutyric acid) and amino alcohols such as Pheol (phenylalaninol) and Leoul (leucinol) seem to be specific to peptaibols produced mainly by fungi. Consequently, some fungal monomers are simply not recognized as no predictive models exist for these substrates. However, predictive models are only applicable if the monomer is directly selected by an adenylation domain and not resulting from the modification of a precursor amino acid. While Aib seems to be directly loaded (Chutrakul et al., 2008), amino alcohols in NRPs result from a modification of the amino precursors during synthesis of the peptide (Du and Lou, 2010; Manavalan et al., 2010). Determining if the monomer is loaded as such or modified during the synthesis is a complex task as no generalized rules can be applied. For example, ε-amino acids are generally the result of an epimerization domain activity on L-amino acids, but, in some cases, ε-Ala was reported to be directly loaded (McErlean et al., 2011).

Even if a few software packages, such as NRPSpredictor2 (Röttig et al., 2011), have previously developed models to predict the substrate specificity of adenylation domains, there is still an important need for devising tools enabling to predict more accurately which monomers are incorporated in fungal NRPs. Therefore, future challenges will be to go further in the elucidation of fungal NRPS mechanisms, and to establish novel tools allowing for increased the accuracy of substrate predictions in specific fungal systems, especially regarding to exotic- and ε- amino acids.

3. Heterologous production of NRPS in fungal hosts

The number of available fungal genomic sequences in databases is continuously expanding thanks to individual or large-scale collaborative efforts, such as the “1000 fungal genomes project” (Araujo and Sampaio-Maia, 2018; Stajich, 2017). These available fungal genomes represent a gold mine for novel NRP compounds discovery. Unfortunately, in many cases, the newly sequenced fungi do not (or only poorly) produce these natural products under usual laboratory conditions in synthetic media, which are not always suitable to develop the filamentous fungi or to produce NRPs (Sung et al., 2017). One option is to activate or enhance the homologous production in the native host through promoter exchange for instance (Bode et al., 2015; Soukup et al., 2016). If this is not possible, due to a lack of molecular tools for example, the heterologous production is one of the best approaches to reach a sufficient yield of production, and enable further characterization of the NRP (Alberti et al., 2017). In this part, the methodology commonly used for heterologous production of NRP is presented, taking into account different parameters such as the choice of the heterologous host, the genetic tools required for the transformation, the different expression systems used and the methods used for large scale heterologous production of NRP.

3.1. Heterologous host choice

The choice of the host is one of the most critical parameters to consider for heterologous gene expression and product synthesis. The selected strain has to present some desirable characteristics, such as accessible and efficient genetic manipulation tools, capabilities to grow at high cell density, ability to secrete the produced compounds, and above all, a codon compatibility with the foreign genes (Beites and Mendes, 2015). The well-known bacterial cellular platforms Escherichia coli, Bacillus sp., and Streptomyces sp., which display most of these characteristics, are commonly used for heterologous expression of NRPs. Nonetheless, bacteria exhibit a different codon usage from eukaryotic genes, which must be considered. One option for addressing this difficulty is to provide additional tRNA sequences for missing and rare codons, such as the AUA codon for Ile in E. coli (Skiba et al., 2018). Alternatively, as they are more closely related to filamentous fungi at a taxonomic level, yeasts constitute appropriate hosts to express foreign fungal NRPS genes. Different studies reporting NRP heterologous production in yeasts are presented in Table 2. Among yeasts, Saccharomycyes cerevisiae displays both a high codon usage compatibility towards foreign fungal genes (Tsunematsu et al., 2013b), and a great potential for the secretion of the produced compounds (Mattanovich et al., 2012). This host also displays a certain tolerance towards foreign bacterial sequences, as was recently demonstrated with the heterologous production of a NRP initially produced by Streptomyces lavendulae (Wehrs et al., 2018). As it is possible in this yeast to express a NRPS-encoding gene coming from bacteria, which is taxonomically very distant to the host, its potential is even more promising for the heterologous production of fungal NRP. Moreover, the poorly developed secondary metabolism of S. cerevisiae is an asset, as it minimizes the competition with homologous SM pathways, and promotes the detection/purification of the targeted compound. Genetic manipulations are also made straightforward by a plethora of molecular tools, and the aptitude of the yeast to efficiently perform transformation-associated recombination (TAR) enabling the cloning of large DNA fragments (Anyagoo and Mortensen, 2015). However, S. cerevisiae lacks a

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Previous studies reporting NRP heterologous production in yeasts.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural product (NRP)</td>
<td>NRPS</td>
</tr>
<tr>
<td>ACV</td>
<td>ACVS (or PchAB)</td>
</tr>
<tr>
<td>Fumiquinazoline F</td>
<td>Afu6g12080</td>
</tr>
<tr>
<td>Tricyclic ardeemin F-Q</td>
<td>ArdA</td>
</tr>
<tr>
<td>Bassianolide</td>
<td>BsSYN</td>
</tr>
<tr>
<td>Brevianamide F</td>
<td>FtmA</td>
</tr>
<tr>
<td>Acetylaslanzolamin</td>
<td>AnaPS</td>
</tr>
<tr>
<td>Asperlicin C/D</td>
<td>AnpA</td>
</tr>
<tr>
<td>Indigoidine</td>
<td>BpeA</td>
</tr>
<tr>
<td>α-Phe-LLeu</td>
<td>TyaA</td>
</tr>
<tr>
<td>2,4-dihydroxy 5,6-dimethyl benzaldehyde</td>
<td>ATEG_03630</td>
</tr>
<tr>
<td>Phenguignardic acid</td>
<td>PgaA</td>
</tr>
<tr>
<td>Atromentin</td>
<td>ATEG_0390</td>
</tr>
<tr>
<td>Aquuplysonine E</td>
<td>Apva &amp; MeaA</td>
</tr>
</tbody>
</table>

* Estimated on the basis of the amino acid sequence length (with an average of 110 Da per amino acids).
functional PPTase encoding gene, which thus must be introduced into the host (Barajas et al., 2017). In accordance with its advantages, and despite the risk of obtaining a low rate of production, S. cerevisiae is by far the most widespread yeast host for NRP heterologous production. Nevertheless, the non-conventional yeast Hansenula polymorpha (also belonging to the Saccharomycetales) can be an alternative choice. This methylophoric yeast is also easily cultivable, even at large-scale, and strongly regulated promoters are available (Saraya et al., 2014). Moreover, H. polymorpha was successfully used to produce the NRP d-\((l\text{-}\alpha\text{-aminoacidyl})_2\text{-cysteinyl}-\gamma\text{-valine}\) (ACV), a precursor of penicillin and cephalosporin (Giudalia et al., 2009). Another Saccharomycetales, Pichia pastoris (now Komagatella pastoris) is also an appropriate host to heterologously express large SM clusters. In this host, multiple examples of large multi-modular PKS expression have been reported (Kasahara et al., 2010; Xue et al., 2017). Nonetheless, to date, no NRP heterologous production has been reported, despite the obvious potential of P. pastoris for SM production. Finally, Schizosaccharomyces pombe is the only yeast reported to produce naturally a NRP (ferri-ọne). This characteristic would establish S. cerevisiae far the most widespread yeast host for NRP heterologous production. Despite the risk of obtaining a low rate of production, S. cerevisiae is by far the most widespread yeast host for NRP heterologous production. Nonetheless, to date, no NRP heterologous production has been reported, despite the obvious potential of P. pastoris for SM production. Finally, Schizosaccharomyces pombe is the only yeast reported to produce naturally a NRP (ferri-ọne). This characteristic would establish S. cerevisiae far the most widespread yeast host for NRP heterologous production.
for cloning large SM clusters via Fungal Artificial Chromosomes (FACs) has been developed (Bok et al., 2015; Cleverger et al., 2017). An equivalent artificial chromosome system is also available in yeasts (YACs) (Hughes et al., 2015). The use of artificial chromosomes has enabled the direct cloning of DNA fragments up to 300 kb (Zhang et al., 2013). The golden gate cloning method can facilitate to a considerable extent the process of cloning multiple transcription units (promoter – gene – terminator) in a single expression vector. This strategy relies on the use of type IIs restriction enzymes that can generate a specific 3′ or 5′ DNA overhangs of 4 nucleotides depending of the enzyme considered. Subsequently to the enzymatic digestion, two fragments with suitable building blocks. These kinds of DNA assembling platforms have been already designed for several potential hosts such as the filamentous fungus Ustilago maydis (Terfrüchte et al., 2014), and the yeasts S. cerevisiae (Agmon et al., 2015; Mitchell et al., 2015), and P. pastoris (Pan et al., 2011; Sauer, 1987). A recycling system is also operational in A. nidulans, where it is possible to replace the marker of a previous transformation by cloning into the same marker (van Dijk et al., 2014; Yeh et al., 2012; Geib et al., 2015). Alternatively, viral 2A peptide sequences can also be used to express several genes under the control of the same promoter in a single plasmid. These sequences, interspaced between each ORF, are cleaved during protein translation, hence releasing the proteins from each other (Kim et al., 2011; Unkles et al., 2014). However, if the use of multiple plasmids and genomic integrations is required, methodologies allowing for recycling of selection markers have been developed in some common fungal hosts. The loss of URA3 selection gene in S. cerevisiae, can be achieved by cultivation on 5-fluoorotic acid plates (Akada et al., 2006; Boeke et al., 1987). The cre-lox system also enables selection marker recycling in yeast such as S. cerevisiae and P. pastoris (Pan et al., 2011; Sauer, 1987). A recycling system is also operational in A. nidulans, where it is possible to replace the marker of a previous transformation by cloning into the first selection gene, a second construction with a different marker (van Dijk and Wang, 2016; Yaegashi et al., 2014).

