Engineering the Polyketide Cell Factory

Mølgaard, Louise; Patil, Kiran Raosaheb; Thykær, Jette; Mortensen, Uffe Hasbro; Eliasson Lantz, Anna

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Engineering the Polyketide Cell Factory

Louise Mølgaard

Ph.D. Thesis

June 2012

Center for Microbial Biotechnology

Department of Systems Biology

Preface

The work presented in my Ph.D. thesis was carried out from March 2008 to June 2012 at Center for Microbial Biotechnology at DTU Systems Biology. The work was supported by a grant from The Danish Council for Independent Research, Technology and Production Sciences (09-064240).

The project was supervised by Associate Professor Uffe Hasbro Mortensen and Assistant Professor Kiran Raosaheb Patil and I am thankful for their guidance and support over the years of my Ph.D. Unfortunately, Kiran Raosaheb Patil had to leave CMB before the completion of my project, but has stayed involved from his work at EMBL, Heidelberg. I owe thanks to Assistant Professor Jakob Blæsbjerg Nielsen for guidance in molecular biology of *A. nidulans*. I have been fortunate to have the invaluable support of Associate Professor Jette Thykaer during the last part of my Ph.D. both in reviewing parts of the thesis as well as guidance in fermentation technology. Furthermore, Associate Professor Anna Eliasson Lantz has provided guidance for the paper on image analysis as a platform for microbial screening and I owe her great thanks as well.

Beyond this I have had invaluable assistance both from technical staff and fellow Ph.D. students and Post Docs. I would especially like to thank Tomas Strucko, Line Due Buron, Lars Poulsen, Hanne Jakobsen, Martin Engelhard Kogle, Martin Nielsen, Jette Mortensen, Elisabeth Krøger and Tina Johansen for their help. I have supervised a number of students

during my Ph.D. of those Stine Prehn Lauritzen, Sebastian Wingaard Thrane and Niels Bjørn Hansen have contributed to the work presented in this thesis. My office mates Ana Rita Brochado and Anna Lena Heinz have also been invaluable for discussions and help through ups and downs. In addition I owe a great thanks to all of the people working at CMB for making it a wonderful place to be.

I am eternally grateful to my friends and family for their support over the years. I would not have finished this work if it were not for the loving support from my son Alexander and husband Mikkel.

Abstract

Natural products constitute one of the largest sources of therapeutics known to mankind. Among the natural products polyketides such as erythromycin (antibiotic) and lovastatin (cholesterol lowering) have long proven their immense value to patients around the world. Polyketides are naturally produced by plants, fungi and bacteria. However, the natural producers often do not achieve commercial titers of the polyketide therapeutic. Thus the natural production must be improved. This can be done by random mutagenesis or heterologous expression of the polyketide gene cluster resulting in production sufficient titers. To improve the production of polyketides biological engineering principles have been applied for the development and engineering of microbial polyketide cell factories.

The two biological hosts used for heterologous polyketide production were *Aspergillus nidulans* and *Saccharomyces cerevisiae*. Both organisms have well-known genetic tools available for gene targeting and heterologous expression. It has been the aim to create a stable expression platform with all genes integrated in the genome. This has been achieved through the use of two advanced genetic engineering systems for *A. nidulans* and *S. cerevisiae.* Both systems have been aided by USER™ cloning vectors that were developed for efficiently generating large amounts of gene targeting substrate. Upon integration the targets should lead to high expression of the polyketide synthase (PKS) as well as the activating phosphopantetheinylase (PPTase). This versatile vector system can easily be used for expression of other polyketides of interest as well as extended to express whole gene clusters.

After achieving proof of principle in terms of expression, the polyketide cell factory must be optimized. The optimization can be achieved through the use of adaptive evolution, random mutagenesis and screening as well as metabolic engineering.

Firstly, *in silico* guided metabolic engineering was used as a tool to direct metabolism towards higher levels of 6-MSA production in *A. nidulans*. 6-MSA was stably expressed in the *A. nidulans* genome and bioreactor cultivations resulted in high titers of 6-MSA. The genome scale model of *A. nidulans* and the optimization algorithm OptGene was used to predict a knockout strategy designed to increase the production of 6-MSA in *A. nidulans*. Among the predicted targets deletion of the NADPH dependent glutamate dehydrogenase (*gdhA*) was selected as it should result in greater availability of NADPH for polyketide production. The deletion resulted in decreased growth rate of *A. nidulans*, which was partially rescued by the insertion of an extra copy of the NADH dependent glutamate dehydrogenase (*gdhB*). Physiological characterization in bioreactors revealed that the yields of 6-MSA on biomass increased albeit not significantly. As a result of this it may be argued that there is still more work to be done in terms of model building in *A. nidulans*.

Utilizing another well-established cell factory *S. cerevisiae* the capabilities of a novel gene amplification system was demonstrated. The system was aimed at creating up to ten copies of a gene integrated in specific targeting sites of the *S. cerevisiae* genome. First, large amounts of gene targeting substrates were generated through the construction of a USER® vector. Through the use of one, two and four copy amplification strains the stable production of 6-MSA was established. The 10.5 kb fragment of genes was successfully amplified. The constructed strains were evaluated in Erlenmeyer flasks. The results showed that the copy number of the genes and the 6-MSA titer correlated well. This indicates that even more copies of the genes for 6-MSA production could yield even higher titers. Thus the acyl-CoA substrates do not appear to be limiting the production of 6-MSA.

Construction of a cell factory and engineering it to increase production is one approach to obtaining an efficient cell factory. To aid the strain development further, it was sought to demonstrate the usefulness of a microtiter plate based cultivation system that uses CCD-flatbed scanners and image analysis as a tool to follow microbial growth and product formation. This CCD-flatbed scanning platform can be used for both process optimization as well as screening libraries of mutants generated through random mutagenesis. The experiments validated the CDD-flatbed scanning platform as a tool for quantifying microbial biomass from both bacteria and yeasts. Furthermore, the platform can be used to detect onset of production as well as volumetric productivities of the colored polyketide actinorhodin in *Streptomyces coelicolor*. It is a system that can be used in industrial settings for optimizing cell factory conditions. The use of microtiter plates makes it high-throughput and inexpensive method.

Thus in conclusion significant steps have been taken towards engineering an effective polyketide cell factory.

Dansk Resumé

Naturstoffer udgør en af de største kilder til lægemidler. Blandt naturstofferne har polyketider såsom erythromycin (antibiotikum) og lovastatin (kolesterolsænkende) længe bevist deres enorme værdi for patienter over hele verden. Polyketider produceres naturligt af planter, svampe og bakterier. Men de naturlige producenter opnår ofte ikke kommercielle koncentrationer af medicinske polyketider. Derfor må den naturlige produktion ofte forbedres ved mutagenese eller heterolog ekspression af polyketid-generne for at opnå en produktion med tilstrækkelige koncentrationer. For at forbedre produktionen af polyketider er bioteknologiske principper blevet anvendt til udvikling og konstruktion af forskellige polyketid cellefabrikker.

Først og fremmest blev de to biologiske værtsstammer *Aspergillus nidulans* og *Saccharomyces cerevisiae* anvendt til heterolog polyketid produktion. Begge organismer har kendte genetiske værktøjer til rådighed til ”gene targeting” og heterolog ekspression. Det har været målet at skabe en stabil ekspressions platform med alle gener integreret i genomet. Dette er opnået ved anvendelse af to avancerede gensplejsningsteknikker brugt på *A. nidulans* og *S. cerevisiae*. Begge systemer er blevet hjulpet af USER ™ kloningsvektorer, som blev udviklet for effektivt at frembringe store mængder af ”gene targeting” substrater og høj ekspression af polyketidsyntase (PKS) samt den aktiverende phosphopantetheinylase (PPTase). Dette alsidige system kan bruges til ekspression af andre interessante polyketider samt udvides til at udtrykke hele genklynger.

Efter at have opnået høj og stabil ekspressions niveau af polyketidet, skal cellefabrikken optimeres yderligere. Optimering kan opnås ved anvendelse af adaptive evolution, mutagenese og screening samt ”metabolic engineering”.

For det første blev *in silico* hjulpen ”metabolic engineering” anvendt som et værktøj til at styre metabolismen mod højere niveauer af 6-MSA produktion i *A. nidulans*. 6-MSA blev stabilt udtrykt i *A. nidulans* genomet og bioreaktor dyrkninger resulterede i høje titre af 6-MSA. Brugen af genom skala modellen af *A. nidulans* og optimerings algoritmen OptGene førte til udvikling af en knockout strategi, der skulle øge produktionen af 6-MSA i *A. nidulans*. Blandt kandidaterne forudsagt af modellen blev det valgt at deletere NADPH afhængig glutamat dehydrogenase (*gdhA*) genet. Dette burde resultere i større tilgængelighed af NADPH til polyketid produktion. Deletionen resulterede i nedsat veksthastighed af *A. nidulans*, som blev delvist reddet ved indsætning af en ekstra kopi af NADH afhænging glutamat dehydrogenase (*gdhB*). Fysiologisk karakterisering viste, at udbytterne af 6-MSA per biomasse blev forøget dog ikke væsentligt. Som et resultat af dette kan det hævdes, at der stadig er mere arbejde at gøre i form af modelkonstruktion i *A. nidulans*.

Anvendelse an en anden veletableret cellefabrik *S. cerevisiae* viser kapaciteten af et nyt gen amplifikationssystem. Systemet sigter mod at skabe op til ti kopier af et gen integreret specifikke steder i *S. cerevisiae* genomet. Først blev store mængder af ”gene targeting” substrater frembragt ved konstruktionen af en USER ® vektor. Ved anvendelse af en, to og fire kopier af amplifikations-stammerne blev stabil produktion af 6-MSA etableret. 10.5 kb fragmentet blev succesfuldt amplificeret. De konstruerede stammer blev evalueret i Erlenmeyerkolber. Resultaterne viste, at kopitallet af gener og 6-MSA titer korrelerede godt.

Dette indikerer, at endnu flere kopier af generne kan give endnu højere titere. Således virker det ikke som om Acyl-CoA substraterne er den begrænsende faktor i biosyntesen af 6-MSA.

Konstruktion af en cellefabrik og ”metabolic engineering” for at øge produktionen, er en metode til at opnå en effektiv cellefabrik. For at hjælpe stammeudviklingen yderligere har man søgt at påvise anvendeligheden af en mikrotiterplade baseret dyrkningssystem, der anvender CCD-flatbedscannere og billedanalyse som et værktøj til at følge mikrobiel vækst og produktdannelse. Denne CCD-flatbedscanning platform kan anvendes til både stamme optimering samt screening af biblioteker af mutanter genereret gennem mutagenese. Forsøgene validerede CDD-flatbedscannings platformen som et redskab til kvantificering af mikrobiel biomasse fra både bakterier og gær. Endvidere kan platformen anvendes til at detektere starten af produktionen såvel som volumetriske produktiviteter af det farvede polyketid actinorhodin i *Streptomyces coelicolor*. Det er et system, der kan anvendes i industrien til optimering af cellefabrik betingelser, hvor ”high-throughput” er afgørende.

Som konklusion er betydelige skridt taget mod konstruktion af en effektiv polyketid cellefabrik.

Table of Contents

Abbreviations

6-MSA	6-Methyl Salicylic Acid
CFP	Cyan Flourescent Protein
DW	Dry weight
FACS	Fluorescence Activated Cell Sorting
FBA	Flux Balance Analysis
gdhA	NADPH dependent glutamate dehydrogenase
gdhB	NADH dependent glutamate dehydrogenase
GSMN	Genome Scale Metabolic Networks
MOMA	Minimization of Metabolic Adjustment
PDA	Potato Dextrose Agar
PK	Polyketide
PKS	Polyketide synthase
PPP	Pentose phosphate pathway
PPTase	Phosphopanteteinyl Transferase
RGB	Red Green Blue (color components)

YFP	Yellow Flourescent Protein

Overview of the Thesis

This thesis represents a wide body of work that has touched upon many disciplines within metabolic engineering. Thus the reader is guided through the many interesting disciplines that have come together to fill this thesis. This overview aims at setting the stage for how the projects that have resulted in this thesis complement each other in providing insights into the process of engineering a polyketide cell factory.

Discovering natural products has long been the goal of the pharmaceutical industry as well as academic research. When new natural product clusters have been discovered the next challenge becomes to establish a production process that enables the eventual purification and formulation of the product to be used as therapeutics. Along this road there are many challenges to be met. First of all, the decision of which organism to use as a production system must be made. Should one rely on the natural host or make the decision to express the polyketide in a heterologous host. If one chooses the heterologous host one is dependent on knowledge and skills in molecular biology to accomplish this goal. It is not a trivial goal as there are many choices to be made on the road to optimality. Beyond that one needs methods to characterize the strains in order to evaluate the productivity and growth behavior and optimize the production even further. On top of that the use of *in silico* model guided approaches can lead to further insights into how metabolism can be adjusted towards a higher productivity. These subjects are covered in chapter 1 and 2.

Among natural products the polyketides are the most marvelous structures. They represent a diversity that continues to grow and that provides us with many fascinating structures.

Among them 6-methyl-salicylic acid constitutes a comparatively small model polyketide that is thus an ideal candidate for scientific endeavors. The polyketides and in particular 6-MSA will be described in chapter 3.

In chapter 4 the construction of a polyketide cell factory in *A. nidulans* for the production of 6-MSA is described. In an effort to guide metabolism towards higher productivity of 6-MSA in *A.nidulans* an *in silico* guided modeling approach was applied.

A novel gene amplification system for *S. cerevisiae* was applied for establishing stable 6-MSA production. The system is useful for high and stable production of many biological cell factory products. The model polyketide 6-MSA was used a way to test the applicability of this novel gene amplification system. The construction and evaluation of this cell factory is described in chapter 5.

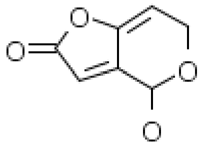
A fascinating microtiter based cultivation platform using CCD-scanning and image analysis is presented in chapter 6. The platform is valuable for discovering novel polyketides through the use of different conditions or combinations of strains and to evaluate mutant libraries. The system can monitor growth and metabolite formation in yeast and bacteria. It provides a flexible and robust framework for testing many cultivation conditions, large mutant libraries and evaluating colored product formation.

Chapter 1

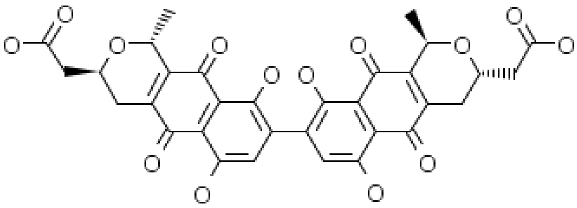
From Product to Production

Many of the compounds that are used in medicine today were originally produced as natural compounds by bacteria, fungi and plants (Newman and Cragg, 2007; Newman, 2008). Some of these compounds such as the immunosuppressant mycophenolic acid are still produced in the original host. One of the most famous examples is the non-ribosomal peptide penicillin that during and after the Second World War saved countless lives in the fight against bacterial infections (Rolinson, 1998). Since then many more compounds have been discovered. Often it is analogs of the natural product that end up as therapeutics with a stronger effect or less toxicity to humans. Examples of natural product analogs include analogs of penicillin, tetracycline and lovastatin (Sheehan and Henery-Logan, 1959; Xie et al., 2006; Pickens and Tang, 2009).

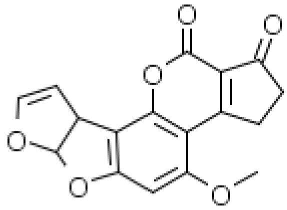
One class of compounds that has yielded many therapeutics is the polyketides (PK). Some examples of polyketides are the antibiotics erythromycin, tetracycline and rifamycin. The immunosuppressants tacrolimus and rapmycin are also polyketides. One polyketide that has a large market value is the cholesterol lowering agent lovastatin (Weissman, 2009). An estimate from 2009 states that on average polyketide derived medicines have annual sales above US$ 20 billion (Weissman, 2009). Some polyketides have also been found to be quite toxic such as the polyketide derived aflatoxin and patulin and they are thus also of some health concern to humans (Weissman, 2009). Structures of common polyketides can be seen in figure 1.



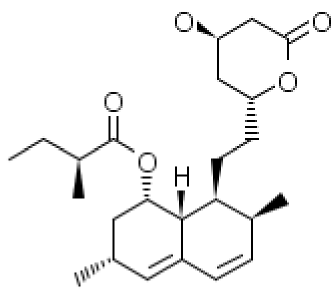
Patulin
(Toxin



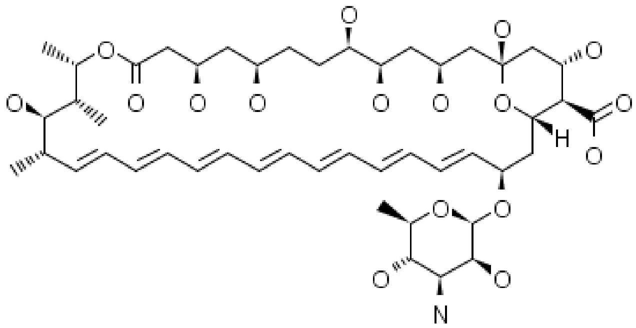
Actinorhodin
(Antibiotic)



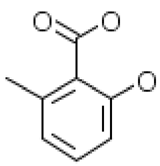
Aflatoxin
(Mycotoxin)



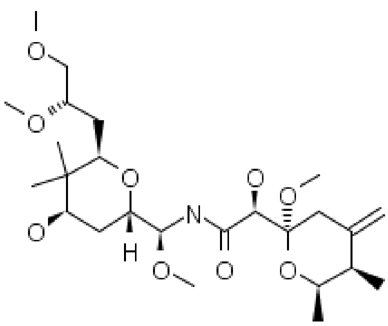
Lovastatin
(Cholesterol lowering agent)



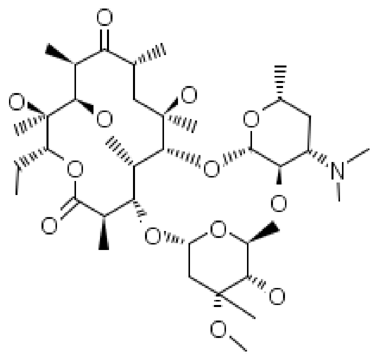
Amphotericin
(Antifungal)



6-MSA
(Antibiotic)



Pederin
(Anti-cancer)



Erythromycin
(Antibiotic)

Figure 1 The structures of different polyketides in parenthesis are given the significant properties of the polyketide (GenomeNet, 2012).

There are various ways of establishing production of polyketides, but there are three initial steps that must be overcome as illustrated in figure 2. First of all the polyketide compound most be discovered and the genes responsible for the production of the polyketide identified. Then through careful analysis possible host strains can be selected and an initial production established. Thereby establishing proof of concept in an initial cell factory.

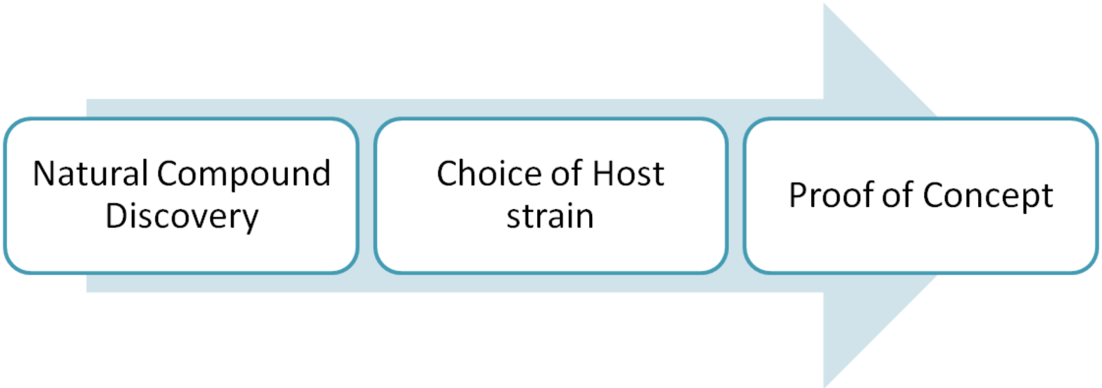


Figure 2 The figure illustrates the typical flow in the first steps of natural product discovery. The first scientific endeavor goes to actually discover a new compound that may be of therapeutic value. Beyond this one most select an appropriate production method and hopefully establish proof of concept in terms of being able to produce the natural product.

Discovery of Polyketides

Due to the many significant medical effects of polyketides the hunt for new polyketide derived compounds continues. One significant reason is the ever increasing resistance of bacteria to antibiotics. There are several recent advances in the technology that allows us to take a better guess at where the next natural compound might be found.

In order to aid the discovery of new secondary metabolites from newly sequenced microorganisms, the bioinformatics tool antiSMASH was developed (Medema et al., 2011). It is designed to identify potential polyketide clusters from DNA sequences. Another approach to the discovery of natural product pathways is a proteomics approach called PrISM (Bumpus et al., 2009). The method takes advantage of the fact that most enzymes found in natural product biosynthesis are quite large. The method analyses the biggest proteins from a gel of culture broths from various organisms (including non-sequenced) via mass spectrometry (Bumpus et al., 2009). Once a protein has been identified the corresponding gene cluster is amplified by PCR and sequenced (Bumpus et al., 2009). Furthermore a database called ThYme with tertiary structures of polyketide synthase domains has been constructed that potentially can help identify the active site and catalytic residues of newly discovered PKSs (Cantu et al., 2011). Furthermore, the advancements in mass spectrometry have made the detection of natural products produced in very low quantities easier.

An impressive amount of polyketide clusters have been found among the *Aspergilli* and streptomyces genomes. In *Aspergillus nidulans* 27 PKS related genes have been identified through sequence analysis, but only a fraction have been linked to a product (Keller et al., 2005; Sanchez et al., 2008; Schroeckh et al., 2009). In addition *Aspergillus oryzae* and *Aspergillus fumigatus* are predicted to contain 30 and 14 PK gene clusters respectively (Keller et al., 2005). Among the streptomycetes *S. coelicolor* contains 21 natural product clusters and *S. avermitilis* contains 25 natural product clusters (Sanchez et al., 2008).

It has been demonstrated that the polyketide clusters are subject to regulation and thus identification of a given polyketide requires activation of the gene cluster. There are generally two types of regulators the ones that act on several genes known as global regulators and local regulators that belong to a specific gene cluster. One example of a global regulator in *A. nidulans* is the LaeA regulator that has been shown to regulate sterigmatocystin, penicillin and lovastatin gene expression (Bok and Keller, 2004). In addition the polyketides aspyridone A and B were discovered based on the inducible expression of a pathway specific regulator that was integrated ectopically into the genome (Bergmann et al., 2007). The induction by a transcription factor of Asperfuranone biosynthesis in *Aspergillus nidulans* is yet another example (Chiang et al., 2009). Furthermore, two new Xanthones were discovered in *A. nidulans* based both on genome sequencing and deletion studies (Sanchez et al., 2011). Another interesting way of triggering polyketide production was demonstrated by Schroeckh *et al*. in the cocultivation of *S. hygroscopicus* and *A. nidulans* that lead to the activation of two polyketide gene clusters (Schroeckh et al., 2009). Lastly, the use of different media to screen for induced polyketide production is another example. This approach has been shown to induce the production of the polyketide containing meroterpenoids (Frisvad and Samson, 2004; Nielsen et al., 2011).

6-Methyl Salicylic Acid

The small model polyketide 6-methyl salicylic acid (6-MSA) has generated significant research attention. To date several natural producers of 6-MSA have been discovered. The production of 6-MSA in various fungal extracts has been studied since the 1950´s (Birch et al., 1955). It was first discovered from the fungus *penicillium patulum*, where it is further converted into patulin. Patulin is a toxic compound that is considered a dangerous food contaminant and is produced by many fungi as listed in table 1. There are thus limits both in the EU and the US on the amount of patulin allowed in food from e.g. apples such as juices and baby food (Puel et al., 2010).

Table 1 Species that are known to produce patulin and thus also 6-MSA (Puel et al., 2010).

Producers of Patulin	
Aspergillus	*A. clavatus, A. giganteus, A. longivesica*
Penicillium	*P. carneum, P. carneum, P. clavigerum, P. concentricum, P. coprobium, P. dipodomyicola, P. expansum, P. glandicola, P. gladioli, P. griseofulvum, P marinum, P. paneum, P. scerotigenum, P. vulpinum.*
Other	*Byssochlamys nivea, Paecilomyces saturates*

The gene encoding the methyl salicylic acid synthase (MSAS) has been identified as the *msas* gene from *P. patulum* (Beck et al., 1990). Later the gene was also sequenced from *P. urticae* (Wang et al., 1991). Recently it has been sequenced from *A. clavatus* and was identified as the *PatK* gene (Artigot et al., 2009). *P. chrysogenum*, *Talaromyces stipitatus* and *Aspergillus terreus* contain clusters similar to the patulin cluster in *P. clavatus*, but have not been identified as patulin producers (Puel et al., 2010). In the case of *A. terreus* the gene *AtX* encodes the MSAS which has been verified by heterologous expression in *A. oryzae* (Fujii et al., 1996; Varga et al., 2005). In *A. terreus* however it has been found that 6-MSA is needed as a precursor for terreic acid biosynthesis and thus this may be the underlying reason behind 6-MSA synthesis in *A. terreus* (Read et al., 1969). A study of inter kingdom transfer of polyketides also identified a homologous gene for 6-MSA synthesis in *Aspergillus niger* XM 001402371 and *Actinomadura madurare* AY271660 (Schmitt and Lumbsch, 2009). Likewise lichenized ascomycetes have through sequence analysis been identified as potential 6-MSA producers (Schmitt et al., 2008). Furthermore a *msas* like gene was found in *Aspergillus westerdijkiae* and showed to be involved in aspyrone and isoasperlactone biosynthesis (Bacha et al., 2009).

When new polyketides have been discovered the next step is to find a suitable production method as well as optimizing the productivity and yields.

Selection of a Cell Factory

Once a polyketide has been discovered picking the right production method is important. As polyketides are large compounds chemical synthesis is often very difficult making the use of microorganisms as biological cell factories the method of choice. In this respect the first challenge is which cell factory to use. One option is to use the native host. Alternatively, when the gene clusters responsible for the production of a specific polyketide are known, they can be expressed in a heterologous host for production. Picking the right heterologous host for polyketide genes presents a challenge. Some examples of industrially used cell factories and the product they produce are given in table 2.

Table 2 Examples of different cell factories that are used for biopharmaceutical and chemical productions in the industry (modified from Papini *et al.*) (Papini et al., 2010).

Product	Cell Factory
Biopharmaceuticals	
Anticoagulant (*tPA*)	CHO Cells
Insulin	*S. cerevisiae*
Insulin	*E. coli*
Human growth hormone	*E. coli*
Recombinant vaccine against Hep B	*S. cerevisiae*
Monoclonal antibody based product	CHO Cells
Industrial Biotechnology Products	
Citric Acid (organic acid)	*A. niger*
L-lysine (amino acid)	*C. glutamicum*
Penicillin (antibiotic)	*P. chrysogenum*
α-Amylases (enzyme)	*A. oryzae*
1,3 Propane diol (polymer)	*E. coli*

The most commonly used cell factories are mammalian cells (e.g. CHO cells), bacteria (e.g. *E. coli*) and fungi (e.g. yeast and *Aspergilli*). Mammalian cells are mainly used for biopharmaceuticals due to their ability to apply human like glycosylation patterns (dos Reis Castilho, 2008). The remaining fungi and bacteria are thus the method of choice for industrial biotechnological applications. The cell factories available today all have advantages and drawbacks that are product and process dependent. The main cell factory properties of bacteria and fungal cell factories are listed in table 3. The modeling properties given in table 3 are important for *in silico* guided metabolic engineering, but are still of more academic than industrial interest.

Table 3 The properties of some industrial cell factories. The table is modified from Papini *et al* (Papini et al., 2010).

	S. cerevisiae	*Aspergillus* spp.	*E. coli*
Genetic Engineering Tools			
Availability of genomic tools	+++	++	+++
Online Resources/databases	+++	++	+++
Advanced molecular biology techniques	+++	++	+++
Strain Construction speed	++	+	+++
Fermentation Properties			
Robust fermentation technologies	+++	++	+++
Tolerance to low pH	++	+++	-
Capability of using complex feedstock	+	++	+
Presence of toxins or endotoxins/viral proteins/LPS	-	+	+
Fast Growth	++	+	+++
Models of Metabolism			
Genome Scale Model Completion	+++	+	+++
Kinetic Models	++	+	++
In silico guided Metabolic Engineering	++	-	+++

Yeasts and *Aspergilli* have been shown to produce polyketides efficiently and will be discussed in greater detail in the following sections.

Yeasts as Cell Factories

Yeasts have a long tradition in biotechnological processes as they have been used for decades in the baking and brewing industry. Examples of industrially relevant yeast species include *S. cerevisiae*, *Pichia pastoris, Yarrowia lipolytica* and *Kluyveromyces lactis*. *S. cerevisiae* is also used for recombinant protein production (e.g. insulin) and commodity chemicals (e.g.

ethanol). As a result of the long history, the knowledge of biochemistry, genetics and physiology of *S. cerevisiae* is great. This has lead to the establishment of advanced molecular biology and fermentation technologies (Papini et al., 2010). *S. cerevisiae* is a unicellular organism, which makes it easy to genetically engineer and cultivate in reactors compared to higher eukaryotes. Furthermore vectors, dominant and auxotrophic markers and efficient transformation protocols are available for *S. cerevisiae* (Papini et al., 2010; Mattanovich et al., 2012). Yeasts also have the added advantage that they are able to due most protein processing similar to other eukaryotic organisms such as folding, assembly and post-translational modifications (Mattanovich et al., 2012).

