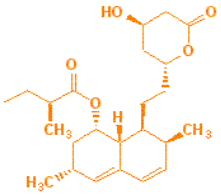
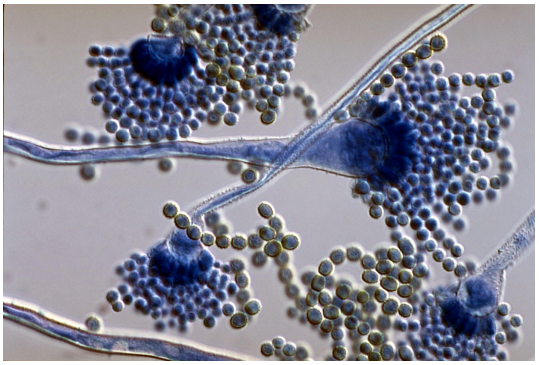
rijksuniversiteit groningen

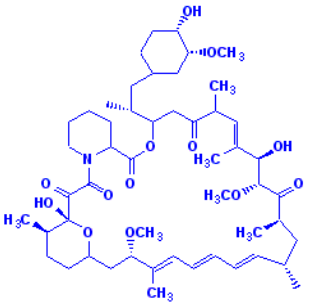
The genomic research of secondary metabolites in fungi

Silent gene clusters and the discovery of new secondary metabolites

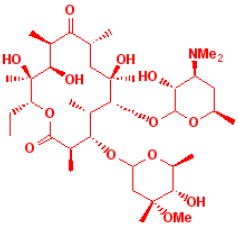


lovastatin

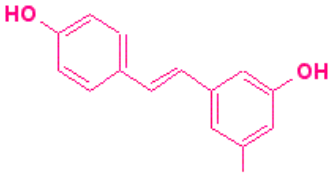




?



erythromycin A



?

Naam: Paula van Mourik
Docent: Prof. A.J.M. Driessen
Onderwerp: Bachelor scriptie
Opleiding: Schakelprogramma Moleculair Biology and Biotechnology
Datum en plaats: 8 febr 2010, Groningen

Rijksuniversiteit Groningen
Studiejaar 2009/2010

TABLE OF CONTENTS

- pictures of title page is a *Aspergillus flavus*, and the unknown structures are rapamycin (blue) resveratrol (purple) [27] [28].

Summary

Fungi produce bioactive natural products, known as secondary metabolites. These secondary metabolites are low- molecular weight compounds and often bioactive. Some of these secondary metabolites are used for pharmaceutical purposes, such as human antitumor therapy, antiviral and antifungal drug or as antibiotic.

These secondary metabolites are synthesized by polyketide synthases (PKSs) and non-ribosomal polyketide synthetases (NRPSs). The genes that are encoding for these enzymes are clustered together in the genome. Genome sequencing on fungi revealed large numbers of secondary metabolite gene clusters, especially PKS and NRPS. From allot of these predicted PKS and NRPS gene clusters the function or metabolic pathway that these enzymes are involved with, remains unknown. Also the predicated metabolites don't match always with the number of secondary metabolites that are expressed in the fungus. These genes that are involved in this prediction are called silent or cryptic genes, and are not expressed under standard laboratory conditions.

Genomic research gives us the ability to discover why these secondary metabolites are made by fungi, and what the products are that are synthesized by the PKSs and NRPSs.

This reports gives a review on the genomic research that already has been done on different PKSs and NRPSs genes that are found in fungus species. It is followed by the genomic research on silent gene clusters in fungi and the discovery of new secondary metabolites.

List of abbreviations

A	adenylation
ACVS	δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase
ASP	acyl carrier domains
AT	acyl transferase
C	condensation
ChIP	chromatin immunoprecipitation
DH	dehydratase
DMATS	dimethylallyl diphosphate tryptophan synthases
ER	enoyl reductase
FAS	fatty acid synthases
HR PKS	highly reduced polyketide synthase
KR	ketoreductase
KS	ketoacyl synthase
ME	methylesterase
MSAS	methylsalicylic acid synthase
MT	methyl transferase
NR PKS	Noninteratively polyketide synthase
NRPS	non-ribosomal polyketide synthetase
ORF	open reading frame
OTA	ochratoxin A
PCP	peptidyl carrier protein
PKS	polyketide synthase
PR PKS	partially reduced polyketide synthases
T	thiolation
TE	thioesterase
TS	terpene cyclases

1. Introduction secondary metabolites in fungi

1.1. Secondary metabolites

Primary and secondary metabolites are two types of microbial metabolism biosynthesized products. Primary metabolites are essential compounds for the growth of the cell and are the building blocks for macromolecules, intermediates for generating energy compounds (adenosine triphosphate), (co)enzymes and vitamins. Several primary metabolic pathways are identified as a source of precursors for the synthesis of secondary metabolites. Examples of these are: amino acid metabolism, fatty acid metabolism, carbohydrate metabolism and purine and pyrimidine metabolism [1].

Secondary metabolites a distinguished from primary metabolites Organisms can grow and reproduce without secondary metabolites, but still these compounds are important for the life cycle of the organism [2].

Secondary metabolites are low- molecular weight compounds and often bioactive. They are produced from related compounds at a particular part at the life cycle at a specific stage of morphological development in the organism [1]. Secondary metabolites are found in different organisms such as bacteria (*Streptomyces*), fungi (*Penicillum notatum*, *Aspergillus nidulans*) and also in plants (*Papaver somnuferum*). Fungal secondary metabolites that are already discoved are: aflatoxins, cephalosporion, compactin, ergot alkaloids, fumonisin, gibberellins, HC toxins, lovastatin, melanin, paxillin, penicillin, sterigmatocystin, sirodesmin, and triochothecenes [2].

1.2. Secondary metabolites in fungi

In 1922 Harold Raistrick started with the study of the function of secondary metabolites in fungi. Later when the antibiotic penicillin was discovered, more interest was focused on fungal metabolism. Since then many secondary metabolites were discovered that are bioactive and have the ability to the growth inhibition of bacteria, parasites, insects, fungi and also human tumour cells. Other biological effects are cytotoxicity, mutagenicity, carcinogenicity, immunosuppressive functions and enzyme inhibition [2].

The classification of the fungal secondary metabolites can be given by the physiological activity were they can be divined by the human interest of the metabolite (pharmaceuticals or growth hormones) or by the impact of the compound for plants and animals by their toxicity. Another classification, that will be used in this review, is the chemical activity of the secondary metabolites. All secondary metabolites arise from precursors from primary metabolites. The enzyme classes that are involved in the biosynthesis are used for this classification are: polyketides synthesized by polyketide synthases (PKSs), non-ribosomal peptides synthesized by non-ribosonal polyketide synthetases (NRPSs) , terpenes or alkaloids synthesized by terpene cyclases (TSs) and dimethylallyl diphosphate tryptophan synthases (DMATSs) [3]. In table 1 there is an overview from the different classes and how many genes these different fungi have for each class. Of course now these days, more genes encoding for these enzyme classes are predicted or found. The focus in this report is on the PKSs and NRPSs in fungi.

1.3. Genomic research on the PKS and NRPS encoding genes and silent genes in fungi

Fungal polyketide synthases (PKSs) are multidomain proteins that synthesize linear and cyclic polyketides. These PKS are classified in different types of groups by their structure. Non-ribosomal polyketide synthetases (NRPSs) synthesize the non-ribosomal peptides in fungi. NRPSs can synthesize linear and cyclic non-ribosomal peptides that can differ in length [2].