### Table 3

Previous studies reporting NRP heterologous production in filamentous fungi.

<table>
<thead>
<tr>
<th>Natural product (NRP)</th>
<th>NRPS</th>
<th>NRPS MW (kDa)</th>
<th>Native producer</th>
<th>Heterologous host</th>
<th>Expression system</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>KK-I</td>
<td>KK-I NRPS</td>
<td>1,430&lt;sup&gt;a&lt;/sup&gt;</td>
<td>A. clavata</td>
<td>A. oryzae</td>
<td>Genomic integration</td>
<td>Yoshimi et al., 2018</td>
</tr>
<tr>
<td>Cycloasp peptide A/E</td>
<td>PscyB</td>
<td>623&lt;sup&gt;a&lt;/sup&gt;</td>
<td>P. sopii</td>
<td>A. oryzae</td>
<td>Plasmid pYGS-arg</td>
<td>De Mattos-Shiple et al., 2018</td>
</tr>
<tr>
<td>Ferrirhodin</td>
<td>FSN1</td>
<td>495&lt;sup&gt;a&lt;/sup&gt;</td>
<td>F. sacchari</td>
<td>A. oryzae</td>
<td>Plasmid pYAVASarg</td>
<td>Munawar et al., 2013</td>
</tr>
<tr>
<td>ACV</td>
<td>ACVS (or PchAB)</td>
<td>450</td>
<td>P. chrysogenum</td>
<td>N. crassa</td>
<td>Cosmid Pcs3.2</td>
<td>Smith et al., 1990</td>
</tr>
<tr>
<td>Bassianolid</td>
<td>BaSYN</td>
<td>350</td>
<td>B. bassiana</td>
<td>A. niger</td>
<td>Genomic integration</td>
<td>Boecker et al., 2018</td>
</tr>
<tr>
<td>CDP hexa-bassianolid</td>
<td>BaSYN-EnSYN TC&lt;sub&gt;3&lt;/sub&gt;</td>
<td>350</td>
<td>B. bassiana</td>
<td>A. niger</td>
<td>Genomic integration</td>
<td>Boecker et al., 2018</td>
</tr>
<tr>
<td>Beauvericin</td>
<td>BeSYN</td>
<td>350</td>
<td>B. bassiana</td>
<td>A. niger</td>
<td>Genomic integration</td>
<td>Boecker et al., 2018</td>
</tr>
<tr>
<td>Octa-beauvericin</td>
<td>BeSYN-BaSYN TC&lt;sub&gt;3&lt;/sub&gt;</td>
<td>350</td>
<td>B. bassiana</td>
<td>A. niger</td>
<td>Genomic integration</td>
<td>Boecker et al., 2018</td>
</tr>
<tr>
<td>Einniatin</td>
<td>EnSYN</td>
<td>350</td>
<td>F. oxyporum</td>
<td>A. niger</td>
<td>Genomic integration</td>
<td>Richter et al., 2014</td>
</tr>
<tr>
<td>Octa-enniatin</td>
<td>EnSYN→C&lt;sub&gt;3&lt;/sub&gt;BaSYN C&lt;sub&gt;3&lt;/sub&gt;</td>
<td>350</td>
<td>F. oxyporum</td>
<td>A. niger</td>
<td>Genomic integration</td>
<td>Boecker et al., 2018</td>
</tr>
<tr>
<td>[PheLac]Beauvericin</td>
<td>PSYN M&lt;sub&gt;3&lt;/sub&gt;→BeSYN M&lt;sub&gt;3&lt;/sub&gt;</td>
<td>350</td>
<td>R. abscindita</td>
<td>A. niger</td>
<td>Genomic integration</td>
<td>Boecker et al., 2018</td>
</tr>
<tr>
<td>Microperforanone</td>
<td>MicA</td>
<td>103&lt;sup&gt;a&lt;/sup&gt;</td>
<td>A. nidulans</td>
<td>A. niger</td>
<td>Genomic integration</td>
<td>Guo et al., 2015</td>
</tr>
<tr>
<td>Aspulvinone E</td>
<td>ApvA</td>
<td>92&lt;sup&gt;a&lt;/sup&gt;</td>
<td>A. terreus</td>
<td>A. niger</td>
<td>Genomic integration</td>
<td>Geib et al., 2016; Geib and Brock, 2017</td>
</tr>
</tbody>
</table>

<sup>a</sup> Estimated on the basis of the amino acid sequence length (with an average of 110 Da per amino acid).
3.3. Fungal expression systems

3.3.1. Yeast heterologous expression systems

The first study reporting a successful heterologous expression of a fungal NRPS in yeast was related to the production of ACV in *H. polymorpha* (Gidijala et al., 2008). This example will be discussed in detail in Section 4. In contrast, for all subsequently reported heterologous expressions of NRPS in yeast, *S. cerevisiae* was selected as a host (Table 2), for example with the fumitremorgin biosynthetic pathway from *A. fumigatus* (Tsunematsu et al., 2013a). The precursor of fumitremorgin is the NRP brevianamide F, resulting from a condensation of L-tryptophan and L-proline by FtmA NRPS. This compound is then prenylated by FtmB to form tryprostatin B, before being further modified by FtmC, FtmD, FtmE, and FtmF leading to spirotryprostatin A or spirotryprostatin B. After the cloning of the ftmA gene were amplified and assembled together using overlap extension PCR methodology. The resulting 11.8-kb fragment was cloned under the control of the strong constitutive *Pfae* promoter in pKW1810, a *S. cerevisiae* expression plasmid, by TAR cloning. The coding sequences for Flag and 6xHis were added, respectively, to the 5’ and 3’ ends of the NRPS gene. The presence of tags attached to protein ends facilitates the further detection and purification of the produced NRPS. In this study, the authors were able to confirm the expression of Afu6g12080 by Western blot using antibodies directed against 6xHis and Flag tags.}

A heterologous expression approach has also been used to experimentally assign a NRPS cluster identified on the basis of genome mining to the synthesis of a given NRP. In *A. fumigatus*, 14 gene clusters were identified to be involved in the biosynthesis of NRPs (Frisvad et al., 2009). Among them, Afu6g12080 was identified and predicted to be involved in fumiquinazoline biosynthesis (Ames et al., 2010). Using a similar approach to that taken with *ftmA*, total mRNA was extracted from *A. fumigatus* and converted into cDNAs by a reverse transcriptase. From this cDNA, multiple overlapping PCR fragments with homologous ends (< 3 kb) of the introns-free Afu6g12080 NRPS gene were amplified and assembled together using overlap extension PCR methodology. The resulting 11.8-kb fragment was cloned under the control of the strong constitutive *Pfae* promoter in pKW1810, a *S. cerevisiae* expression plasmid, by TAR cloning. The coding sequences for Flag and 6xHis were added, respectively, to the 5’ and 3’ ends of the NRPS gene.

3.3.2. Yeast heterologous expression systems

**Fig. 4.** Description of the four main cloning methods enabling the reconstitution, from multiple amplified PCR fragments, of long-length NRPS sequences into a targeted vector: A. Overlap extension PCR approach, B. Gibson assembly, C. USER Cloning, and D. TAR Cloning (homologous recombination).
production, Gao and coworkers have introduced the npgA PPTase encoding gene into the genome of S. cerevisiae (Strain BJ5464-NpgA). Meanwhile, the tqaA cDNA, devoid of introns, was obtained by reverse transcription on mRNA extracted from P. aerophilum. Subsequently, TAR cloning enabled the integration of the 5'-Flag modified NRPS gene in the yeast expression plasmid pWX55 under the control of the strong constitutive ADH2 promoter. The heterologous expression of tqaA was demonstrated through successful purification of the 450-kDa synthetase with anti-Flag antibody affinity chromatography followed by gel filtration. Afterward, the activity of the synthetase was confirmed with the detection of fumiquinazoline (Gao et al., 2012).