Homologous recombination occurs at a high efficiency in *S. cerevisiae* and it is thus easy to integrate genes stably in the genome and perform gene targeted deletions or overexpressions. A recent study in *S. cerevisiae* identified the *TEF1* and *PGK1* promoters as having the highest expression levels under most growth conditions making them good candidates for cell factory construction (Partow et al., 2010). Furthermore, there are already libraries of single and double deletion mutants and the effects of metabolism are thus well studied for genetic engineering strains. The ability to cross yeast strains also makes combining engineered traits easier.

Today many processes involving *S. cerevisiae* as a cell factory have obtained GRAS (Generally Regarded as Safe) status thus making the approval of other processes involving this organism easier (Papini et al., 2010). In addition, yeasts can use a wide range of carbon and energy sources for production (e.g. glucose, lactose, maltose, starch, alkanes and fatty acid) (Mattanovich et al., 2012). In terms of polyketide production S. cerevisiae is ideally suited as it does not produce any polyketides itself and there is thus not a strong competition for precursors. The possibility of the polyketide product being transformed through other polyketide modifying enzymes is thus also limited. From a purification perspective it is also simpler to work a non-producer as a cell factory.

S. cerevisiae however also has some drawbacks as a production host especially for pharmaceutically relevant proteins. The strong fermentative metabolism and limited recombinant protein productivity presents a challenge (Mattanovich et al., 2012). Furthermore, *S. cerevisiae* has a tendency to hyper-glycosyalte proteins and they may be kept in the periplasmic membrane leading to degradation products that are difficult to remove (Mattanovich et al., 2012).

Aspergilli as Cell Factories

Filamentous fungi are used today to produce a wide range of products including organic acids (e. g. citric acid) and enzymes (e.g. α-Amylase).). The most commonly used industrial *Aspergilli* are *A. niger and A. oryzae* (Lubertozzi and Keasling, 2009). Especially the production of enzymes from filamentous fungi is a large market constituting about half of a total enzyme market of nearly US$ 5 billion (Lubertozzi and Keasling, 2009). On top of this fungi have capabilities of utilizing a wide range of natural organic substrates (Papini et al., 2010). They can tolerate a wide range of temperature, pH and salt concentrations (Lubertozzi and Keasling, 2009). The multicellular nature and tendency to aggregate and grow on wall surfaces makes it more challenging to cultivate *Aspergilli* than yeast, however it results in much easier purification of extracellular enzymes and metabolites as cells can be removed through simple filtration (Lubertozzi and Keasling, 2009).

In terms of genetic engineering methods filamentous fungi are still far behind *S. cerevisiae*. Thus the availability of dominant and auxotrophic markers and vectors is only a fraction of what can be found for *S. cerevisiae*. Although it is not used in the fermentation industry, *A. nidulans* molecular biology is well studied. The development of host strains for genetic transformation deficient in the non-homologous end-joining (NHEJ) pathway, e.g. the *nkuA*Δ strain, has significantly increased the frequency of homologous recombination paving the way for efficient gene targeting (Nielsen et al., 2008). However, the minimal size of homologous gene fragments needed to perform gene targeting is still larger than for *S. cerevisiae*. There are not as many well-studied promoters available for *Aspergillus nidulans*, but one study concluded that among the three promoters tested, the p*gpdA* promoter gave the best stable expression of a heterologous gene (Lubertozzi and Keasling, 2006).

In terms of polyketide production *Aspergilli* are well-suited as hosts. They are already capable of producing quite large amounts of polyketides and their metabolism is thus geared to providing the necessary acyl-CoA precursors for polyketide production. However, there is a risk of cross-reactions in the polyketide pathway leading to the production of altered polyketides from the originally intended.

Production of Polyketides

During the design of a process for production of polyketides several factors need to be considered. There are two main options. One is using the native host of the polyketide synthase as a production strain. The other involves transferring the gene clusters to a heterologous host. Both aspects involve challenges from a bioinformatics, molecular biology and process point of view.

Native Producers of Polyketides

The natural production of polyketides occurs in many different organisms amongst these are fungi, bacteria and plants. Examples include the production of 6-MSA by *Penicillium griseofulvum* and the production of actinorhodin by several streptomycete species (Sanchez et al., 2008). If the genes responsible for polyketide production are not known using the natural host may be the only option.

Successful industrial production of the polyketide lovastatin in *A. terreus* was originally set up in 1980 at Merck (Manzoni and Rollini, 2002). Later improvements have been made and the Metkinen group have reported titers up to 7-8 g/L by using random mutagenesis procedures (Metkinen Oy, 2012). In addition the production of erythromycin by *Saccharopolyspora erythraea* occurs in the natural host and the annual production of erythromycin amounts to several thousand tons (Minas et al., 1998). Another example is mycophenolic acid that is produced quite efficiently from *P. brevicompactum* with final titers up to 5.7 g/L (Xua and Yang, 2007; Ardestani et al., 2010). Only recently the genes encoding the PKS that leads to mycophenolic acid production were partially annotated and heterologous expression of central elements made possible (Hansen et al., 2011; Regueira et al., 2011).

However, the natural producers sometimes have low production rates of polyketides and are not suited for industrial scale production. For example *P. griseofulvum* produces only 0.2 mg/L of 6-MSA on minimal medium (Peace et al., 1981). Beyond low yields it can often be very difficult to cultivate secondary metabolite producing bacteria and fungi. Low growth rate and limited knowledge of nutritional requirements can lead to the use of a heterologous production platform.

Heterologous Expression of Polyketide Synthase Genes

The main cell factories used for heterologous polyketide production are bacteria, fungi and seldom plants (Pfeifer and Khosla, 2001). Figure 3 illustrates some of the challenging aspects that need to be considered when selecting a polyketide cell factory.

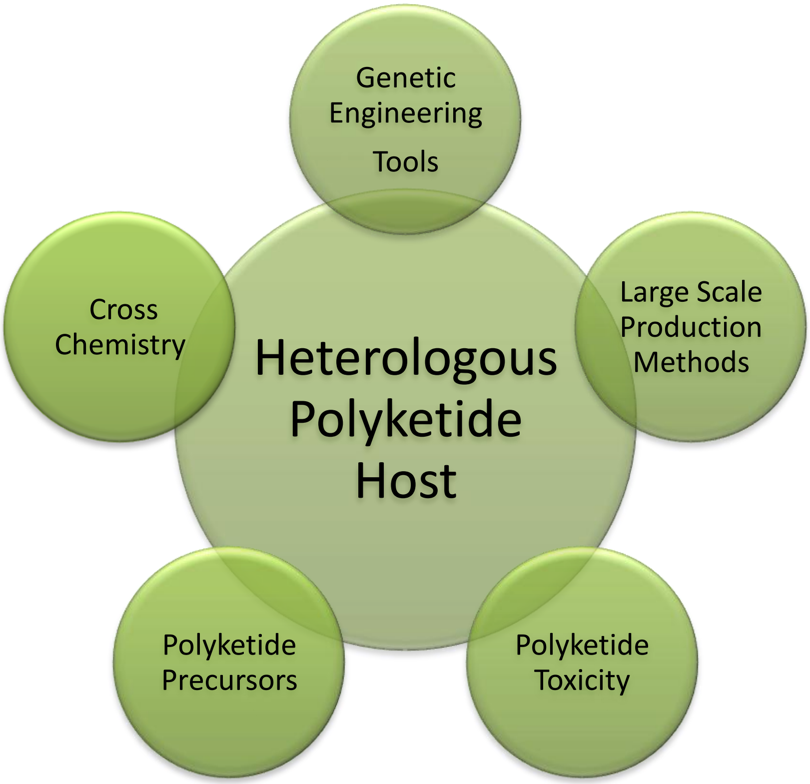


Figure 3 The circles illustrate some of the central elements that one needs to be considering when choosing a polyketide host organism.

First of all the molecular biology tools need to be available for the host strain. Depending on the choice of host strain there are going to be a more or less established method for genetic engineering. There may be a trade of between selecting an organism that is capable of producing high amounts of the desired polyketide versus an organism that is easy to manipulate and engineer. Furthermore, knowledge of large scale production methods by bioreactor cultivations also has to be well established. If an efficient production process is to be established it is key to have knowledge of process factors that influence productivity and yields.

An additional challenge is the fact that the genes of many polyketide synthases are very GC rich. Thus it may be of interest to codon optimize the genes for heterologous expression. The price of whole gene synthesis has fallen dramatically and it represents a true alternative to using the original polyketide gene (Carlson, 2009). The size of genes that can be synthesized has also increased rapidly (Carlson, 2009). However, codon optimization does not always lead to the desired expression of a gene.

Another complicating factor of heterologous expression is the large size of some of the PKSs (100 to 10,000 kDa) (Pfeifer and Khosla, 2001). The genes thus are ranging from 20 to more than 100 kb in size (Murakami et al., 2011). Furthermore the genes often contain introns that may have to be removed before the gene is introduced in a heterologous host.

Furthermore the natural PK host often has resistance mechanisms that allow it to tolerate its own toxin production. This problem also has to be overcome in a heterologous host. One example is the selfresistance to mycophenolic acid that is observed in *penicillium* *brevicompactum* (Regueira et al., 2011). It is a result of an extra copy of IMP dehydrogenase (IMPDH) located in the polyketide cluster (Hansen et al., 2011; Regueira et al., 2011).

In addition, all the precursors used by the polyketide synthase such as methyl-malonyl-CoA or propionyl-CoA need to be present in the host or the pathways engineered. Alternatively the precursor can be fed to the heterologous host. An example of an engineered precursor in *S. cerevisiae* and *E. coli* was the engineering of methylmalonyl-CoA Mutase-Epimerase Pathway for the production of the polyketide precursor methyl-malonyl-CoA (Dayem et al., 2002; Mutka et al., 2006). Methyl-malonyl-CoA is used for biosynthesis of triketide lactone and the polyketide part of erythromycin (Dayem et al., 2002; Mutka et al., 2006).

Lastly, heterologous expression in a host that already has extensive secondary metabolism may result in the polyketide product being altered as a result of modifying enzymes from other polyketide clusters. Thus the possibility of cross chemistry between the heterologous polyketide and native polyketide modifying enzymes must be carefully considered.

Examples of Heterologous Polyketide Production

Several polyketides have been expressed heterologously in research labs to study their structure and function. Some of different polyketides that have been produced in a heterologous host are shown in table 4. The heterologous expression of the *msas* gene has been attempted several times within the last 20 years (see table 4). The first heterologous expression was achieved in *S. coelicolor* (Bedford et al., 1995) and the authors report a titer of 20 mg/L 6-MSA. Beyond that there have been reports of up to 1.7 g/L in *S. cerevisiae* and 445 mg/L in *A. nidulans* (Kealey et al., 1998; Panagiotou et al., 2009).

Table 4 Examples of the heterologous expression of 6-MSA and other polyketides in different microorganisms.

Polyketide	**Heterologous host**	**Natural Producer**	**Expression system**	**Reference**
Fungal type I PK				
6-MSA	*S. coelicolor*	*P. patulum*	Plasmid	(Bedford et al., 1995)
6-MSA	*E. coli*	*P. patulum*	Plasmid	(Kealey et al., 1998)
6-MSA	*S. cerevisiae*	*P. patulum*	Plasmid	(Kealey et al., 1998)
6-MSA	*S. cerevisiae*	*P. patulum*	Plasmid	(Wattanachaisaereekul et al., 2008)
6-MSA	*A. nidulans*	*A. terreus*	Ectopic Integration	(Panagiotou et al., 2009)
6-MSA	*Nicotiana tabacum (tobacco)*	*P. patulum*	Integration	(Yalpani et al., 2001)
Bacterial type I PK				
6-deoxyerythronolide (antibiotic)	*E. coli*	*S. erythraea*	Plasmid	(Pfeifer et al., 2001)
Epothilone (anticancer)	*Myxococcus xanthus*	*Sorangium cellulosum*	Integration	(Julien and Shah, 2002)
Picromycin/methymycin	*S. lividans*	*S. venezuelas*	Plasmid	(Tang et al., 1999)
Bacterial type II PK				
Actinorhodin	*S. coelicolor*	*S. parvulus*	Plasmid	(Malpartida and Hopwood, 1984)
Oleandomycin	*S. lividans*	*S. antibioticus*	Plasmid	(Shah et al., 2000)
Type III PK (often plant)				
Naringenin	*E. coli*	*R. idaeus* (raspberry)	Plasmid	(Zheng et al., 2001)
Stilbene	*E. coli*	*P. strobes* (pine)	Plasmid	(Raiber et al., 1995)

References

Ardestani, F., Fatemi, S.S., Yakhchali, B., Hosseyni, S.M., Najafpour, G., 2010. Evaluation of mycophenolic acid production by *Penicillium brevicompactum* MUCL 19011 in batch and continuous submerged cultures. Biochem. Eng. J. 50, 99-103.

Artigot, M.P., Loiseau, N., Laffitte, J., Mas-Reguieg, L., Tadrist, S., Oswald, I.P., Puel, O., 2009. Molecular cloning and functional characterization of two CYP619 cytochrome P450s involved in biosynthesis of patulin in *Aspergillus clavatus*. Microbiology 155, 1738.

Bacha, N., Dao, H.P., Atoui, A., Mathieu, F., O'Callaghan, J., Puel, O., Liboz, T., Dobson, A.D.W., Lebrihi, A., 2009. Cloning and characterization of novel methylsalicylic acid synthase gene involved in the biosynthesis of isoasperlactone and asperlactone in *Aspergillus westerdijkiae*. Fungal Genetics and Biology 46, 742-9.

Beck, J., Ripka, S., Siegner, A., Schiltz, E., Schweizer, E., 1990. The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum* Its gene structure relative to other polyketide synthases. European Journal of Biochemistry 192, 487-98.

Bedford, D.J., Schweizer, E., Hopwood, D.A., Khosla, C., 1995. Expression of a functional fungal polyketide synthase in the bacterium *Streptomyces coelicolor* A3 (2). J. Bacteriol. 177, 4544.

Bergmann, S., Schuemann, J., Scherlach, K., Lange, C., Brakhage, A.A., Hertweck, C., 2007. Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. Nature Chemical Biology 3, 213-7.

Birch, A., Massy-Westropp, R., Moye, C., 1955. Studies in relation to biosynthesis. VII. 2-Hydroxy-6-methylbenzoic acid in *Penicillium griseofulvum* Dierckx. Aust. J. Chem. 8, 539-44.

Bok, J.W., Keller, N.P., 2004. LaeA, a regulator of secondary metabolism in Aspergillus spp. Eukaryotic cell 3, 527-35.

Bumpus, S.B., Evans, B.S., Thomas, P.M., Ntai, I., Kelleher, N.L., 2009. A proteomics approach to discovering natural products and their biosynthetic pathways. Nat. Biotechnol. 27, 951-6.

Cantu, D.C., Chen, Y., Lemons, M.L., Reilly, P.J., 2011. ThYme: a database for thioester-active enzymes. Nucleic Acids Res. 39, D342.

Carlson, R., 2009. The changing economics of DNA synthesis. Nat. Biotechnol. 27, 1091-4.

Chiang, Y.M., Szewczyk, E., Davidson, A.D., Keller, N., Oakley, B.R., Wang, C.C.C., 2009. A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in Aspergillus nidulans. J. Am. Chem. Soc. 131, 2965-70.

Dayem, L.C., Carney, J.R., Santi, D.V., Pfeifer, B.A., Khosla, C., Kealey, J.T., 2002. Metabolic engineering of a methylmalonyl-CoA mutase-epimerase pathway for complex polyketide biosynthesis in *Escherichia coli*. Biochemistry (N. Y.) 41, 5193-201.

dos Reis Castilho, L., 2008. Animal Cell Technology: From Biopharmaceuticals to Gene Therapy. Taylor & Francis.

Frisvad, J.C., Samson, R.A., 2004. Polyphasic taxonomy of *Penicillium* subgenus *Penicillium*. A guide to identification of food and air-borne terverticillate Penicillia and their mycotoxins. Stud. Mycol. 49.

Fujii, I., Ono, Y., Tada, H., Gomi, K., Ebizuka, Y., Sankawa, U., 1996. Cloning of the polyketide synthase gene atX from *Aspergillus terreus* and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression. Molecular and General Genetics MGG 253, 1-10.

GenomeNet, 2012. www.genome.jp. Accessed 25th of April 2012.

Hansen, B.G., Salomonsen, B., Nielsen, M.T., Nielsen, J.B., Hansen, N.B., Nielsen, K.F., Regueira, T.B., Nielsen, J., Patil, K.R., Mortensen, U.H., 2011. A versatile gene expression and characterization system for Aspergillus: heterologous expression of the gene encoding the polyketide synthase from the mycophenolic acid gene cluster from *Penicillium brevicompactum* as a case study. Appl. Environ. Microbiol. .

Julien, B., Shah, S., 2002. Heterologous expression of epothilone biosynthetic genes in *Myxococcus xanthus*. Antimicrob. Agents Chemother. 46, 2772-8.

Kealey, J.T., Liu, L., Santi, D.V., Betlach, M.C., Barr, P.J., 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proceedings of the National Academy of Sciences 95, 505.

Keller, N.P., Turner, G., Bennett, J.W., 2005. Fungal secondary metabolism—from biochemistry to genomics. Nature Reviews Microbiology 3, 937-47.

Lubertozzi, D., Keasling, J.D., 2006. Marker and promoter effects on heterologous expression in *Aspergillus nidulans*. Appl. Microbiol. Biotechnol. 72, 1014-23.

Malpartida, F., Hopwood, D., 1984. Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host.

Manzoni, M., Rollini, M., 2002. Biosynthesis and biotechnological production of statins by filamentous fungi and application of these cholesterol-lowering drugs. Appl. Microbiol. Biotechnol. 58, 555-64.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., Breitling, R., 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39, W339.

Metkinen Oy, F., 2012. http://www.metkinen.fi/technologies/lovastatin. Accessed 15th of March 2012.

Minas, W., Brünker, P., Kallio, P.T., Bailey, J.E., 1998. Improved erythromycin production in a genetically engineered industrial strain of *Saccharopolyspora erythraea*. Biotechnol. Prog. 14, 561-6.

Murakami, T., Burian, J., Yanai, K., Bibb, M.J., Thompson, C.J., 2011. A system for the targeted amplification of bacterial gene clusters multiplies antibiotic yield in *Streptomyces coelicolor*. Proceedings of the National Academy of Sciences 108, 16020-5.

Mutka, S.C., Bondi, S.M., Carney, J.R., Da Silva, N.A., Kealey, J.T., 2006. Metabolic pathway engineering for complex polyketide biosynthesis in Saccharomyces cerevisiae. FEMS yeast research 6, 40-7.

Newman, D.J., 2008. Natural products as leads to potential drugs: an old process or the new hope for drug discovery? J. Med. Chem. 51, 2589-99.

Newman, D.J., Cragg, G.M., 2007. Natural Products as Sources of New Drugs over the Last 25 Years⊥. J. Nat. Prod. 70, 461-77.

Nielsen, M.L., Nielsen, J.B., Rank, C., Klejnstrup, M.L., Holm, D.K., Brogaard, K.H., Hansen, B.G., Frisvad, J.C., Larsen, T.O., Mortensen, U.H., 2011. A genome-wide polyketide synthase deletion library uncovers novel genetic links to polyketides and meroterpenoids in Aspergillus nidulans. FEMS Microbiol. Lett. .

Panagiotou, G., Andersen, M.R., Grotkjaer, T., Regueira, T.B., Nielsen, J., Olsson, L., 2009. Studies of the production of fungal polyketides in *Aspergillus nidulans* by using systems biology tools. Appl. Environ. Microbiol. 75, 2212-20.

Papini, M., Salazar, M., Nielsen, J., 2010. Systems biology of industrial microorganisms. Biosystems Engineering I , 51-99.

Partow, S., Siewers, V., Bjørn, S., Nielsen, J., Maury, J., 2010. Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. Yeast 27, 955-64.

Peace, J., Bartman, C., Doerfler, D., Campbell, I., 1981. 6-Methylsalicylic Acid Production in Solid Cultures of *Penicillium patulum* Occurs Only When an Aerial Mycelium Is Present. Appl. Environ. Microbiol. 41, 1407.

Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., Khosla, C., 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science 291, 1790-2.

Pfeifer, B.A., Khosla, C., 2001. Biosynthesis of Polyketides in Heterologous Hosts. Microbiol. Mol. Biol. Rev. 65, 106-18.

Puel, O., Galtier, P., Oswald, I.P., 2010. Biosynthesis and Toxicological Effects of Patulin. Toxins 2, 613-31.

Raiber, S., Schröder, G., Schröder, J., 1995. Molecular and enzymatic characterization of two stilbene synthases from Eastern white pine (Pinus strobus) A single Arg/His difference determines the activity and the pH dependence of the enzymes. FEBS Lett. 361, 299-302.

Read, G., Westlake, D., Vining, L., 1969. Quinone epoxides. V. 1 The biosynthesis of terreic acid. Can. J. Biochem. 47, 1071-9.

Regueira, T.B., Kildegaard, K.R., Hansen, B.G., Mortensen, U.H., Hertweck, C., Nielsen, J., 2011. Molecular basis for mycophenolic acid biosynthesis in *Penicillium brevicompactum*. Appl. Environ. Microbiol. 77, 3035-43.

Rolinson, G.N., 1998. Forty years of beta-lactam research. J. Antimicrob. Chemother. 41, 589.

Sanchez, J.F., Chiang, Y.M., Wang, C.C.C., 2008. Diversity of polyketide synthases found in the Aspergillus and Streptomyces genomes. Molecular Pharmaceutics 5, 226-33.

Sanchez, J.F., Entwistle, R., Hung, J.H., Yaegashi, J., Jain, S., Chiang, Y.M., Wang, C.C.C., Oakley, B.R., 2011. Genome-based deletion analysis reveals the prenyl xanthone biosynthesis pathway in *Aspergillus nidulans*. J. Am. Chem. Soc. .

Schmitt, I., Kautz, S., Lumbsch, H.T., 2008. 6-MSAS-like polyketide synthase genes occur in lichenized ascomycetes. Mycol. Res. 112, 289-96.

Schmitt, I., Lumbsch, H.T., 2009. Ancient horizontal gene transfer from bacteria enhances biosynthetic capabilities of fungi. PLoS One 4, e4437.

Schroeckh, V., Scherlach, K., Nützmann, H.W., Shelest, E., Schmidt-Heck, W., Schuemann, J., Martin, K., Hertweck, C., Brakhage, A.A., 2009. Intimate bacterial–fungal interaction triggers biosynthesis of archetypal polyketides in Aspergillus nidulans. Proceedings of the National Academy of Sciences 106, 14558-63.

Shah, S., Xue, Q., Tang, L., Carney, J.R., Betlach, M., McDaniel, R., 2000. Cloning, characterization and heterologous expression of a polyketide synthase and P-450 oxidase involved in the biosynthesis of the antibiotic oleandomycin. J. Antibiot. 53, 502.

Tang, L., Fu, H., Betlach, M.C., McDaniel, R., 1999. Elucidating the mechanism of chain termination switching in the picromycin/methymycin polyketide synthase. Chem. Biol. 6, 553-8.

Varga, J., Toth, B., Kocsube, S., Farkas, B., Szakács, G., Teren, J., Kozakiewicz, Z., 2005. Evolutionary relationships among *Aspergillus terreus* isolates and their relatives. Antonie Van Leeuwenhoek 88, 141-50.

Wang, I.K., Reeves, C., Gaucher, G., 1991. Isolation and sequencing of a genomic DNA clone containing the 3'terminus of the 6-methylsalicylic acid polyketide synthetase gene of *Penicillium urticae*. Can. J. Microbiol. 37, 86-95.

Wattanachaisaereekul, S., Lantz, A.E., Nielsen, M.L., Nielsen, J., 2008. Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. Metab. Eng. 10, 246-54.

Weissman, K.J., 2009. Introduction to Polyketide Biosynthesis. Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates 459, 3-16.

Xua, Z., Yang, S., 2007. Production of mycophenolic acid by *Penicillium brevicompactum* immobilized in a rotating fibrous-bed bioreactor. Enzyme and Microbial Technology 40, 623-8.

Yalpani, N., Altier, D.J., Barbour, E., Cigan, A.L., Scelonge, C.J., 2001. Production of 6-Methylsalicylic Acid by Expression of a Fungal Polyketide Synthase Activates Disease Resistance in Tobacco. The Plant Cell 13, 1401-9.

Zheng, D., Schröder, G., Schröder, J., Hrazdina, G., 2001. Molecular and biochemical characterization of three aromatic polyketide synthase genes from *Rubus idaeus*. Plant Mol. Biol. 46, 1-15.

Chapter 2

Strain Improvement

The enhancement of biological cell factories for the production of chemicals and therapeutics with increasing market value has long been a goal of the biotech industry (Otero and Nielsen, 2010). In general there are three main strategies employed for strain improvement. They are referred to as metabolic engineering, adaptive evolution and random mutagenesis (see figure 1). The use of one method of strain improvement does not necessarily exclude the others however the fundamental approach is different.

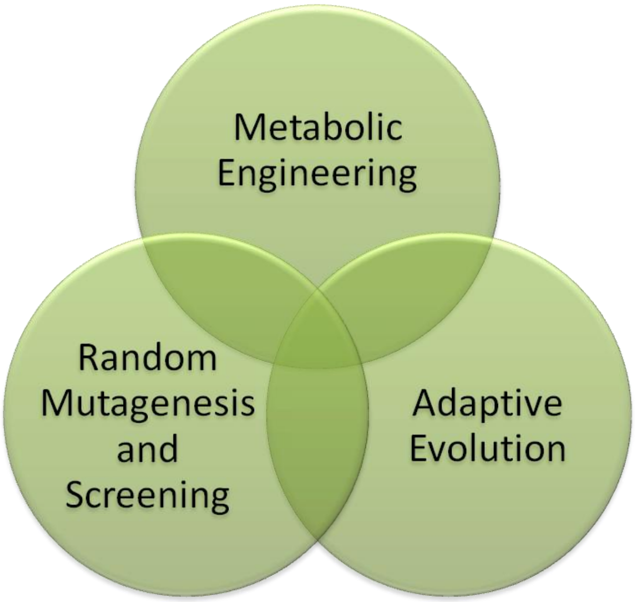


Figure 1 The three main strategies used for strain improvement

Adaptive evolution relies on the inherent rate of mutation occurring in a population of microorganisms as a result of a selective pressure. Often adaptive evolution is used to increase the growth rate, increase tolerance to process conditions or byproducts as well as reduce the need for media supplementation (Ibarra et al., 2002). An example of successful adaptive evolution is the two fold increased uptake of lactose in a recombinant *S. cerevisiae* that was subjected to prolonged growth on a lactose containing media (Guimarães et al., 2008). Another example where adaptive evolution was used after gene targeted deletions was the evolution of a *S. cerevisiae* mutant that was initially dependent on 500 mg/L glycine in the media to in the end be able to grow without glycine (Otero, 2009). Furthermore growth rate was increased from 0.03 h^{-1} to 0.14 h^{-1} through repeated subcultivation in shake flasks (Otero, 2009). Other examples can be found in the review by Wright, but will not be discussed in detail as adaptive evolution has not been used in the work performed in this thesis (Wright, 2004).

Random Mutagenesis and Screening

Historically, random mutagenesis and selection has been widely used in strain improvement programs and still is (Thykaer and Nielsen, 2003; Adrio and Demain, 2006). One example is the production of erythromycin in *Saccharopolyspora erythraea* that through random mutagenesis and selection has been increased 50-100 times to produce 8 g/L of erythromycin (Pfeifer and Khosla, 2001). Another example is the improvement obtained in the production of Penicillin by *Penicillium chrysogenum* (Thykaer and Nielsen, 2003).

There are various methods of inducing random mutations with the aim of strain improvement. The most commonly used mutagens are the physical e.g. ultraviolet, gamma and Xirradiation and chemical e.g. ethyl methane sulphonate (EMS), nitrosomethyl guanidine (NTG) and mustards such as ICR170 (Rowlands, 1984). When employing these mutagens, it is important to be aware of the dose and which type of mutation they induce dependent on the outcome of the strain development that is desirable (Rowlands, 1984).

In recent years however the production of a plethora of candidate improved strains has become possible through even more sophisticated methods of generating random mutations and recombination of different traits between different strain backgrounds. Examples of different methods have been reviewed by Santos and Stephanopoulos (Santos and Stephanopoulos, 2008). They provide examples of how synthetic promoter libraries, random knock-out and overexpression libraries, artificial transcription factors and genome shuffling can be used to generate a large variety of mutations (Santos and Stephanopoulos, 2008). Some of these methods may even allow for probing a larger phenotypic space as some mutations may not be as easily obtained by classical mutagenesis. In addition, it has been demonstrated that plasmids carrying mutation inducers can be used to generate fast mutations in a strain background and then the strains can be cured of the plasmid leading to a stable production strain (Selifonova et al., 2001).

In continuation of these procedures of randomly generating mutations and selecting for improved phenotypes, it is possible to take the improved strains and compare them to the original wildtype strains by e.g. sequencing of central elements. This has lead to inverse metabolic engineering where the beneficial mutations of an observed improvement can be engineered back into the original strain background thereby potentially avoiding deleterious mutations (Bailey et al., 1996; Santos and Stephanopoulos, 2008).