Genome sequencing on the fungus genome uncovered a large number of secondary metabolite gene clusters, especially PKSs and NRPSs [13]. The secondary metabolic genes are clustered together in the fungal genomes, whereas the pathway-specific regulatory genes are located next to the gene clusters [2]. However, from allot of these predicted secondary gene clusters it is not know for what kind of PKS or NRPS they encoding for. Also the function and the product that these enzymes synthesize in a specific metabolic pathway in the fungus are in allot of cases still not known. Another thing that has been discovered is that the predicated metabolites don't match with the number of secondary metabolites that are expressed in the fungus. These genes remain silent and are not expressed in the fungus. These genes are called silent or cryptic genes, and remain silent under standard fermentation conditions [2] [24]. These silent genes are very interesting, because these genes can be involved in new secondary metabolite pathway products. Maybe some of these, yet unknown, secondary metabolites can be used for pharmaceutical goals, for example as a antifungal drug, antiviral drug, antitumor drug or as a antibiotic [21]. Furthermore this can helps us understand more about the function of these secondary metabolites in fungi.

Genomic research can helps us to find new secondary metabolite gene clusters. A technique that is very successful in finding new gene clusters and predicting the role or function of these enzymes is genome mining [25] [26].

The focus in this report is on the genomic research of PKSs and NRPSs and especially the silent gene clusters in fungi.

Table 1. Secondary metabolism genes in fungal genomes [17].

Fungus	PKS	NRPS	PKS-NRPS hybrid	Sesqui-/di-terpene cyclases	DMATS	Total
Magnaporthe grisea	22	8	10	1/2	3	45
Fusarium graminearum	13	12	2	3/0	0	30
F. oxysporum	11	8	3	0/0	2	24
F. verticillioides	12	10	4	1/0	1	28
Haematonectria haematococca	12	8	1	1/0	1	28
Trichoderma reesei	11	8	2	1/1	0	23
Chaetomium globosum	20	11	6	0/1	1	39

Neurospora crassa	6	3	1	0/1	1	12
Podospora anserina	17	9	3	0/1	1	31
Botrytis cinerea	17	8	5	1/4	1	36
Scleretinia sclerotiorum	16	5	2	0/0	1	24
Stagonospora nodorum	22	10	2	1/1	2	38
Mocosphearella graminicola	11	6	2	0/1	0	20
Cochliobolus heterosporus	29	9	2	0/0	0	34
Aspergillus flavus	7	15	4	1/1	5	33
A. oryzae	30	15	3	1/1	4	54
A. terreus	28	18	3	4/1	4	58
A. niger	15	12	5	2/2	0	32
A. fumigatus	12	14	2	1/1	3	33
A. clavatus	13	12	7	0/0	2	34
A. nidulans	26	10	2	2/3	2	45

2. Polyketide syntases (PKS) in fungi

2.1. The function of polyketide synthetases (PKSs) and their products synthesis in fungi

In the fungal secondary metabolism polyketides are the most common made secondary metabolites. Fungal polyketides are synthesized by type I polyketide synthases (PKSs), which have several domains with different catalytic activities [6]. The polyketide synthase genes are clustered together on the chromosomal locus. The clustering of these genes can have advantage in selection and evolution of the fungus. [21].

The polyketides that are synthesized by the PKSs are diverse; this results from how many iteration reactions occur, the number of reduction reactions, and whether the polyketide is synthesized as a aromatic, cyclic or linear peptide [2]. From different fungi species there are already PKSs analyzed by their structure and function in secondary metabolism biosynthesis (see table 2) [11].

Table 2: Proven relationships between PKS genes and compounds in fungi [11].

organism	gene	protein	Final product
Aspergillus parasiticus	*pksA*	NSAS	Aflatoxin B1
Aspergillus nidulans	*pksST*	NSAS	Sterigmatocystin
Dothistroma septosporum	*pksA*	NSAS	Dothistromin
Aspergillus nidulans	*wA*	WAS	YWA1
Aspergillus fumigatus	*alb1*	alb1p	YWA1
Colletotrichum lagenarium	*PKS1*	THNS	Tetrahydroxy naphthalene
Wangiella dermatitidis	*WdPKS1*	THNS	Tetrahydroxy naphthalene
Gibberella zeae	*PKS13*	ZS–B	Zearalenone
Monascus purpureus	*pksCT*	CitS	Citrinin
Penicillium patulum	*MSAS*	MSAS	6-MSA
Aspergillus terreus	*atX*	MSAS	6-MSA
Glarea lozoyensis	*pks2*	MSAS	6-MSA
Aspergillus terreus	*lovB*	LNKS	Lovastatin
Aspergillus terreus	*lovF*	LDKS	Lovastatin
Penicillium citrinum	*mlcA*	CNKS	Compactin
Penicillium citrinum	*mlcB*	CDKS	Compactin
Giberella zeae	*PKS4*	ZS–A	Zearalenone
Phoma sp.	*PhPKS1*	SQTKS	Squalestatin tetraketide
Cochliobolus heterostrophus	*pks1*	TTS1	T-toxin

Cochliobolus heterostrophus	*pks2*	TTS2	T-toxin
Giberella fujikuroi	*fum1*	FUMS	Fumonisin B1
Fusarium moniliforme	*ORF3*	FUSS	Fusarin
Fusarium heterosporum	*eqiS*	EQS	Equisetin
Beauveria bassiana	*ORF4*	TENS	Tenellin

However, from allot of these PKS it is now known what their specific role is that these carries out within the fungus and the product they synthesize. Some PKSs can be involved in the synthesis of fungal pigments (heptaketide naphthopyrone) or production of peptides with antiviral or antifungal function [7] [17]. Because these peptides are bioactive they can be used as pharmaceutical drugs for humans. A famous fungal metabolite is penicillin G (Fig 1), which is used as a antibiotic against several bacteria infections. Cyclosporin and ergotamine are peptides used as a immunosuppressant in humans, and enediyne is a anti-tumor agent (Fig 1). A aromatic peptide that is synthesized by a polyketide synthase in *Aspergillus terreus* is lovatatin. This peptide is used as a antihypercholestemic drug [21] [17] [7].

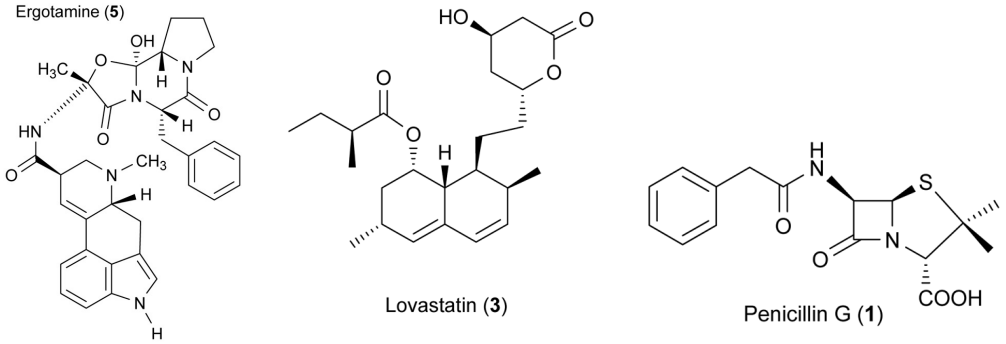


Figure 1: Peptides that are synthesized by polyketide synthases (PKS) in fungi.
Number 5 is ergotamine, number 3 is lovastatin and number 1 is penicillin G [7].

2.2. The PKS genes and proteins in fungi

2.2.1. The domain structures of PKSs

The polyketide biosynthesis has similarities with the fatty acid biosynthesis. The polyketide synthases (PKSs) and fatty acid synthases (FASs) are multi-domain enzymes with a single covalent linked polypeptide chain.
Short carboxylic acids chains (Acetyl CoA and malonyl CoA) are condensed to form a carbon chain of varityping lengths. The polyketides and fatty acids are difference in the full length of the beta-carbon reduction. Both enzymes contain the similar domain structures: ketoacyl synthase (KS), acyl transferase (AT), ketoreductase (KR), dehydratase (DH), enoyl reductase (ER), and acyl carrier domains (ASP). For the fungal

polyketide synthases the KS, AT and ACP are essential for polyketide synthesis, whereas for the fatty acid synthases the KR DH and ER are present, but some are all present or absent in the PKSs: methyl transferase (MT), methylesterase (ME), and thioesterase (TE) [6] [7]. Short carboxylic acids chains (Acetyl CoA and malonyl CoA) are condensed to form a carbon chain of varityping lengths [7].