The strategy developed to express tqaA was repeated several times to heterologously express NRPS genes in the BJ5464-NpgA strain with the same plasmid. The arlA NRPS gene from Neosartorya fischeri, was in such manner expressed in this strain, leading to tricyclic ardeemin FQ production (Haynes et al., 2013). From the same filamentous fungi, anaPS, involved in the biosynthesis of a precursor of acetylaslanol, was heterologously expressed to investigate the role of the terminal condensation domain in NRP cyclization (Gao et al., 2012). The expression of the 276-kDa bimodal synthetase AspA from Aspergillus alliaceus in S. cerevisiae revealed the iterative mode of synthesis of the first module as it led to the synthesis of a tri-peptide Ant-Ant-Trp (Gao et al., 2013). Likewise, the production of fungal anticancer cyclooligomer depsipeptides, initially produced by B. bassiana, was reconstituted in S. cerevisiae through cloning of the gene encoding the bassianolide synthetase (348 kDa) and the beauvericin synthetase (352 kDa) (Yu et al., 2017; Yu et al., 2013).

Despite the opportunities offered by TAR cloning methodology, the insertion and correct expression of large sequences, such as NRPS genes, remain difficult in numerous cases. One potential approach to address this challenge is to dissect and clone the synthetase gene into separate modules. The feasibility of module dissection was first been highlighted with the aspyridone synthetase AdpA, encompassing a single NRPS module fused to the C-terminal part of a PKS. The sequence encoding this module was isolated from A. nidulans and expressed in S. cerevisiae. Subsequently the heterologous AdpA module was purified, prior to confirming the correct enzymatic activity by in vitro production of acyltetramic acid in co-incubation with the PKS part (Xu et al., 2010). Conversely, several mono-module NRPS-encoding genes from Aspergillus terreus were expressed in the multi-copy pESC-URA plasmid: apvA-melA, pgnA, and ATEG_03090, respectively involved in the biosynthesis of aspyridine A, phenguinaric acid, and atromentin (Hühner et al., 2018). The cryptic mono-module NRPS encoding gene ATEG_03630 from the same filamentous fungi was similarly awakened by heterologous production, leading to the production of 2,4-dihydroxy-5,6-dimethyl benzaldehyde (Wang and Zhao, 2014).

NRPSs encompassing more than one module can be dissected and each module cloned separately. Nonetheless, some key factors must be considered. Primarily, the position where the NRPS can be dissected is crucial. In the case of the bimodal AspA, the A1-T1-C2 tridomain can be solubly expressed unlike the bidomain A1-T1. Similarly, the second module was soluble in the four domains C2-A2-T2-Ct form and not in the A2-T2-Ct form (Gao et al., 2013). Recently, Bozhiyuk et al. (2019) have suggested an alternative fusion site area between the two sub-domains of the C-domain, where NRPSs can be dissected. Authors postulated that the self-contained catalytically active unit can be redefined as CAsub-A-T-CDubh where CAsub is the sub-domain accepting the growing peptide chain in the N-terminal C-domain and CDubh is the sub-domain providing the peptide chain in the C-terminal C-domain. Genetic engineering enables, by exchanging these functional units between different NRPSs with similar domain composition, to generate chimeric NRPSs assembling novel NRPs. Up to now, these units have been employed only to reconstitute entire chimeric NRPSs, and there is no evidence that they can be expressed separately in a soluble form (Bozhiyuk et al., 2019). A second point to consider for the dissection of NRPSs is related to the presence of intermodular linkers, which are short amino acid regions connecting covalently subsequent modules. The importance of these short sequences has recently been highlighted between modules one and two of the beauvericin and bassianolide synthetases (Yu et al., 2013). The dissection of these synthetases by cloning module 1 and module 2+3 into two distinct plasmids without disrupting the pre-identified linker sequence led to an effective production of beauvericin and bassianolide. In contrast, the same experiment without the linker did not lead to any production of cyclooligomer depsipeptides. In addition to linkers, there are also short amino acid sequences, called COM domains, enabling noncovalent association of the different proteins when the NRPS is composed of several proteins such as in the tyrocidine synthetase or the surfactin synthase. These COM domains are essential when different fractions of the NRPSs are expressed separately but can also be exploited to generate chimeric NRPSs systems. This was demonstrated with the co-expression of the first module of the tyrocidine NRPS (TycA) from Brevibacillus parabrevis with the third module of the surfactin synthase (SrFAc) from Bacillus subtilis (Siewers et al., 2010). TycA was expressed with its native COM domain, whereas SrFAc was fused with the COM domain, compatible with COM, isolated from TycB module. Co-expression in S. cerevisiae of these two complexes into distinct pESC plasmids lead to the expected production of deoxy-Phe-Leu.

3.3.2. Filamentous fungi heterologous expression systems

With the exception of the pchAB NRPS gene (involved in ACV synthesis), expressed in Neurospora crassa (Smith et al., 1990), Aspergillus spp. were predominantly used to heterologously express NRP compounds (Table 3). One of the earliest reported NRP heterologous expressions in filamentous fungi was related to brevianamide F from A. fumigatus. The brevianamide NRPS gene, also later expressed in S. cerevisiae as described above, had previously been identified through heterologous expression in A. nidulans (Maiya et al., 2006). In order to prevent mis-amplification, the 8.7-kb fmaA gene encoding the bimodal NRPS was amplified from A. fumigatus as two distinct fragments (5' and 3'). The fragments were assembled together, cloned under the control of the alcohol dehydrogenase promoter (PalcA), and integrated in the genome of the host strain using pyrG as a selection marker. In A. nidulans, PalcA is a very strong promoter induced by ethanol or γ-threonine, and repressed in the presence of glucose (Felenbok et al., 2004). Cultivation under induction conditions led to brevianamide F synthesis with a titer of 45 mg·l−1, which represents a 3,500 fold increase as compared to native strain A. fumigatus A2936. Later, fmaA gene, assembled through ExRec overlap-extension PCR-yeast homologous recombination (Fig. 4), was also expressed in A. niger (Tsunematsu et al., 2013a). The FmaA synthetase from N. fischeri, displaying 88% identity at the amino acids level with that from A. fumi-
gatus, was also successfully expressed in A. nidulans (Wunsch et al., 2015). FmaA gene under the control of the constitutive glyceraldehyde-3-phosphate dehydrogenase promoter (PgalA) was inserted into the genome of A. nidulans, resulting in a production of brevianamide F ranging between 9.7 and 36.9 mg·l−1, depending on the mutant considered. Interestingly, the productivity from one positive mutant to another displayed large variations (from 1 to 4 magnitudes) (Maiya et al., 2006; Wunsch et al., 2015). This can be due to the ectopic integration of the gene in variable regions of the genome.

Numerous fungal heterologous expression studies were conducted to decipher the mode of synthesis of NRPSs. For example, the typical mode-of-operation of two siderophore synthetases has been highlighted by heterologous expression in filamentous fungi. For this purpose, FSN1 from Fusarium sacchari and CsNPS2 from Ceriporiopsis subvermispora were respectively expressed in A. oryzae and A. niger (Brandenburger et al., 2017; Munawar et al., 2013). The FSN1 gene is encoding a NRPS involved in the synthesis of ferrirhodin, whereas CsNPS2 is related to the biosynthesis of a type VI siderophore. In both cases, the NRPSs were operating through an iterative mechanism with three additional thio-
lation-condensation domains receiving the same monomer (N-acetyl-
N$_5$-hydroxy-$\alpha$-ornithine) from a single upstream located adenylation domain. Likewise, through heterologous expression in *A. nidulans*, the synthesis of asperphenamate has been demonstrated to be dependent on the joint action of two distinct NRPSs: ApmA and ApmB. Indeed, the NRP produced by ApmA, N-benzoylphenylalaninol, binds onto the monomer attached to ApmB to generate subsequently the asperphenamate compound (Li et al., 2018). On the other hand, there are some studies reporting identification and confirmation of NRP biosynthetic pathways through heterologous expression. In the case of gliotoxin, a toxin synthesized by *A. fumigatus*, the related gene cluster was known but the role of each gene had not been elucidated (Chang et al., 2013). NRPS gene *glpE* expressed under the control of P$_{lac}$ in *A. nidulans* was identified to be involved in the first step of gliotoxin biosynthesis with the assembly of NRPS-type precursors: cyclo-$\alpha$-phenylalaninyl$_1$-seryl) and cyclo-$\alpha$-tryptophanyl$_1$-seryl). Likewise, by heterologous expression in *A. oryzae*, psycyA and psycyB genes from *Penicillium sojini* were demonstrated to be the only two genes required for the biosynthesis of cycloaspeptide. Methylated amino acids constituting this NRP are encoded by *pscyA* and *pscyB* genes, while the peptide assembly is performed by a 5-modules NRPS encoded by psycyB (De Mattos-Shipley et al., 2018). Another reported example was the case of KK1, a NRP natively produced by *Carvularia clavata*, displaying anti-fungal activity against the plant pathogen *Botrytis cinerea*. The 40-kb gene cluster potentially involved in KK1 biosynthesis was split into two fragments and cloned into *A. oryzae*. The transformed strain was able to produce KK1, demonstrating the role of this gene cluster in the NRP synthesis (Yoshimi et al., 2018).