Regardless of how the large number of strains have been generated it is essential to be able to screen for improved properties such as growth and titer in a reproducible and scalable manor. Thus the development of microtiter plate screening programs has proven indispensible for large scale industrial screening programs.

Microtiter Plate Screening

During the last decade the need for miniature bioreactors has dramatically increased within industry and research. The demand has been elevated due to the need for fast and efficient strain characterization. Micro-bioreactors have gradually replaced shake flasks as being the preferred testing vessel of large screening programs. The success of such random screenings again is also largely dependent on the number of strains that can be screened simultaneously. In this case micro-titer plates offer the possibility to culture many strains at the same time. As the field has received increasing attention quite a few reviews of the different platforms have been written (Betts and Baganz, 2006; Duetz, 2007).

The main parameters evaluated for growth of microorganisms in a bench-scale and production scale bioreactor is stirring, pH, aeration and temperature. When moving towards a smaller scale it becomes increasingly difficult to measure and control these parameters. However technology in this field is moving fast and it is becoming increasingly possible to measure and control these parameters even at 0.1-100 mL. Something that is not possible from the traditionally used shake-flasks. Although some very advanced shake-flaks systems have been developed (Wittmann et al., 2003). Of course ideally any microbioreactor should be able to produce results that are scale-able thus making the selection of production strains more efficient.

The test systems need to be reliable and scalable to be comparable to large scale production facilities. One of the key issues with micro-bioreactors is achieving adequate mixing and oxygen transfer rates. When dealing with microtiter plates several investigations have been made on the influence of well geometry and size (Duetz and Witholt, 2004; Funke et al., 2009). Experiments have also determined the effects of the amplitude and rpm of the shaker during orbital shaking of the plates (Duetz et al., 2000; Duetz and Witholt, 2001). It has been demonstrated that when using an orbital shaker with an amplitude of 50 mm gives much better oxygen transfer then lower amplitudes (Duetz et al., 2000).

Another issue with scaling down reactor volumes is the limited amount of samples that can be withdrawn from a given vessel. As a rule of thumb only 10-20% of the volume should be extracted. This of course puts pressure on the analysis instruments such as HPLCs in terms of the volume of sampler required per measurement.

Furthermore the systems are mainly amenable to singe celled organisms such as yeast and bacteria whereas filamentous organisms such as *A. nidulans* are very difficult to cultivate in microtiter plates. Filamentous fungi have a tendency to form thick mycelium broth or pellets that limit the flow of nutrients to all cells. Furthermore they tend to grow on surfaces of the cultivation vessels and require adequate stirring to avoid this. One example of successful cultivation in 24 deepwell microtiter plates involves the use of the filamentous bacterium *S.*

coelicolor (Sohoni et al., 2012). Sohoni *et al.* successfully established this system through the use of square well plates and glass beads increasing shearing considerable and enabling more dispersed cultivations (Sohoni et al., 2012).

Random mutagenesis and adaptive evolution methods are however being challenged by more targeted genetic modifications made possible by genome sequencing as well as method development in molecular biology. Secondly, over the past decade the development of genome scale metabolic networks (GSMN) and optimization algorithms has made a model guided approach to strain improvement possible.

In silico guided Metabolic Engineering

The definition of metabolic engineering was first given by James E. Bailey in 1991. He stated that "Metabolic engineering is the improvement of cellular activities by manipulation of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology" (Bailey, 1991). As technologies have matured metabolic engineering has evolved to use systems biology tools, where information at a systems level such as fluxome, transcriptome and metabolome are now being used as tools to guide beneficial modifications of a desired production system. The next step that is still only in its emerging tool building stage will be to use synthetic biology. Instead of modifying the existing cell factory synthetic biology aims to build the cellular parts and put them into a cell factory. This is thus a more bottom up oriented approach compared to the top down modifications coming from systems biology. The technologies that support each step of the metabolic engineering cycle have seen tremendous improvements over the years. There is a continuous strive to push each field forward to the point where designing cellular systems from scratch may be the ultimate goal.

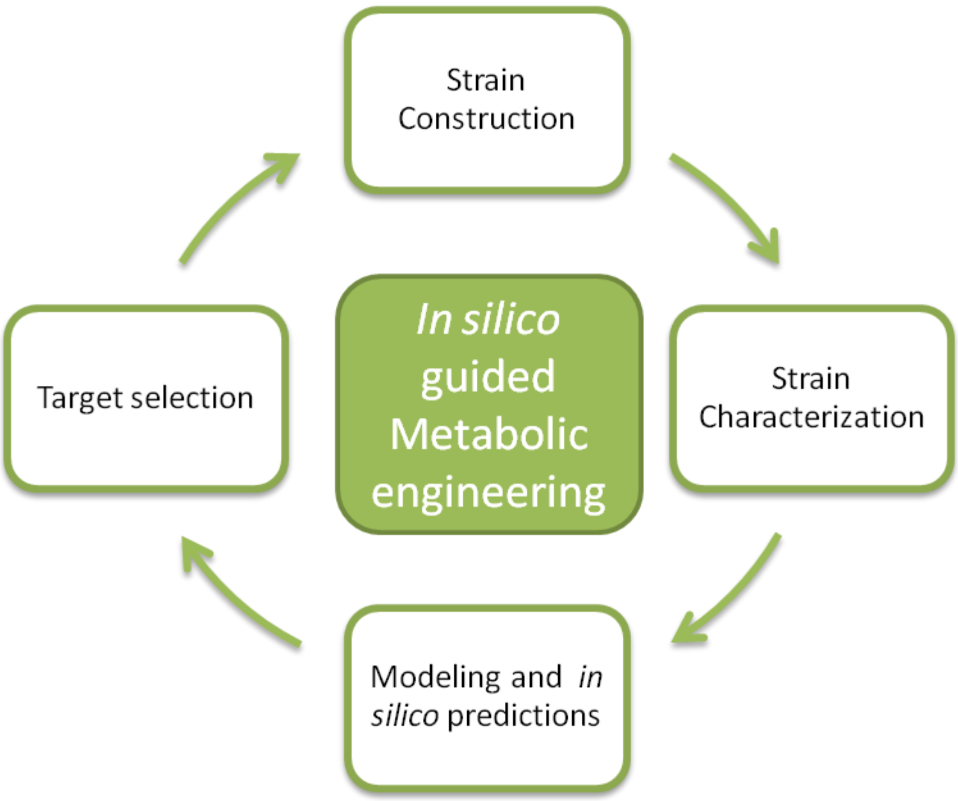


Figure 2 The steps needed to accomplish *in silico* guided metabolic engineering. The iterative process can be started at any point in the circle.

Strain Construction

The process of constructing a heterologous production host and subsequently optimizing it based on different metabolic engineering strategies is largely dependent on recombinant DNA technology. The first step in heterologous polyketide production involves the construction of vectors or gene targeting substrates for the genetic modification of the host cells or production strain. As this thesis involves the study of *Saccharomyces cerevisiae* and *Aspergillus nidulans* as polyketide hosts the focus of this section shall mainly be on the recent advances in tools available for these species.

For the expression of a specific polyketide gene cluster in a host organism, there are principally two ways to go. Either the polyketide cluster is integrated in the host organism genome or the strain is transformed with plasmids or autonomously replicating elements that will multiply within the host. The decision of which way to go is largely dependent on the tools at hand. However, ultimately the goal is to have a stable high expression of the polyketide gene clusters. Keeping in mind that extremely high expression levels will eventually limit cellular growth. Thus there is potentially a trade of between growth and productivity.

Vector based System

The availability of suitable cloning vectors is increasing rapidly. There is a research community effort to deposit many of these vectors and make them readily accessible in databases such as www.addgene.org and www.partsregistry.org (Vick et al., 2011). Several companies also sell cloning vectors and will synthesize any gene for you and clone it into your vector. Thus creating a vector albeit pricy can be done by a simple order with companies such as DNA 2.0 and GeneArt® (Invitrogen) (DNA 2.0, 2012). Some standardized cloning systems have also been established such as the Gateway™ system from Invitrogen that is commercially available.

One of the problems with using vectors in yeast is that the ability to sustain high copy numbers of the plasmids as cells that are not burdened by plasmids will tend to outgrow the plasmid carrying producer cells. In order to counteract this of course antibiotic resistance markers or autotrophic strains are used. A further problem with yeast 2µ plasmids is that despite having reported copy number from 50-100 in each cell the population often has only a few cells with high copy numbers and many with low copy numbers (Futcher and Cox, 1983; Futcher and Cox, 1984; Albertsen et al., 2011) . For a production process the addition of antibiotics and use of defined media compared to the cheaper complex media further makes plasmid based expression a more expensive method.

Integrative System

Vectors can also be used to simply generate the gene targeting substrate of interest. USER™ cloning is an excellent fast cloning technique for constructing gene targeting substrates of large genes (Nour-Eldin et al., 2006; Frandsen et al., 2008; Hansen et al., 2011).

USER™ cloning has been used successfully for the construction of a gene targeting substrates for the production of 5-methyl-orsolinic acid that is the polyketide part of mycophenolic acid produced in *Penicillium brevicompactum* (Hansen et al., 2011). The design of USER cloning fragments for inserting several genetic elements at once can be quite complicated. To this end an automatic primer generation program called PHUSER (Primer Help for USER) has been developed by Olsen *et al*. (Olsen et al., 2011).

Beyond the gene targeting substrate, it is important to select the right sites for insertion into the genome as some regions are more highly expressed then others. This was shown in a reporter beta-gal assay in *S. cerevisiae* where the difference between the lowest and highest expressed site was 8.7 fold (Bai Flagfeldt et al., 2009). In *A. nidulans* a study showed that there was no correlation between the gene copy number and expression of bovine chymosin, strongly indicating that integration site has an influence on expression level (Cullen et al., 1987). In *Aspergillus paraciticus* and *A. niger* it has also been demonstrated that there is a large variation in gene expression dependent on the integration site (Verdoes et al., 1995; Chiou et al., 2002).

To achieve high expression levels and insertion of entire biosynthetic clusters the ability to sequentially insert several genes is very important. However the insertion of many pathway genes into the genome is often laborious and time-consuming. Especially if more than one copy of the gene is needed for sufficient production levels. Therefore the development of new methods to meet this end is imperative. It is important to be able to insert several copies of the genes of interest into the genome. Beyond this if a full pathway is inserted, it will be important to check if any enzyme constitutes a rate limiting step and its copy number should be amplified.

There are a few recent examples of gene amplification in both bacteria and yeast. In *S. coelicolor* the gene cluster of 22 genes for the production of the polyketide actinorhodin was amplified to 4-12 tandem copies by recombination (Murakami et al., 2011). This resulted in a 20-fold increase in the actinorhodin producing strain compared to the parental strain (Murakami et al., 2011). Another example is the DNA assembler that allows for the in vivo assembly of pathways in yeast or plasmids (Shao et al., 2009). The strategy named “Reiterative Recombination” shows another excellent example of how not only to insert a single gene into *S. cerevisiae*, but efficiently insert an entire pathway while allowing for the construction of pathway libraries (Wingler and Cornish, 2011). In addition, the stable integration in 10-11 copies from an integrative plasmid was achieved in the *Yarrowia lipolytica* genome using retrotransposons as integration sites (Juretzek et al., 2001).

From a production perspective the ability to integrate genes present the advantage of avoiding applying auxotrophic or antibiotic selection pressure on the system. The antibiotics represent a problem in purification processes as well as adding to the expense of the media. The use of recyclable markers such as *URA* (*S. cerevisiae*) and *PyrG* (*A. nidulans*) that allow marker excision by direct repeat recombination 5-Fluoroorotic acid (5-FOA) makes it further possible to repeatedly apply new modifications to the production host. Both *S. cerevisiae* and *A. nidulans* have the added advantage that it is possible to cross the strains. This makes the interchange of markers or traits among strains readily achievable.

Strain Characterization

After the construction of production strains they must be tested for productivity and yields in a reproducible growth environment that is also scale-able. Furthermore, the conditions under which the optimal production occurs must be found through testing of several possible options (Xie, 2012).

The typical flow in a strain evaluation program is illustrated in figure 3 and can be described as follows. First large scale screening occurs of thousands of strains in microtiter plates or test tubes. Beyond this there is a screening of hundreds of strains in shake flask or microfermentors. Afterwards dozens of strain candidates are evaluated in bench-scale fermentors. In the end the best strain will be evaluated in pilot or commercial-scale fermentors. The strains selected will also be optimized in different media as well as optimum pH and pO_2 (Xie, 2012).

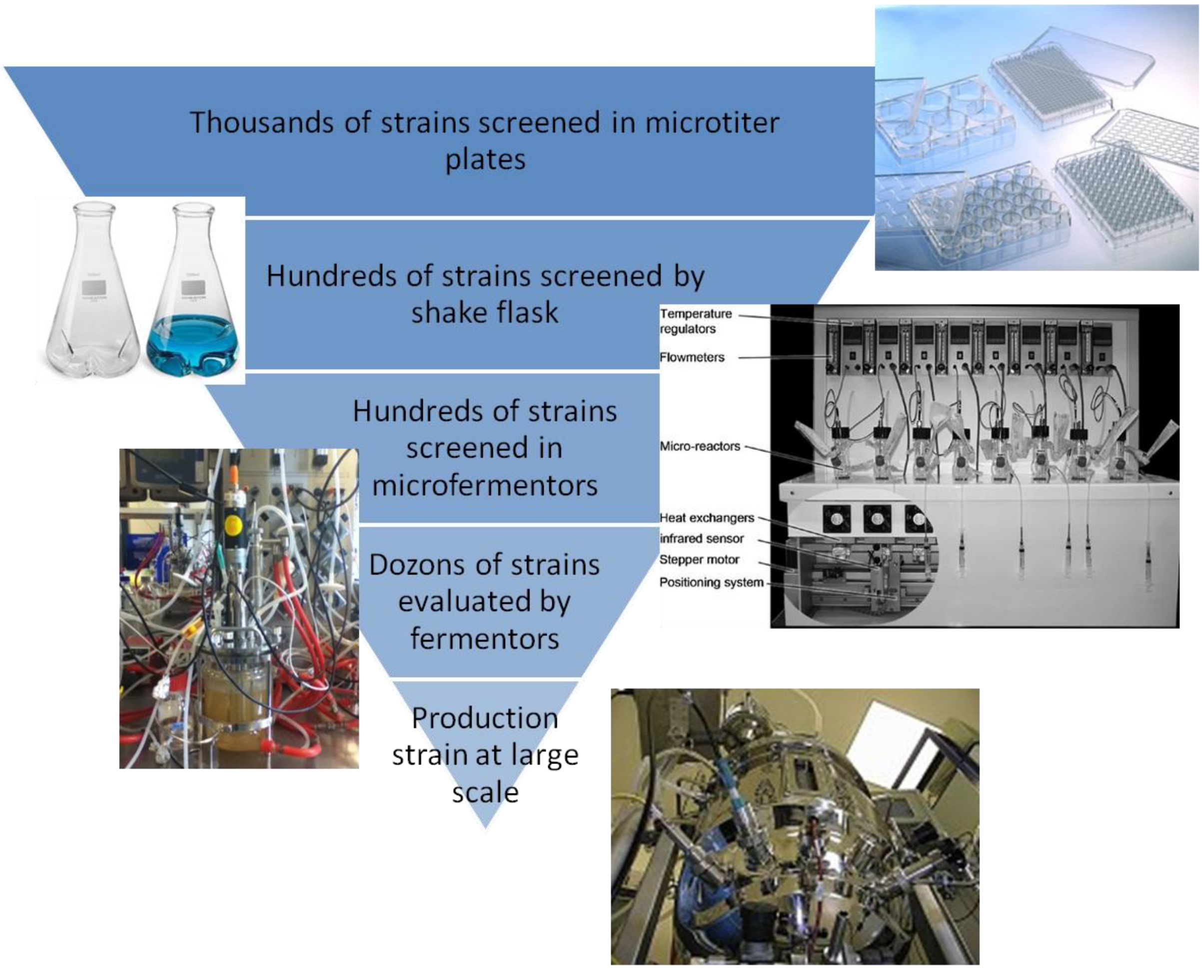


Figure 3 The different steps of a strain evaluation program typically used. The figure is modified from (Xie, 2012) and microbioreactors are from Frachon *et al.* (Frachon et al., 2006).

In silico guided Target Identification

The field of bioinformatics has generated many strain modeling frameworks to predict cellular behavior and aid our understanding of biology. The main focus of this text is the algorithms build to aid metabolic engineers in their quest for the construction of the best possible production strain. *In silico* guided identification of targets for genetic engineering involves several steps as outlined in figure 4. The steps will be discussed in the following sections.

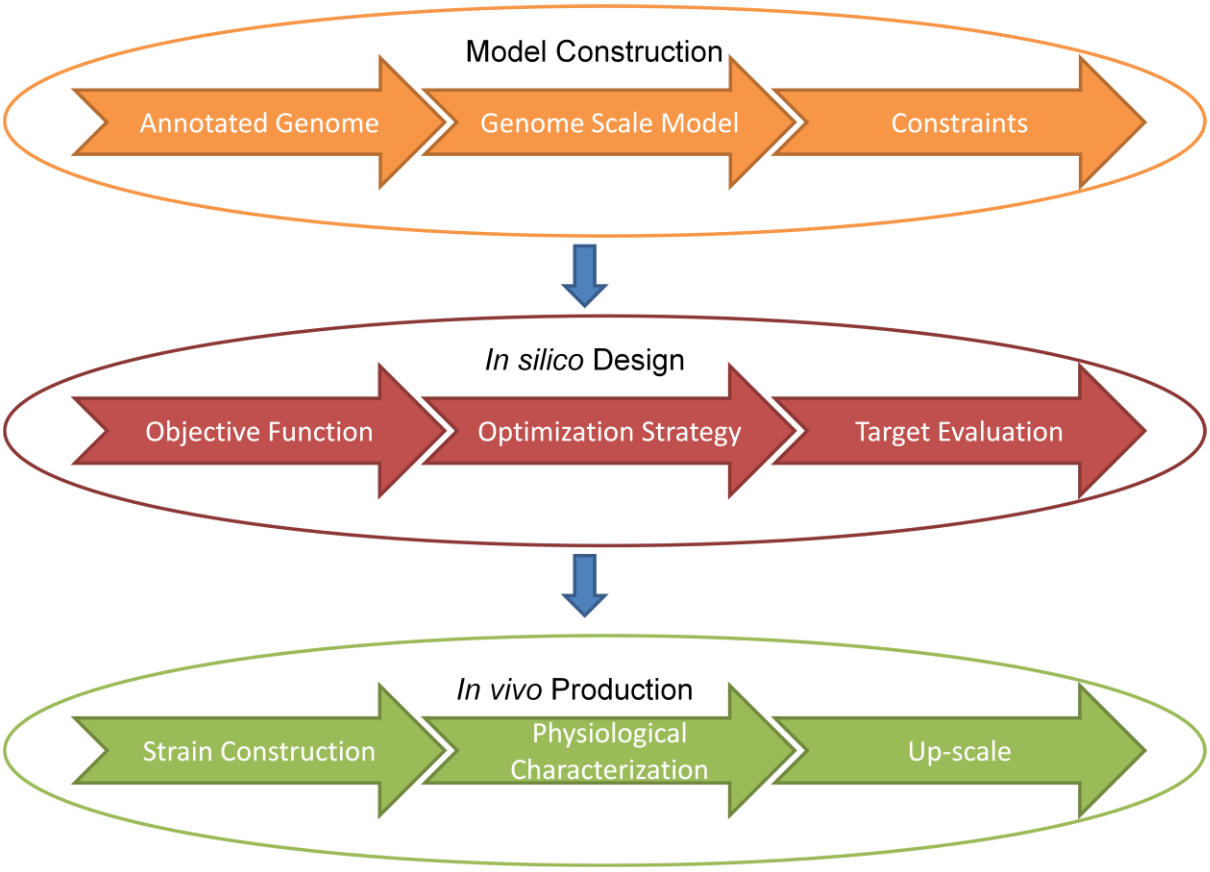


Figure 4 The steps that are needed for a model guided metabolic engineering approach. First one must construct a genome scale model where it is helpful to have an annotated genome. Then one must chose the objective such as biomass or product that one needs to obtain. Beyond this an optimization algorithm is chosen for predictions and the relevant targets evaluated. Lastly the improvements are constructed and tested *in vivo*.

Model Construction

The first step towards model construction has been the genome sequencing of many industrially relevant microorganisms. To date more than 3079 genomes have been sequenced and published (Genomes Online Database, 2012). Furthermore comparative genomics has made it possible to annotate full genomes very efficiently using e.g. BLAST and FASTA programs (Covert et al., 2001). With the full coverage of a species annotated genome and thus to a great extent enzymes and reactions, the next step has been to build genome scale metabolic networks (GSMN). An early review on how to build a genome scale model was by Covert et *al.* in 2001 describing the basic build-up of a model (Covert et al., 2001). It is interesting to see the tremendous development of modeling that has succeeded this paper in only a decade. The construction of GSMN has increased quite rapidly and has been completed for more than 35 organisms (Orth et al., 2010). In 2010, Thiele and Palsson designed a protocol that describes how to build a metabolic reconstruction and gives guidelines for its refinement (Thiele and Palsson, 2010). This procedure includes an algorithm called metaSHARK that automatically generates a draft model thus saving considerable time and effort compared to building a model with only literature as a guide (Pinney et al., 2005). Of course researchers building these models in general have also looked at the models that already existed and used them as an inspiration and frame upon which to build new models of an organisms metabolism. One must also take the time to manually curate the model as not all genes have homologs in other organisms and substrate specificities might be different. Thereby experimentally determined data sets such as C-13 fluxomics and transcriptomics are needed. Examples of genome scale models of microorganisms of which most are industrially relevant are given in table 1.

Table 1 Genome scale models. When more than one genome scale model exists for the same strain the first and most recent models are given.

	Total Reactions (unique)	Total Metabolites	Total genes	Model Id	Reference
E. coli (2000)	627	438	660	*i*JE660	(Edwards and Palsson, 2000)
E.coli (2011)	2,251	1,136	1,366	*i*JO1366	(Orth et al., 2011)
S. cerevisiae (2003)	1,175 (842)	-	708	*i*FF708	(Förster et al., 2003)
S. cerevisiae (2008)	1,761	1,168	-	-	(Herrgård et al., 2008)
A. nidulans (2008)	1,095 (681)	738	666	*i*HD666	(David et al., 2008)
A. niger (2008)	2,240 (1,190)	1,045 (782)	871	*i*MA871	(Andersen et al., 2008)
A. oryzae (2008)	(1,679)	1,040	1,184	*i*WV1184	(Vongsangnak et al., 2008)

The degree at which these models have actually been studied and validated varies in part due to the timeline in which they were developed. It is clear that the unicellular *E. coli* and *S. cerevisiae* are the most widely used microorganisms in research labs and thus modelling and validation of these far succeeds that of the *Aspergilli* models. The models of *E. coli* and *S. cerevisiae* have been updated and extended quite a lot (Herrgård et al., 2008; Orth et al., 2011). Furthermore, it must be assumed that the number of genes and reactions occurring in *A. nidulans*, *A. niger* and *A. oryzae* exceeds that of *E. coli* and *S. cerevisiae*. However, due to less time afforded the build up and extension of *Aspergilli* models, the number of reactions and compartments covered in the *Aspergilli* models is still less than those for *E. coli* and *S. cerevisiae*. Thus *E. coli* and *S. cerevisiae* GSMN represent a larger part of metabolism than *Aspergilli* models. An interesting approach to further validation and optimization of a model as well as identifying species-specific reactions has been the comparative systems analysis of two different models (in this case *Pseudomonas*) to correct and verify the reactions in the individual models (Oberhardt et al., 2011).

***In silico* Guidance**

A model that does nothing more than be a quantitative detailed description of a system does not give us more knowledge than what was used to build the model. It may put the knowledge together, but it must be combined with skilled predictive power that can tell us not only what is in the system, but also how it will behave if perturbed. The model is not more than the index of a book as were the parts of it the chapters. In engineering the goal is not only to build a model of a system, but to push the system to its limits. Metabolic engineering occurs through the observation of a system and predicting its potential to aid in improving the productivity of substance. Thus genome scale models have to be combined with predictive algorithms to aid the construction of improved production strains.

To this end quite a lot of software suites have seen the light of day. Among the most well established whole cell analysis tools available are the CellNetAnalyzer (Klamt and von Kamp, 2011) , the BioMet Toolbox (Cvijovic et al., 2010) and the COBRA Toolbox (Schellenberger et al., 2011) . A recent extension that adds expression, transcription and regulatory information to the COBRA toolbox is TIGER (Jensen et al., 2011).

Prediction of Biomass Formation

A timeline of the development of the different algorithms currently used for *in silico* guided metabolic engineering is presented in figure 5. Common to all metabolic engineering algorithms is an underlying mathematical optimization problem (Banga, 2008). The basic part of most metabolic engineering algorithms involves designing an objective function upon which the quality of the solution is decided. In most metabolic engineering strategies the objective is growth or product formation. Secondly, the model is confined by a set of constraints that limit the solution space. First of all the constraints are based on stoichiometry and direction of a given reaction as well as reversibility. In metabolic models there are then further constraints or bounds on the uptake of substrates such as glucose and oxygen.

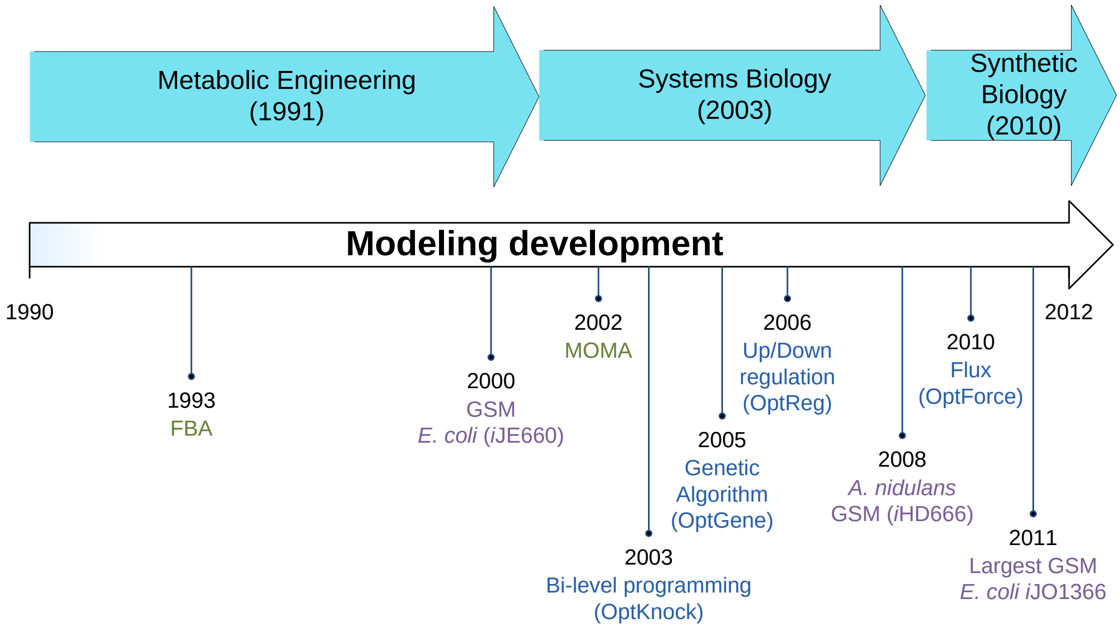


Figure 5 The figure illustrates the metabolic engineering algorithm development over the last 20 years. In purple are examples of when *E. coli* and *A. nidulans* genome scale models were constructed. In green are given the most common algorithms for prediction of growth behavior. In blue are given some of the algorithms that can be used to predict product formation and guide metabolic engineering strategies.

The most commonly used method is flux balance analysis (FBA) (Varma and Palsson, 1994; Orth et al., 2010). FBA is a linear programming based method that allows prediction of flux distribution in a given cell based on mass balance constraints (often at the GSMM scale) and a biological objective function. FBA is often used in conjunction with constraints on the uptake of nutrients and maximization of biomass formation (or growth) as the objective function. Thus, one can get a quantitative description of predicted fluxes in cellular metabolism. FBA has been very effective in accurately predicting the lethality of single gene knockouts (i.e. no biomass formation) (Förster et al., 2003; David et al., 2008). It has also been applied quite effectively for metabolic engineering along with several complementary algorithms (Curran et al., 2012). However, when it comes to predicting the growth rate and behavior of knockout strains, the later developed algorithm Minimization of metabolic adjustment (MOMA) has been found to be more effective in some systems including yeast (Segre, 2002; Brochado, 2010). MOMA uses quadratic programming to minimize the Euclidean distance between the fluxes observed at steady state before and after deletion of a specific gene (/s). It is basically a least squares minimization problem, in some regards similar to linear regression, only applied to a multidimensional space. The biological hypothesis behind MOMA is that an organism will try to counteract the effects of a gene knockout by adjusting cellular metabolic fluxes back towards the initial unperturbed state (Segre et al., 2002). Lastly, an algorithm called regulatory On/Off Minimization (ROOM) of metabolic flux changes tries instead to minimize the number of significant flux changes resulting from a knockout, but is otherwise quite similar to MOMA (Shlomi et al., 2005).