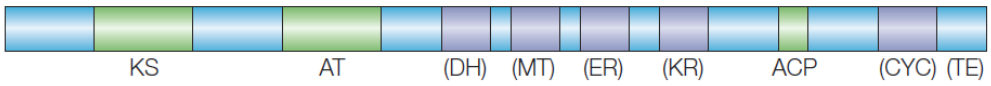


Figure 2: Fungal polyketide synthase (PKS) domain structure. KS-AT-ACP is the minimum structure for the protein. The acetyl an malonyl coenzyme A (CoA) are attached to a 4'-phosphopantotheine of the ACP domain, in this case it is the AT domain for fungi. Another thioester intermidiate is bound to the ketoacyl CoA synthase (KS) domain. Decarboxylation occurs with the ACP-bound intermediate. This results in a beta-ketothioesterthat isthen reduced by the ketoreductase (KR) domain, and dehydration by the dehydrate (DH) domain follows. In some fungi an ER domain is also in the enzyme present, this domain can form a unsaturated intermediate. The methyl transferase (MT) domain, if present in the PKS, can methylate the alpha-carbon of the thioester. CYC, cyclase; TE, thioesterase [2].

2.2.2 The PKS subclasses

2.2.2.1. The type I, II and III subclasses of PKS

The PKS are organized in different types according to their structure:

- Type I polyketide synthases are large, highly modular proteins
- Type II poltketide synthases are aggregates of mono-functional proteins
- Type III polyketide synthases which have no ACP domains [8].

Type II PKSs are found exclusively in bacteria and are multi-enzyme complexes where each individual enzyme is used iteratively for chain elongation for each cycle of synthesis aromatic polyketides. These PKSs are only found in bacteria and synthesizing aromatic polyketides [8].

Type III PKSs are homodimeric enzymes and are know as calcone synthases. These PKSs are also used iteratively [8]. Type I PKS genes have single, large open reading frames (ORFs) and contain often multiple introns. The proteins have a size between 180 to 250 kDa with domains with individual functions for the polyketide synthesis [10]. These enzymes have several modulas; a module is a complete set of domains required for one round of chain elongation and catalyzation of the associated reduction steps in the polyketide biosynthesis [8].

Fungal type I PKSs have only one module, whereas bacteria have several modules to carry out chain extension and modification. The fungal type I PKSs are mono-modular

enzymes; many are iterative type I PKSs and can catalyze the biosynthetic reactions of condensation of subunits by a adding two-carbon molecule (CoA ester) into the polyketide backbones repeatedly [2].

The noninteratively (NR) PKSs are encoded by gene clusters (genes that are located near each other in the chromosome), alongside a PKS gene that encodes for a iterative PKS. The fungal (NR) PKSs have only one condensation cycle and produces diketiden [2].

2.2.2.2. The subclades of the reducing and non-reducing polyketetide synthases

The iteratively PKSs are subdivided into two classes: reducing and non-reducing. The non-reducing (NR) PKSs, also called WA-type PKSs, repeating condensation cycles to synthesize usually aromatic polyketides which are precursors to toxins (melanin, aflatoxin), but don't reduce the chemical structure. Furthermore these enzymes are involved in the synthesis of fungal pigments [8].The C-terminal region of the protein has a thioesterase domain which can do a Claisen catalyzation. It not know how the condensation cycles are stopped or controlled for these type I PKS enzymes [6] [8].
These PKSs are subdivided into four subclades by their typical domain structure:

- Non-reducing PKS subclade I [KS-AT-PP-(PP)-CYC] haven't got ME-domain. (e.g.sterigmatocystin, aflatoxin, bikaverin).
- Non-reducing PKS subclade II [KS-AT-PP-(PP)-CYC] lost of ME-domain. (e.g.melanins).
- Non-reducing PKS subclade III [KS-AT-PP-(PP)-(CYC)] contains a ME-domain located after the PP-domain.
  (e.g. no characterized genes)
- Non-reducing PKS subclade IV [KS-AT-PP-(PP)-ME-(CYC)] haven't got always the ME domain [6].
  (e.g. no characterized genes)

The reducing PKSs synthesize various chemical reductions in structures and contain the domains KR, DH and/or ER. These domains catalyze the b-keto reduction, enoyl reduction and dehydration, which gives various of polyketide structures.
The partially reduced (PR) PKSs, also classified as methylsalicylic acid synthase (MSAS) -type, contain the domains KR and DH and are involved in the MSA synthesis [8].
The other group of reducing PKSs are involved in the highly reduced (HR) PKs and contain all three domains.
These PKSs are subdivided into four subclades by their typical domain structure:

- reducing PKS subclade I [KS-AT-DH-(ME)-ER-KR-PP] haven't got always the ME domain.
  (e.g. lovastatin, citrinin diketide, T-toxin (diketide)?)
- reducing PKS subclade II [KS-AT-DH-(ME)-KR-PP-(CON)-(AMP-PP)]lost the ER domain. The polyketides lack reduced alkyl groups or the alkyl groups are reduced by an external ER domain-containing gene.
  (e.g. lovastatin, citrinin nonaketide)

- reducing PKS subclade III [KS-AT-DH-ER-KR-PP-(PP)] lacked a ME domain. (e.g. no characterized genes)
- reducing PKS subclade IV [KS-AT-DH-(ME)- ER-KR-PP] haven't got always the ME domain [6].
  (e.g. no characterized genes)

2.3. The gene activation of the polyketide synthases in fungi

Transcription factors that act specifically on genes within the cluster are often found in secondary metabolic gene clusters in fungi and are positively regulated for gene expression. These are often zinc binuclear cluster proteins, for example these are found in the mycotoxin biosynthesis and particularly in the aflatoxin and sterigmatocystin biosynthesis cluster found in Aspergillus spp. The gene *aflR* (aflatoxin regulator) encodes for the binuclear zinc cluster (Zn(II)2Cys6) that activates the aflatoxin cluster genes in *A. parasiticus* and *A. flavus* and sterigmatocystin cluster genes for *A. nidulans* [3]. AflR is a DNA-binding protein and binds to the palindromic sequence, 5'-TCG(N5)GCA, of the promoters of the aflatoxin and sterigmatocystin genes, but also a second binding site has been found in these species which are important for autoregulation of the transcription of *aflR* [5]. The aflatoxin pathway regulator AflR induces gene transcription inside but also outside of the aflatoxin biosynthetic cluster [3].

The other transcription factors that also play a role in the biosynthetic gene clusters are Cys2His2 zinc finger proteins (MlcR for compactin biosynthesis) and an ankyrin repeat protein (ToxE for HC-toxin production). A two-peptide forkhead complex (CPCR1 for cephalosporin production), and a HAP-like transcriptional complex (penicillin production) are two examples of cluster regulators that are not found the cluster itself [5]. The biosynthesis of secondary metabolites responses to the environmental signals and global transcription factors encoded by unlinked genes than from the secondary gene clusters in the fungal species. This includes the carbon and nitrogen source, pH, light and temperature. The responses of the environmental signals are mediated trough Cys2His2 zinc finger global transcription factors for transport of nitrogen (AreA), carbon (CreA) an pH (PacC) signalling. The gene clusters aflatoxin, sterigmatocystin and gibberellin are negatively or positively regulated by the zinc finger global transcription factor proteins [2].