Some mono-modular NRPSs have also been heterologously expressed in filamentous fungi. The *melA* gene from *A. terreus* was cloned into *A. nidulans* (Guo et al., 2015) and *A. niger* (Geib et al., 2016; Geib and Brock, 2017) resulting in the production of aspilavinone E. Similarly, following the cloning of *miaA* obtained from *A. nidulans*, the transformed *A. niger* strain was able to produce microperfunnone (Yeh et al., 2013). *A. niger* was also used to heterologously express several cyclodepsipeptides, high value NRP compounds, with anticancer, anti-microbial and antihelminthic properties (Sivanathan and Scherkenbeck, 2014). Enniatin, beauvericin, and bassianolide belonging to this group of compounds are assembled by iterative bimodular synthetases able to link $\alpha$-$\omega$-hydroxy acid with $\alpha$ amino acids. The *esyn1* NRPS gene from *Fusarium oxysporum* was expressed in *A. niger*, leading to the production of enniatin at initial un-optimized titer of 1 mg.l$^{-1}$ (Richter et al., 2014). Beauvericin and bassianolide were also heterologously produced after the cloning of the corresponding *bbBeas* and *bbBals* synthetase genes from *B. bassiana* (Boecker et al., 2018). In these two studies, a promising *A. niger* heterologous expression host was developed based on the Tet-on bacterial-fungal hybrid expression system. Tet-on system encompassed an rtTAS2-M2-dependent promoter controlling the expression of the NRPS gene and a tetracycline-dependent transactivator rtTAS2-M2. Addition of a tetracycline derivative, doxycycline (Dox), in the medium activated the system and induced the expression of the gene. The expression strength is directly correlated to the Dox concentration added, thus making Tet-on tunable and very efficient, with expression levels equivalent to those obtained with the widely used strong constitutive promoter P$_{gpdA}$ (Meyer et al., 2011). In addition, the impact of the number of *bbBeas* gene copies integrated in *A. niger* genome on the beauvericin productivity was also investigated (Boecker et al., 2018). The strain harbouring a tandem gene copy produced approximately two-fold higher titers of beauvericin than the strain harbouring a single copy, under similar culture conditions (628 versus 294 mg.l$^{-1}$).

The typical iterative mechanism and the high homology of bi-modular cyclodepsipeptide synthetases were also exploited to assemble hybrid NRPSs. In the case of closely related synthetases, inter-module swapping is one of the most effective approaches for altering the incorporated monomer, and thus generating a novel putatively bioactive compound (Brown et al., 2018). For this purpose, the first module of a synthetase is combined with the second module of another system in order to generate new non-natural nonribosomal peptides. Studies reporting this inter-module swapping between cyclodepsipeptide synthetases have employed the above described Tet-on expression system in *A. niger* (Steiniger et al., 2017; Zobel et al., 2016). Module 1 of PSYN NRPSs (from *Rosellinia abscondita*), activating either $\delta$-Lac or $\delta$-Phe, was fused to module 2 of EnSYN (from *F. oxysporum*) and BeSYN (from *B. bassiana*) activating $\gamma$-Val and $\gamma$-Phe, respectively. The subsequently transformed *A. niger* colonies were able to produce six novel compounds, but with lower yields than for native cyclodepsipeptides. The lower yields observed can also be related to a bottleneck in substrates for different derivatives (Zobel et al., 2016). Using a similar approach, after screening hybrid NRPS resulting from several combinations of modules from different cyclodepsipeptide NRPSs (i.e., EnSYN, BeSYN, and BaSYN) in *E. coli*, the most promising combinations were expressed in *A. niger* (Steiniger et al., 2017). The combinations EnSYN-BaSYN-T.C$_3$, BeSYN-BaSYN-T.C$_3$, BaSYN-EnSYN-T.C$_3$, and BaSYN-BeSYN-T.C$_3$ were successfully expressed, leading to high titers of hybrid cyclodepsipeptides: octa-enantiom B (4 mg.l$^{-1}$), octa-beauvericin (10.8 mg.l$^{-1}$), and hexa-bassianolide (1.3 g.l$^{-1}$). Module and domain swapping between different PKS/NRPS is an innovative approach to generate new compounds, difficult to obtain by chemical synthesis, and displaying interesting biological activity (Süssmuth and Mainz, 2017). With this approach, 6 novel compounds were also produced in *Fusarium heterosporum*, after 57 swapping combinations with 34 modules from various fungal PKS-NRPSs (Kakule et al., 2014). Other notable examples of chimeric compounds include: niduchimaerin A/B in *A. nidulans* by swapping the PKS-NRPS CsaA from *Aspergillus clavatus* with the Syn2 from *Magnaporthe oryzae* (Nielsen et al., 2016), and pre-desmethylbassianin A/$\gamma$-pretenellin A in *A. oryzae* by swapping between the PKS-NRPS DMBS and TENS from *B. bassiana* (Fisch et al., 2011).

Likewise, hybrid PKS-NRPS genes have also been heterologously expressed in filamentous fungi to produce compounds of interest. For example, tenellin was produced in *A. oryzae* after expression of *tesyn* PKS-NRPS from *B. bassiana* under the control of the starch-inducible *amyB* promoter (P$_{amyB}$), leading to up to 5-fold higher yields than in the native fungus (Halo et al., 2008; Heneghan et al., 2010). Under the control of the same promoter, a PKS-NRPS from *Chaetomium globosum* was expressed in *A. nidulans*, enabling the production of the potential anti-HIV Sch 210972 (Sato et al., 2015). Once again, P$_{amyB}$ was employed to express the 12.4-kb *ACE1* PKS-NRPS from *M. oryzae* in *A. oryzae*, leading to magnaporphyrynec production (Song et al., 2015). Similarly, the aspirdone biosynthetic pathway (adpA gene cluster) was reconstituted in *A. oryzae* by cloning the genes of *AdpA* cluster from *A. nidulans* (Wasil et al., 2013). Additionally, two successive PKS and NRPS genes (glpk3-gnrps7) in one cluster of *Glarea losoyensis* were also successfully expressed as a polycistronic mRNA in *A. nidulans* to produce xelorlozoyenone (Yue et al., 2015). The biosynthetic pathways of astechrome and macrolactone valactamide were reconstituted also in *A. nidulans* through fungal artificial chromosomes.

### 3.4. Key factors to scale up heterologous production of NRPs

Multiple key factors have to be considered to succeed in NRP heterologous production and to increase the productivity. Each step, from the cloning and the transcription of the long heterologous gene to the functional synthetase, can be a barrier to the production (Fig. 5). Some of these limiting steps, and the possible solutions to overcome them, will be discussed in this section.

#### 3.4.1. Incompatibility of the host with foreign NRPS nucleic and amino acid sequences

The incompatibility of the foreign nucleic acid sequences with the host machinery is a major cause of failure of heterologous production. This incompatibility is more pronounced when the native producer is
taxonomically distant from the host organism. It can mainly be related to a difference in codon usage slowing down or ending the mRNA translation. As mentioned above, codons of the foreign sequence can be optimized by replacing rare codons by more common ones for the host (Ongley et al., 2013). This approach also enabled, in Aspergillus species, a stabilization of the transcripts of heterologous genes by preventing premature polyadenylation and subsequent degradation of mRNA (Tanaka et al., 2014). However, this methodology is time consuming in the design of the optimized sequence and relies on still expensive DNA synthesis technology.

At the protein level, the large size of NRPSs increases the possibility to display recognition sequences for host proteases. In particular, the amino acid sequences of inter-module sequence regions, which are largely unfolded (i.e., limited number of α-helix and β-sheet) and display a high degree of flexibility, are therefore more vulnerable to protease activity (Samel et al., 2007; Tarry et al., 2017). Yeasts have been described to express multiple vacuolar proteases, such as endoproteinases A and B, potentially released during cell lysis. To prevent the degradation of the newly-synthesized NRPS, a S. cerevisiae strain has been mutated in pep4 and prb1 genes encoding the previously mentioned proteases (Jones, 1991; Kealey et al., 1998). The resulting BJ5464 strain has been widely used for heterologous production of various NRPs (Gao et al., 2013, 2012; Haynes et al., 2013; Yu et al., 2013).