Product Optimization Strategies

FBA, MOMA and ROOM however mainly focus on describing the growth potential of a specific network (wild-type or perturbed). For the application in metabolic engineering the next level of programming frameworks needed to take into account the formation of the desired product competing with biomass formation for the substrate (/s). Thus, bilevel optimization algorithms had to be developed that could take into account both growth and product formation as these are often competing factors. The first development in this area was the OptKnock framework (Burgard et al., 2003). In this framework the cellular objective (growth) is optimized within an industrial objective (product formation) optimization problem. The algorithm was initially applied to *E. coli* and identified several strategies for overproducing succinate, lactate and 1,3 propanediol.

Industrially relevant objectives for strain improvement are often non-linear and linear programming based algorithms are of limited use in such cases. To address this problem, by using FBA or MOMA as scoring algorithms, a genetic algorithm using evolutionary search procedures for solving the combinatorial optimization problems was developed called OptGene (Patil et al., 2005). In the OptGene routine, the model is first preprocessed removing all lethal deletions from the solution space. Then for an initial random population of deletion strains the fitness, i.e. the design objective function (often productivity or biomass product coupled yield (BPCY)), is evaluated by using FBA or MOMA as a biological objective function. The best individuals are then crossed and a new population generated. In this way OptGene searches for optimal solution. Although not guaranteed, OptGene has been proven by a limited number of iterations to converge towards a global optimal solution in a number of cases.

Another algorithm useful for *in silico* metabolic engineering is OptStrain that can be used to find the optimal pathway for the conversion of a substrate to a new product in a host microorganism (Pharkya et al., 2004). This is achieved by searching online databases for all possible reactions to be inserted while minimizing the use of foreign reactions.

The next step in terms of modeling has been to expand the algorithms not only to include a present/absent state of the reaction in a sort of on/off manor, but also to attempt predicting effects of up or down regulations of different genes. One example of such algorithms is OptReg (Pharkya and Maranas, 2006). Currently, such methods are limited due to the lack of comprehensive knowledge and models for *in vivo* enzyme kinetics and regulation. On the experimental front, it should be noted that controlled regulation of enzyme expression requires well-established tools, e.g. promoter libraries that exist for *S. cerevisiae* and *E. coli* (Nevoigt et al., 2006; Maertens and Vanrolleghem, 2010; Babiskin and Smolke, 2011). Such tools have not yet been published in the case of *A. nidulans*.

The most recent addition to the prediction algorithms family is OptForce that relies on exploitation of complex flux couplings in the network rather than kinetic models (Ranganathan et al., 2010). OptForce has several advantages over gene knockout prediction algorithms in that it predicts combinations of gene knockout and flux overexpression targets, which are likely to result in more improvements per modification.

An interesting approach to predict the yield of chemical biosynthesis that is not based on genome scale model has been published (Varman et al., 2011). The predictions are based on a first order mathematical model containing parameters such as cultivation mode, oxygen, nutritional sources and length of the pathway all relevant to biological production (Varman et al., 2011). The model used data from 40 papers with metabolic engineering of *S. cerevisiae* and can be used for an initial yield prediction (Varman et al., 2011). They conclude that the influence of the model parameters such as pathway length and cultivation conditions should be kept in mind when new metabolic engineering strategies of *S. cerevisiae* are designed.

Examples of *in silico* guided Metabolic Engineering

The optimization algorithms OptKnock and OptGene along with minimization of metabolic adjustment (MOMA) have recently been used to predict the improvement of the production of vanillin and the sesquiterpene cubebol. Brochado *et al.* used a model-guided approach with Optgene and MOMA to predict the target deletion of *pdc1, gdh1* for over-production of vanillin-glucoside. When this was combined with an overexpression of *GDH2*, the yield of vanillin glucoside on glucose was improved 1.5 times (Brochado et al., 2010). In the case of cubebol the deletion of *gdh1* let to an 85% increase in cubebol titers (Asadollahi et al., 2009). Another similar study for the production of lycopene in *E. coli* predicted the deletion of *gdhA, gpmA* and *gpmB* (Alper et al., 2005a). In case of the deletion of *gdhA* the study revealed an increase in lycopene production of 13% albeit a decrease in growth rate to 82% of the wildtype growth rate (Alper et al., 2005a). The predicted deletions have been patented for carotene production in bacteria (Stephanopoulos et al.,). In a further study they combined their modeling targets with transposon mutagenesis generating a total of 64 mutants one of which ended up with an 8.5 times increase in lycopene production compared to the wildtype strain (Alper et al., 2005b). The authors note that combining the randomly generated transposon targets yields a decrease in lycopene production thus making these targets more of a one shot solution. Whereas the model guided approach with a total of three deletions shows a stepwise increase. Thus leading to the conclusion that for a continuous improvement of strains, a model guided approach has an advantage (Alper et al., 2005b).

Improvement of the Polyketide Cell Factories

There have been many attempts to improve the production of polyketides in different microorganisms. The tetracycline pathway has been improved various ways both in industrial settings with random mutagenesis and by gene targeting through a metabolic engineering approach (Pickens and Tang, 2009). Another example is the 5.6 fold titer improvement of the polyketide fredericamycin in *Streptomyces griseus* through the overexpression of a transcription factor (Chen et al., 2008).

The algorithm OptForce was used to increase the flux towards malonyl-CoA in order to improve the production of the flavanone naringenin (Xu et al., 2011). The model data is very well documented on several mutants in *E. coli*.

Another study modeled the production of the polyketide part of erythromycin known as 6-deoxyerythronolide B (6-dEB) in three different organisms *E. coli*, *Bacillus subtilis* and *S. cerevisiae* using MoMA (Boghigian et al., 2010). In *E. coli* the top scoring targets for deletion were the succinate dehydrogenase complex (isdhABCD), succinyl-CoA synthetase (*sucCD)* and the glutamate dehydrogenase gene *(gdhA*). In *B. subtilis* the highest scoring candidates were α-ketoglutarate dehydrogenase (*citK*) and *sucCD.* In *S. cerevisiae* the authors find that succinyl-CoA ligase (LCS1, LCS2) were the top candidates for knockout studies. However, these results remain to be validated *in vivo*.

As seen from the above examples many of the targets for the improvement of polyketide production through model predictions are focused on the improved generation of NADPH and malonyl-CoA. As precursors of polyketides these target metabolites are important for the improved polyketide production.

Improvement of 6-MSA Production

The production of 6-MSA has been increased in the heterologous host *S. cerevisiae* by overexpression of the *ACC1* gene responsible for the conversion of acetyl-CoA to the polyketide precursor malonyl-CoA (Wattanachaisaereekul et al., 2008). It was also found that the use of ethanol as a carbon source increased the production of 6-MSA, which is expected to be a result of its conversion into acetyl-CoA that also generates NADPH (Wattanachaisaereekul et al., 2007). Another study on 6-MSA production in *A. nidulans* tested the different carbon sources xylose, glycerol and ethanol, but did not find any increase in productivity compared to glucose (Panagiotou et al., 2009).

In conclusion the construction of cell factories and optimization of these is a complex procedure. It involves the knowledge of molecular biology, bioreactor operations and evaluation as well as the use of large scale metabolic models for evaluating and predicting cellular behavior.

References

Adrio, J.L., Demain, A.L., 2006. Genetic improvement of processes yielding microbial products. FEMS Microbiol. Rev. 30, 187-214.

Albertsen, L., Chen, Y., Bach, L.S., Rattleff, S., Maury, J., Brix, S., Nielsen, J., Mortensen, U.H., 2011. Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes. Appl. Environ. Microbiol. 77, 1033-40.

Alper, H., Jin, Y.S., Moxley, J.F., Stephanopoulos, G., 2005a. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab. Eng. 7, 155-64.

Alper, H., Miyaoku, K., Stephanopoulos, G., 2005b. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat. Biotechnol. 23, 612-6.

Andersen, M.R., Nielsen, M.L., Nielsen, J., 2008. Metabolic model integration of the bibliome, genome, metabolome and reactome of Aspergillus niger. Molecular systems biology 4.

Asadollahi, M.A., Maury, J., Patil, K.R., Schalk, M., Clark, A., Nielsen, J., 2009. Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through *in silico* driven metabolic engineering. Metab. Eng. 11, 328-34.

Babiskin, A.H., Smolke, C.D., 2011. A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. Molecular systems biology 7.

Bai Flagfeldt, D., Siewers, V., Huang, L., Nielsen, J., 2009. Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. Yeast 26, 545-51.

Bailey, J.E., 1991. Toward a science of metabolic engineering. Science 252, 1668.

Bailey, J.E., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W.A., Tsai, P.S., 1996. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. Biotechnol. Bioeng. 52, 109-21.

Banga, J., 2008. Optimization in computational systems biology. BMC systems biology 2, 47.

Betts, J., Baganz, F., 2006. Miniature bioreactors: current practices and future opportunities. Microbial Cell Factories 5, 21-.

Boghigian, B.A., Lee, K., Pfeifer, B.A., 2010. Computational analysis of phenotypic space in heterologous polyketide biosynthesis--Applications to *Escherichia coli*, *Bacillus subtilis*, and *Saccharomyces cerevisiae*. J. Theor. Biol. 262, 197-207.

Brochado, A.R., Matos, C., Møller, B.L., Hansen, J., Mortensen, U.H., Patil, K.R., 2010. Improved vanillin production in baker's yeast through in silico design. Microbial Cell Factories 9, 84.

Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. Biotechnology and Bioengineering, 84, 647-57.

Chen, Y., Wendt-Pienkowski, E., Shen, B., 2008. Identification and utility of FdmR1 as a Streptomyces antibiotic regulatory protein activator for fredericamycin production in *Streptomyces griseus* ATCC 49344 and heterologous hosts. J. Bacteriol. 190, 5587-96.

Chiou, C.H., Miller, M., Wilson, D.L., Trail, F., Linz, J.E., 2002. Chromosomal location plays a role in regulation of aflatoxin gene expression in *Aspergillus parasiticus*. Appl. Environ. Microbiol. 68, 306-15.

Covert, M.W., Schilling, C.H., Famili, I., S.Edwards, J., I.Goryanin, I., Selkov, E., Palsson, B.,O., 2001. Metabolic modeling of microbial strains in silico. TRENDS in Biochemical Sciences 26, 179-86.

Cullen, D., Gray, G.L., Wilson, L.J., Hayenga, K.J., Lamsa, M.H., Rey, M.W., Norton, S., Berka, R.M., 1987. Controlled expression and secretion of bovine chymosin in *Aspergillus nidulans*. Bio/Technology 5, 369-76.

Curran, K.A., Crook, N.C., Alper, H.S., 2012. Using flux balance analysis to guide microbial metabolic engineering. Methods in molecular biology (Clifton, NJ) 834, 197.

Cvijovic, M., Olivares-Hernandez, R., Agren, R., Dahr, N., Vongsangnak, W., Nookaew, I., Patil, K.R., Nielsen, J., 2010. BioMet Toolbox: genome-wide analysis of metabolism. Nucleic Acids Res. 38, W144-9.

David, H., Özçelik, İ.Ş., Hofmann, G., Nielsen, J., 2008. Analysis of *Aspergillus nidulans* metabolism at the genome-scale. BMC Genomics 9, 163.

DNA 2.0, 2012. www.dna20.com. Accessed 3rd of February 2012.

Duetz, W.A., Witholt, B., 2004. Oxygen transfer by orbital shaking of square vessels and deepwell microtiter plates of various dimensions. Biochem. Eng. J. 17, 181-5.

Duetz, W.A., 2007. Microtiter plates as mini-bioreactors: miniaturization of fermentation methods. Trends Microbiol. 15, 469-75.

Duetz, W.A., Ruedi, L., Hermann, R., O'Connor, K., Buchs, J., Witholt, B., 2000. Methods for Intense Aeration, Growth, Storage, and Replication of Bacterial Strains in Microtiter Plates. Appl. Environ. Microbiol. 66, 2641-6.

Duetz, W.A., Witholt, B., 2001. Effectiveness of orbital shaking for the aeration of suspended bacterial cultures in square-deepwell microtiter plates. Biochem. Eng. J. 7, 113-5.

Edwards, J., Palsson, B., 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proceedings of the National Academy of Sciences 97, 5528.

Förster, J., Famili, I., Fu, P., Palsson, B.Ø., Nielsen, J., 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 13, 244-53.

Frachon, E., Bondet, V., Munier-Lehmann, H., Bellalou, J., 2006. Multiple microfermentor battery: a versatile tool for use with automated parallel cultures of microorganisms producing recombinant proteins and for optimization of cultivation protocols. Appl. Environ. Microbiol. 72, 5225-31.

Frandsen, R.J.N., Andersson, J.A., Kristensen, M.B., Giese, H., 2008. Efficient four fragment cloning for the construction of vectors for targeted gene replacement in filamentous fungi. BMC molecular biology 9, 70.

Funke, M., Diederichs, S., Kensy, F., Müller, C., Büchs, J., 2009. The baffled microtiter plate: increased oxygen transfer and improved online monitoring in small scale fermentations. Biotechnol. Bioeng. 103, 1118-28.

Futcher, A., Cox, B., 1984. Copy number and the stability of 2-micron circle-based artificial plasmids of *Saccharomyces cerevisiae.*. J. Bacteriol. 157, 283-90.

Futcher, A., Cox, B., 1983. Maintenance of the 2 microns circle plasmid in populations of *Saccharomyces cerevisiae*. J. Bacteriol. 154, 612-22.

Genomes Online Database, 2012. http://www.genomesonline.org/cgi-bin/GOLD/index.cgi?page_requested=Complete+Genome+Projects. Accessed 24th of January 2012.

Guimarães, P.M.R., François, J., Parrou, J.L., Teixeira, J.A., Domingues, L., 2008. Adaptive evolution of a lactose-consuming *Saccharomyces cerevisiae* recombinant. Appl. Environ. Microbiol. 74, 1748-56.

Hansen, B.G., Salomonsen, B., Nielsen, M.T., Nielsen, J.B., Hansen, N.B., Nielsen, K.F., Regueira, T.B., Nielsen, J., Patil, K.R., Mortensen, U.H., 2011. A versatile gene expression and characterization system for Aspergillus: heterologous expression of the gene encoding the polyketide synthase from the mycophenolic acid gene cluster from *Penicillium brevicompactum* as a case study. Appl. Environ. Microbiol. .

Herrgård, M.J., Swainston, N., Dobson, P., Dunn, W.B., Arga, K.Y., Arvas, M., Büthgen, N., Borger, S., Costenoble, R., Heinemann, M., 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat. Biotechnol. 26, 1155-60.

Ibarra, R.U., Edwards, J.S., Palsson, B.O., 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. Nature 420, 186-9.

Jensen, P., Lutz, K., Papin, J., 2011. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. BMC Systems Biology 5, 147.

Juretzek, T., Le Dall, M.T., Mauersberger, S., Gaillardin, C., Barth, G., Nicaud, J.M., 2001. Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. Yeast 18, 97-113.

Klamt, S., von Kamp, A., 2011. An application programming interface for CellNetAnalyzer. BioSystems .

Maertens, J., Vanrolleghem, P.A., 2010. Modeling with a view to target identification in metabolic engineering: A critical evaluation of the available tools. Biotechnol Progress 26, 313-31.

Murakami, T., Burian, J., Yanai, K., Bibb, M.J., Thompson, C.J., 2011. A system for the targeted amplification of bacterial gene clusters multiplies antibiotic yield in *Streptomyces coelicolor*. Proceedings of the National Academy of Sciences 108, 16020-5.

Nevoigt, E., Kohnke, J., Fischer, C.R., Alper, H., Stahl, U., Stephanopoulos, G., 2006. Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 72, 5266-73.

Nour-Eldin, H.H., Hansen, B.G., Nørholm, M.H.H., Jensen, J.K., Halkier, B.A., 2006. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res. 34, e122-.

Oberhardt, M.A., Puchałka, J., dos Santos, V.A.P.M., Papin, J.A., 2011. Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. PLoS computational biology 7, e1001116.

Olsen, L.R., Hansen, N.B., Bonde, M.T., Genee, H.J., Holm, D.K., Carlsen, S., Hansen, B.G., Patil, K.R., Mortensen, U.H., Wernersson, R., 2011. PHUSER (Primer Help for USER): a novel tool for USER fusion primer design. Nucleic Acids Res. 39, W61.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.Ø., 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. Molecular systems biology 7.

Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28, 245-8.

Otero, J.M., Nielsen, J., 2010. Industrial systems biology. Biotechnol. Bioeng. 105, 439-60.

Otero, J.M., 2009. Industrial Systems Biology and Metabolic Engineering of *Saccharomyces cerevisiae* A case study in succinic acid production.

Panagiotou, G., Andersen, M.R., Grotkjaer, T., Regueira, T.B., Nielsen, J., Olsson, L., 2009. Studies of the production of fungal polyketides in *Aspergillus nidulans* by using systems biology tools. Appl. Environ. Microbiol. 75, 2212-20.

Patil, K.R., Rocha, I., Förster, J., Nielsen, J., 2005. Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinformatics 6.

Pfeifer, B.A., Khosla, C., 2001. Biosynthesis of Polyketides in Heterologous Hosts. Microbiol. Mol. Biol. Rev. 65, 106-18.

Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. OptStrain: a computational framework for redesign of microbial production systems. Genome Res. 14, 2367-76.

Pharkya, P., Maranas, C.D., 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab. Eng. 8, 1-13.

Pickens, L.B., Tang, Y., 2009. Decoding and engineering tetracycline biosynthesis. Metab. Eng. 11, 69-75.

Pinney, J.W., Shirley, M.W., McConkey, G.A., Westhead, D.R., 2005. metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. Nucleic Acids Res. 33, 1399-409.

Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS computational biology 6, e1000744.

Rowlands, R., 1984. Industrial strain improvement: mutagenesis and random screening procedures. Enzyme Microb. Technol. 6, 3-10.

Santos, C.N.S., Stephanopoulos, G., 2008. Combinatorial engineering of microbes for optimizing cellular phenotype. Curr. Opin. Chem. Biol. 12, 168-76.

Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. Nature Protocols 6, 1290-307.

Segre, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. PNAS Proceedings of the National Academy of Sciences of the United States 99, 15112-7.

Selifonova, O., Valle, F., Schellenberger, V., 2001. Rapid evolution of novel traits in microorganisms. Appl. Environ. Microbiol. 67, 3645-9.

Shao, Z., Zhao, H., Zhao, H., 2009. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 37, e16-.

Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory On/Off Minimization Of Metabolic Flux Changes Following Genetic Perturbations. Biological Sciences: Microbiology PNAS 102, 7695-700.

Sohoni, S.V., Bapat, P.M., Lantz, A.E., 2012. Robust, small-scale cultivation platform for *Streptomyces coelicolor*. Microbial Cell Factories 11, 9.

Stephanopoulos, G., Alper, H.S., Jin, Y., New genetically engineered bacterium having a disruption in one or more gdhA, gpmB aceE, ppc, talB, fdhF, yjiD, hnr or yjfP genes, useful for overexpressing genes impacting carotene biosynthesis and enhancing carotenoid production.

Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature protocols 5, 93-121.

Thykaer, J., Nielsen, J., 2003. Metabolic engineering of beta-lactam production. Metab. Eng. 5, 56-69.

Varma, A., Palsson, B.,O., 1994. Metabolic flux balancing: Basic Concepts, Scientific and Practical Use. Nature Biotechnology 12, 994-8.

Varman, A.M., Xiao, Y., Leonard, E., Tang, Y.J., 2011. Statistics-based model for prediction of chemical biosynthesis yield from *Saccharomyces cerevisiae*. Microbial cell factories 10, 45.

Verdoes, J.C., Punt, P., Hondel, C., 1995. Molecular genetic strain improvement for the overproduction of fungal proteins by filamentous fungi. Appl. Microbiol. Biotechnol. 43, 195-205.

Vick, J.E., Johnson, E.T., Choudhary, S., Bloch, S.E., Lopez-Gallego, F., Srivastava, P., Tikh, I.B., Wawrzyn, G.T., Schmidt-Dannert, C., 2011. Optimized compatible set of BioBrick™ vectors for metabolic pathway engineering. Appl. Microbiol. Biotechnol. , 1-12.

Vongsangnak, W., Olsen, P., Hansen, K., Krogsgaard, S., Nielsen, J., 2008. Improved annotation through genome-scale metabolic modeling of *Aspergillus oryzae*. BMC Genomics 9, 245.

Wattanachaisaereekul, S., Lantz, A.E., Nielsen, M.L., Nielsen, J., 2008. Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. Metab. Eng. 10, 246-54.

Wattanachaisaereekul, S., Lantz, A.E., Nielsen, M.L., Andresson, O.S., Nielsen, J., 2007. Optimization of Heterologous Production of the Polyketide 6-MSA in *Saccharomyces cerevisiae*. Biotechnology and Bioengineering 97, 893-900.

Wingler, L.M., Cornish, V.W., 2011. Reiterative Recombination for the in vivo assembly of libraries of multigene pathways. Proceedings of the National Academy of Sciences 108, 15135-40.

Wittmann, C., Kim, H.M., John, G., Heinzle, E., 2003. Characterization and application of an optical sensor for quantification of dissolved O_2 in shake-flasks. Biotechnol. Lett. 25, 377-80.

Wright, B.E., 2004. Stress-directed adaptive mutations and evolution. Mol. Microbiol. 52, 643-50.

Xie, D., 2012. Using an advanced microfermentor system for strain screening and fermentation optimization. Methods in molecular biology (Clifton, NJ) 834, 217.

Xu, P., Ranganathan, S., Fowler, Z.L., Maranas, C.D., Koffas, M.A.G., 2011. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab. Eng. .

Chapter 3

Structure and Function of Polyketide Synthases

In this chapter the group of natural compounds known as polyketides will be described. This chapter does not aim to be a complete description of these fascinating compounds, but to give the reader a fundamental knowledge that will aid the understanding of the underlying mechanisms that are important to the research conducted in this thesis. Thus it will mainly focus on the polyketide building blocks derived from primary metabolism and how these form a final polyketide. Information that is essential for any optimization strategy.

Polyketides

Polyketides are natural products made from acyl-CoA monomers that are connected through several rounds of condensation. The monomers typically include acetyl-, malonyl- and propionyl-CoA. Polyketides are produced naturally in fungi, bacteria and plants. The term polyketide was first coined in 1907 by the British scientist John Norman Collie (1859-1942) (Collie, 1907; Bentley and Bennett, 1999; Bentley, 1999). He actually used the word ketene in his paper, but later used the word ketide in his lectures. He discovered that the compounds were made of units by the general formula $H[CH_2\text{-}CO]_x$. . Polyketides come in many different sizes from the very small triacetic lactone with only 6-8 carbons to the very big maitotoxin with 164 carbons in the chain (Weissman, 2009).

Biosynthesis of Polyketides

Polyketides are produced by large enzymes known as polyketide synthases (PKS). The polyketide synthases are made up of a set of domains that each serves a specific catalytic function. The minimal PKS contains an acyl transferase domain (AT) responsible for selecting the acyl-CoA starter unit for incorporation into the polyketide chain (Tsai and Ames, 2009). The tethering domain (T) or acyl carrier protein (ACP) is the subunit where the growing PK chain is attached. The ketosynthase (KS) or condensation domain is responsible for the Claisen condensation that connects two acyl-CoA substrates tethered to the T and KS-domain by use of a conserved cysteine residue (Tsai and Ames, 2009). Beyond this many PKSs also contain ketoreductase (KR), dehydratase (DH), enoyl reductase (ER), cyclase (Cyc), thioesterase (TE), C-methyl transferase (C-MT) activities among others (Evans et al., 2011).

In order to function, the PKS needs to be activated by a phosphopantetheinyl transferase (PPTase) that covalently tethers the cofactor 4'phosphopantetheine to the polyketide synthase (figure 1) (Evans et al., 2011). This causes the transition of the enzyme from its apo to its holo state. There are several PPTase's discovered to date. Two examples include Sfp from *Bacillius subtilis* and NpgA from *Aspergillus nidulans* that have been shown to be quite promiscuous in their activation and can activate an array of different PKSs (Cox and Simpson, 2009). The NpgA gene from *A. nidulans* has a functional allele known as CfwA and it has been suggested that these are the ones responsible for all PKS activation in *A.* nidulans (Márquez-Fernández et al., 2007).

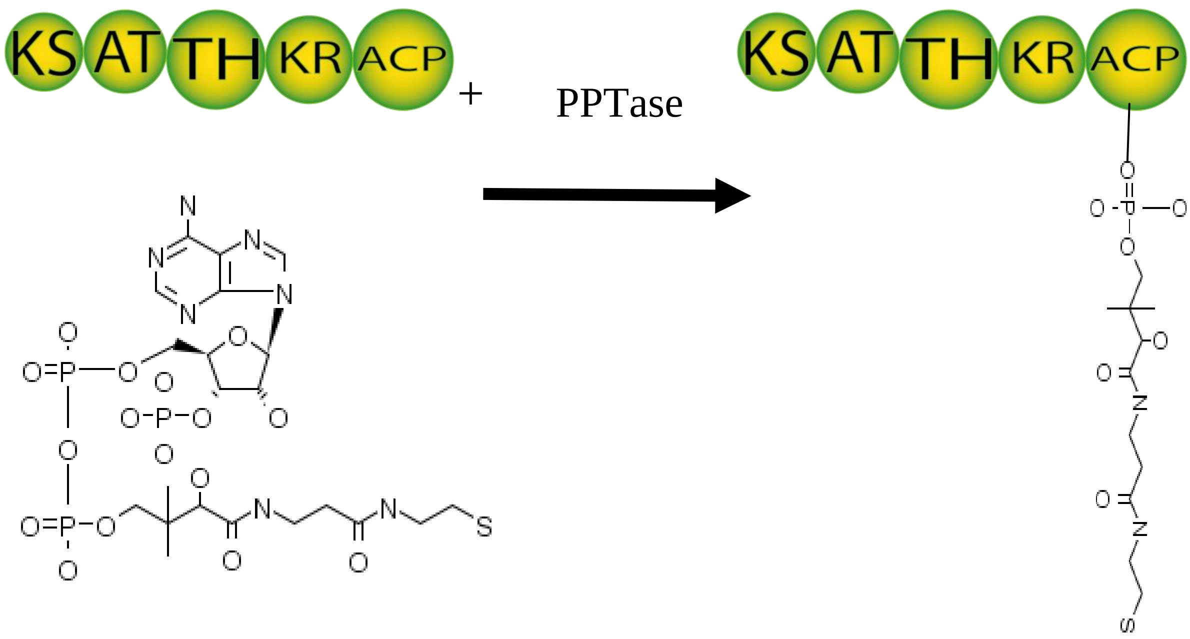


Figure 1 Polyketide synthases have to be activated by a phosphopantetheinyl transferase (PPTase) in order to function. The action of the PPTase results in the covalent tethering of the CoA part of Coenzyme A to the polyketide synthase as illustrated above.

Polyketides are grouped into different types depending on their method of elongation and degree of reduction of the polyketide chain. Type I PKSs can be subdivided into either modular or iterative type I polyketide synthases. The modular type I bacterial PKS can be exemplified by the DEBS polyketide synthase that eventually leads to the production of erythromycin in *Sacchalospora erythraea* (Katz, 2009). In the modular PKS there is one domain that is responsible for each modification to the extending polyketide chain (Katz, 2009). Alternatively the polyketide structure is constructed by the iterative use of the catalytic sites referred to as Type I iterative PKSs typically found in fungi. Type II PKSs use single ACP-domains. On top of this ACP-domain they are made up of a KS and a chain-length factor (CLF) that are typically encoded by separate genes. The chain length is determined by the cavity size of the CLF (Tang et al., 2003). Type III polyketide synthases use acyl-CoAs but have no ACP-domains and only a single KS which initiates, tethers, extends and terminates the polyketide chain. In order to function they have to adopt a homodimer conformation.

The fungal iterative type I PKSs can be further divided into three categories depending on the degree of reduction of the polyketide chain. An example of a non-reducing polyketide synthase (NR-PKS) is the aflatoxin/sterigmatocystin pathway found in *A. nidulans* (Sanchez et al., 2008). A prime example of the partially reducing (PR-PKS) fungal polyketide synthase genes is the *atX* gene form *Aspergillus terreus* encoding the 6-MSA synthase (Fujii et al., 1996; Fujii, 2010). The last category is the highly reducing polyketide synthase (HR-PKS) of which the prime example is lovastatin produced by *A. terreus* (Cox and Simpson, 2009). Lovastatin and its derivatives are well known as potent inhibitors of the HMG CoA reductase and thus have cholesterol lowering effects that renders them block-buster drugs (Sanchez et al., 2008).

Another interesting class of compounds are the hybrid NRPS-PKS genes that on top of the Acyl-CoA incorporations of the PKS also incorporates amino acids from the NRPS part of the enzyme (Cox and Simpson, 2009). An example of a combined NRPS-PKS product is tenillin.

Biosynthesis of 6-MSA

In 1967 further details on the synthesis of many polyketides including 6-MSA were discovered (Birch, 1967). The 6-Methyl-Salicylate Synthase (MSAS EC 2.3.1.165) belongs to the type I PR-PKS (Cox and Simpson, 2009). The structure of the domains in the 6-MSA synthase is illustrated in figure 2 (Fujii, 2010).