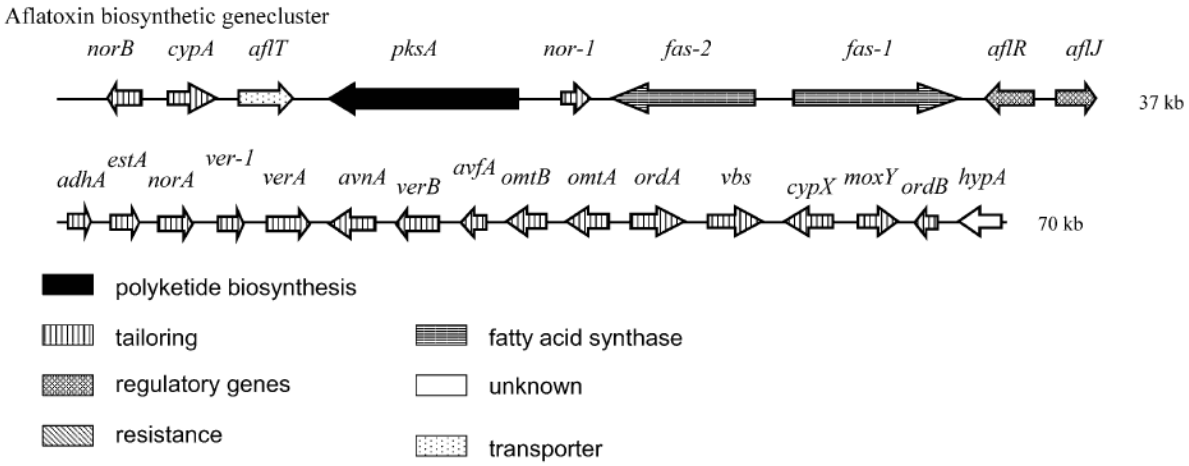


Figure 3: **The organisation of selected fungal PKS gene cluster (aflatoxin)** [20].

Another important regulator in secondary metabolism production in the Aspergillus spp is LaeA. When the *leaA* gene is disrupted it was show that lower levels of secondary metabolites were produced in the strains. In *Aspergillus Flavus* genomes of the wild-type and mutant it has been shown that LaeA controls 10% of the gene transcription. LaeA can play a role in the support of the chromatin immunoprecipitation (ChIP), because it is possible that this protein acts via chromatin remodelling [3].

2.4. Genomics research on polyketide synthases in fungi

Fungi genomes are sequenced for medical purpose and to understanding more about the micotoxin products. These toxins can have an influence on the food safety and thus also on human health. So the secondary metabolites, which are synthesized by polyketide synthases, can also be used for pharmaceutical and medical aspects. For instance these secondary metabolites from fungi can be used as a antifungal drug (or also antiviral), and as antitumor compounds [21].

Filamentous fungi are rich in genetic material that is related to the secondary metabolism production. Especially the Aspergillus spp is rich in polyketide synthase genes, for example *A. nidulans* harbors 26 genes for polyketides synthases, and *A. terreus* has even 28 genes for PKSs (see table 1). But it is possible that even more PKS genes are discovered these days in these fungi [17].

We still understand not that much about the mechanism of the starter unit selection and the chain length control mechanism that PKS carries out. Also the designation of methylation of the polyketide and the reduction are not yet completely understood [21].

One of the first completely genome sequences were published for *Aspergillus nidulans*, *A. fumigatus* and *A. oryzae*. The number of genetic loci that are related to the secondary metabolism production and revealed by genomics, often don't predict the compounds that are related to these genes [21]. A study that recently was done for

characterization of an PKS gene was on the ochratoxin A (OTA) production in *Aspergillus carbonarius*. OTA is a nephrotoxinand can be a possible human carcinogen. PKSs are responsible for OTA biosynthesis in Aspergillus and Penicillin spp. By doing the genetic characterization of the OTA biosynthetic genes it is possible to do assays on the expression profile and also to determine the triggers for the biosynthesis control [22].

Trough genomics it is also possible to find out more why these secondary metabolites are produced and what the function and ecological relevance is for the fungus [21].

For doing genomic research it is still necessarily to isolate single genes for genome sequencing. In a recent publication an alternative approach was presented that is less time-consuming. By addressing and isolating the transcript of the gene it is possible to do also genomic research. The technology that was used is called the SMART-rapid amplification of cDNA ends (RACE), where the full-length coding region of the PKS is cloned and characterisized by using RNA extract from pure mycobionts cultures. In this study it was successfully done from PKS transcript gene from the *Xanthoria elegans* cultures [23].

BLAST search is often used to find related proteins that are already available in the NCBI database. Also in this study it was used to find out what the resulted amino acid sequence encodes for. In figure 5 an alignment is shown of the predicted PKS gene in *Xanthoria elegans* and compared with five other fungi spp. For example it was found that the N-terminus has a conserved AA sequence (LFGDQ) that is typically for the fungal type I PKS enzymes. So by comparing PKS gene sequences it becomes easier to find out what kind of domains the unknown PKS gene transcribes [23].

As mentioned before, secondary metabolites can also play a role as a drug; for instance antibiotics. Genome analyses can help to find genes that can synthesize potential drug targets by pathogen genome screening to analyze homologues in other fungi and that are absent in the human and mammalian genome. Also researches can now do transcriptome-wide studies to see what kind of response the fungi have after drug treatment. An example of this kind of study is done on the genome of *A. fumigatus*, which was compared with other analysis of fungal allergens or the interactions that they had with IgE. *A. fumigatus* was screened for putative allergens tend it was predicted that it might encode 58 allergens [21]