In addition, non-native large multi-modular complexes may undergo a complete or partial mis-folding, leading to a loss of enzymatic
activity due to rapid degradation of the complex or an inappropriate tertiary structure of catalytic sites. This can be solved or attenuated by co-expressing foreign chaperone encoding genes, which are likely to be involved in the folding of the synthetase in the native host (Ongley et al., 2013). Alternatively, it was reported that a lower growth temperature could promote the folding of heterologously produced proteins. For example, an increase in the stability of pcbAB synthetase was demonstrated in H. polymorpha grown at 25°C instead of 37°C (Gidijala et al., 2008). Nonetheless, this effect was not observed in S. cerevisiae with the same NRPS with the native PPTase gene (Siewers et al., 2009). Moreover, in the enantiomer heterologous producer A. niger, the effect of the temperature was also investigated among other parameters. It appears that reducing the temperature from 30°C to 26°C did not significantly improve the productivity, especially compared to the impact of the culture medium composition (Richter et al., 2014).

3.4.2. Required enzymes for the biosynthesis of active nonribosomal peptides

In most cases, obtaining the properly folded synthetase is not sufficient to enable the complex to synthesize the compound of interest. Indeed, to be active, NRPSs have to be converted from their apo form into their holo forms through the addition of a flexible phosphopantetheine arm on each thiolation domain by a phosphopantetheinyl transferase. This cofactor is crucial because it carries the monomers previously specifically recognized by the different A-domains. In contrast to filamentous fungi, yeasts (except S. pombe) lack a functional PPTase (Schwecke et al., 2006). An endogenous lys5 PPTase gene, involved in lysine biosynthesis, is expressed in several yeasts but the corresponding enzyme was demonstrated to be unable for NRPS activation (Gidijala et al., 2009; Moote et al., 2002). To complement the lack of a functional PPTase, sfp from B. subtilis or npgA from A. nidulans, are generally co-expressed with NRPS genes. NpgA and Sfp have been reported to display a very broad specificity on various NRPSs in diverse fungal and bacterial organisms (Ishiszchi et al., 2012; Mofid et al., 2002; Quadri et al., 1998; Tobias et al., 2016). The efficiency of these two enzymes was demonstrated in S. cerevisiae, and the same order of magnitude of NRPS titers was obtained with both of them (Siewers et al., 2009). Conversely, in E. coli, the productivity in terms of cyclodepsipeptides could be increased by 4.8- and 14-fold with npgA gene instead of sfp and endogenous PPTase, respectively (Steiniger et al., 2017). The npgA PPTase gene has even been integrated in the genome of a S. cerevisiae strain to generate a cellular platform ready to produce NRPs compounds (Gao et al., 2013, 2012; Haynes et al., 2013; Yu et al., 2013).

NRPs synthesized by functional synthetases are not necessarily active compounds, and are often only early intermediates of the targeted bioactive product. Therefore, to obtain the final functional product, addition of gene(s) encoding the tailoring enzyme(s) responsible for NRPS modification step(s) is required. This is the case for example, for ACV, which is sequentially modified by three tailoring enzymes to give isopenicillin G (Fig. 6B). To heterologously produce isopenicillin G in yeasts, and not only ACV, addition of the isopenicillin N synthase, isopenicillin N acyl transferase and phenylacetyl CoA ligase genes was required (Awan et al., 2017; Gidijala et al., 2009, 2008). Moreover, the two last above mentioned tailoring enzymes are located in the peroxisome in the native producer Penicillium chrysogenum. This parameter can be crucial, as the efficiency of these enzymes is impacted by their location and environment. An absence of compartmentalization of these enzymes in peroxisomes decreased the productivity by more than 60% in heterologous hosts (Gidijala et al., 2009). This trend was confirmed later in S. cerevisiae with an increased productivity gained by tagging these enzymes with PTS1 peroxisome addressing signal (Awan et al., 2017).

Gliotoxin produced by A. fumigatus, is also a bioactive product resulting from multiple modifications of a NRPS diketopiperazine encoded by gliP synthetase gene. The first step of the biosynthetic pathway from the NRPS to gliotoxin was elucidated through a heterologous expression in A. nidulans (Chang et al., 2013; Gardiner and Howlett, 2005). Two tailoring enzymes candidates, namely gliC and gliF were co-expressed separately with gliP synthetase gene. The disappearance of the precursor NRPS in favor of an L-Phe hydroxylated derivative, was only observed in the strain harbouring the cytochrome P450 monoxygenase encoded by gliC. This result demonstrated the role of this tailoring enzyme in the first step of gliotoxin biosynthesis.

Similarly, brevianamide F nonribosomal compound, assembled by FtmA (or FtmS) in A. fumigatus and N. fischeri, is a platform molecule, precursor to multiple compounds. Wunsch and co-workers have co-transformed ftmS in A. nidulans with three different prenyltransferase genes from the two previously mentioned fungi. The resulting strains were able to produce, in addition to brevianamide F, up to three supplementary derivative compounds, such as deoxybrevianamide E (Wunsch et al., 2015). With co-expression of ftmA with ftmB, ftmC, ftmD, ftmE, ftmF tailoring enzyme-encoding genes, production of derivatives of brevianamide F was also observed. The NRPS-encoding gene and five tailoring enzyme genes were assembled successively in the same single plasmid, through multiple homologous recombinations in S. cerevisiae. All ORFs were positioned in the plasmid under the control of distinct promoters and terminators. After transformation in both S. cerevisiae and A. nidulans, the enzymatic activity of tailoring enzymes enabled the synthesis of multiple precursors of spirottryprostatin starting from brevianamide F: tryprostatin B, tryprostatin A, and fumitremorgin C (Tsunematsu et al., 2013a).

3.4.3. How to address NRPS precursors deficiency?

In the classical ribosomal pathway, only 20 proteinogenic α-amino acid residues are incorporated into proteins. In addition to the proteinogenic α amino acid residues, up to 500 distinct monomers were reported to constitute NRPs building blocks (Cabocche et al., 2010). Among this tremendous biodiversity of monomers, some are directly recognized and incorporated as such, while others are modified during the peptide synthesis by the activity of a NRPS domain. This incredible variety of precursor monomers is the root of the wide diversity of structure and activity of NRPs (Challis and Naismith, 2004; Keller and Schauwecker, 2003). On the other hand, this feature raises a serious issue regarding heterologous production, as hosts lack the ability to synthesize most of these non-common monomers. A possible approach to complement monomers deficiency relies on the cloning of the genes involved in the monomer biosynthesis pathway. The monomer ω-hydroxyisovaleric acid (ω-Hiv) is for instance not synthetized by A. niger and poorly produced by S. cerevisiae. ω-Hiv is a precursor engaged in the biosynthesis of cyclodepsipeptides, such as enniatin, beauvericin and bassianolide. It can be obtained by the modification of α-ketovaleric acid by α-ketoisovalerate reductase encoded by kivR gene in B. bassiana or P. oxysporum, among others (Richter et al., 2014; Yu et al., 2013). A two-fold increase of productivity was obtained with the expression of kivR gene in a S. cerevisiae strain heterologously producing beauvericin (61.7 and 33.8 mg.l⁻¹ with or without kivR, respectively) (Yu et al., 2013). In A. niger, the enniatin heterologous production has also been made possible with the co-expression of the missing kivR encoding α-ketoisovalerate reductase with esyn1 NRPS gene (Richter et al., 2014).

On the other hand, a precursor deficiency can also be solved by feeding the corresponding compound directly into the culture medium. This approach avoids time consuming gene cloning steps but relies on the host strain being able to take up the precursor from the medium. The supplementation of the above mentioned A. niger, expressing foreign enniatin synthetase, was also achieved by feeding it with ω-Hiv instead of co-expressing kivR. The feeding experiment was quite successful, as it yielded a high enniatin B titer of 950 mg.l⁻¹ after optimization of feeding conditions and culture medium composition. Under the same conditions, the strain fed with 20 mM ω-Hiv was produced 25% more than the strain transformed with kivR (Richter et al., 2014). Nonetheless, the cloning approach remains more cost effective due to
the high cost of D-HIV (Yu et al., 2013). Moreover, some adenylation domains display a broader specificity that can be leveraged to generate alternative final products by feeding with alternative precursors. For instance, enniatin B is preferably produced by feeding L-valine in the medium, whereas the alternative enniatin A production is enhanced by feeding with L-leucine or L-isoleucine (Richter et al., 2014). In another filamentous fungus, A. nidulans, asperphenamate production was knocked down by mutating the apmA gene responsible for N-benzoyl-phenylalaninol biosynthesis. This production was subsequently restored by feeding the same strain with the deficient precursor (Li et al., 2018).

In yeast, the relevance of precursor supplementation in the culture medium has been highlighted in H. polymorpha expressing pcbAB NRPS gene. The production of ACV was not possible due to the incapacity of H. polymorpha to synthesize the precursor aminoadipic acid. Supplementation of the culture medium with this monomer enabled to unlock the production of ACV (Gidijala et al., 2009). Nonetheless, the use of a precursor feeding strategy is not always a guarantee of success as some negative feedback mechanisms can occur and hinder the compound production. For instance, the feeding of the NRP-precursor 2-amino-butyric acid in the culture medium of C. asteris, was recently demonstrated to negatively impact the production of the astin C compound as well as the development of the fungus (Vassaux et al., 2018).