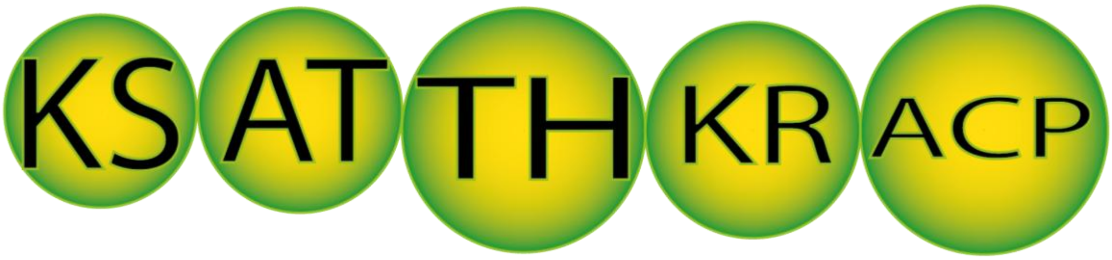


Figure 2 The structure of the domains in the 6-MSAS. It consists of a ketosynthase (KS), Acyltransferase (AT), Thioester hydrolase (TH), Ketoreductase (KR) and Acyl Carrier Protein (ACP).

In 1969 the first proposed reaction mechanism of the enzyme was given by Dimroth et al. (Dimroth et al., 1970). Later other reaction mechanisms have been proposed by Staunton and Weissman (Staunton and Weissman, 2001). The latest proposed mechanism for 6-MSA synthesis is from Fujii and involves a thioester hydrolase domain involved in the release of the 6-MSA. This mechanism was observed based on the 6-MSA synthase gene found in *A. terreus*, however the polyketide synthases are almost identical and it can be assumed that the *P. patulum* derived synthase has the same function. Figure 3 illustrates the proposed reaction mechanism for the production of 6-MSA by the continuous condensation of acetyl and then malonyl-CoA monomers (Fujii, 2010). First the acetyl-CoA is loaded onto the ACP domain. Then follows two Claisen condensations with malonyl-CoA. The reduction by the keto-reductase is done at the expense of 1 NADPH that is reduced to $NADP^+$. Then follows another condensation with a malonyl-CoA and finally the cyclisation and release by the thioesterase domain.

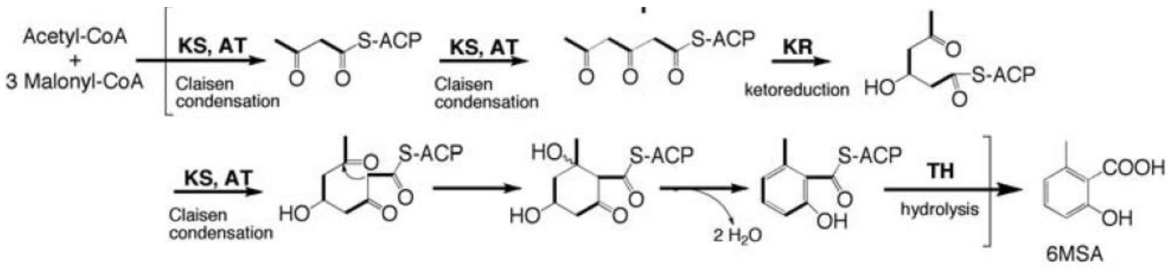


Figure 3 The mechanism by which 6-MSA is produced by the MSAS PKS from *A. terreus*. Reproduced from Fujii *et al.* (Fujii, 2010).

References

Bentley, R., 1999. John Norman Collie: Chemist and Mountaineer. J. Chem. Educ. 76, 41-.

Bentley, R., Bennett, J.W., 1999. Constructing polyketides: From Collie to combinatorial biosynthesis. Annu. Rev. Microbiol. 53, 411-.

Birch, A.,J., 1967. Biosynthesis of polyketides and related compounds. Science 156, 202-6.

Collie, J.N., 1907. CLXXI.Derivatives of the multiple keten group. J. Chem. Soc. , Trans. 91, 1806-.

Cox, R.J., Simpson, T.J., 2009. Fungal Type i Polyketide Synthases. Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates 459, 49-78.

Dimroth, P., Walter, H., Lynen, F., 1970. Biosynthese von 6-Methylsalicylsäure. European Journal of Biochemistry 13, 98-110.

Evans, B.S., Robinson, S.J., Kelleher, N.L., 2011. Surveys of non-ribosomal peptide and polyketide assembly lines in fungi and prospects for their analysis in vitro and in vivo. Fungal Genetics and Biology 48, 49-61.

Fujii, I., Ono, Y., Tada, H., Gomi, K., Ebizuka, Y., Sankawa, U., 1996. Cloning of the polyketide synthase gene atX from *Aspergillus terreus* and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression. Molecular and General Genetics MGG 253, 1-10.

Fujii, I., 2010. Functional analysis of fungal polyketide biosynthesis genes. J Antibiot 63, 207-18.

Katz, L., 2009. The Debs Paradigm for Type i Modular Polyketide Synthases and Beyond. Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates 459, 113-42.

Márquez-Fernández, O., Trigos, Á., Ramos-Balderas, J.L., Viniegra-González, G., Deising, H.B., Aguirre, J., 2007. Phosphopantetheinyl transferase CfwA/NpgA is required for *Aspergillus nidulans* secondary metabolism and asexual development. Eukaryotic cell 6, 710-20.

Sanchez, J.F., Chiang, Y.M., Wang, C.C.C., 2008. Diversity of polyketide synthases found in the Aspergillus and Streptomyces genomes. Molecular Pharmaceutics 5, 226-33.

Staunton, J., Weissman, K.J., 2001. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 18, 380-416.

Tang, Y., Tsai, S.C., Khosla, C., 2003. Polyketide chain length control by chain length factor. J. Am. Chem. Soc. 125, 12708-9.

Tsai, S., Ames, B.D., 2009. Structural Enzymology of Polyketide Synthases. Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates 459, 17-47.

Weissman, K.J., 2009. Introduction to Polyketide Biosynthesis. Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates 459, 3-16.

Chapter 4

An *in silico* Guided Approach to Metabolic Engineering of Heterologous Polyketide Production in *A. nidulans*

L. Mølgaard[a], J. B. Nielsen[a], S. P. Lauritzen[a], N. B. Hansen[a], S. W. Thrane[a], U. H. Mortensen[a], J. Thykær[a*], K. R. Patil[b]

[a]Department of Systems Biology

Technical University of Denmark

Kgs. Lyngby, Denmark

[b] The European Molecular Biology Laboratory

Heidelberg, Germany

*Corresponding Author

Abstract

Polyketides constitute a large source of compounds used for therapeutics today (Newman and Cragg, 2007). Often the natural polyketide host does not display sufficient productivity needed for large-scale production. Furthermore, if there is no prior knowledge of large scale production in the natural host organism, setting up an industrial scale process can be very challenging. Thus, having an efficient heterologous production system is highly desirable. To address this possibility, we investigated the production of the model polyketide 6-methyl salicylic acid (6-MSA) in *Aspergillus nidulans*. In an attempt to improve production we identified NADPH dependent glutamate dehydrogenase (*gdhA*) as a potential deletion target using the genome scale metabolic model of *A. nidulans* (*i*HD666) and *in silico* metabolic engineering algorithm OptGene (Patil et al., 2005; David et al., 2008). To alleviate the growth inhibition observed in the *gdhA*Δ mutant an extra copy of NAPH dependent glutamate dehydrogenase B (*gdhB*) was inserted. The engineered *A. nidulans* strains were characterized in batch reactors. The growth rate of the *gdhA*Δ strains decreased significantly with only a marginal increase in 6-MSA yield. The extra copy of *gdhB* resulted in increased growth rate. In addition, we have shown that we can establish model predicted deletions in *A. nidulans*. Thus an efficient polyketide cell factory has been generated for 6-MSA production and the first steps towards model guided optimization have been taken.

Introduction

Natural products are an important source of new drugs and natural product derived medicines constitute one of largest groups of therapeutics (Newman and Cragg, 2007). One class of medically significant compounds is the polyketides. Some examples of medically relevant polyketides are erythromycin (antibiotic), lovastatin (cholesterol lowering) and tacrolimus (immunosuppressant) and polyketide derived medicines generally have annual sales above US$ 20 billion (Weissman, 2009). Other polyketides are quite toxic e.g. aflatoxin, fuminisin and patulin.

Polyketides are produced by large enzyme complexes known as polyketide synthases (PKS). Most PKSs are made up of different domains serving specific catalytic functions such as condensation, chain elongation, ketoreduction, dehydration and cyclization (Tsai and Ames, 2009). In order for the polyketide synthase to function it must be converted from its apo form to its holo form. This is done by a phosphopantetheinyl transferase (PPTase) that tethers a phosphopantetheinyl group from coenzyme A onto to the catalytic domain carrying the growing polyketide chain.

6-methyl salicylic acid (6-MSA) is a model polyketide with minor antibiotic activity. The polyketide synthase gene is relatively small making it ideal as a model system as it simplifies the molecular biological challenge of heterologous expression. The first report of 6-MSA production was from *P. patulum* (Spencer and Jordan, 1992). However, the very low productivity of the natural host has lead to the establishment of a heterologous production platform that further enabled the study of polyketide synthesis. 6-MSA production has been attempted in both bacteria and fungi. The production of 6-MSA in *S. coelicolor* was measured based on cultivation from plates to be 20 mg/L (Bedford et al., 1995). Later the use of a glycerol medium and a plasmid based system titers improved to 75 mg/L in *E. coli* (Kealey et al., 1998). By using *S. cerevisiae* and a multicopy plasmid the titer in YPD media was increased to 1.7 g/L (Kealey et al., 1998). Another study of 6-MSA production in *S. cerevisiae* obtained a final titer of 200 mg/L in a minimal medium with a multi-copy plasmid (Wattanachaisaereekul et al., 2007).

Another well-suited host for polyketide production is the ascomycete *A. nidulans.* It is already renowned for its polyketide productivity and has been found to contain around 27 polyketide gene clusters (Galagan et al., 2005). One study of heterologous 6-MSA production in *A. nidulans* with a random integration of one or more copies of the *A. terreus msas* gene was reported to produce 445 mg/L in a batch on glucose (Panagiotou et al., 2009). Moreover the targeted genetic engineering of *A. nidulans* is well established making it a suitable host for heterologous expression (Nielsen et al., 2006).

Beyond the heterologous expression of the polyketide encoding gene several optimization steps must be taken. The adjustments of media or process conditions can have profound effects on the final yield. In addition, genetic modifications that can divert the carbon flow towards the polyketide precursors are also important. One example of metabolic engineering for polyketide production is the overexpression of Acetyl-CoA carboxylase (*ACC1*) that converts acetyl-CoA to malonyl-CoA in *S. cerevisiae* (Wattanachaisaereekul et al., 2008). The overexpression of *ACC1* is an obvious target for 6-MSA production as the PKS requires three malonyl-CoAs to produce 6-MSA. As an alternative genome scale metabolic models and optimization algorithms can be used to predict less obvious targets that can lead to increased yields of 6-MSA.

Metabolic Engineering Through *in silico* Design

By using FBA or MOMA as scoring algorithms a genetic algorithm using evolutionary search procedures for solving the combinatorial optimization problems was developed called OptGene (Patil et al., 2005). In OptGene the model must first be preprocessed removing all lethal deletions from the solution space. Then from an initial population of deletion strains the fitness is evaluated based on an objective function often productivity or biomass product coupled yield (BPCY) is evaluated with FBA or MOMA. The best individuals are then crossed *in silico* and a new population generated. In this way OptGene searches for local optima, but has been proven by a limited number of iterations to converge towards a global optimal solution.

Examples of successful improvements obtained through *in silico* guided metabolic engineering include vanillin (Brochado et al., 2010), lycopene (Alper et al., 2005), sesquiterpenes (Asadollahi et al., 2009) and succinic acid (Otero, 2009) production in *S. cerevisiae* and *E. coli*. However, it remains to be shown whether these methods are also applicable for higher eukaryotes. There are still no examples of *in silico* guided approaches to metabolic engineering in filamentous fungi.

Thus the aim of the study was to establish a platform for the heterologous production of polyketides in *A. nidulans* and explore the possibilities of *in silico* guided optimization. As a result the following steps were taken. The genetic changes are all targeted to specific genomic loci making it a reproducible system amenable to other polyketide synthases. The fact that different *A. nidulans* strains can be crossed makes it possible to exchange markers and efficiently incorporate changes into newly constructed strains. The possibility to achieve marker excision by direct repeat recombination enables repeated rounds of genetic modifications to metabolism. The engineered *A. nidulans* strains were subjected to physiological characterization in bioreactors evaluating the cellular performance in terms of growth rate, yields and productivities.

Materials and Methods

The metabolic engineering approach was achieved by first modelling 6-MSA production followed by genetic engineering of the predicted strains. Finally the strains were physiologically characterized in bioreactors.

Model Predictions

The genome scale model *i*HD666 was used as a starting point for modelling 6-MSA production in *A. nidulans*. The reaction for biosynthesis of 6-MSA was added to the model. It involves the use of one acetyl-CoA, three malonyl-CoAs and one molecule of NADPH for production of one molecule of 6-MSA. On top of this a transport reaction for 6-MSA allowing secretion of 6-MSA from the cell, as observed in previous studies (Panagiotou et al., 2009) was added to the model. The growth was constrained to be at least 0.01 hr^{-1} as mutants with growth below this limit are likely to be experimentally non-viable and/or biotechnologically irrelevant. The glucose uptake rate was fixed at 3 mmole/gDW/hr since unlimited uptake of substrate is not biologically meaningful and this value is within the range of observed rates for fungi. OptGene predictions are insensitive to the actual value of this constraint unless it is set too low. The entry flux into the pentose phosphate pathway (PPP) via glucose-6-phosphate-1-dehydrogenase (AN2981) was constrained to be between 1 and 2 as it was experimentally observed through C-13 flux measurements on glucose with the same strain background that a maximum of 60% of the glucose taken up is directed to the PPP in this step (Panagiotou et al., 2009). The FBA simulations were performed using the GLPK. MOMA simulations were performed by using a custom function that calls OOQP as a quadratic solver. The deletion strategy was devised based on OptGene using Biomass-Product Coupled Yield for ranking the best candidates (Patil et al., 2005).

Strain Construction

The *A. nidulans* strains used for physiological characterization were constructed using the gene targeting methods of Nielsen *et al.* followed by sexual crossing (Nielsen et al., 2006). For the insertion of the *msas* the gene targeting substrates were cloned in a USER^{TM} vector (Geu-Flores et al., 2007).

Strains

E. coli DH5α cells were used for cloning the USER^{TM} vectors and propagating plasmids. Fungal transformations were performed with the strain IBT29539 for the knockouts and IBT28738 (Nielsen et al., 2008) for 6-MSA production. The control strain NID210 was constructed as described by Hansen *et al.* (Hansen et al., 2011). The *A. nidulans* strain IBT27263 was used for amplification of genomic DNA fragments for gene targeting. The strain is derived from G051 of the Glasgow strain collection (Clutterbuck, 1974).

Media

The Fungal minimal medium (MM) was as described in Cove (1966), but with 1% glucose, 10 mM $NaNO_3$ and 2% agar (Cove, 1966). The media were supplemented with 10 mM uridine, 10 mM uracil and 4 mM L-arginine, 0.02 mg/L biotin when required. After crossing the *gdhA*Δ the spores were spread on plates with 10 mM glutamine to increase the growth rate of the spores harboring the mutation.

For the batch cultivations minimal media containing $(NH_4)SO_4$ 7.5 g/L, KH_2PO_4 1.5 g/L, $MgSO_4$ $7H_2O$ 1 g/L, NaCl 1 g/L, $CaCl_2$ 0,1 g/L, antifoam 0,05 g/L and 1 mL of a trace metals solution containing $CuSO_4 \cdot 5\,H_2O$ 0.4 g/L, $Na_2B_2O_7 \cdot 10\,H_2O$ 0.04 g/L, $FeSO_4 \cdot 7\,H_2O$ 0.8 g/L, $MnSO_4 \cdot H_2O$ 0.8 g/L, $Na_2MoO_4 \cdot 2\,H_2O$ 0.8 g/L, $ZnSO_4 \cdot 7\,H_2O$ 8.0 g/L was used.

Vector Construction

Ectopic integration of the *msas* gene was achieved by first constructing the gene targeting substrate by USER^{TM} fusion (Geu-Flores et al., 2007). The 5.3 kb *msas* gene was amplified from the vector pRS306CRUDTMSA-PP using the primers AGAgcgaUatgcattccgctgcaacttc and TCTgcgaUttaatggtgatggtgatga (Wattanachaisaereekul et al., 2007) and cloned into BGHAp71. The resulting plasmid was named BGHAp75. A list of the plasmids used in this work can be seen in table 1. All plasmids were sequenced (StarSEQ, Germany).

Table 1 The plasmids used for construction of a fragment for the insertion of the *msas* gene in the site IS1 (Hansen et al., 2011). A map of the plasmid p75 can be seen in supplementary S1.

Plasmid	**Characteristics**	**Reference**
BGHAp1	USER cassette	(Hansen et al., 2011)
BGHAp71	p1 with a p*gdp*::*ArgB2::trpC*	(Hansen et al., 2011)
pRS306CRUDTMSA-PP	Yeast vector with the *msas* gene from *P. patulum*	(Wattanachaisaereekul et al., 2007)
BGHAp75	BGHAp71 with an inserted *msas* gene	This study

Recombinant Strain Construction

In table 2 are given the genotypes of all of strains constructed in this study. The details of the construction will be given in the following text.

Table 2 The *A. nidulans* strains constructed that were not used for batch cultivations.

Strain	Genotype*	Source
	Gene targeting and marker elimination strains	
IBT29539	*argB2, pyrG89, veA1, nkuA*Δ	(Nielsen et al., 2008)
IBT28738	*argB2, pyrG89, veA1, nkuA-trS::AFpyrG*	(Nielsen et al., 2008)
IBT25456	*biA1, veA1*	Glasgow strain collection
NID596	*argB2, nkuAΔ, wAΔ, yAΔ*	In house strain collection
	Gene targeting	
NID63	6MSAS::*argB, nkuAΔ*	IBT29539
NID58	*gdhA::AFpyr:: gdhB, nkuAΔ*	IBT29539
NID61	*gdhAΔ::AfpyrG, nkuAΔ*	IBT29539
NID210	IS1::PgpdA::TtrpC::*argB*	IBT28738
	Removal of *nkuA*Δ	
NID137	IS1::PpgdA-6MSAS-TtrpC::*argB*	NID63 X IBT25456
	Combining 6-MSA with the deletion strains	
NID164	IS1::PpgdA-6MSAS-TtrpC::*argB*, gdhAΔ::gdhBtr::AFpyrG	NID58 X NID137
NID180	IS1::PpgdA-6MSAS-TtrpC::*argB*, gdhAΔ::AFpyrG	NID61 X NID137
	Removal of *pyrG* on 5-FOA	
NID393	IS1::PpgdA-6MSAS-TtrpC::*argB, gdhAΔ::gdhB*	NID164 pop
NID392	IS1::PpgdA-6MSAS-TtrpC::*argB, gdhAΔ*	NID180 pop
	Final Strains**	
NID521	IS1::PgpdA-TtrpC::*argB*	NID210 X IBT25456
NID875	IS1::PpgdA-6MSAS-TtrpC::*argB*	NID137 X NID596
NID898	IS1::PpgdA-6MSAS-TtrpC::*argB, gdhAΔ*	NID392 X NID596
NID605	IS1::PpgdA-6MSAS-TtrpC::*argB, gdhAΔ::gdhB*	NID393 X IBT25456

*All strains except IBT25456, NID521, NID875, NID898 and NID605 are *veA1, argB2, pyrG89*. **The final strains are all *argB2* and *veA1*.

The gene targeting substrate from BGHAp75 was excised by NotI digestion and transformed into IBT28738 using the *A. nidulans argB* selectable marker as described by Nielsen *et al.* (Nielsen et al., 2008). Transformants were streak purified and were verified for insertion into IS1 (situated between AN6638 and AN6639 on chromosome I) by diagnostic PCR using the primers given in table 7 supplementary S1. The strain selected for further work was named

NID63. As a control strain for the insertion site the strain NID210 was used (Regueira et al., 2011). The primers used can be seen supplementary S1.

The knockout of *gdhA* (AN4376) was carried out as previously described by Nielsen *et al.* (Nielsen et al., 2006). The gene targeting substrates were amplified using the primers described in supplementary S1. The substrates were purified using illustra® DNA and Gel band purification kit (GE Healthcare) and transformed into IBT29539.The transformants were screened by PCR and positive candidates were streak purified. The selected strain was named NID61.

In case of the *gdhB* replacement strain, parts of the *gdhB* gene was used as direct repeats thus a recombination event on 5-FOA results in the elimination of the *pyrG* marker and reconstitution of a functional *gdhB* gene in the AN4376 locus as shown in supplementary S1 figure 3. The strain was constructed by the same methods as the *gdhA* strain. The replacement strain was named NID58.

In the first round the *nkuA*Δ was eliminated by crossing NID63 with IBT25456 to increase genetic strain stability. Based on PCR verification of *msas* insertion and the loss of the *nkuA* mutation along with growth on selective media the progeny strain NID137 was chosen.

The 6-MSA producing strain NID137 was crossed with NID58 and NID61, respectively resulting in the generation of the two progeny strains NID164 and NID180. The *Aspergillus* *fumigatus pyrG* marker was eliminated from NID164 and NID180 by direct repeat recombination on 5-fluororotic acid (5-FOA) medium resulting in the strains NID393 and NID392.

The *gdhA* deletions of NID392 and NID393 were verified by southern blotting. Genomic DNA from NID392 and NID393 was purified by using the FastDNA® SPIN for Soil Kit (MP Biomedicals, LLC) and restricted with the enzymes *Afl*II and *pme*I. The probe was PCR amplified as described in supplementary S4. The deletion of *gdhA* was verified by southern blotting. Blotting was performed as described by Sambrook and Russel 2001 (Sambrook and Russell, 2001). Labeling of the probes was performed according to the manufacturer's protocol using the Biotin DecaLabel Kit, #0652 from Fermentas. Detection was done by the Biotin Chromogenic detection kit, #K0662 according to the manufacturer's description.

In the case of NID393 the reconstituted *gdhB* gene including the upstream region was PCR amplified with the primers GACTGCCGAAGTAAGAGCGCGG and TTATGCTTTGGACTGTGCAAGTC and sequenced to insure that there were no mistakes in the direct repeat recombination (StarSEQ, Germany).

In order to avoid any differences in metabolism and expression levels of some genes due to auxotrophies the mutation *pyrG89* was eliminated by crossing with a strain that had the wildtype allele of *pyrG*. The strains NID210 and NID393 were crossed with GO51 and the resulting progeny were tested on minimal media. NID137 was crossed with NID596 and the resulting progeny were tested on minimal media. Due to the mutations leading to white (*wAΔ*) and yellow (*wAΔ* , *yAΔ*) color of some of the progeny in NID596 a visual selection for green progeny was undertaken. The spores were grown on minimal media and selected colonies were PCR verified for insertion in IS1 and the deletions. Based on sequence alignment it was established that the *argB2* harbors a point mutation at basepair 245 in the coding sequence. Thus in order to confirm the presence *argB2* mutation the first 600 bp of the gene was sequenced in all strains (StarSEQ, Germany). The resulting strains can be seen in table 2 under the heading final strains. All primers used can be seen in supplementary S1.

Physiological Characterization

The four strains NID875, NID521, NID898 and NID605 were all characterized in batch fermentations performed as described below.

Inoculum Preparation

The spores for inoculation of the batch reactor were generated by growing the *A. nidulans* strains on potato dextrose agar (PDA) plates for 4-5 days at 37 °C. The spores were harvested by addition of 5 mL of sterile saline water (0.9 % NaCl) per plate. The spores were counted in a counting chamber and the bioreactor was inoculated with 2 *10^9 spores per L.

Bioreactor Cultivation Conditions

The cells were grown in 2 L Braun bioreactors with a working volume of 1.8 L. The temperature was controlled at 30 °C. Stirring and airflow was set as a ramp starting at 100 rpm and 0.1 vvm during germination. The pH was maintained at pH 3 for the first 8 hours of the fermentation after which it was slowly increased to pH 5.5 over the following 2 hours by addition of 2 M NaOH. Stirring was gradually increased to 300 rpm after 5 hours then 600 rpm after 10 hours and 800 rpm after 11 hours. The airflow was increased to .6 vvm/min after 9 hours and then to 1 vvm/min after 10 hours and pH was set at 5.5 by addition of 2 M NaOH. During the cultivation pH was kept at 5.5 by the addition of 2 M NaOH and 2 M HCl. The concentrations of carbon dioxide in the exhaust gas were monitored with a gas analyzer throughout the fermentation (Innova AirTech Instruments 1313 Fermentation monitor was used for NID605 and NID521. The PrimaPro Process MS from Thermo Scientific was used for NID875 and NID898).

Biomass measurements

First the sampling port was emptied by removing approximately 3 mL of the broth and discarding it. Then approximately 5 mL of fermentation broth was pulled from the broth in the reactor. The syringe was then weighed before the liquid was filtered through a preweighed, dried filter (Sartorius AG, Goettingen, Germany) with a poresize of 0.45 µm. The filter was washed twice with the same volume sterile saline water (0.9 % NaCl). The filter was then dried at 150 W for 20 minutes in a microwave oven. The filters were cooled in a desiccator for at least 2 hours before the filters were weighed to determine the amount of biomass in the broth.

Metabolite Measurements

To determine the metabolite concentrations in the broth samples were filtered through a 0.45 µm filter (Sartorius AG, Goettingen, Germany) and stored at -20 °C until further analysis.

Glucose was measured using an Agilent HPLC series 1100 with a RI (refractive index) detector. The samples were run on an ion-exclusion column with the dimensions 300 mm by 7.8 mm (Aminex HPX-87H from Biorad). The column was eluted at 60 °C with 5mM H_2SO_4 at flow rate of 0.6 mL/min. The metabolites were detected with an RI detector.

6-MSA (SC-274880, Santa Cruz Biotechnology Inc., California, USA) was measured on an Agilent 1100 series HPLC with a degasser, binary pump, column oven and DAD. The Luna C18(2) column (100*2 mm with 3µ particles, Phenomenex (Torrence, CA)) was heated at 40 °C during analysis. The samples were run on a gradient of milliQ water with 50 ppm TCA (Solvent A) and 50 ppm TFA in acetonitril (solvent B). The gradient of the solvents was as mix of A and B. The amount of B was changed in gradient from 20% to 60% in 10 minutes and then 20% B for the next 2 minutes.

Results

In silico optimization of 6-MSA production

First the *i*HD666 model of *A. nidulans* metabolism was modified by the addition of the reactions for biosynthesis and export of 6-MSA. Then the production of 6-MSA was modeled using FBA and MOMA for prediction of a single deletion mutant with improved yields of 6-MSA. The targets were sorted based on the maximum biomass product coupled yield (BPCY) giving the highest combined yield of biomass and 6-MSA. The list of predicted targets can be found in table 8 supplementary S3. The predicted targets were evaluated in terms of biological impact and significance. This inspection led to the removal of several unfeasible targets. Hence, several transport reactions were removed. Many of the predicted targets were aimed at increasing the availability of the cofactor NADPH for production of 6-MSA. As the pentose phosphate pathway (PPP) is the largest source of NADPH many of the targets were related to the PPP (red). However, the split between the Embden-Meyerhof-Parnas pathway (EMP) and PPP is highly regulated and difficult to engineer (Hankinson and Cove, 1974), so targets in the PPP were not prioritized. Several genes encoding enzymes acting in the TCA cycle (blue) were also suggested as targets. With the same argument of tight regulation, these candidates fore gene deletion were not prioritized either. Lastly, genes that were proven to be lethal in deletion studies of *S. cerevisiae* or *A. nidulans* (Giaever et al., 2002; Roumelioti et al., 2010) were removed from the list (green). A reduced list of targets can be seen in table 3.

Table 3 The final list of possible candidates in a deletion strategy for improved 6-MSA production. Hydrogen ions (H+) are not balanced in the *i*HD666 model and thus not given in the reactions. The targets are sorted based on the Biomass Product Coupled Yield (BPCY).

	Gene ID	BPCY	Reaction
1	AN8866.2 (putative phosphoglycerate dehydrogenase)	0.00513	3-phospho-D-glycerate + NAD+ -> 3-phosphonooxypyruvate + NADH
2	AN3223.2b *pfkA*	0.003093	ATP + sedoheptulose-7-phosphate-> ADP + seduheptulose-1,7-bisphosphate
3	AN7459.2a (putative 6-phosphofructokinase)	0.002952	ATP + Mannose -> ADP + D-Mannose-6-Phosphate
4	AN3432.2a (aldose-1-epimerase)	0.00284	UDP-galactose <-> UDP-glucose
5	AN3058.2 (putative glycine hydroxymethyltransferase)	0.002723	tetrahydrofolate + L-serine <-> glycine + 5, 10 methylenetetrahydrofolate
6	AN3741.2c *alcB*	0.002465	glycerol + NAD -> D-glyceraldehyde + NADH
7	AN4684.2a (putative triacylglycerol lipase)	0.001991	triacylglycerol + H2O -> diacylglycerol + 0.0821 C120ACP + 0.0444 C140ACP + 0.0407 C141ACP + 0.0081 C150ACP + 0.5161 C160ACP + 0.0681 C161ACP + 0.0276 C162ACP + 0.0039 C170ACP + 0.0860 C180ACP + 1.2429 C181ACP + 1.0870 C182ACP + 0.1051 C183ACP + 0.0200 C200ACP (*fatty acids*)
8	AN4323.2/AN7878.2/AN5957.2d (Putative branched chain amino acid aminotransferase)	0.001877	(R)-2-oxoisovalerate + glucose <-> 2-oxoglutarate + L-valine
9	AN4376.2 *gdhA*	0.001756	Oxoglutarate+NADPH -> glutamate and NADP+

After screening the initial targets both based on the level of characterization of the genes as well as the potential effect on metabolism one of the main targets was the NADPH-GDH (glutamate dehydrogenase, AN4376), gdhA. It was chosen for experimental validation, as previous studies strongly suggest that its deletion also may lead to dramatic reorganization of metabolic fluxes in other fungi (Thykaer et al., 2008). The main argument for choosing gdhA as a target is that it consumes NADPH. By eliminating GDHA activity, potentially m ore NADPH will be available fore 6-MSA production. A more detailed overview of the reactions in *A. nidulans* involving GDHA and nitrogen metabolism can be seen in supplementary S2.