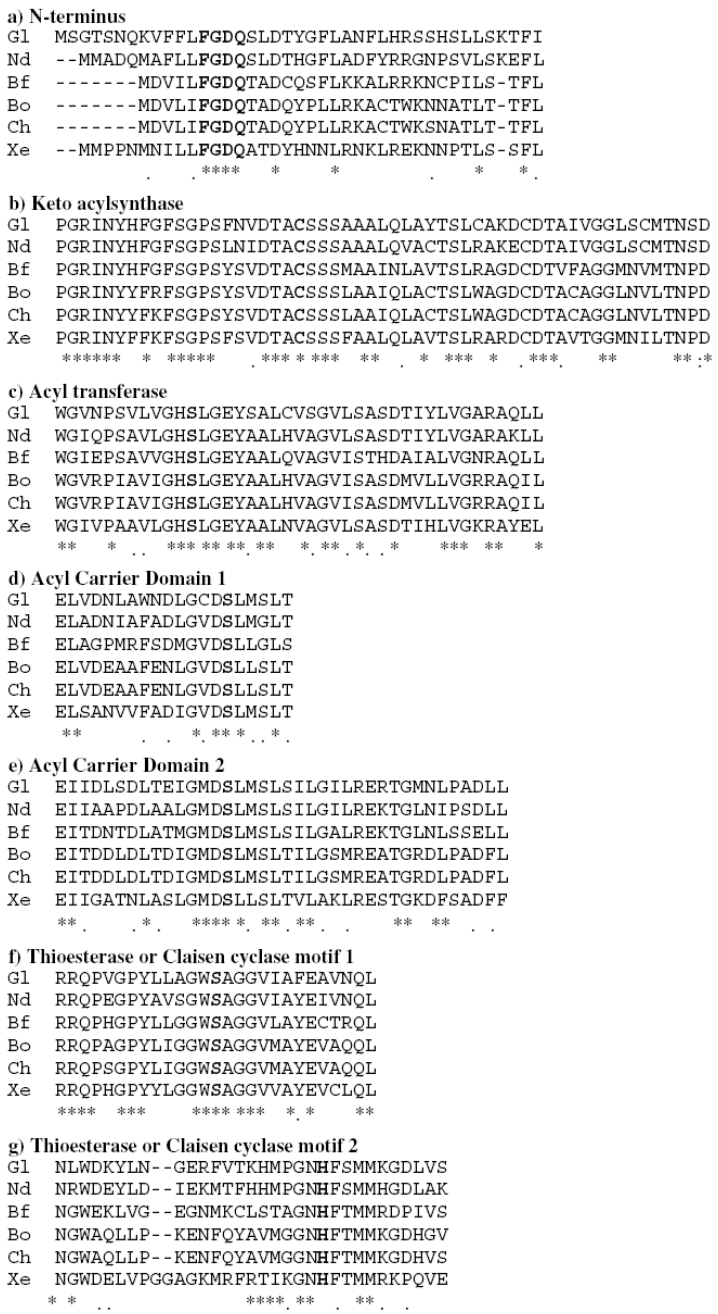


Figure 4: **A transcribed polyketide synthase gene from *Xanthoria elegans*.**
The alignment of predicted PKS CDS in *Xantoria elegans*, and five close related fungal PKSs. Active sites are bold. Alignment of the putative catalytic motifs in the sequence of *Xanthoria elegans* PKS (Xe, ABG91136*); Bipolaris oryzae* (Bo, BAD22832); *Botryotinia fuckeliana* (Bf, AAR90249); *Cochliobolus heterosphorus* (Ch, AAR90272); *Glazea lozoyensis* (Gl, AAN59953); *Nodulisporium sp*. (Nd, AAD38786).Amino acid residues conserved the sequence are marked with an asterisk, variability of two amino acids are marked with a dot [23].

3. Non-ribosomal polyketide synthetases (NRPS) in fungi

3.1. The function of non-ribosomal polyketide synthetases (NRPSs) and their products

Non-ribosomal peptides synthetases (NRPSs) are responsible for the biosynthesis of non-ribosomal peptides in fungi and bacteria. The non-ribosmal peptides difference in length, whether the peptide is cyclized and the domain variety. NRPSs can synthesize linear and cyclic non-ribosomal peptides [2]. For example the fungus *trichoderma virens* has a NRPS, called peptaibol synthetase, that produces a 18-residue lineair peptide [18]. In *Tolypocladium niveum* a cyclosporine NRPS complex was identified that can synthesize the cyclic peptide cyclosporin, which is used as a immunosuppresive drug [2].

The specific function and the product that is synthesized by the NRPS is not known yet for all the NRPSs that are screened in different fungi. But for a few it is known what the function (and product synthesis) is of the NRPS in some fungi, here are a few examples of NRPSs which the functions are already discovered. For example the *Aspergillus nidulans* genome carries 27 genes that are encoding for NRPSs, but only five of these NRPS structures have been identified now. One of these NRPS is Pes1, which is relevant for the resistance to oxidative stress in *A. nidulans* [13]. The NRPS that is encoded by the gene NRP6 is involved in the virulence of maize (*Zea mays*) pathogen *Cochliobolus heterostrophus* and resistance to oxidative stress [13].*Aspergillus fumigatus* has a NRPS, called GliP, that is involved in the biosynthesis of gliotoxin, which is a toxic metabolite [14]. Also a NRPS product is found in *A. fumigatus* that can modify gliotoxin and can specifically block the respiratory burst in humans by inhibition of assembled NADPH oxidase in isolated polymorphonuclear leukocytes(12). NRPSs are involved in the production of peptide toxins that are important for resistance of the fungi and as a pigment product. These peptides can be used for pharmacological goals, such as antibiotics, cytostatics and immunosuppressors. [12] [13] [14].

3.2. The NRPS genes and proteins in fungi

The NRPS genes are located in gene clusters over the fungal chromosome. The NRPS genes in fungi contain ORFs [12].

Genomic libraries of genomics of fungi are these days screened for research to find NRPS encoding gene clusters. In the fungus *Alternaria brassicae* a nonribosomal peptide synthetase (NRPS) gene named *AbrePsy1* was screened. The sequence is 22 kbp long and encodes for a 792 kDa protein (with 7,191 amino acid residues) and has typical modular NRPS. The gene might be responsible of the pathogenicity of the fungus. [15]. The NRPSs have molecular mass up to 2.3 MDa and are multimodular enzymes. Each NRPS module contains several domains: adenylation (A), thiolation (T) or also called thioesterase (TE) and peptidyl carrier protein (PCP) and condensation (C) domains [12]. These domains recognize, activates and binds a module-specific amino acid as a thioester to 4'-phosphopantetheine cofactor, which is fasten to a serine amino acid. Then the amino acids are linked together by a peptide bond and then released by the thioesterase-like domain, which is located at the C-terminal of the last module [2].

The function of each domain is explained further:

The adenylation (A) domain (~550 aa), which is located within the NPRSs, does the selection of the amino acids to control the primary sequence. The amino or carboxy acid substrate is activated as an amino acyl adenylate and ATP is used for caring out the reaction. (16) Also this domain controllers the specificity of amino acid connection to the peptidyl carrier protein (PCP) [12].

The peptidyl carrier protein (PCP) domain (~80–100 aa) is the transporter that accepts the amino acid from the A domain that is covalently tethered to its 4_PP cofactor as thioester. The amino acid is then attached to a serine of the carrier protein (CP) and allows the protein to travel between the different catalytic centres [16].

The condensation (C) domains (~450 aa) are responsible for catalysation of the formation of the peptide bonds [12]. The amino group is catalyzed by the nucleophilic attack into the acyl group of the amino tethered module [16].

The thiolation (T) (~250 aa) catalyzes product cleavage by hydrolysis or complex macrocyclization [2] [12] [16].

NRPS genes differ in the length and arrangement of the modulas that are transcripted to form the enzyme. δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine synthetase (ACVS) is a trimodular NRPS from *Penicillium chrysogenum* and is encoded by the 11 kb *pcbAB* gene. The enzyme catalyzes the first step in the bio-synthesis of β-lactam antibiotics penisillion and cephalosporin in *P. chrysogenum*. The gene cluster consist of three repeated modules that are involved in the activation of condense the amino acids L-α-aminoadipic acid, L-cysteine and L-valine. Also L-valine is epimerized to D-valine at the end of the TE domain. The linear AVC peptide is formed and is cyclized to isopenicillin by the enzyme isopenicillin N synthase (Fig 5) [2].

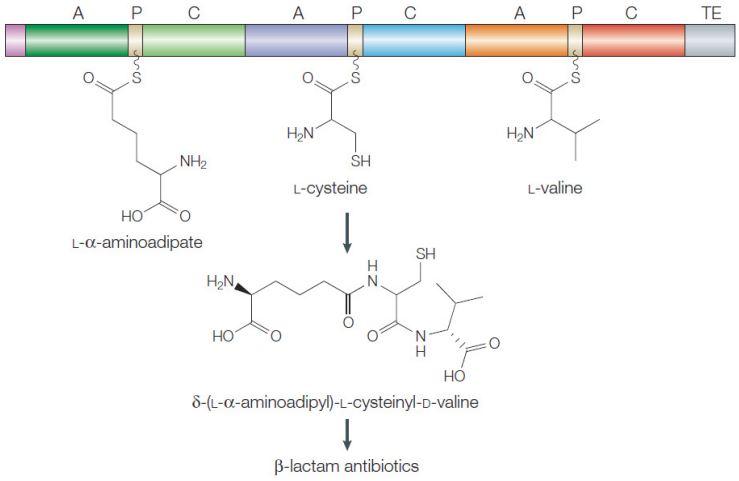


Figure 5: **AVC synthetase, a trimodular non-ribosomal peptide synthetase.**
AVSC is involved in the biosynthesis of penicillin and cephalosporin. The adenylation (A) domain recognize and activates the amino acid and is attached to the peptidyl carrier domain (P). Then the condensation (C) domain carries out the forming of the peptide bonds. The tri-peptide that is formed at the C-terminal end of the P domain is then released by the thioesterase (TE) domain. (The L-valine is isomerized to D-valine [2].