In yeast, the relevance of precursor supplementation in the culture medium has been highlighted in H. polymorpha expressing pcbAB NRPS gene. The production of ACV was not possible due to the incapacity of H. polymorpha to synthesize the precursor aminoadipic acid. Supplementation of the culture medium with this monomer enabled to unlock the production of ACV (Gidijala et al., 2009). Nonetheless, the use of a precursor feeding strategy is not always a guarantee of success as some negative feedback mechanisms can occur and hinder the compound production. For instance, the feeding of the NRP-precursor 2-amino-butyric acid in the culture medium of C. asteris, was recently demonstrated to negatively impact the production of the astin C compound as well as the development of the fungus (Vassaux et al., 2018).

A further challenge facing NRP heterologous production is related to the toxicity of some compounds towards the selected host. Native producers have developed various strategies to prevent this toxicity, such as specific secretion mechanisms, compartmentalization of the toxic products, or expression of genes encoding enzymes able to induce a resistance (Bond et al., 2016; Chanda et al., 2009; Martín et al., 2005). These mechanisms are often specific to the toxic NRP product and therefore may be only present in the native producer and not in the chosen host (Nielsen and Nielsen, 2017). The most straightforward strategy to overcome this issue consists to express the NRPS gene under the control of an inducible promoter. This approach enables the choice of the period of production of the toxic compound and to dissociate it from the cell growth period. Hence, GAL1-10, amyB and Tet-on inducible promoters were used to express NRPS genes, respectively in S. cerevisiae (Siewers et al., 2009), A. nidulans (Fujii et al., 1996), and A.
niger (Richter et al., 2014).

Another common approach consists of transferring into the host not only the NRP biosynthesis related genes, but also the specific resistance genes pre-identified into the genome of the native producer. The mechanisms of self-resistance to a produced toxic compound can be classified into three categories of mode-of-operation: (1) by inducing the compound efflux, (2) through enzymatic deactivation of the toxic product, or (3) with a targeted alteration of the drug receptors (Cundliffe and Demain, 2010). Considering that the latter are quite uncommon, and that the deactivation of the toxic product can also impair the targeted activity of the compound, the implementation of an efflux system is often the most straightforward approach to confer a resistance (Keller, 2015). NRPs are likely to be small structures and accordingly possibly naturally secreted by some organisms. Nonetheless, numerous studies reported a poor secretion of NRPs by heterologous strains such as H. polymorpha (Gidijala et al., 2009), S. cerevisiae (Awan et al., 2017), and A. niger (Zobel et al., 2016). Therefore, expression of a foreign NRP specific transport protein can be required to improve the secretion procedure (do Valle Matta et al., 2001). However, the identification of resistance genes (regardless of their mode-of-operation) in the native producer is not always easy. In filamentous fungi, genes conferring resistance to toxic secondary metabolites have been highlighted multiple time within the corresponding gene cluster. For instance, in the gliotoxin BGC of A. fumigatus, the glfA and glfT genes respectively encode an efflux pump and a gliotoxin oxidoreductase which can reduce the toxicity of the compound (Dolan et al., 2015; Sharma and Chowdhary, 2017). Nonetheless, these resistance genes are not systematically located within the corresponding gene cluster, and the annotation methods are often far to be sufficiently efficient to identify precisely resistance genes, especially in fungal organisms (Tran et al., 2019).

There are further options to protect the heterologous host from the toxicity of the foreign compound, such as extracting the product from the culture medium. The toxic compound can for instance be extracted by introducing a layer of organic solvent on the top of the culture, where the product can diffuse (Bond et al., 2016; Rodriguez et al., 2014). Some absorber resins are also able to bind the toxic compound when they are added to the culture broth. This was for instance demonstrated with the Diaion HP20 resin binding the mycotoxin cecoramic acid contained in a fungal fermentation broth, thus enabling to increase by 100-fold the final product yield (Singh et al., 2010).

3.4.5. From the heterologous host to the in vitro production assays: a viable approach?

3.4.5.1. Approach and methodologies. The in vitro NRP synthesis constitutes a cell-free production system where the enzymes involved in the biosynthetic pathway (NRPS gene, tailoring enzymes) are purified and subsequently incubated with the precursor monomers to generate the desired compound. The in vitro approach is a suitable solution to address most of the issues described above. Indeed, it addresses the deficiency in NRP precursors, but also circumvents issues of NRP toxicity towards the host strain. In order to perform in vitro assays, the synthetase and/or associated tailoring enzymes need to be purified in soluble form and in sufficient quantity from the heterologous host. Addition of a tag sequence, either on the C-terminal or on the N-terminal part of the heterologously expressed synthetase, is highly recommended to facilitate further extraction-purification steps. The use of 6xHis tag (Brandenburger et al., 2017; Gao et al., 2013; Haynes et al., 2013) or Flag tag (Gao et al., 2012), respectively purified through metal or antibody affinity chromatography, have been reported in previous in vitro NRPS studies. Alternatively, NRPSs were also purified by size exclusion chromatography as their high molecular weight facilitates their separation from other host proteins (Fickers et al., 2008; Lawen and Zocher, 1990).

Thereafter, it is possible to assess the presence of the phosphopantetheine arm on thiolation domains through three approaches. The most recently developed method consists to generate a peptide mixture by a trypsin digestion of the proteins contained in the cell extract, before to separate fractions from the peptide mixture through a strong cation exchange chromatography. Subsequently, a MS/MS analysis enables to detect into the different fractions the presence of unique marker ions derived from the NRPS-bonded phosphopantetheine arm (Bumpus et al., 2009). Alternatively, the presence of the phosphopantetheine arm on the holo-NRPS can also be evaluated through the thioester-binding assay measuring the consumption of radiolabeled [14C]-amino-acids incorporated in the synthesized NRP (Schwarzer et al., 2001). An older technic consists to perform an alkaline phosphate treatment to hydrolyze the phosphopantetheine part from the holo-NRPS. This treatment releases a pantetheinic acid which can be quantified through a microbiological assay against Lactobacillus plantarum (Lawen and Zocher, 1990). Regarding to A-domains, the most common approach to confirm their enzymatic activity consists to quantify the incorporation of 32P into ATP in presence of the precursor amino acids (ATP pyro-phosphate exchange radioassay) (Brandenburger et al., 2017; Fickers et al., 2008). Alternatively, a nonradioactive method has also been developed enabling to evaluate the A-domains activity through a MS/MS analysis. This technic relies on the measurement of the isotopic back exchange of the pyrophosphate into a [18O]-labeled ATP (Phelan et al., 2009). Recently, Duckworth et al. (2016) have developed another method based on the measurement of the pyrophosphate released during the adenylation reaction. The pyrophosphate released subsequently to the action of an active A-domain is first cleaved into an inorganic phosphate by an inorganic pyrophosphatase. Subsequently, a purine nucleoside phosphorylase uses this inorganic pyrophosphate in a reaction converting 7-methylthio- guanosine into 7-methylthioguanine which can be monitored with a spectrometric-based method at 360 nm (Duckworth et al., 2016). Finally, the functionality of the holo-NRPS complex can be confirmed with the identification of the final NRP by using appropriate analytical method (LC-MS and/or NMR).

3.4.5.2. Successful in vitro assays for reconstitution/production of fungal NRPs. Genome mining and in silico approaches have enabled, in C. subvermispora, to assign the basidioferrin siderophore synthesis to the CnNPS2 NRPS-encoding gene. Experimental validation of these predictions has been performed through heterologous production coupled with in vitro assays. CnNPS2 was expressed in A. niger and subsequently purified through metal affinity chromatography. The single A-domain of CnNPS2 displayed specificity for two substrates: N6-acetyl-N5-cis-hydroxy-ornithine (L-AHO) and N5-cis-acyethylvalonvaloyl-N5-hydroxy-ornithine (AMHO). Nonetheless, basidioferrin was only detected after 24 h in vitro assay with 1 mM L-AHO and not with L-AMHO. These results confirmed the role of CnNPS2 in the basidioferrin biosynthesis (Brandenburger et al., 2017).

Several NRPSs were also heterologously expressed and purified from S. cerevisiae for the purpose of in vitro assays: AnaPS (Gao et al., 2012), TqaA (Gao et al., 2012), AspA (Gao et al., 2013), ArdA (Haynes et al., 2013). The presence of the ngA gene in the genome of the yeast was compulsory to produce the PPTase enabling to convert the synthetase from the apo to the holo activated form. In vitro assays with 10 μM pure AspA with its monomer Ant (1 mM) and L-Trp (1 mM) generated a tripeptide, the iterative mode-of-operation of this bimodular NRPS (Gao et al., 2013). The TqaA synthetase is to date the largest NRPS protein (450 kDa) whose activity has been reconstituted in vitro in its active form (Gao et al., 2012). The ArdA synthetase from N. fischeri, displaying high homology with TqaA, was also used to perform in vitro assays after being purified from a heterologous S. cerevisiae producer. This 430-kDa bimodular NRPS was demonstrated in vitro to sufficiently assemble Ant, L-Ala, and L-Trp into ardeemin FQ (Haynes et al., 2013).