Strain Construction

The strains were constructed through several rounds of gene targeting, crossing and marker elimination by direct repeat recombination as illustrated in figure 1.

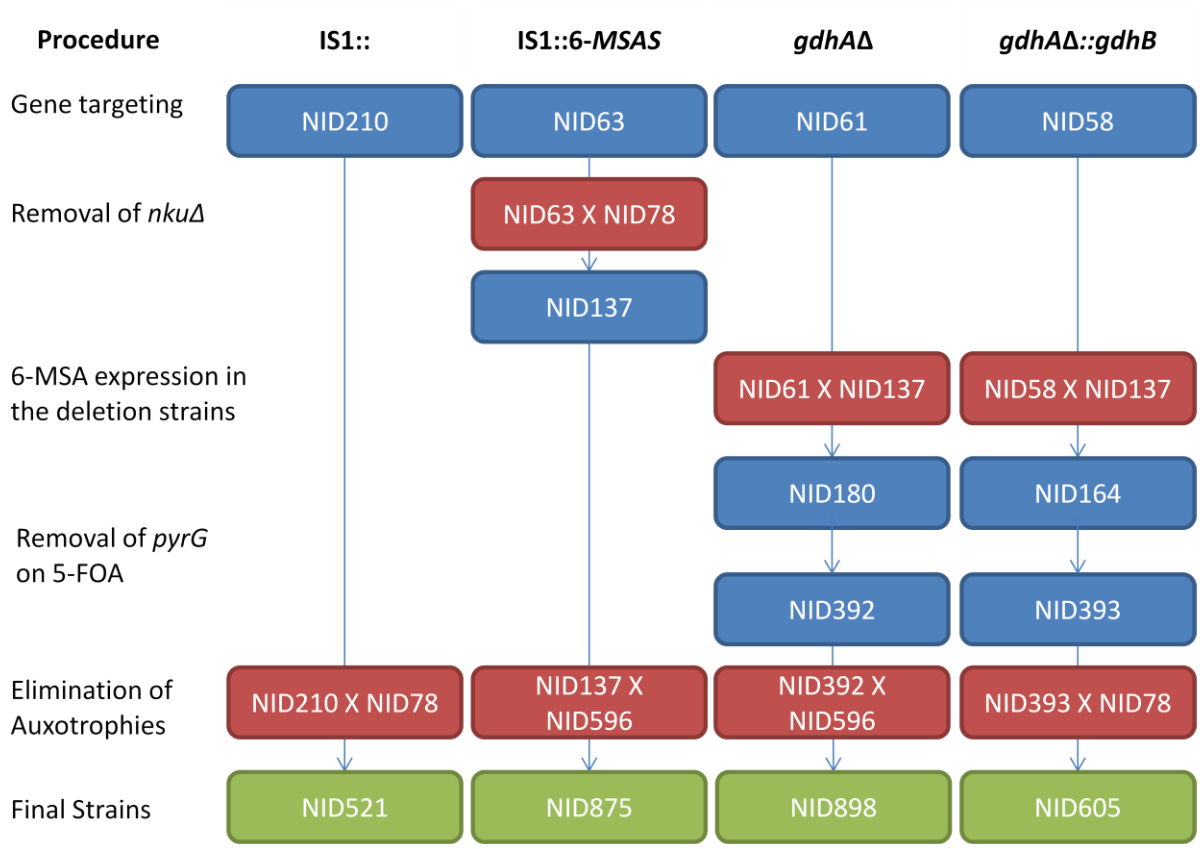


Figure 1 Overview of the strain construction. All intermediate strains used for the construction are given in blue boxes. The red boxes represent the transient diploid state of a cross. The strains in the top row were constructed through gene targeting. Through several rounds of crossing (red boxes) and counterselection on 5-FOA the final strains (green boxes) given in the last row were obtained.

First we established the production of 6-MSA through genomic integration resulting in the strain NID63 that was then put through several crosses in order to obtain the final strain NID875. We then constructed the strain harboring the model predicted *gdhA* deletion (NID61) and crossed it with the 6-MSA producing strain and after crossing and marker elimination had the final strain NID898. In order to improve the growth rate of the *gdhA* deletion strain a strain with the deletion and an extra copy of *gdhB* was constructed (NID58) and crossed to a 6-MSA producing strain where the marker was eliminated by crossing resulting in the strain NID605. Furthermore, southern blot analysis in supplementary S4 confirms the deletion of *gdhA* and the deletion of *gdhA* with the insertion of *gdhB* when compared to the reference strain IBT29539. All the final strains were able to grow in cheap non-selective media. The markers were eliminated such that the only difference to the reference was the intended deletions. The genotype of the strains constructed for evaluation of the metabolic engineering strategies can be seen in table 2 in the materials and methods section.

Physiological Characterization

The four strains NID875, NID521, NID898 and NID605 were all characterized in 2 L batch fermentations to determine the effect of the *gdhA* deletion on 6-MSA production. The growth and metabolite profiles are presented in figure 7 supplementary S6.

The maximum specific growth rates of the strains were calculated based on dry weight measurements in the exponential phase of the cultivations. As expected, the growth rate of the *gdhA*Δ strain (NID898) was reduced to one third of the reference strain expressing 6-MSA. Although the growth rate of the *gdhA*Δ*::gdhB* strain (NID605) is still severely affected, it doubled when compared to the *gdhA*Δ strain presumably due to the extra copy of GDHB. Furthermore, the lag phase of NID605 and NID898 increased by more than 10 hours, possibly due to a redox in-balance.

Table 4 The maximum specific growth rate (μ_{max}) and yields (Y) calculated from the fermentation data. Y_{sx}, Y_{xp} were all calculated in the exponential growth phase. Dry weight (dw)

Strain	NID521	NID875	NID898	NID605
Genotype	IS1::PgpdA-TtrpC::*argB*	IS1::PgpdA-6MSAS-TtrpC::*argB*	IS1::PgpdA-6MSAS-TtrpC::*argB, gdhA*Δ	IS1::PgpdA-6MSAS-TtrpC::*argB, gdhA::gdhB*
μ_{max} (h^{-1})	0.24 ±0.01	0.21 ±0.02	0.07 ±0.01	0.14 ±0.03
[2]$Y_{sx\ overall}$ (g dw/g glucose)	0.56 ±0.03	0.56 ±0.01	0.4 ±0.07	0.38 ±0.05
[1]Y_{xp} (mg 6-MSA/g dw)		57.66 ±11.19	74.32 ±31.69	42.98 ±22.24
r_p (mg 6-MSA/g dw/h)		10.7 ±4.35	5.44 ±1.89	5.32 ±1.22
[2]$Y_{xp\ overall}$ (mg 6-MSA/g dw)		48.66 ±2.39	34.82 ±11.41	30.04 ±7.89

[1] Y_{xp} was calculated in the exponential growth phase.

[2] $Y_{xp\ overall}$ and $Y_{sx\ overall}$ are calculated at the time point of maximum biomass and 6-MSA titer.

The 6-MSA and metabolite concentrations were measured in order to estimate the yields. The yield of biomass in the strains NID521 and NID875 are as expected quite similar. The yield of biomass decreases in the *gdhA*Δ strains probably as a result of reduced nitrogen assimilation.

The model assumes a steady state, which is only present in the exponential phase. Thus the exponential phase yield should provide a better fit to the model predictions. When estimating the yield of 6-MSA per biomass (Y_{xp}) in the exponential phase the mutant strain NID898 showed increased yields. A more robust characteristic is the overall yield of 6-MSA per biomass. The overall yield of 6-MSA ($Y_{xp\ overall}$) was however larger in NID875 than in the engineered strains NID605 and NID898. When taking biomass into account by calculation the productivity, the reference strain NID875 still has a higher productivity (r_p) than the mutant strains. In addition, we measured the polyols glycerol, erythritol and arabitol, but did not find significant amounts.

Discussion

Based on the results presented in the previous section, we have documented that the engineered strains were able to produce 6-MSA. To further optimize the 6-MSA production an *in silico* guided approach was taken. Model predictions are only as valid as the data on which the model was built and thus must be evaluated carefully to guide any optimization process. Thus the *in silico* predicted metabolic engineering targets were carefully evaluated before an experimental course was set. Firstly, transport reactions were removed as they are often artificial reactions added to the model allowing for the diffusion of products in and out of the cell. Moreover, many transporters are able to transport more than one type of metabolite making the outcome more unpredictable. In addition, targets in the pentose phosphate pathway were not prioritized as the pathway is highly regulated and thus difficult to engineer (Hankinson and Cove, 1974). Furthermore, the model tends to maximize the flux through the pentose phosphate pathway (PPP), which leads to the PPP potentially having a presumed higher impact on metabolism than is true *in vivo*.

Eventually, *gdhA* was selected as a deletion target due to the consumption of NADPH that upon deletion would be available for 6-MSA production. Studies in the yeast *S. cerevisiae* (Brochado et al., 2010) as well as in *E. coli* (Alper et al., 2005) have shown that this target can have quite beneficial effects as a metabolic engineering target. When modeling 6-MSA production in *S. cerevisiae* this gene was also reported as a top scoring candidate (Wattanachaisaereekul, 2007).

The efficient generation of gene targeting substrates by USER™ cloning combined with a well established gene targeting method lead to the stable production of 6-MSA in *A. nidulans*. The deletion of *gdhA* using the counter-selectable marker of *pyrG* made it possible to create a clean strain that was only affected by the genetic modifications. At the same time, it is still possible to further engineer the strain. This could be used to increase the copy number of the 6-MSA synthase or other metabolic engineering strategies aimed at increasing production of 6-MSA.

The insertion of an extra copy of *gdhB* resulted in a significant increase in the growth rate of the *gdhA*Δ strain (NID898). This result is comparable to results obtained in yeast where the growth of *gdh1*Δ (*gdhA*) was increased by the overexpression of *gdh2* (*gdhB*) (Nissen et al., 2000; Brochado et al., 2010).

Physiological characterization showed that the yield of 6-MSA in the exponential phase increased in NID898, although the high standard deviation should be kept in mind. One reason that the study is not entirely conclusive could be the severe growth retardation observed in the *A. nidulans* mutant, which may be overcome by inserting extra copies of *gdhB* or supplementing with glutamine in the media. Supplementation with glutamine could alleviate the growth inhibition of the *gdhA*Δ mutation and thus lead to results that show if metabolism is shifted due to the *gdhA* deletion. Alternatively, running chemostat cultivations of the strains would remove the effect of growth rate and give a better insight into the shifts in metabolism resulting from the *gdhA* deletion. As the growth rate is low the chemostats should be run at a low dilution rate, which is challenging to maintain with filamentous fungi. Alternatively, the growth rate of the mutant may be increased by the use of adaptive evolution where one continuously selects for the faster growing strain.

A challenge for further applications of *in silico* approaches to guide metabolic engineering is improvement of the genomes-scale model of *A. nidulans* as well as optimization algorithms that may be tailored into taking regulation into account. The fact that the *A. nidulans* model (*i*HD666) only contains 666 genes and 1095 reactions whereas the newest *E. coli* model (*i*JO1366) contains 1366 genes and 2251 reactions (Orth et al., 2011) indicates that there is still room for improvement of the *A. nidulans* model. From the higher complexity of *A. nidulans* it may be argued that the model should contain even more reactions than the *E. coli* model. The lack of reactions is probably also due to the fact that the *A. nidulans* genome is less annotated. New modelling information may also be gained from the knowledge assembled in the *A. niger* and *A. oryzae* genome scale models (Andersen et al., 2008; Vongsangnak et al., 2008). In addition the *S. cerevisiae* and *E. coli* models have been trained on a lot of physiological evidence from mutant strains. The lesser amount of data available for *A. nidulans* results from the fact that mutants of *A. nidulans* have to be constructed whereas *S. cerevisiae* single mutants can be ordered from EUROSCARF (EUROSCARF, 2012) as well as *E. coli* strains from Keio collection (Baba et al., 2006) and the the Coli Genetic Stock Center (CGSC) at Yale. On top of these collections 184.624 double knockout mutants from *S. cerevisiae* metabolism have been constructed and tested (Szappanos et al., 2011). In *A. nidulans* there are only very few studies of mutants that systematically target metabolism. Thus the addition of experimentally determined flux and transcriptome data to the *A. nidulans* model may help shed new light on the connections in *A. nidulans* metabolism.

These results exemplify the first use of OptGene as a guide for metabolic engineering of higher eukaryotes such as *A. nidulans.* The higher complexity of such a biological system makes it more difficult to predict the outcome of genetic modifications through the use of algorithms and genome-scale models. This can be ascribed to the highly developed regulatory mechanisms that are more prevalent in multi-cellular organisms compared to the single celled *E. coli* and *S. cerevisiae*. The ability of *A. nidulans* to counteract the modifications by changes in metabolism shows that there is still more to be done on the model building of this fungus.

Often the deletion of a gene in an organism causes severe defects that could have been avoided by up/down regulation of gene expression. Thus an optimization strategy that involves fine tuning expression as opposed to eliminating reactions may yield other interesting results. To predict up/down regulations of gene expression for metabolic engineering an alternative metabolic engineering algorithm called OptForce was developed (Ranganathan et al., 2010). It is an extension to the OptFlux metabolic engineering platform (Ranganathan et al., 2010; Gonçalves et al., 2012). OptForce was used to increase malonyl-CoA supply in order to improve the production of the flavanone naringenin (Xu et al., 2011). The authors chose to use the precursor malonyl-CoA as the optimization target and not the final product naringenin. The use of the polyketide precursor malonyl-CoA as an optimization target may be another way of getting new metabolic engineering targets that will give higher yields of 6-MSA.

A. nidulans already has a secondary metabolism requiring a large pool of acyl-CoA that could be made available for polyketide production. Detailed studies of acyl-CoA availability and cofactor balances through advanced HPLC analysis could shed light on existing precursor pools and the effects of the deletion of *gdhA* on those. Along those lines precursor availability could also be increased by creating knockouts of the major polyketides in *A. nidulans* such as sterigmatocystin that are known to be expressed under the cultivation conditions (Yu and Leonard, 1995). This should ultimately free acetyl-CoA for novel polyketide production. A similar approach includes the deletion of the actinorhodin cluster in *S. lividans* to increase expression of heterologous polyketides (Tang et al., 1999). As three units of malonyl-CoA are required for the production of 6-MSA, the overexpression of *ACC1* that converts acetyl-CoA to malonyl-CoA could have a beneficial effect in *A. nidulans* as it has been shown for *S. cerevisiae* (Wattanachaisaereekul et al., 2008).

In terms of optimizing polyketide production, it has been shown that there is an effect of the pH on the level of sterigmatocystin biosynthesis in *A. nidulans* (Keller et al., 1997). Thus controlled studies of different pH levels in bioreactors could lead to higher 6-MSA productivity.

To summarize there are still many points of entry to further optimize the polyketide production using *A. nidulans* as a cell factory.

Conclusion

We established efficient production of the model polyketide 6-MSA in the heterologous host *A. nidulans.* For the first time ever the model *i*HD666 has been used with the algorithm OptGene to predict targets for genetic modifications that could improve chemical production in *A. nidulans*. The target *gdhA*Δ was successfully engineered in the *A. nidulans* strain and crossed with the 6-MSA producing strain. The *gdhA*Δ led to severe growth retardation, which was partially rescued by insertion of an extra copy of *gdhB.* This study shows that the *A. nidulans* cell factory can produce significant amounts of heterologous polyketides and further metabolic engineering may pave the way for harnessing this enormous potential.

Acknowledgements

The research was supported by a grant from the Danish Research Council for Technology and Production grant nr. 09-064240 to Kiran Raosaheb Patil and Uffe Hasbro Mortensen.

References

Alper, H., Jin, Y.S., Moxley, J.F., Stephanopoulos, G., 2005. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab. Eng. 7, 155-64.

Andersen, M.R., Nielsen, M.L., Nielsen, J., 2008. Metabolic model integration of the bibliome, genome, metabolome and reactome of Aspergillus niger. Molecular systems biology 4.

Asadollahi, M.A., Maury, J., Patil, K.R., Schalk, M., Clark, A., Nielsen, J., 2009. Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through *in silico* driven metabolic engineering. Metab. Eng. 11, 328-34.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H., 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular systems biology 2.

Bedford, D.J., Schweizer, E., Hopwood, D.A., Khosla, C., 1995. Expression of a functional fungal polyketide synthase in the bacterium *Streptomyces coelicolor* A3 (2). J. Bacteriol. 177, 4544.

Brochado, A.R., Matos, C., Møller, B.L., Hansen, J., Mortensen, U.H., Patil, K.R., 2010. Improved vanillin production in baker's yeast through in silico design. Microbial Cell Factories 9, 84.

Clutterbuck, A.J., 1974. *Aspergillus nidulans*. Handbook of genetics 1, 447-510.

Cove, D., 1966. The induction and repression of nitrate reductase in the fungus Aspergillus nidulans. Biochimica et Biophysica Acta (BBA)-Enzymology and Biological Oxidation 113, 51-6.

David, H., Özçelik, İ.Ş., Hofmann, G., Nielsen, J., 2008. Analysis of *Aspergillus nidulans* metabolism at the genome-scale. BMC Genomics 9, 163.

Droste, P., von Lieres, E., Wiechert, W., Nöh, K., 2010. Customizable Visualization on Demand for Hierarchically Organized Information in Biochemical Networks. Computational Modeling of Objects Represented in Images , 163-74.

EUROSCARF, 2012. http://web.uni-frankfurt.de/fb15/mikro/euroscarf/col_index.html. Accessed 20th of April 2012.

Galagan, J., Calvo, S., Cuomo, C., Ma, L., Wortman, J., Batzoglou, S., Lee, S., Basturkmen, M., Spevak, C., Clutterbuck, J., Kapitonov, V., Jurka, J., Scazzocchio, C., Farman, M., Butler, J., Purcell, S., Harris, S., Braus, G., Draht, O., Busch, S., D'Enfert, C., Bouchier, C., Goldman, G., Bell-Pedersen, D., Griffiths-Jones, S., Doonan, J., Yu, J., Vienken, K., Pain, A., Freitag, M., Selker, E., Archer, D., Penalva, M., Oakley, B., Momany, M., Tanaka, T., Kumagai, T., Asai, K., Machida, M., Nierman, W., Denning, D., Caddick, M., Hynes, M., Paoletti, M., Fischer, R., Miller, B., Dyer, P., Sachs, M., Osmani, S., Birren, B., 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with A-fumigatus and A-oryzae. Nature 438, 1105-15.

GenomeNet, 2012. www.genome.jp. Accessed 25th of April 2012.

Geu-Flores, F., Nour-Eldin, H.H., Nielsen, M.T., Halkier, B.A., 2007. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Res. 35, e55.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature 418, 387-91.

Gonçalves, E., Pereira, R., Rocha, I., Rocha, M., 2012. Optimization Approaches for the In Silico Discovery of Optimal Targets for Gene Over/Underexpression. Journal of Computational Biology 19, 102-14.

Hankinson, O., Cove, D.J., 1974. Regulation of the pentose phosphate pathway in the fungus *Aspergillus nidulans*. J. Biol. Chem. 249, 2344-53.

Hansen, B.G., Salomonsen, B., Nielsen, M.T., Nielsen, J.B., Hansen, N.B., Nielsen, K.F., Regueira, T.B., Nielsen, J., Patil, K.R., Mortensen, U.H., 2011. A versatile gene expression and characterization system for Aspergillus: heterologous expression of the gene encoding the polyketide synthase from the mycophenolic acid gene cluster from *Penicillium brevicompactum* as a case study. Appl. Environ. Microbiol. .

Kealey, J.T., Liu, L., Santi, D.V., Betlach, M.C., Barr, P.J., 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proceedings of the National Academy of Sciences 95, 505.

Keller, N.P., Nesbitt, C., Sarr, B., Phillips, T.D., Burow, G.B., 1997. pH regulation of sterigmatocystin and aflatoxin biosynthesis in Aspergillus spp. Phytopathology 87, 643-8.

Morozov, I., Galbis-Martinez, M., Jones, M., Caddick, M., 2001. Characterization of nitrogen metabolite signalling in Aspergillus via the regulated degradation of areA mRNA. Mol. Microbiol. 42, 269-77.

Newman, D.J., Cragg, G.M., 2007. Natural Products as Sources of New Drugs over the Last 25 Years⊥. J. Nat. Prod. 70, 461-77.

Nielsen, J.B., Nielsen, M.L., Mortensen, U.H., 2008. Transient disruption of non-homologous end-joining facilitates targeted genome manipulations in the filamentous fungus *Aspergillus nidulans*. Fungal Genetics and Biology 45, 165-70.

Nielsen, M.L., Albertsen, L., Lettier, G., Nielsen, J.B., Mortensen, U.H., 2006. Efficient PCR-based gene targeting with a recyclable marker for *Aspergillus nidulans*. Fungal Genetics and Biology 43, 54-64.

Nissen, T.L., Kielland-Brandt, M.C., Nielsen, J., Villadsen, J., 2000. Optimization of Ethanol Production in *Saccharomyces cerevisiae* by Metabolic Engineering of the Ammonium Assimilation. Metab. Eng. 2, 69-77.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.Ø., 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. Molecular systems biology 7.

Otero, J.M., 2009. Industrial Systems Biology and Metabolic Engineering of *Saccharomyces cerevisiae* A case study in succinic acid production.

Panagiotou, G., Andersen, M.R., Grotkjaer, T., Regueira, T.B., Nielsen, J., Olsson, L., 2009. Studies of the production of fungal polyketides in *Aspergillus nidulans* by using systems biology tools. Appl. Environ. Microbiol. 75, 2212-20.

Patil, K.R., Rocha, I., Förster, J., Nielsen, J., 2005. Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinformatics 6.

Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS computational biology 6, e1000744.

Regueira, T.B., Kildegaard, K.R., Hansen, B.G., Mortensen, U.H., Hertweck, C., Nielsen, J., 2011. Molecular basis for mycophenolic acid biosynthesis in *Penicillium brevicompactum*. Appl. Environ. Microbiol. 77, 3035-43.

Roumelioti, K., Vangelatos, I., Sophianopoulou, V., 2010. A cryptic role of a glycolytic-gluconeogenic enzyme (aldolase) in amino acid transporter turnover in *Aspergillus nidulans*. Fungal Genetics and Biology 47, 254-67.

Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual. CSHL press.

Spencer, J., Jordan, P., 1992. Purification and properties of 6-methylsalicylic acid synthase from *Penicillium patulum*. Biochem. J. 288, 839.

Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M.J., Jelasity, M., Myers, C.L., 2011. An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat. Genet. 43, 656-62.

Tang, L., Fu, H., Betlach, M.C., McDaniel, R., 1999. Elucidating the mechanism of chain termination switching in the picromycin/methymycin polyketide synthase. Chem. Biol. 6, 553-8.

Thykaer, J., Rueksomtawin, K., Noorman, H., Nielsen, J., 2008. NADPH-dependent glutamate dehydrogenase in *Penicillium chrysogenum* is involved in regulation of β-lactam production. Microbiology 154, 1242-50.

Tsai, S., Ames, B.D., 2009. Structural Enzymology of Polyketide Synthases. Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates 459, 17-47.

Vongsangnak, W., Olsen, P., Hansen, K., Krogsgaard, S., Nielsen, J., 2008. Improved annotation through genome-scale metabolic modeling of *Aspergillus oryzae*. BMC Genomics 9, 245.

Wattanachaisaereekul, S., 2007. Production of Polyketides by *Saccharomyces cerevisiae*. Ph. D. Thesis Technical University of Denmark.

Wattanachaisaereekul, S., Lantz, A.E., Nielsen, M.L., Nielsen, J., 2008. Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. Metab. Eng. 10, 246-54.

Wattanachaisaereekul, S., Lantz, A.E., Nielsen, M.L., Andresson, O.S., Nielsen, J., 2007. Optimization of Heterologous Production of the Polyketide 6-MSA in *Saccharomyces cerevisiae*. Biotechnology and Bioengineering 97, 893-900.

Weissman, K.J., 2009. Introduction to Polyketide Biosynthesis. Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates 459, 3-16.

Xu, P., Ranganathan, S., Fowler, Z.L., Maranas, C.D., Koffas, M.A.G., 2011. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab. Eng. .

Yu, J.H., Leonard, T.J., 1995. Sterigmatocystin biosynthesis in *Aspergillus nidulans* requires a novel type I polyketide synthase. J. Bacteriol. 177, 4792.

Supplementary

S1 Primers and Plasmids for strain construction

The primers AGAgcgaUatgcattccgctgcaacttc and TCTgcgaUttaatggtgatggtgatga were used to generate the 6-MSAS gene fragment for insertion into BGHAp71 to create BGHAp75 with USER® cloning (see figure 2).

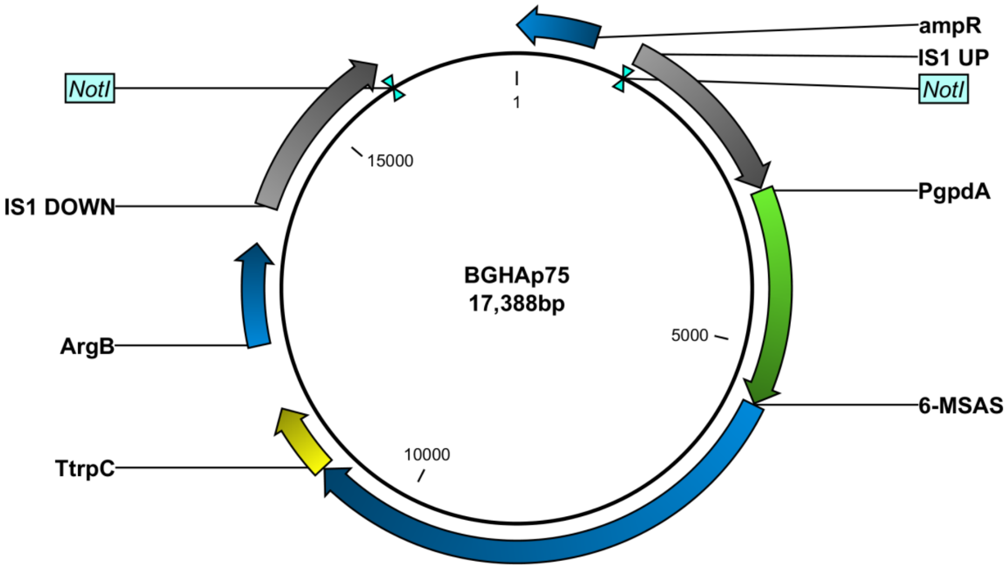


Figure 2 The plasmid BGHAp75. The plasmid contains the *msas* polyketide synthase gene with the gpdA promoter in front of the gene and the trpC terminator after the gene. The selectable marker used is ArgB. Furthermore the gray sequences are complementary to the insertion site 1 on the *A. nidulans* genome. The *NotI* restriction sites were used for excision of the gene targeting substrate.

In table 5 and 6 the primers used for amplification of gene targeting substrates are given. The pyrG fragment flanked by direct repeats was amplified from the plasmid pDel2 for the *gdhA* deletion (Nielsen et al., 2008). Whereas *pyrG* was amplified without direct repeats from the plasmid PCR2.1 and used for the *gdhA::gdhB* strain. All other fragments were amplified from the *A. nidulans* strain IBT27263 (*argB2, pyrG89, veA1*). The strain is derived from G051 of the Glasgow strain collection (Clutterbuck, 1974). The assembly of the different fragments and the gene targeting method is illustrated in figure 3. Briefly the primers ANgdhA-dl-Up-F and ANgdhAΔ::gdhB-Up-R were used to generate a sequence homologous to the upstream region of the *gdhA* gene. The primers ANgdhAΔ::gdhB-Dw-F and ANgdhA-dl-Dw-R were used to generate the corresponding downstream homologous sequence. Then the upper 2/3 and lower 2/3 of *pyrG* from *A. fumigatus* was amplified using the primers pDEL-Up-F-Ad (5A) and AFpyrG-int-F3 (4Q) and AFpyrG-int-R (2K) and pDEL-Dw-R-Ad (2B). Finally, the bipartite substrate was constructed by fusing the upstream Up-gdhA with the upstream *pyrG* fragment and the downstream targeting fragment gdhA-Dw to the downstream fragment of *pyrG* using the primers ANgdhA-dl-Up-F and AFpyrG-int-F3 (4Q) and AFpyrG-int-R (2K) and ANgdhA-dl-Dw-R, respectively. The *pyrG* is surrounded by direct repeats. Thus on 5-FOA plates *pyrG* is excised by direct repeat recombination.

Table 5 Primers used for construction of the gene targeting substrates for the deletion of *gdhA*.