3.3. The activation of the NRPSs in fungi

Bacteria and fungi NRPSs are required to post-translational 49-phosphopantetheinylationto ease secondary metabolism production. The NRPSs are activated by 49-phosphopantetheinylation of serine residues by the 49-phosphopantetheinyl transferases (49-PPTases) which are present in the PCP domain op the enzyme. 49-phosphopantetheine is transferred from coenzyme A to a conserved serine residue within the thiolation (T) domains of cognate apo-NRP synthetases that can a activate holo-NRP synthetases (Fig 6). Non-ribosomal peptide synthetases, which catalyse production of cyclic peptides including numerous toxins [12].

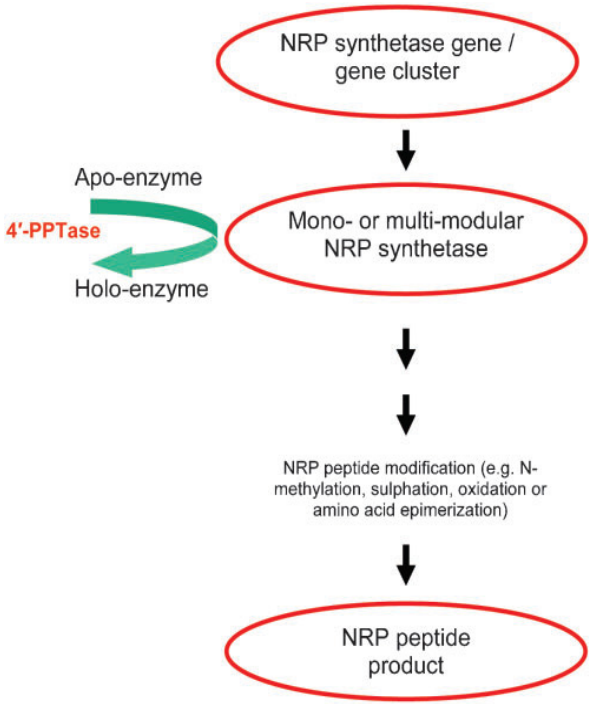


Fig. 6. Schematic representation of the enzyme functions involved in NRPS. NRP synthetases are generally found within transcriptionally co-regulated gene clusters whereby all genes required for NRP peptide product synthesis are co-expressed. Apo-NRP synthetases are converted to the active holo-form by 49-phosphopantetheinylation of thiolation (T or PCP) domains within the NRP synthetase. NRP peptide synthesis is further facilitated by the action of additional enzyme activities – either components of, (e.g. amino acid epimerization) or distinct from (e.g. oxidation), the NRP synthetase [12].

3.4. Genomic research on the NRPSs in fungi

Nonribosomal peptide synthetases in fungi are very important for biomedical and biotechnological research. The discovery of these enzymes and their products have a major impact on the human health and food production in our life. But also for the fungi itself these NRP synthetases are important to make toxins for resistance or protection against other micro-organisms, animals and plants. From different fungi the genomes are

already sequenced and revealed different genes and gene clusters that are predicted to encode for NRPS synthetases (see table 1). For the most of these NRPSs it is not known in what kind of secondary metabolism production they are involved with. This is still under investigation to find new secondary metabolites.

A study was done on the genome of *Aspergillus nidulans* that has 27 genes that are encoding for NRPS structures. Until now five NRPSs have been identified. The focus was on the domain structures, substrate binding pockets, related gene (clusters) and the prediction what these NRPS systems producing of secondary metabolites.

In Aspergilli there were conserved gene clusters analyzed and it was shown that these type of NRPS clusters are connected with an ABC transporter. By this data it was predicted that cyclic peptides that are produced by NRPSs are secreted [13].

In *A. nidulans* a conserved putative fusarinine type siderosphore biosynthesis gene cluster restrained of the NRPS (**AN6236**), called SidD, and also a ABC transporter (**AN6237**) and putative siderophore transporter (**AN6236**) were located within the gene clsuter. This was also found in *Aspergillus fumigatus*, *Aspergillus oryzae*, *Aspergillus terreus*, *Aspergillus clavatus* and *Aspergillus fischeri* these gene clusters were found [12] [13].

An 2560 AA NRPSs, AN9226, was discovered as a putative N-methyl-cyclodipeptide synthetase in *A. nidulans*, that is flacked by a P450 enzymes and dioxygenase. In the second A-domain of the structure (**ATCAMTC**, were **A** is adenylate domain, **T** is a peptididyl carrier domain, **C** a condensation domain,) a region was find that has 45% similarities with a cyclosporin syntase N-methyltransferase. An ortholog was also found in *N. fischeri* (**NFIA043670**, 2662 AA) [13].

In *A. nidulans* a PKS-NRPS gene cluster was found; **AN2545.3** encoding for NRPS 7047AA (member of emericellamide cluster) with structure **TECATECATCAT CATCATC** (were E corresponds to an epimerisation domain) and **AN2547.3** encoding for PKS 2534 AA (acyl-CoA synthetase), and an acylcarrier protein AN2548.3. By the gene knockout technology it has been shown that both the acyl-Coa ligase and acyl carrier protein are necessarily for the emiricellmide formation. Both NRPS and PKS are involved in the emericellmide production, and are needed to make emericellmide [13].

In secondary metabolite production also monomodular NRPS enzymes are involved. These enzymes are not well studied yet. The domain structures are predicted, but still it is not known what the precise functions are of these enzymes. For example the **AN2064.3** encodes for a protein that has an acyl-CoA- related adenylate domain and a NAD-domain. Related proteins are found in all Aspergilli with identity of 52-53% at protein level [13].

In *Aspergillus fumigatus* three NRPSs encoding genes have been identified: *sidD*, *sidC* and *sidE*. These genes are involved in the siderophore biosythesis. These siderophore synthases can play a role in the antimicrobial chemotherapy. In a study it was shown by expression analysis that SidD and SidC were reduced up to 90% under iron-replete conditions. But in the case of SidE it was not that sensitive for different iron levels. The genome of *A. fumigatus* was analyzed were several ORFs encoding NRPS were identified by BLAST analysis. Molecular characterization and phylogenetic analysis was done on the NRPS-encoding genes *sidD*, *sidC* and *sidE*. Siderophore production was measured as well the expression of SidD, SidC and SidE. This study shows that the environment where the fungi grows has an effect on the secondary metabolite production.

In this case an high concentration of iron in the medium causes that the expression of SidD and SidE are lower. The optimal siderophore production occurred in the absence of iron in *A. fumigatus* [19].

By genomics new NRPSs are identified and helps to find out more about the structures and functions of these enzymes and their synthesized products. By analysis gene clusters, as done by Aspergillus spp, and compare them with each other it is easer to predict the function or structure of the NRPS.

4. The silent gene clusters in fungi

4.1. The silent gene clusters in fungi

By genomic studies it is possible to find genes and gene clusters that are encoding for a PKS or NRPS. For some of these it is known what their function is, and their synthesized product. It is found out that more putative secondary metabolite gene clusters are present in the fungal genome than the account of needed products. And some of these gene clusters are not expressed under standard laboratory conditions [2]. These gene loci remain silent because the particular trigger is absent. These silent, or cryptic genes can be involved in the secondary metabolism pathway of metabolites that can code for important virulence factors, antibiotics or toxins. It is important to find out how these silent genes can be activated, so new bioactive compounds can be used for pharmaceutical, as immunosuppressant's and for the toxic activity that they have on other organisms [2] [24].

4.2. Genomics research on silent gene clusters in fungi

By genomics now databases are available that have DNA sequences of a wide variety of organisms. It makes it possible to discover new secondary metabolites in fungi, and helps us to understand the genetics and enzymology of the gene clusters and the transcribed enzymes [24].