In another in vitro study aiming at obtaining NRP derivatives, Tsumatsu et al. (2013a,b) first produced and purified the different intermediates of the spirotryprostatin biosynthetic pathways in a
heterologous S. cerevisiae. In the meantime, tailoring enzymes FtmG and FqgB were also purified from heterologous yeast. With the purpose of deciphering the role of these tailoring enzymes in the final step of spirotryprostatin synthesis, in vitro assays were performed with the purified enzymes mentioned above. As a result of this work, ftmG encoding a cytochrome P450 has been shown to be involved in spiro-tryprostatin formation in spirotryprostatin B, while the FAD-dependent monoxygenase encoded by fqgB was shown to be responsible for the spiro-carbon formation in spiroprostatin A (Tsunematsu et al., 2013a).

The in vitro NRP assembling strategy is especially useful as a complement to heterologous production by facilitating the elucidation of synthetase mode-of-operation or tailoring enzymes function. Nonetheless, due to the very low amounts of NRPs obtained through affinity chromatography purification, this approach is not promising for increasing NRP yields. Moreover, the cost incurred by the addition of precursor monomers and ATP largely prevents its applications for large scale NRP production.

4. Penicillin heterologous production: a NRP model system

Due to the clinical and commercial importance of β-lactam antibiotics, researchers have attempted for years to develop their heterologous production in hosts renowned for being readily cultivate at industrial scale, and easy to handle through genetic engineering. This multi-decade work has positioned the NRP-derived penicillin compounds as a model system regarding NRP heterologous production. These efforts have enabled both to better understand the ins and outs of the penicillin biosynthetic pathway and to develop various strategies to implement heterologous production in fungal cell factories.

Penicillin β-lactam antibiotics are naturally produced by some Penicillium species, especially P. chrysogenum and to a lesser extent by some Aspergillus species such as A. nidulans (Newbert et al., 1997; Suárez and Peñalva, 1996). In P. chrysogenum, the elucidated penicillin biosynthetic pathway relies on the joint action of several enzymes, whose corresponding genes are mainly located within the same gene cluster (Fig. 6A). The pcbA gene encodes a 3-module NRPS that is responsible for assembling ACV (Fig. 6B). This nonribosomal peptide precursor of penicillin, is then modified into isopenicillin (IPN) by an isopenicillin N acyl transferase (IAT) and a phenylacetyl CoA ligase (PCL) into penicillin G (Gidijala et al., 2009). The penicillin V in which the benzyl group is replaced by a phenoxymethyl group is a common analog of the penicillin G displaying an enhanced stability in acidic environment. This analog is preferably produced by P. chrysogenum when phenoxyacetic acid is added in the fermentation medium (Ball et al., 1978). This appears reasonable because the production of low amounts of penicillin compounds, and the presence of enzymes sharing a high level of sequence identity with those from P. chrysogenum, had been reported in (Suárez and Peñalva, 1996). An associated positive impact of these native enzymes cannot be ruled out to explain the success of this heterologous production.

The first transfer of the penicillin biosynthetic pathway in a heterologous host was achieved into N. crassa and A. niger. The cosmids pCX3.2 containing the ACVS encoding gene cluster was cloned into both heterologous hosts, leading directly to production of penicillin. The presence of the produced antibiotic was successfully assessed through a bioactivity assay against B. subtilis. A. niger was reported to produce much more penicillin V than N. crassa (up to 2.3 μg.ml⁻¹ versus 0.170 μg.ml⁻¹) (Smith et al., 1990). One explanation could be the lack of ACV precursor availability in N. crassa. A second hypothesis rests on the fact that A. niger and P. chrysogenum are closely related species both belonging to the order Eurotiales, unlike N. crassa which is associated to Sordariales.

A few years later, the penicillin biosynthetic pathway was also reconstituted into the methylotrophic yeast H. polymorpha, which is well-known for its superior fermentation properties even at industrial scale (Stöckmann et al., 2009). The 12-kb pcbAB NRPS gene was inserted, under the control of the strong methanol-inducible promoter from alcohol oxidase gene (Paox1) in pHlp24 plasmid. H. polymorpha was transformed with the linearized plasmid in order to integrate the gene at the AOX locus. Subsequently, heterologous expression of ACV synthetase (ACVS) was assessed directly at the protein level by Western blot with antibodies directed to ACVS amino-acid sequence. As described in the Section 3, the homologous PPTase Lys5 of the methylotrophic yeast was not adapted to activate ACVS, and an additional sfp PPTase encoding-gene was inserted. Subsequently, IPNS, IAT and PCL-encoding genes from P. chrysogenum were integrated into the genome step by step. Penicillin G productivity of the resulting yeast was similar to those obtained in the native producer NRRL1951 (1.1 and 1.0 μg.ml⁻¹, respectively). In the same yeast, deletion of the native PEX3 gene, responsible for the peroxisome proliferation, led to a 64% reduced yield (0.4 μg.ml⁻¹) (Gidijala et al., 2009). The importance of the peroxisomal location of IAT and PCL had previously been highlighted in P. chrysogenum and A. nidulans (Spröte et al., 2009).

In subsequent work, S. cerevisiae was also selected to heterologously produce ACV (Siewers et al., 2009). pcbAB was inserted into a replicative multi-copy pESC by TAR cloning under the control of the galactose inducible promoter GAL1/GAL10 (Pgal1/gal10). The resulting construct bearing an URA3 selection gene was used to transform a S. cerevisiae Δura3 strain. One of the assets of pESC is the presence of the bi-directional Pgal1/gal10 enabling the cloning of a second gene. This allows the cloning of the PPTase encoding gene in the same vector as the NRPS gene (if the NRPS is encoded by a single ORF). Under galactose induction, the transformed strain carrying this plasmid was able to express functional ACVS (transcription assessed by qPCR and detection of the NRPS by LC-MS analysis). A codon optimization of the 5’ portion of the ACVS gene was also performed leading to a slight, but not significant, increase of the productivity. On the other hand, pcbAB was also integrated into a long terminal repeats (LTR) retrotransposons locus of S. cerevisiae. LTR retrotransposons are transposable elements present in high number and spread all over the yeast genome, thus making these sites suitable for multi-copy genomic integration (Maury et al., 2016). Yields obtained with genomic integration were 6-fold lower than in the multi-copy PESC plasmid, demonstrating the benefits of the latter expression system. A few years later, another ACV producing S. cerevisiae was obtained with the cloning of pcbAB and npgA genes into a low-copy centromeric plasmid (Awan et al., 2017). Multiple pairs of promoters were tested to enhance the expression of both the pcbAB and npgA genes. The combination of pcbAB and npgA, respectively under the control of PPGK and PPGK constitutive promoters, led to a 4-fold higher ACV productivity than with Pgal1/gal10 in PESC plasmid (=0.28 versus ≈0.07 μg.ml⁻¹). This demonstrated the importance of promoter choice for each gene to be expressed, and the underestimated requirement for compatibility of each promoter/gene pair. Subsequently, the complete biosynthetic pathway to the penicillin G was reconstituted by addition of tailoring enzymes IPN, IAT and PCL. Expression of the tailoring enzymes were also evaluated under multiple combinations of promoters for the sake of optimizing the production. The resulting strain was able to efficiently produce and secrete penicillin G, but with yields still two fold less than in heterologous H. polymorpha (Gidijala et al., 2009).

With the purpose of transferring the multiple genes involved in the penicillin synthesis from P. chrysogenum into A. nidulans SAA248 lacking a native penicillin gene cluster, Unkles and coworkers have exploited the previously described properties of the viral 2A peptide. By inserting viral 2A peptide between each single ORF, it is possible to express from the same plasmid a polycistronic mRNA under the control of a single strong/inducible promoter (Kim et al., 2011). Viral 2A sequences were inserted between pcbAB, IPNS gene (pBch), and IAT gene (pENDE). The resulting 15.5-kb sequence was cloned by in vivo homologous recombination in S. cerevisiae, into pJET1.2 plasmid under the
control of xylose inducible promoter. The obtained plasmid was then transferred to A. nidulans, leading to the transcription of the 15.5-kb polycistronic mRNA. Subsequently, each individual enzyme and especially the massive NRPS were cleaved and their joint action enabled the synthesis of penicillin G. Notably, the 331-T25 mutant having integrated multiple copies of the genes, was able to produce 3-fold more penicillin G than the wild type A. nidulans strain 0760 (0.687 versus 0.234 μg.ml⁻¹) (Unkles et al., 2014). Interestingly, in addition to the heterologous IAT enzyme, a second native IAT was identified in the transformed SAA248 strain. Even if A. nidulans SAA248 has lost its ability to synthesize penicillin like other A. nidulans strains, the presence of genes encoding enzymes involved in the penicillin biosynthetic pathway seems to persist in its genome (Suárez and Peñalva, 1996). An associated positive impact of these native enzymes cannot be ruled out to explain the success of this heterologous production.