***gdhA*Δ**	**Primer Sequence 5´-> 3'**	**Fragment**
ANgdhA-dl-Up-F	GACTGCCGAAGTAAGAGCGCGG	Up-gdhA
ANgdhA-dl-Up-Rad	gatccccgggaattgccatgGAATGCGCTGAGCCCGCG	
ANgdhA-dl-Dw-Fad	aattccagctgaccaccatgGTGGTGAATTAGCCCCGTCTC	gdhA-Dw
ANgdhA-dl-Dw-R	CCTGGACTATCAGCTGATGGAG	
pDEL2-*pyrG* (AF)		
pDEL-Up-F-Ad (5A)	catggcaattcccggggatcTGGATAACCGTATTACCGCC	5´pyrG
AFpyrG-int-F3 (4Q)	TGATACAGGTCTCGGTCCC	
AFpyrG-int-R (2K)	GGAAGAGAGGTTCACACC	3´ pyrG
pDEL-Dw-R-Ad (2B)	catggtggtcagctggaattTGCCAAGCTTAACGCGTACC	

The deletion of *gdhA* combined with an insertion of an extra copy of *gdhB* was achieved in a similar manner to the one described for *gdhA* and with the primers given in table 5. The main difference is that the *pyrG* gene does not contain direct repeats. Instead part of the *gdhB* gene is used as direct repeats as indicated by the shading in figure 3.

Table 6 The primers used for the construction of the gene targeting substrates for the deletion of *gdhA* with the immediate insertion of *gdhB*.

gdhAΔ::gdhB	**Primer Sequence 5´-> 3'**	**Fragment**
ANgdhA-dl-Up-F	GACTGCCGAAGTAAGAGCGCGG	Up-*gdhA*
ANgdhAΔ::gdhB-Up-R	cgatgcgggagacacCATTTTTGCGCGAGAAGCTTATTGCTGA	
ANgdhAΔ::gdhB-Dw-F	gacttgcacagtccaaagcataAATTAGCCCCGTCTCCTAATTTTGAT	*gdhA*-Dw
ANgdhA-dl-Dw-R	CCTGGACTATCAGCTGATGGAG	
ANgdhB-start-F	ATGGTGTCTCCCGCATCG	5´gdhB
ANgdhB::pyrG-5-R	gtgaagagcattgtttgaggcCACCGTCAGGTCCACCAG	
ANgdhB::pyrG-3-F	gcctcctctcagacagaaTCCATCTGCGCTTCCGAGAC	gdhB 3´
ANgdhB-end-R	TTATGCTTTGGACTGTGCAAGTC	
pyrGX		
pDEL-Up-F-Ad (5A)	catggcaattcccggggatcTGGATAACCGTATTACCGCC	5´pyrG
pyrG_A.f-5'-F (10J)	GCCTCAAACAATGCTCTTCAC	
pyrG_A.f-3'-R (1K)	ATTCTGTCTGAGAGGAGGC	3´ pyrG
pDEL-Dw-R-Ad (2B)	catggtggtcagctggaattTGCCAAGCTTAACGCGTACC	

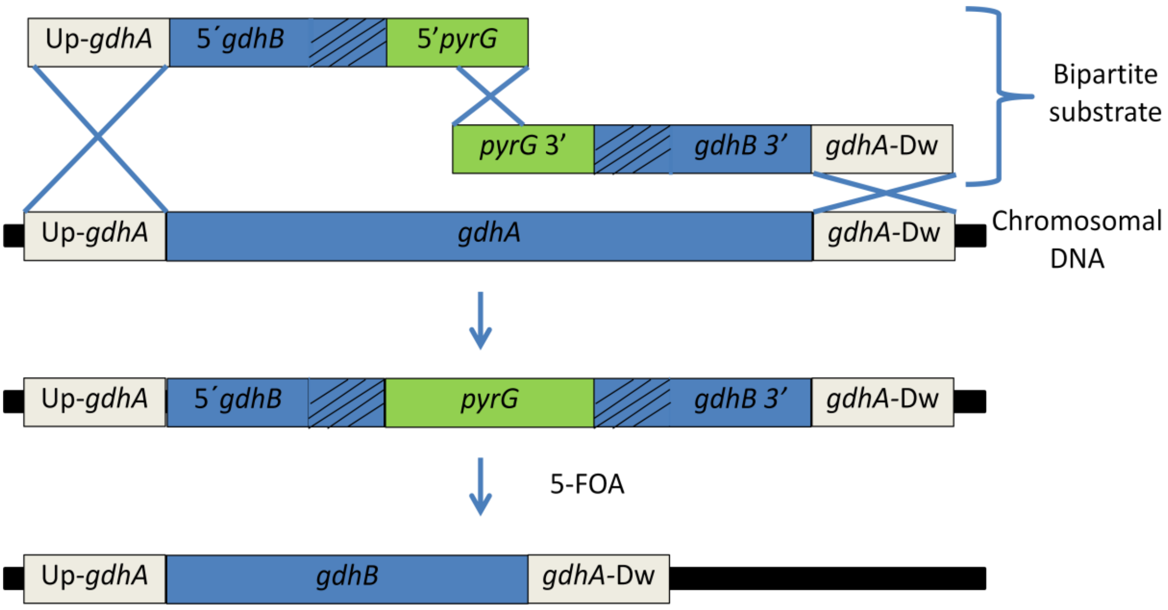


Figure 3 The construction of the deletion of *gdhA* along with the insertion of *gdhB* in *A. nidulans* by the gene targeting method of Nielsen *et al*. (Nielsen et al., 2006). The deletion of *gdhA* was performed in an analogous manner with only *pyrG* flanked by direct repeats and up and down stream regions of the *gdhA* gene in the targeting substrates. Up-gdhA is a sequence homologous to the upstream region of the *gdhA* gene and gdhA-Dw is a sequence homologous to a region downstream of the *gdhA* gene. The shaded regions of *gdhB* are used to generate a direct repeat for excision of *PyrG* by homologous recombination on 5-FOA. This results in the generation of a single copy of *gdhB* that was validated by sequencing.

The primers used for diagnostic PCRs and verification during crossing are given in table 7. For the verification of gene targeting into insertion site 1 (IS1) a primer pair consisting of a primer upstream or downstream of the insertion fragment and a primer within the gene targeting substrate was used. This generated both an up- and down- stream verification of the insertion. The negative control gave a small band of 500 bp if no insertion took place. In the case of *nkuA* the primers were located on either side of the locus and thus the wildtype produced a large band and the mutant a small band. The same was true for the *gdhA* locus verification.

Table 7 The primers used for verification of gene targeting and *nkuA* elimination.

Gene	Primers	Fragment size	
		Wildtype	**Recombinant**
IS1::*msas* (upstream)	CTCACTCGCCTCTCGTTGC AAGCTGTTGGCAGCCTTAAA	-	2600 bp
IS1::*msas* (downstream)	GGTTTCGTTGTCAATAAGGGAA GAGGTGGCGGCTTCGGAG	-	2499 bp
IS1:: (negative control)	GGACAACGGGAAGAGGCTCAG GGAGAGGGAGAGAAGAAGAAGGG	500 bp	> 10,000 bp (not observed)
gdhA	GGTCGTGTGCCTCTCTTGG GCCCTGCATTGAGAACATCT	3512 bp	2026 bp
gdhA::gdhB	GGTCGTGTGCCTCTCTTGG GGAATCATCATCATGCAACATC	1689 bp	3670 bp
nkuA	GAGGTTACCTCAGATCTTG CGAGTGCACAGCACAGCTG	2200 bp	550 bp

S2 Nitrogen metabolism in *A. nidulans*

Nitrogen assimilation is a central process in most bacteria and lower eukaryotes. It is an essential anabolic reaction eventually leading to the build up of amino acids. In figure 4 some of the central reactions involved in nitrogen assimilation in *A. nidulans* can be seen. In essence nitrogen is assimilated in a tightly regulated fashion depending on the available nitrogen source (Morozov et al., 2001). In the batch cultivations run with *A. nidulans* ammonia is used as a nitrogen source.

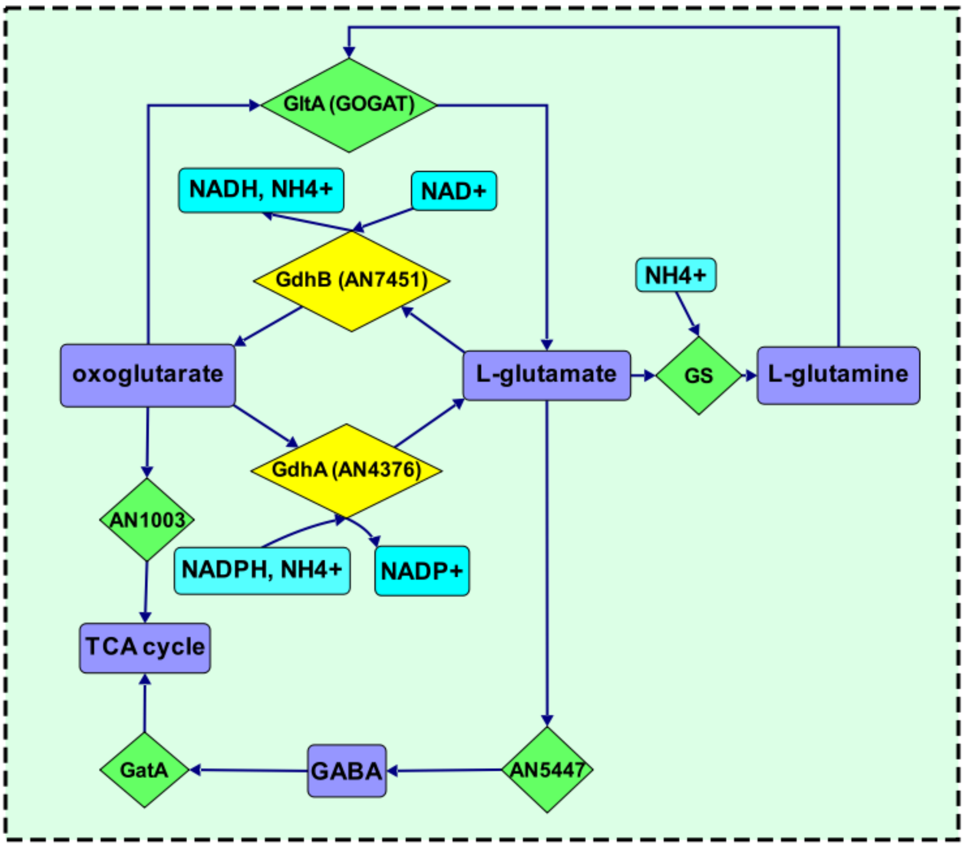


Figure 4 The assimilation of nitrogen in *A. nidulans* connected to the two glutamate dehydrogenases (Droste et al., 2010; GenomeNet, 2012) . The central part of the figure shows the reactions involving the two glutamate dehydrogenases NADP-GDH (*gdhA)* and NAD-GDH (*gdhB*) (yellow). NADP-GDH utilizes NADPH in order to form L-glutamate. The reaction is primarily anabolic. On the other hand NAD-GDH (*gdhB)* uses the cofactor NADH to form L-glutamate albeit the reaction most often runs in the opposite direction generating NADH in catabolic reactions. GS: glutamine synthetase, GltA: glutamate synthase, GatA: 4 amino butyrate transaminase, AN5447: putative glutamate decarboxylase, AN1003: putative isocitrate dehydrogenase. Cofactors are only shown for the GdhB and GdhA reactions.

The glutamate dehydrogenase known as GdhA (AN4376, Chromosome III) is responsible for incorporating nitrogen in amino acids and thus serves an important function in nitrogen metabolism. GdhB (AN7451, Chromosome IV) is primarily responsible for the break-down of glutamate into oxoglutarate that is part of the TCA cycle. GdhA and GdhB are distinguished by their use of either NADH (GdhB) or NADPH (GdhA) as cofactors and are thus also referred to as NAD-GDH or NADP-GDH respectively.

S3 Model Predictions

Table 8 Model predictions with OptGene for knockout targets to improve 6-MSA production. The targets were sorted based on maximum biomass product coupled yield (BPCY). Pentose phosphate pathway (red), TCA cycle (blue), *S. cerevisiae*/*A. nidulans* inviable (green). *This gene is no longer annotated in the *A. nidulans* genome (AspGD).

Gene	BPCY	Reaction
AN2440.2;	0.007445	D-Ribulose 5-phosphate <-> D-Ribose-5-phosphate
AN0688.2a;	0.005996	D-Ribose-5-phosphate+ D-Xylulose-5-phosphate <-> seduheptulose-7- phosphate + D-glyceraldehyde-3-phosphate
AN0688.2b;	0.005204	D-erythrose 4-phosphate + Xylulose-5-phosphate <-> beta-D-fructose-6-phosphate + D-glyceraldehyde-3-phosphate
AN2409.2*;	0.00513	3-phosphonooxypyruvate + glucose -> 2-oxoglutarate + 3-phosphoserine
AN8866.2;	0.00513	3-phospho-D-glycerate + NAD+ -> 3-phosphonooxypyruvate + NADH
AN6717.2;	0.00495	S-malate (mitochondrial) + NAD(mitochondrial) <-> Oxaloacetate (mitochondrial) + NADH (mitochondrial)
AN6900.2;	0.004801	Glycerone phosphate <-> D-glyceraldehyde 3-phosphate
AN8707.2a;	0.004521	Fumarate (mitochondrial) + H2O (mitochondrial) <-> S-malate (mitochondrial)
AN7588.2;	0.004269	Ribulose-5-phosphate <-> Xylulose-5-phosphate
AN2981.2;	0.004097	Alpha-D-glucose-6-phosphate + NADP -> D-glucono-1,5-lactone 6-phosphate + NADPH
AN5629.2/AN6077.2/ AN1728.2/AN2414.2;	0.003945	NADH (mitochondrial) + Ubiquinone (mitochondrial) + 4 H+ (proton mitochondrial) -> NADm + Ubiquinol (mitochondrial) + 4 H+ (proton)
AN2916.2/AN2332.2/ AN8793.2;	0.003727	Succinate (mitochondrial) + Ubiquinone (mitochondrial) <-> fumarate (mitochondrial) + Ubiquinol (mitochondrial)
AN2875.2b;	0.003461	Seduheptolose 1, 7 bisphosphate <-> glycerone phosphate + D-erythrose-4-phosphate
AN3223.2b;	0.003093	ATP + sedoheptulose-7-phosphate-> ADP + seduheptulose-1,7-bisphosphate
AN7459.2a;	0.002952	ATP + Mannose -> ADP + D-Mannose-6-Phosphate
AN3432.2a;	0.00284	UDP-galactose <-> UDP-glucose
AN3954.2;	0.002737	6-phospho-D-gluconate + NADP -> D-ribulose- 5-phosphate + CO2 + NADPH
AN3058.2;	0.002723	tetrahydrofolate + L-serine <-> glycine + 5, 10 methylenetetrahydrofolate
AN3741.2c;	0.002465	glycerol + NAD -> D-glyceraldehyde + NADH
AN4684.2a;	0.001991	triacylglycerol + H2O -> diacylglycerol + 0.0821 C120ACP + 0.0444 C140ACP + 0.0407 C141ACP + 0.0081 C150ACP + 0.5161 C160ACP + 0.0681 C161ACP + 0.0276 C162ACP + 0.0039 C170ACP + 0.0860 C180ACP + 1.2429 C181ACP + 1.0870 C182ACP + 0.1051 C183ACP + 0.0200 C200ACP (*fatty acids*)
AN4323.2/AN7878.2/ AN5957.2d;	0.001877	(R)-2-oxoisovalerate + glucose <-> 2-oxoglutarate + L-valine
AN4376.2;	0.001756	Oxoglutarate+NADPH -> glutamate and NADP+

S4 Southern Blot

Table 9 Primers used for generating southern blot probe Up-*gdhA*.

Name	Primer Sequence 5´-> 3’	Probe
ANgdhA-dl-Up-F	GACTGCCGAAGTAAGAGCGCGG	Up-*gdhA*
ANgdhAΔ::gdhB-Up-R	cgatgcgggagacacCATTTTTGCGCGAGAAGCTTATTGCTGA	

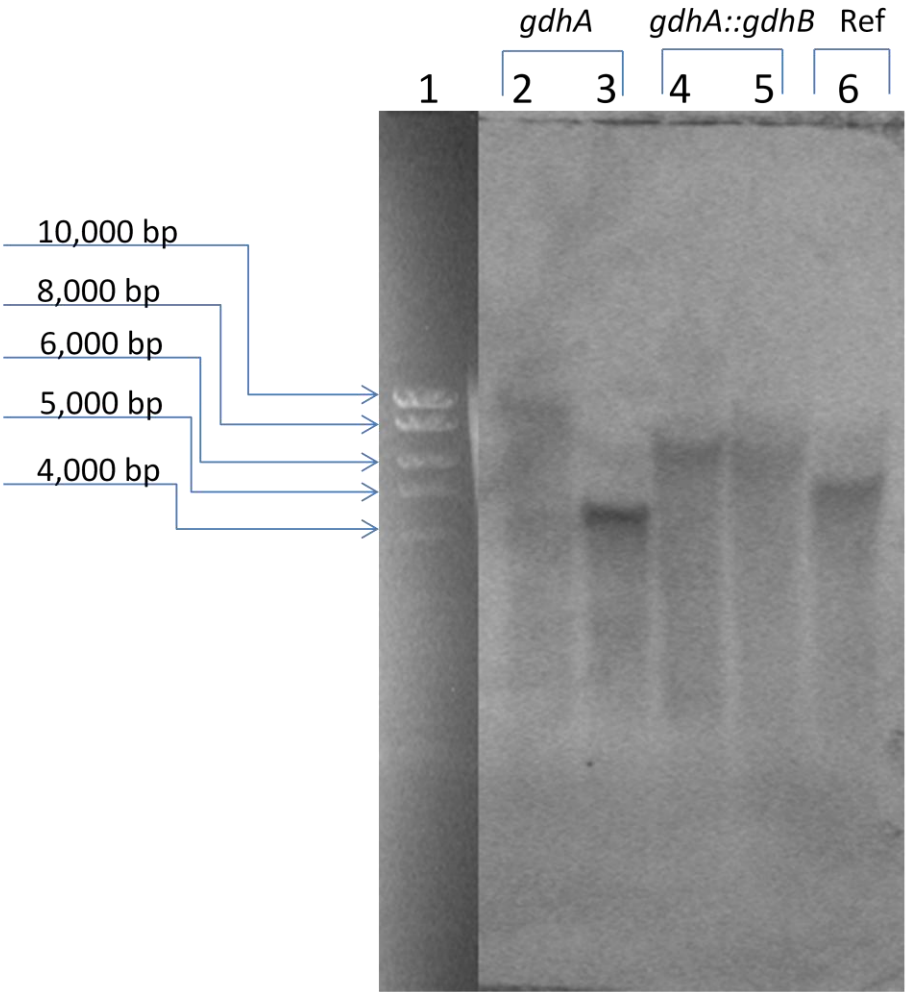
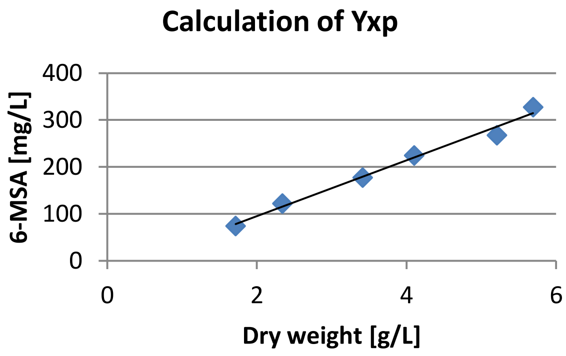


Figure 5 Southern blot of the gdhA and gdhB mutants. In parenthesis is genotype followed by the expected band size. Lane: 1. Ladder (Hyper Ladder 1 (Bioline)), 2. NID391 (*argB2*, *pyrG89*, *veA1*, IS1::PpgdA-6MSAS-TtrpC::*argB, gdhAΔ*) (4300 bp – no band observed), 3. NID392 (*argB2*, *pyrG89*, *veA1*, IS1::PpgdA-6MSAS-TtrpC::*argB, gdhAΔ*) (4300 bp), 4. NID393 (*argB2*, *pyrG89*, *veA1*, IS1::PpgdA-6MSAS-TtrpC::*argB, gdhAΔ::gdhB*) (7400 bp), 5. NID394, *argB2*, *pyrG89*, *veA1*, IS1::PpgdA-6MSAS-TtrpC::*argB, gdhAΔ::gdhB*) (7400 bp), 6. IBT29539 (*argB2, pyrG89, veA1, nkuA*Δ) (5600 bp).

The Southern blot in figure 5 verified the deletion of *gdhA* and the deletion of *gdhA* with the insertion of *gdhB* in the strains NID392 and NID393.

S5 Calculation of Yields and Growth Rate

The yields and growth rates were calculated based on the slope of a least squares approximation in the exponential phase as depicted in figure 6.



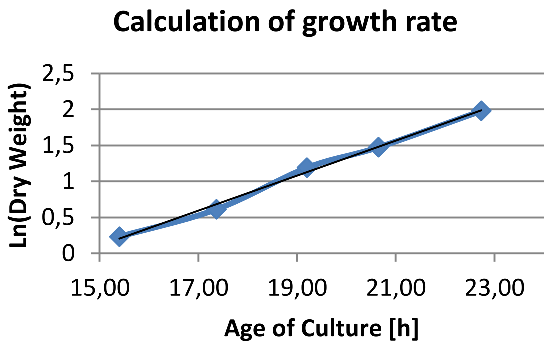


Figure 6 The growth rate was calculated based on the best linear fit of at least 5 data points of the age of the culture vs. ln of the dry weight in the exponential phase (the example is from NID521f4). The yield of 6-MSA on dry weight was calculated based on the slope of a linear fit between biomass concentration and 6-MSA production in the exponential growth phase (the example is from NID875 f4).

S6 Fermentation profiles

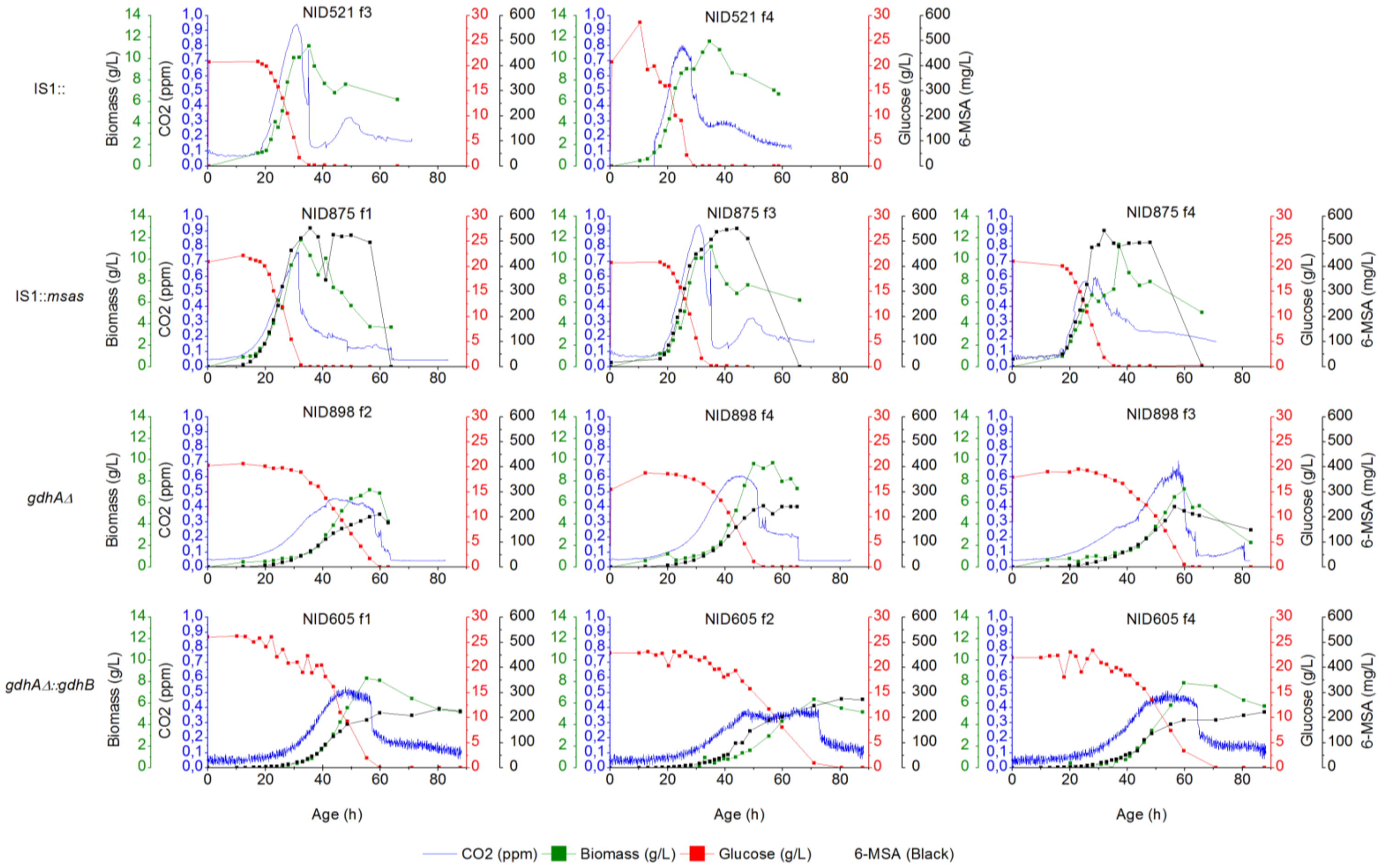


Figure 7 Fermentation graphs from 2 L batch fermentations with 20 g/L glucose. In row 1 (column 1 and 2) are the profiles for NID521(IS1::PgpdA-TtrpC::*argB),* in row two are the profiles NID875 (IS1::PpgdA-6MSAS-TtrpC::*argB*), in row three are the profiles of NID898 (IS1::PpgdA-6MSAS-TtrpC::*argB, gdhAΔ*) . In row four are the profiles for NID605 (IS1::PpgdA-6MSAS-TtrpC::*argB, gdhAΔ::gdhB*). All of the axes have been set to the same scale.

In this chapter a novel gene amplification system is presented that is under development in our lab. The people that have been involved with construction of the system and should be rightfully acknowledged are Christina Spuur Nødvig, Line Due Buron, Tomas Strucko and Zofia Jarczynska.

Chapter 5

Novel gene amplification system for 6-MSA production in *S. cerevisiae*

Abstract

The use of *Saccharomyces cerevisiae* as a platform for the production of biopharmaceuticals and chemicals is well-established. The need to achieve high titers in respect to production processes is an everlasting demand. The first challenge in construction of a production host is having high expression of all the genes involved in the biosynthetic pathway. This has long been achieved through the use of 2µ based multicopy plasmid systems that are fast to construct and engineer. However, some important drawbacks exist in the 2µ based plasmid system. Firstly, the plasmid based systems achieve high copy numbers, but are challenged in terms of stability. If the number of different plasmids needed to achieve expression increases, the stability issue becomes an even greater challenge. Secondly, the ability to adjust levels of different enzymes in a pathway to avoid bottlenecks is not easily achievable in a plasmid based system. Thirdly, the population of cells in a plasmid based system is heterogeneous which complicates physiological analysis of the individual cells. In this paper the use of a gene amplification system that allows for the stable integration of genes in multiple copies is presented. The systems applicability is demonstrated on the polyketide synthase MSAS that must be activated by the PPTase NpgA leading to the production of the polyketide 6-MSA. The genes were amplified in one, two and four copies. Cultivation of the *S. cerevisiae* strains in shake flasks resulted in 6-MSA titers that correlated well with gene copy number.

Introduction

Polyketides constitute a large group of natural products with importance in medicine (e.g. antibiotics) as well as undesirable properties (e.g. toxins) (Weissman, 2009). Polyketides are produced by large enzymes known as polyketide synthases (PKSs). In order for the polyketide synthases to function they have to be converted from their apo form into their holo form. This occurs by the action of a phosphopantetheinyl transferase (PPTase) that tethers a phosphopantetheinyl group onto the PKS (Tsai and Ames, 2009). Commonly used PPTases for heterologous expression are Sfp from *bacillus subtilis* and NpgA from *Aspergillus nidulans* as they are able to activate an array of different PKSs.

The production of polyketides with therapeutic properties is quite a challenge. If the genetics of the natural polyketide producers is not well known, one is often left with time-consuming random mutagenesis and screening for improving production. Even so productivity can still be quite low. Alternatively, there have been efforts to chemically synthesize polyketides, but so far production has not yet achieved quantities that have made the process relevant for commercial production. A different method is to express the PKS and PPTase in a heterologous host that can offer the benefit of being a well known production organism. One well-studied production organisms is the yeast *Saccharomyces cerevisiae*. The heterologous expression of polyketides in yeast has been accomplished in large by the use of plasmid based expression systems. A well-studied model polyketide is 6-methyl salicylic acid (6-MSA). Kealey *et al.* and Wattanachaisaereekul *et al.* have both achieved high titers of 6-MSA by the use of 2μ plasmid expression in *S. cerevisiae* (Kealey et al., 1998; Wattanachaisaereekul et al., 2007).

In the study performed by Wattanachaisaereekul *et al.* expression of *msas* and *npgA* was achieved using a system of two 2µ based plasmids. The *npgA* gene was located on a 2µ plasmid with the *TRP1* marker and the *msas* gene was located on a 2µ plasmid with *URA3* marker. Upon expression in *S. cerevisiae* on a minimal medium they achieved a final titer of 150 mg/L of 6-MSA with 20 g/L glucose in the media in controlled bioreactors (Wattanachaisaereekul et al., 2008).