Secondary metabolites are produced under certain conditions. By genomics new secondary metabolites are discovered that are the products of secondary metabolite gene clusters in fungi, which are synthesized by PKSs and NRPSs [26]. PKSs and NRPSs consist of a number modules that are corresponding to a number metabolite compounds that make the final product, but there are some exceptions. In some cases the building blocks are modified at the end of domains with a tailoring function the presence or absence of this domain can helps us predict the natural product that is synthesized but the PKS or NRPS. Also by sequence analyses of the domains in each module of PKS and NRPS, substrate specificity can help to predict the product that is made [24] [26].

Genome mining is an approach that is used to find new secondary metabolites in fungi and bacteria. Genes encoding for enzymes that are involved in the secondary metabolite biosynthesis are searched in the sequenced genome of the fungi by using computer programs and comparison tools [26].

The are different strategies used to find new secondary metabolites in fungi and bacteria. One method is to do bioinformatics analyses that can be done when the structure of the metabolite product can be deduced of the cryptic gene cluster (Fig 7). This helps us to predict the physiochemical properties of the metabolite, and can simplify the identification of the metabolic product of the cryptic biosynthetic gene cluster [26].

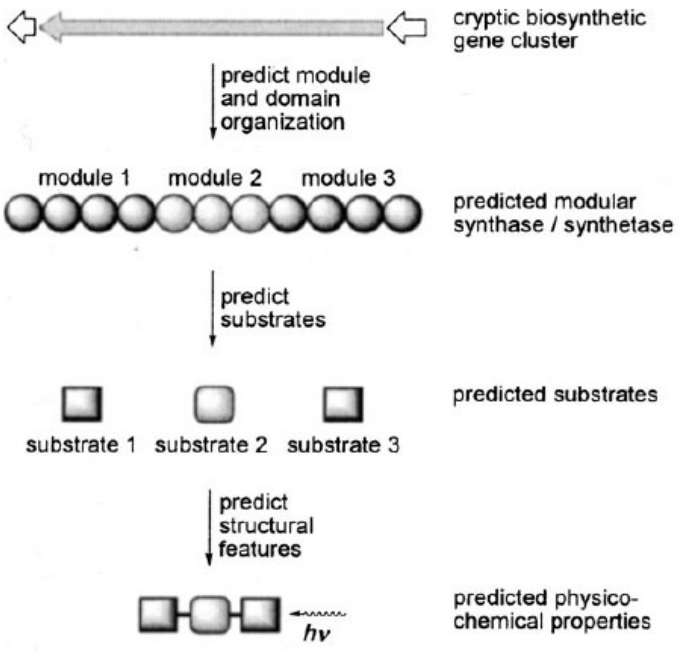


Figure 7: A method to predict the physicochemical properties of a putative secondary metabolite or a cryptic gene cluster [26].

Under differed growth conditions silent cryptic gene clusters can be activated. An alternative is to manipulate the regulatory genes to activate the silent gene clusters. An example of this approach is the investigation on LaeA, which is a protein that has a global regulatory function in *Aspergillus genus*. When LaeA is deleted several gene clusters are not expressed. Overexpression of LaeA shows that by some gene clusters the transcription is increased in the strain. The wild-type metabolite profile and that of the mutants can be compared in the case of LaeA, which is a pleiotropic regulator. LaeA can be overexpressed or deleted and the metabolic profiles can differ from each other. This difference can help identify the products that are expressed by the gene clusters. Sometimes cryptic silent gene clusters are expressed by putative pathway-specific regulators under the control of an inducible promoter [26].

A method to find the metabolic product is to use target gene activation to find new secondary metabolites [26]. This technique can only be used when the metabolite is produced by the strain. It was used in the research of the fungus *Aspergillus nidulans* that led to a discovery of novel PKS-NRPS hybrid metabolites that was induced by a silent metabolic pathway. The genome *A. nidulans* was sequenced for cryptic (or orphan) gene clusters that can encode of PKSs or NRPSs. A single putative hybrid polyketide synthase-nonribosomal synthetase (PKS-NRPS)-encoding gene, called *apdA* (**AN8412.3**) was noted. Also the PKS-NRPS gene cluster contains a putative activator gene, called *apdR* (**AN8414.3**). It was shown that under standard fermentation conditions this gene remained silent. The *apdR* gene was amplified and cloned into a expression vector (pAL4), that contains a dehydrogenase promoter *alcAp* of *A. nidulans*, to activate the

silent gene cluster by induction in the *A. nidulans* strain. The *apdA* gene that encodes for the PKS-NRPS hybrid synthase was transcribed under the inducing conditions two new cytotoxic metabolites were isolated: aspyrodones A and B, that belong to the pyrodone metabolism products (Fig 8) [24].

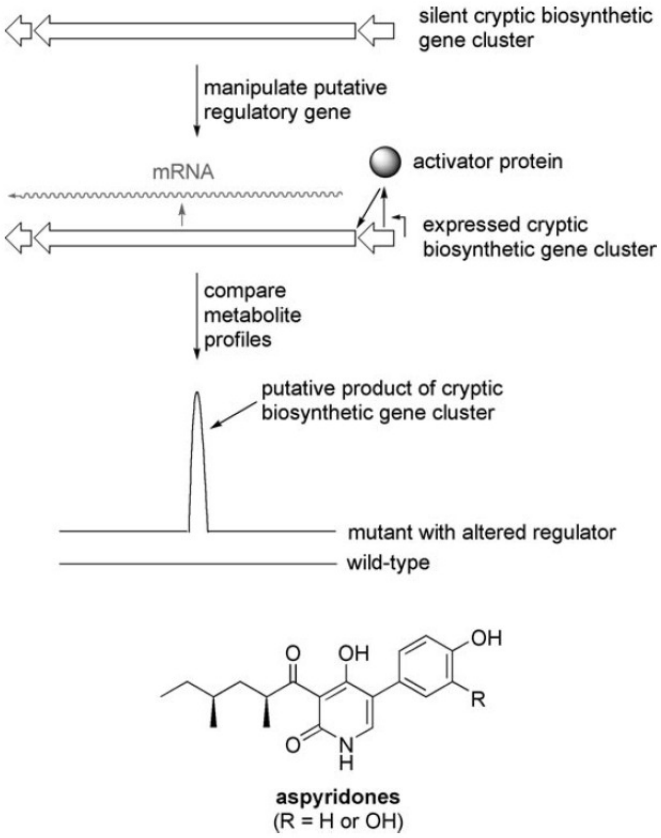


Figure 8: **A method to manipulate the regulatory genes to activate the expression of silent genes clusters**. For example in the aspyridones discovery this method was used [26].

In another study on *Aspergillus nidulans* two cryptic gene clusters (**AN1034.3** and **AN1036.3**) were investigated to find the polyketide(s) that is/are synthesized.
The NR-PKS (**AN1034.3**) and HR-PKS (**AN1036.3**) are located nearby each other in the chromosome. Under standard culture conditions no metabolite profile difference was found under the wild-type and mutant strain. Nearby both PKSs genes there was an gene found, **AN1029.3**, witch has homology to *CtnR* that encodes for a citrinin biosynthetic transcriptional activator. By expressing the transcription factor the silent gene cluster was expressed. The pathway was identified and produces a novel polyketide, asperfuranone [25].

5. Conclusion/Discussion

Secondary metabolites are low-molecular weight compounds and are often bioactive. The secondary metabolites are synthesized by polyketide synthases (PKS) or non-ribosomal polyketide synthetases (NRPS). The genome of different fungi are sequenced for analysing the gene clusters that are involved in the secondary metabolite biosynthesis. From allot of these predicted secondary metabolite gene clusters it is not know for what kind of PKS or NRPS they encoding for and what specific function the PKS or NRPS carry out has and the product that is synthesized. The predicted metabolites that might be synthesized by these enzymes don't match with the number of produced secondary metabolites. These genes that lack transcription are called silent or cryptic genes, and are not switched on because the trigger is absent.

Different genomic studies were reported and compared about the predicted PKS, NRPS and silent gene cluster function and their products that might be synthesized.