It is not straightforward to compare the penicillin production yields obtained with the different heterologous strains described in this section. Nonetheless, authors have each time compared these yields with those observed either in the P. chrysogenum NRRL1951 or A. nidulans 0760 native penicillin producers, enabling some cross-study comparisons. Regarding to yeast heterologous systems, H. polymorpha was reported to produce similar penicillin G amount to P. chrysogenum NRRL1951 strain (~1 μg.ml⁻¹) (Gidijala et al., 2009). In contrast, S. cerevisiae was half efficient, probably due to the considerably less developed peroxisomal system compared to the methylotrophic yeast (Awan et al., 2017). Regarding to filamentous fungi heterologous systems, the 331-T25 mutant of A. nidulans SAA248 expressing penicillin biosynthetic genes coming from P. chrysogenum was 3-fold more efficient than the native penicillin G producer A. nidulans 0760 (~0.69 μg.ml⁻¹), but still 30% less productive than P. chrysogenum NRRL1951 (Unkles et al., 2014). In terms of heterologous penicillin V production, wide variations have been observed between the two described hosts with up to 13-fold higher productivity in A. niger compared to N. crassa. The highest penicillin V concentrations reported in the heterologous A. niger strain was likely due to the close taxonomic relatedness of this strain with P. chrysogenum (Smith et al., 1990).

Over many years, the heterologous production of penicillin has been a long-standing test case of study. It enabled researchers to deal with most of the key factors limiting NRP heterologous production and to unlock some of them. Unfortunately, in terms of productivity, the heterologous production approach is decades behind on current industrial penicillin G production systems. Indeed, several studies were conducted to generate and screen thousands of P. chrysogenum mutants, with the aim of improving production yields. Industrial production, with these selected strains, has been optimized for years in terms of strain productivity and fermentation process (Newbert et al., 1997; Vournakis and Elander, 1983). The analysis of the genome of the strains with an improved penicillin production revealed the presence of multiple copies of the penicillin biosynthetic genes, which is likely to contribute in the increased productivity (Fierro et al., 2006; Müller et al., 2018; Newbert et al., 1997). Another characteristic of high- penicillin producing strains is the loss of several gene clusters involved in secondary metabolite synthesis which were detected in the wild-type P. chrysogenum NRRL195. For instance, the improved penicillin producer P. chrysogenum Wisconsin 54-1255 lacks the ability of producing sorbicillin and penitriinic acid which are present in the genome of the parental NRRL195 strain. The deletion of secondary metabolite biosynthetic pathways also contributed to improve the strain productivity for β-lactam antibiotics (Salo et al., 2016, 2015). Although the current penicillin G industrial production yields are not available in the literature, these yields were reported few years ago to be already more than 1000-fold higher than those obtained with the wild-type P. chrysogenum NRRL1951 (Harris et al., 2009).

5. Concluding remarks

The recent development of high-throughput sequencing technologies combined with substantial drop in sequencing costs have driven an impressive increase in the number of available fungal genomes. This factor coupled with the improvement of genome annotating methodologies has partially unlocked the SM production potential of fungi. Hence, efficient bioinformatics tools now enable a more accurate analysis in terms of nonribosomal pathway identification and substrate specificity predictions. Nonetheless, the prediction robustness is lowered because already elucidated fungal NRPs and NRP genes remain underrepresented in dedicated databases as compared to bacterial ones. Consequently, the predictions based on genomic sequences need to be validated by structural elucidation of the NRPs which implies its efficient production.

The production without genetic manipulation of NRP compounds identified in silico is not systematically possible in native hosts due to the existence of cryptic biosynthetic gene clusters and/or incompatibility between the strain and the laboratory cultivation conditions. Therefore, the heterologous production approach is often the most appropriate strategy to deal with this issue. Multiple fungal cellular platforms have been developed to heterologously produce foreign fungal nonribosomal peptides in particular. Among others, the yeast S. cerevisiae and filamentous fungi related to Aspergillus spp have been widely used. Heterologous production of NRPs is particularly challenging due to a plethora of potential limiting steps. Combining multiple strategies to optimize each of them can greatly increase the chances of successful production. Especially, the transfer of large NRPS sequences encompassing tens of kb is one of the main hurdles to overcome. Several methodologies, such as TAR cloning, ExRec, Gibson assembly, USER cloning, and overlap extension PCR, have been developed to enable the reconstruction of single large NRPs genes or gene clusters. This laborious and time-consuming engineering of large NRPs biosynthetic pathways can be avoided by de novo DNA synthesis, but this remains prohibitively costly in many cases. However, considering that costs of gene synthesis are continuously decreasing, an intensification of the use of this rapid methodology for transferring foreign NRPS genes and gene clusters can be expected in the upcoming years.

The correct transfer of the targeted gene in the heterologous host is not a guarantee of success for the production of the related NRP. For instance, Harvey et al. (2018) have transferred 41 cryptic fungal gene clusters to S. cerevisiae but only 54% have been processed by the yeast to lead to a detectable compound. This result, while promising, demonstrates the challenges related to the heterologous expression of large fungal gene clusters. Nonetheless, the efficiency of the heterologous production of NRPs can be greatly improved by taking into account some key factors. First, the transcription rate can be improved by using an appropriate promoter, and sometimes with adequate initiation factors. Translation and proper folding of the protein are also crucial factors to consider. Optimization of this can be achieved by expression of foreign chaperones enzymes or with different cultivation conditions. It is also important to protect the foreign NRPS from the proteolytic activity of the selected host. Once the properly folded NRPS is obtained and activated via a homologous or heterologous phosphopantetheinyll transferase, the assembly of the NRP can be considered. At this last stage, NRP precursor deficiency or NRP toxicity against the selected hosts can also become limiting factors. The first one can be overcome by feeding or cloning of the missing genes related to precursor biosynthesis, while the second may be prevented by conferring secretion ability to the host with the cloning of transporter-associated genes, for example.

Despite these drawbacks, heterologous production strategies were successfully employed to identify novel NRP compounds, to decipher NRPS modes-of-operation and in some cases, to improve productivity. This approach is especially relevant to producing compounds for which the corresponding genes are cryptic or weakly expressed in native
organisms for which no genetic tools are available yet. Conversely, transferring an NRP biosynthetic pathway to a heterologous host is often not sufficient to overcome production yield obtained with a native producer already optimized and well industrially established strain as in the penicillin case. Overall, the NRP production rates currently obtained with heterologous expression do not yet reach levels suitable for industrial production. Chemical synthesis remains in most of the cases, the more competitive way to obtain peptides and especially regarding linear peptides including only proteinogenic amino acids. A possible exception concerns certain high-value compounds, which are cyclic, or include unusual building block and/or monomers which are further modified by tailoring enzyme. Indeed, these specific characteristics are often difficult to obtain via chemical synthesis and should be obtained more easily through a heterologous expression of the corresponding NRP-encoding gene.

Declaring of Competing Interest
Authors of the manuscript declare there is no conflict of interests.

Acknowledgments
This work was supported by the CPER Alibiotech project; the INTERREG V France-Wallonie-Vlaanderen Project SmartBioControl/ BioScreen; the Fonds pour la Formation à la Recherche dans l’Industrie et l’Agriculture (FRILA); and the international laboratory Microbial Secondary Metabolites MOM (University of Lille/Gembloux Agro-Bio Tech University of Liege). Grant PhD co-funding of the University of Liege (Welcome Grant Philippe Jacques) and the University of Lille. The work was also supported by the INTERREG V France-Wallonie-Vlaanderen Project SmartBioControl/BioScreen; the Fonds pour la Formation à la Recherche dans l’Industrie et l’Agriculture (FRILA); and the international laboratory Microbial Secondary Metabolites MOM (University of Lille/Gembloux Agro-Bio Tech University of Liege). Grant PhD co-funding of the University of Liege (Welcome Grant Philippe Jacques) and the University of Lille.

References


A. Vassaux, et al.

Biotecnology Advances 37 (2019) 107449


Ongley, S.E., Bian, X., Neilan, B.A., Müller, R., 2013. Recent advances in the heterologous

Pan, R., Zhang, J., Shen, W.L., Tao, Z.Q., Li, S.P., Yan, X., 2011. Sequential deletion of


Plant, A., Beldman, S., 2007. Microbial production of a class Ib nonribosomal peptide synthetase


Pride, S., Rapoport, I., 2015. Heterologous expression of the avrRpt2 gene from the fungal rice