Genomics helps to find new gene clusters encoding for PKSs and NRPSs. Complete ORFs can be sequenced to predict the number, size of introns, and the product that is synthesized by the PKS or NRPS. By comparing gene clusters of different fungi, domains of PKSs can be predicted so the function and gene trigger can be revealed. Also the NRPS systems can be predicted by the domain and substrate binding pockets of the unknown NRPS and related gene comparison of other known gene clusters.

For genomic research it is still necessarily to isolate single genes for genome sequencing, which is time-consuming. An alternative approach was announced; the SMART-rapid amplification of cDNA ends (RACE). RNA extract was used to clone and characterizes PKS gene. This a successful new method the find the coding sequence of a transcribed PKS gene. However, this method is only useful when the gene is transcribed in the genome.

Genome mining is a good approach to find new PKS/NRPS encoding gene cluster; differed computer programs can be used for genome comparison of fungi spp. This is also a good method to identify cryptic gene clusters and to discover new secondary metabolites. By manipulating the regulatory genes cryptic gene clusters can be switched on. This is only a effective method when the regulatory gene are identified and the manipulation of the host is possible.

Comparison the metabolic profile of wild-type and mutant strains it can helps to identify the particular product that is synthesized by a specific silent gene cluster. Another strategy that can be used is different growth conditions to activate silent gene clusters. However, this is a time-consuming method and not always successful, because there can be more secondary metabolites expressed by one silent gene cluster or other cryptic gene clusters can be switched on with the same growth conditions.

The method that can work to express the cryptic genes is to do an in vitro reconstruction approach. By most of the cryptic gene clusters it is not known what the trigger is that activates the silent gene cluster. By removing the gene cluster and reconstruct it in a vector with an inducible promoter, it might be possible to switch on the silent gene cluster.

Although genomic research helps us to find new PKS/NRPS gene clusters and the predict pathway products, and even maybe say something about the gene expression, it is still important to do studies on molecular basis as well. For example when a predicted

secondary metabolite has toxic activities, this has to be tested on cells or plants. This can be done by using LC/MS (Liquid Chromatography/Mass Spectrometry) analysis of cultures from PKS–NRPS over-expressing transformants or deletion mutants. Another method is to over-express the cluster-specific regulator.

The focus should not only be on the unknown PKSs and NRPSs and their function, but also on the already studied PKSs and NRPSs. These proteins can be modified to synthesize new secondary metabolites. It is also important to do research on the mechanism that decides what the length is of the peptide. This has opportunities to make new bioactive secondary metabolites.

The secondary metabolites are used as antibiotics, antitumor drugs, antiviral drugs or antifungal drugs. So it is clear that these compounds are very important for humankind and helps us to survive. It is important to continue the research to find new secondary metabolites. Especially the silent gene clusters are very interesting, because these genes can be involved in, yet unknown, new secondary metabolite products.

Literature

[1] Isaac, S. Fungi naturally form many diverse biochemical products, some of which are now commercially important; how an why do they do this? **Mycologist** 1997; **11**: 182-183.

[2] Keller, N.P., Geoffrey Turner, Bennett J.W. Fungal secondary metabolism- from biochemistry to genomics. **Nature rev. microbial** 2005; **3**: 937-947.

[3] Fox, E.M., Howlett, B.J. Secondary metabolism: regulation and role in fungal biology. **Current Opinion in Microbiology** 2008; **11**: 481–487.

[4] Hicks, J.K., Shimizu, K., Keller, N.P. Genetics and biosynthesis of aflatoxins and sterigmatocystin. **The Mycota XI** 2002; **11**: 55–69.

[5] Yu, J.H., Keller, N. Regulation of secondary metabolism in filamentous fungi. **Annu. Rev. Phytopathol** 2005; **43**: 437–58.

[6] Kroken, S. et al. Phylogenomic analysis of type I polyketide synthasegenes in pathogenic and saprobic ascomycetes. **PNAS** 2003; **100**: 15671-15675.

[7] Amnuaykanjanasin, A. et al. Diversity of type I polyketide synthase genes in the wood-decay fungus *Xylaria* sp. BCC 1067. **FEMS Microbiology Letters** 2005; **251**: 125–136.

[8] Uwe, J. et al. Novel Insights into Evolution of Protistan Polyketide Synthases through Phylogenomic Analysis. **Protist** 2008; **159**: 21-30.

[9] Weissman, K.J., Leadlay, P.F. Combinatorial biosynthesis of reduced polyketides. **Nature rev. Micriobiology** 2005; **3**: 925-936.

[10] Cox, R.J., Simpson, T.J. Fungal Type I Polyketide Synthases. **Methods in Enzymology** 2009; **459**: 49-78.

[11] Cox, R.J. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. **Org. Biomol. Chem** 2007; **5**: 2010 – 2026.

[12] Stack, D., Neville, C., Doyle, S. Nonribosomal peptide synthesis in *Aspergillus fumigatus* and other fungi. Microbiology 2007; **153:** 1297-1306.

[13] Dohren von, H. A survey of nonribosomal peptide synthetase (NRPS) genes in *Aspergillus nidulans*. **Fungal Genetics and Biology** 2009; **46**: 45–52.

[14] Cramer, Jr, R.E. Disruption of a Nonribosomal Peptide Synthetase in *Aspergillus fumigatus* Eliminates Gliotoxin Production. **Eukaryotic cell** 2006; **5**: 972-980.

[15] Guillemette, T., Sellam, A., Simoneau, P. Analysis of a nonribosomal peptide synthetase gene from *Alternaria brassicae* and flanking genomic sequences. **Curr Genet** 2004; **45**: 214–224.

[16] Finking, R. Mohamed M.A. Biosynthesis of nonribosomal peptides 1. **Annu. Rev. Microbiol**. 2004; **58**: 453–88.

[17] Collemare, J., Billard, A., Bohnert, H., Lebrun, M.H. Biosynthesis of secondary metabolites in the rice blast fungus *Magnaporthe grisea*: the role of hybrid PKS-NRPS in pathogenicity. **Mycological research** 2008; **112**: 207-215.

[18] Wiest, A., Grzegorski, D., Xu, B.W. Identification of Peptaibols from *Trichoderma virens* and Cloning of a Peptaibol Synthetase. **The journal of biological chemistry** 2002; **277**: 20862–20868.

[19] Rieber, k., Reeves, E.P. The expression of selected non-ribosomal peptide synthetases in *Aspergillus fumigatus* is controlled by the availability of free iron. **FEMS Microbiology Letters** 2005; **248**: 83–91.

[20] Schumann, J., Hertweck, C. Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes. **Journal of Biotechnology** 2006; **124**: 690–703.

[21] Misiek, M., Hoffmeister, D. Fungal genetics, genomics, and secondary metabolites in pharmaceutical sciences. **Planta Med** 2007; **73**: 103-115.

[22] Gallo, A., Perrone, G. et al. Characterisation of a pks gene which is expressed during ochratoxin A production by *Aspergillus carbonarius*. **International Journal of Food Microbiology** 2009; 129: 8–15.

[23] Brunauer, G., Muggia, L., Stocker-Worgotter, A., Grube, M. A transcribed polyketide synthase gene from *Xanthoria elegans*. **Micrological research** 2009; 113: 82-92.

[24] Bergmann, S. et al. Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. **Nature** 2007; 3: 213-217.

[25] Chiang, Y.M. A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in *Aspergillus nidulans*. **J. AM. CHEM. SOC.** 2009; **131**: 2965–2970.

[26] Zerikly, M., Challis, G.L. Strategies for the discovery of new natural products by genome mining. **ChemBioChem** 2009; 10: 625 – 633.

[27] http://www-users.york.ac.uk/~drf1/body_lecture_8.htm

[28] http://www.clt.astate.edu/mhuss/Aspergillus%20flavus%20pict.jpg