Plasmid based expressions systems offer a quick way to provide proof of principle in terms of expression of polyketide genes in a heterologous host. Furthermore, high titers can be achieved by multicopy plasmids such as the 2µ based systems in yeast. The overexpression of genes is one of the most commonly used genetic tools both in academic research as well as industrial production strains. From a production perspective high expression levels often leads to higher titers of the product. In academia high expression levels can be used to study the physiological effects of overexpression of a specific gene. If the 2µ based system does not provide equal stable expression in all cells the effects observed from such an analysis may not be correct. Furthermore, when testing if genes have a dominant negative (e.g. toxic) effect after being overexpressed, it will also be difficult to access if only a proportion of the cellular population harbors the desired overexpression of a gene.

Several studies have documented that 2µ based expression systems are not stable and plasmids are readily lost (Futcher and Cox, 1983; Futcher and Cox, 1984; Albertsen et al., 2011). Plasmid stability has also been shown to be dependent on the media used, the selection markers applied, size of the plasmid and process conditions (Zhang et al., 1996). In addition, Albertsen *et al.* demonstrated that individual cells in a culture expressing a heterologous protein tagged with either yellow fluorescent protein (YFP) or cyan fluorescent protein (CFP) as an N-terminal fusion showed a great variability in expression level (figure 1) (Albertsen et al., 2011). Figure 1 clearly demonstrates that only a proportion of the yeast cells shown in the brightfield image (DIC) are fluorescing (CFP, YFP) and thus expressing protein at a detectable level. On top of this there is a great variability between the dimmest and brightest fluorescing cells indicating that protein expression is not even within the cellular population.

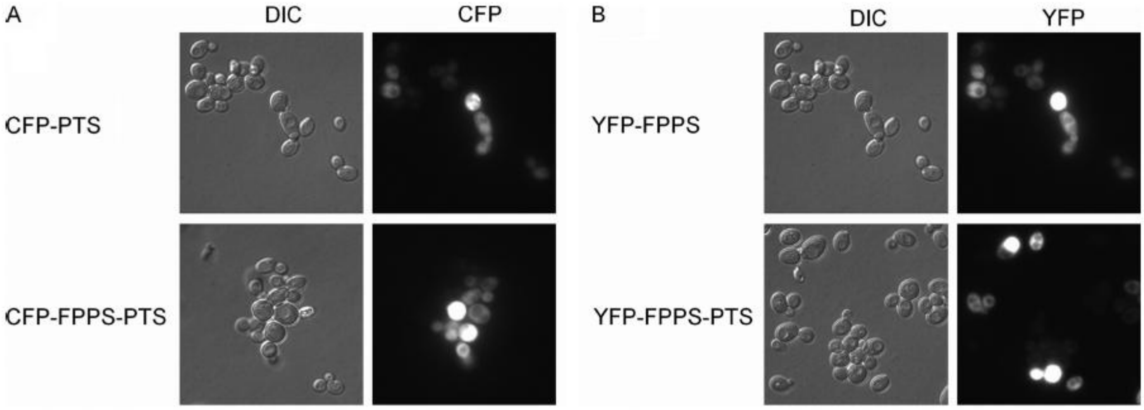


Figure 1 The figure shows the fluorescence of CFP and YFP of cells expressing proteins for the production of sesquiterpenes. In the left panel is pictures obtained through differential interference contrast microscopy (DIC) and in the right panel CFP (figure A) or YFP (figure B) fluorescence of the 2µ based protein expression. The figure is reproduced from Albertsen *et al.* (Albertsen et al., 2011).

The expression of the CFP and YFP fusion proteins were analyzed by FACS and the results are illustrated in figure 2 (Albertsen et al., 2011). The results of Albertsen *et al.* further substantiate the claim that there is a large variation in CFP and YFP expression levels within a population of cells when the proteins are expressed from a 2µ plasmid. The lower left corner shows that for about half of the cells, the fluorescence signal was below the detection level. The top right corner of the figure shows the linear correlation expected as the CFP and YFP are expressed from the same plasmid. However, as the scale is logarithmic there is a 100 times difference between the level of expression in the lowest and highest expressing cells of the counted population.

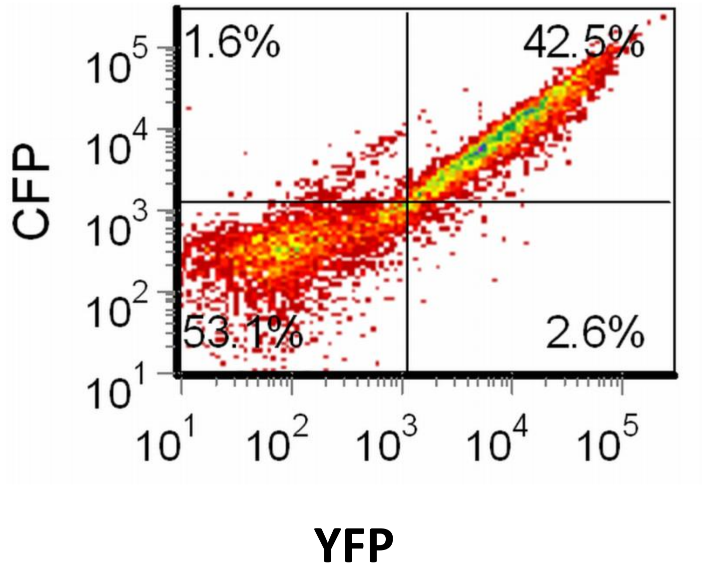


Figure 2 Cells expressing YFP and CFP on a 2µ plasmid were analyzed by FACS and the amount of fluorescence was measured. In 53.1 % of the cells the fluorescence is very low and can to some degree be attributed to noise in the measurement. The remaining 42.5 % of the cells show a linear expression as CFP and YFP are expressed in the same plasmid. However there is a 100 fold difference in expression level between the lowest and highest expressing cells.

As described, there are problems of 2µ plasmid instability and high variability of expression level within a cellular population. These problems can be overcome by stably integrating the genes of interest in the yeast genome through homologous recombination. However, one integrated copy of a PKS gene does not suffice to achieve significant amounts of polyketide product (Wattanachaisaereekul, 2007). Thus, one needs to incorporate several copies of the gene in the genome in order to achieve titers as the ones observed when using the 2µ based plasmid systems. If homologous recombination is achieved through the use of recyclable markers, it would take considerable time to integrate the new genes and achieve marker excision by direct repeat recombination making the strain ready for another round of transformation. An alternative strategy to repeated rounds of integration could be to achieve high copy numbers via gene amplification.

Gene amplification of large pieces of DNA occurs naturally in many microorganisms as well as mammalian cells (Haber and Debatisse, 2006). This occurs in *S. cerevisiae* and other yeasts as a result of Ty retrotransposons that can spread genes throughout the genome (Lesage and Todeschini, 2005). Taking advantage of the natural ability to amplify genes in microorganisms is a method that could potentially give rise to high expression in production strains. For example, targeting of repetitive elements such as rDNA and δ-sequences can be used for multiple integration of genes in *S. cerevisiae* (Lopes et al., 1989; Sakai et al., 1990). However, targeting genes to these natural elements often results in unstable tandem repeats and it is difficult to detect where integration has occurred.

One successful example is the gene amplification of the bacterial gene cluster for the production of the antibiotic actinorhodin to improve the yield in *Streptomyces coelicolor* (Murakami et al., 2011). The cluster was amplified in 4-12 tandem copies resulting in a 20-fold increase in actinorhodin production compared to the parental strain (Murakami et al., 2011). Another example is the gene amplification through multicopy integrative plasmids in *Yarrowia lipolytica* (Juretzek et al., 2001). The gene amplifications resulted in a stable expression of 10-11 copies of the integrative plasmids. Measurement of β-galactosidase activity in strains with 5 to 13 copies of *LacZ* resulted in a 10-11 fold linear increase (Juretzek et al., 2001). This method takes advantage of evenly dispersed retrotransposons in the *Y. lipolytica* genome. However, when this method is used, it is difficult to detect where the gene copies are located as there are many retrotransposons in the genome.

Alternatively, the use of random mutagenesis and screening over several years in pharmaceutical companies has resulted in gene amplification of entire biosynthetic clusters. One example is the penicillin biosynthesis cluster (57 kb) that was amplified several times (Fierro et al., 1995; Newbert et al., 1997). Another example is the kanamycin-over-producing strain *Streptomyces kanamyceticus* where the 145 kb DNA cluster has been shown to be amplified up to 36 tandem copies (Yanai et al., 2006). However, random mutagenesis and screening is a quite time consuming process and strains are often hard to characterize with respect to copy number. Moreover, tandem repeats are prone to gene loss by homologous recombination.

To develop a plasmid free gene expression system that allows high copy gene expression we set out to construct a gene amplification system *S. cerevisiae* (see figure 4). By exploiting one of the cells natural DNA repair mechanisms, genes were amplified by a process that assembles transposon migration. The gene amplification system uses homologous recombination to multiply a single copy of a gene and insert it in up to ten well-defined integration sites.

First gene amplification strains were constructed that contained one, two, four and up to ten copies of a specific gene targeting cassette in a selected genomic loci (figure 3 top). The sites chosen for integration were selected based on previous studies showing expression of heterologous genes in those sites (Mikkelsen et al., 2012). The gene targeting cassettes in each site were constructed to contain two up and downstream targeting sequences called A and B that were taken from the *Neurospora crassa* genome to avoid homology with the *S. cerevisiae* genome. Between the sequences A and B an I-*Sce*I cut site and the counter-selectable marker *URA3* was inserted. The genomic targeting site is illustrated in figure 3. I-*Sce*I is a homing endonuclease with an 18 bp recognition sequence. Thus the recognition sequence exists in approximately one out of $7*10^{10}$ bp. In comparison the yeast genome is approximately $12*10^{6}$ bp making it unlikely that a cut site will exist anywhere else than in the gene amplification cassettes. Thus restriction with I-*Sce*I will result in double stranded breaks that can be repaired from the gene targeting substrate that has replaced one of the cassettes after transformation.

In order to take advantage of the gene amplification cassette sites we constructed a gene targeting platform with USER™ vectors that were designed for integration into the evaluated sites. Thus the vectors are amenable to any gene of interest (your favorite gene (YFG)). In this case we wanted to integrate multiple copies of *msas* and npgA as illustrated in figure 3.

Yeast Genomic DNA with the gene amplification cassette

DNA Substrate

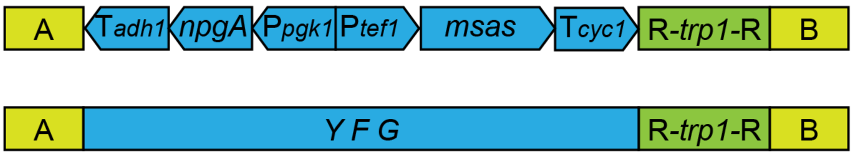


Figure 3 At the top of the figure is an illustration of the gene targeting cassette in the genomic DNA of the *S. cerevisiae* strains. The site is present in one, two or four copies depending on the strain. The gene targeting substrates are illustrated with the gene targeting substrate used for expression of the polyketide synthase *msas* and the PPTase *npgA*. The bottom figure illustrates the general set-up of the method with your favorite gene (YFG) (courtesy of Tomas Strucko).

In this study we used a system with up to four sites. In order to accomplish successful gene amplification the five simple steps illustrated in figure 4 must be followed. First the gene targeting substrate excised from the USER™ vectors is transformed into the receptor cell harboring a number of copies of the amplification cassette with an I-*Sce*I cut site and the *URA3* marker (step 1). After selection and streak purification the cells are plated on 5-FAA media, which results in the elimination of the *TRP1* marker by direct repeat recombination (step 2). The isolated strains are then transformed with a plasmid containing the I-*Sce*I endonuclease under the control of a galactose inducible promoter (step 3). Then the transformants are streaked on SC-gal-trp resulting in the induction of the expression of the I-*Sce*I endonuclease (step 4). This results in double stranded breaks in the gene amplification cassettes where no gene targeting has occurred. Subsequently, the inherent double strand break repair system repairs the breaks using your favorite gene sequence as a template, hence resulting in the desired gene amplification. The gene amplification cassettes contain the *URA3* marker, so if all sites have recombined to contain Your Favorite Gene the *URA3* marker will be lost. Thus, the cells where the gene amplification cassette has been replaced by Your Favorite Gene at all sites can be selected based on their ability to grow on 5-FOA plates (step 5). In order to prevent the loss of gene amplification cassettes by homologous recombination between adjacent integration sites on the yeast chromosome, the sites were chosen such that essential genes are present between the sites. Thus if homologous recombination should occur between adjacent sites, the yeast cells will not survive.

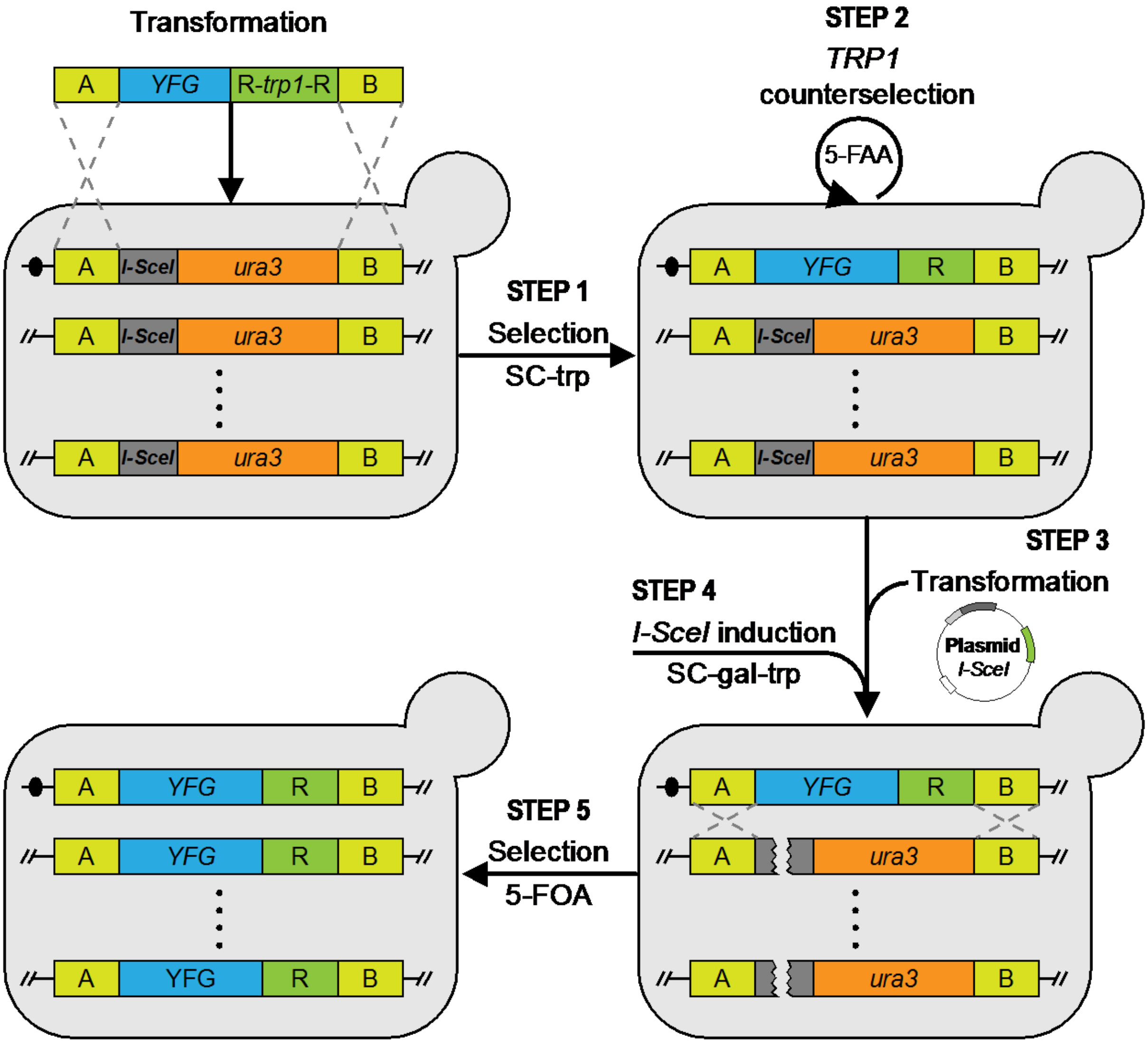


Figure 4 The steps involved in the gene amplification used in *S. cerevisiae* (courtesy of Tomas Strucko).

In order to investigate the capability of our gene amplification system we set out to amplify the MSAS PKS and the NpgA PPTase in *S. cerevisiae*. Firstly, we wanted to challenge the system by investigating if the large gene targeting fragment composing a total of 10.5 kb could be amplified. Secondly, we sought to investigate if more gene copies would result in increased 6-MSA titers. Thirdly, we wanted to check if it was the PKS enzymes or acyl-CoA substrates that were limiting the production of 6-MSA. Thus by increasing the copy number of the genes and hence enzyme concentration we wanted to examine if we would reach a point of substrate limitation. This is essential in order to determine what strategies to use for metabolic engineering of polyketide production in *S. cerevisiae.* If the Acyl-CoA substrates are not the limiting factor strategies to improve substrate availability may not result in improved 6-MSA titers.

Materials and Methods

The construction and evaluation of the *S. cerevisiae* strains was performed as described in the following.

Media

The media used for genetic manipulation in yeast were prepared as previously described (Sherman et al., 1986) with the only exception that the synthetic medium contained twice the amount of leucine (60 mg/L). The strains with the *TRP1* marker excision were obtained on synthetic complete media containing 500 mg/L of 5-Fluoroanthranilic acid (5-FAA). The media containing 5-FOA for counterselection of the *URA3* marker was prepared as described (Mikkelsen et al., 2012). The rich media for the shake flask experiments (YPD) was prepared as previously described (Mikkelsen et al., 2012).

Plasmid Construction

E. coli DH5α cells were used for cloning and propagating plasmids. Plasmids were constructed by USER® cloning (Geu-Flores et al., 2007). The promoters were amplified from the vector pSP-GM2 (Partow et al., 2010). The *msas* gene and *npgA* gene were amplified from the plasmids pRS426CTMSA-PP and pRS424CTnpgA respectively (Wattanachaisaereekul et al., 2008). The primers used are given in table 4 of the supplementary. All constructed plasmids were verified by sequencing (StarSEQ GmbH, Germany).

Table 1 The plasmids used in the construction of gene targeting substrates and gene amplification.

Plasmid	Genotype or relevant features	Source
DNA Fragment Vectors		
pSP-GM2	p*TEF1* and *pPGK1* bidirectional promoter	(Partow et al., 2010)
pRS426CTMSA-PP	6-MSAS gene from *P. patulum*	(Wattanachaisaereekul et al., 2008)
pRS424CTnpgA	npgA gene from A. nidulans	(Wattanachaisaereekul et al., 2008)
Gene Amplification Vectors		
PWJ1320	I-*Sce*I endonuclease expression plasmid	(Lisby et al., 2003)
pCSN1	Sites A, B, DR::*TRP1*::DR, USER cassette, *ccdB*	Gene Amplification System
Constructed Vectors		
pCSN1_LM1	A- pPGK1-npgA-TEF-msas-DR::TRP1::DR-B	This study

Strain Construction

Yeast transformation was performed with the lithium acetate/polyethylene glycol/single carrier DNA transformation method (Gietz and Schiestl, 2007) with the *Not*I excised substrate of the pCSN1_LM1 vector.

The gene amplification strains were isolated on SC-trp plates and streak purified. Then the *TRP1* marker was eliminated by direct repeat recombination on 5-FAA. The isolated strains were transformed with the plasmid PWJ1320-trp and expression of I-*Sce*I was induced by streaking on SC-gal-trp plates. The strains were streak purified and selected on 5-FOA to find the colonies where amplification was successful at all sites. In table 2 is a list of the used and constructed strains are shown

Table 2 The gene amplification strains used in this study.

Name	**Genotype**	**Amplification Cassettes**	**Source**
Strains from the Gene Amplification System			
CEN.CSN-XII	MATa MAL2-8C SUC2 his3Δ1 trp1-289 ura3-52 XII-5::A-Ics-URA3-B	1	Gene Amplification System
CEN.CSN-X-XII-9C	MATa MAL2-8C SUC2 his3Δ1 trp1-289 ura3-52 X-3::A-Ics-URA3-B XII-5::A-Ics-URA3-B	2	Gene Amplification System
CEN.CSN-VII-X-XI-XII-39A	MATa MAL2-8C SUC2 his3Δ1 trp1-289 ura3-52 VII-1::A-Ics-URA3-B X-3::A-Ics-URA3-B XI-2::A-Ics-URA3-B XII-5::A-Ics-URA3-B	4	Gene Amplification System
Strains with gene amplification of the *msas* and *npgA* genes			
CEN.CSN-XII_LM1	MATa MAL2-8C SUC2 his3Δ1 trp1-289 ura3-52 XII-5::A- pPGK1-npgA-TEF-msas-B	1	This study
CEN.CSN-X-XII-9C_LM1	MATa MAL2-8C SUC2 his3Δ1 trp1-289 ura3-52 X-3:: pPGK1::npgA-TEF::msas-B XII-5::A-pPGK1::npgA-TEF::msas-B	2	This study
CEN.CSN-VII-X-XI-XII-39A_LM1	MATa MAL2-8C SUC2 his3Δ1 trp1-289 ura3-52 VII-1::A- pPGK1::npgA-TEF::msas-B X-3::A-pPGK1::npgA-TEF::msas-B XI-2::A-pPGK1::npgA-TEF::msas-B XII-5::A-pPGK1::npgA-TEF::msas-B	4	This study

After successful amplification the correct integration of the substrate was verified by colony PCR with one substrate specific primer and one outlying primer as illustrated in figure 5.

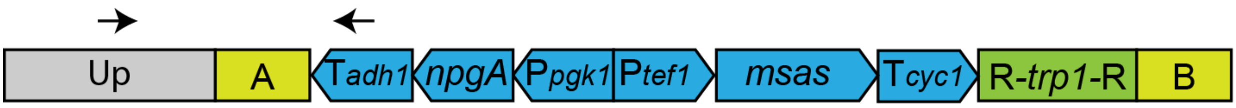


Figure 5 The figure illustrates the placement of the primers for verification of the insertion of the gene targeting fragment in the integration sites. Thus the region marked in gray is the upstream region of the insertion at either site VII, X-3, XI-2 or XII-5.

The verification primers are given in table 3.

Table 3 Primers used for verification of the insertion of the gene targeting fragment at the targeted integration sites.

Integration site	Forward Primer	Reverse Primer	Fragment length
VII	VII-UP-M-Fwd CAACGTGAGCTGGTTGTTGT	Tadh-Rv CTGGCGTAATAGCGAAGAGG	862 bp
X-3	X-3-UP-M-fwd CGAGGGAAGGGAAATAAGGT	Tadh-Rv CTGGCGTAATAGCGAAGAGG	714 bp
XI-2	XI-2-UP-M-Fwd TTGCTCACCTTCCTGGACTT	Tadh-Rv CTGGCGTAATAGCGAAGAGG	909 bp
XII-5	XII-5-UP-M-Fwd GCTCTATTGGAGGTGGCAGA	Tadh-Rv CTGGCGTAATAGCGAAGAGG	721 bp

Cultivations

To determine the effect of gene copy number on 6-MSA titers the constructed strains were grown in 500-mL baffled Erlenmeyer flasks containing 100 mL YPD media. The flasks were incubated with shaking at 150 rpm and constant temperature at 30°C. The growth of the cultures was followed by measuring A_{600} during growth on glucose.

Quantification of Metabolites

To determine the metabolite concentrations the samples were centrifuged at maximum speed for 2 minutes and the supernatant was transferred to HPLC vials and stored at -20 °C until further analysis.

Glucose was measured using an Agilent HPLC series 1100 with a RI (refractive index) detector. The samples were run on an ion-exclusion column with the dimensions 300 mm by 7.8 mm (Aminex HPX-87H from Biorad). The column was eluted at 60 °C with 5mM H_2SO_4 at flow rate of 0.6 mL/min. The metabolites were detected with an RI detector.

6-MSA (SC-274880, Santa Cruz Biotechnology Inc., California, USA) was measured on an Agilent 1100 series HPLC with a degasser, binary pump, column oven and DAD. The Luna C18(2) column (100*2 mm with 3µ particles, Phenomenex (Torrence, CA)) was heated at 40°C during analysis. The samples were run on a gradient of milliQ water with 50 ppm TCA (Solvent A) and 50 ppm TFA in acetonitril (solvent B). The gradient of the solvents was as mix of A and B with 20% to 60% B in 10 minutes and then 20% B for the next 2 minutes.

Results

In order to test whether the large gene fragments or even small pathways of several enzymes could be amplified in our gene amplification system we set out to express the 6-MSA synthase gene and the PPTase in multiple copies. The two genes needed to establish 6-MSA production were cloned into the vector pCSN1 resulting in the vector pCSN1_LM1. The gene targeting substrates were excised and transformed into three strains (1, 2, 4 sites) allowing for the amplification of the genes in two and four copies. The successful transformants were tested for gene insertion and transformed with the ISce-I plasmid to allow for gene amplification in the case of the two and four copy strains. The transformation colonies were streaked on SC-gal-trp plates to induce the amplification. From these plates 20 colonies were picked and replica plated on 5-FOA and SC-URA plates. In the case of the two copy strain 17 out of the 20 selected colonies grew on 5-FOA plates and not on –URA plates indicating that the genes had been amplified. In case of the four copy strains 9 out of the 20 selected colonies grew only on 5-FOA and not on –URA. Thus the amplification frequency appears to decrease as a result of increased copy numbers.

The strains CEN.CSN-XII_LM1, CEN.CSN-X-XII-9C_LM1, CEN.CSN-VII-X-XI-XII-39A_LM1 containing one, two and four copies of the *msas* and *npgA* gene were cultivated in Erlenmeyer flasks with 100 mL YPD (four replicates per strain). The concentration of 6-MSA and primary metabolites was measured in all samples. The titer of 6-MSA at the time of glucose depletion is given in figure 6. The concentration of 6-MSA increases with the number of copies of the genes in an almost linear fashion. The final biomass measured as the optical cell density A_{600} was comparable for all strains indicating that growth was not effected by 6-MSA production.

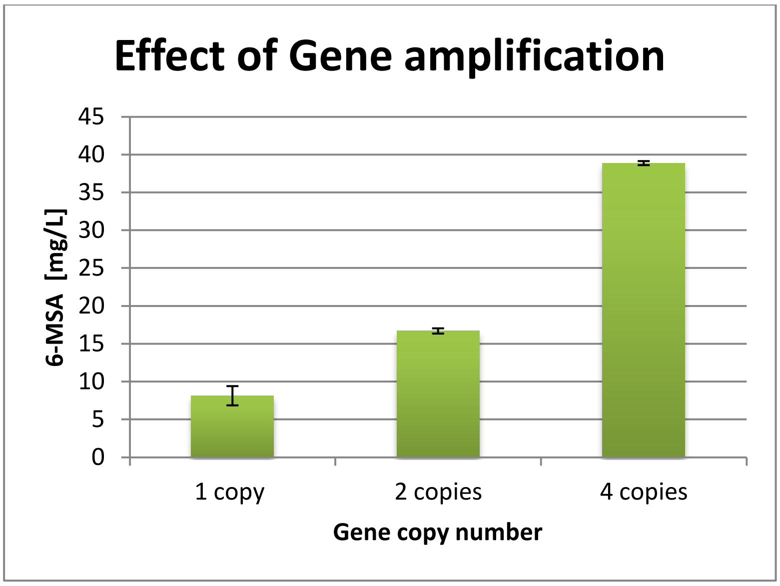


Figure 6 The effect of the copy number of the *msas* and *npgA* gene on 6-MSA titer at the time of glucose depletion.

Discussion

The gene amplification system was clearly validated for the production of 6-MSA through the use of the two enzymes the PKS and the PPTase. The system is still under development, but it is expected to be able to amplify as much as ten copies of the pathway potentially making expression levels and thus titers more than competitive with a 2µ based plasmid system.

When dealing with enzyme pathways one often finds that one or more steps in the pathway are rate limiting. By the use of this novel gene amplification system the limiting step could be identified by amplification of one enzyme in the pathway and subsequent insertion of the other enzyme in one copy either through integration in the pathway strains or by generation of diploid strains. In the case of polyketide synthesis, it could be used to evaluate what level of the PPTase is sufficient to activate all PKSs. It could also be used to study the efficiency of different PPTases. Furthermore, in terms of metabolic engineering it could be used to establish if the acyl-CoA substrates become limiting to product formation at high enzyme concentrations resulting from high gene copy numbers.

Conclusion

We successfully amplified the 10.5 kb gene fragment for production of the polyketide 6-MSA in the yeast *S. cerevisiae*. Thus we have shown that the gene amplification system can be used for amplification of large gene fragments and even small pathways of enzymes that need to interact to form a product. Furthermore 6-MSA titers correlated well with gene copy number. Thus the acyl-CoA substrates do not appear to limit production of 6-MSA in our system. This method can be used to pave the way for studying the relationship between different enzyme pathways and generating knowledge on pathway bottlenecks that can further aid the construction of advanced microbial cell factories.

References

Albertsen, L., Chen, Y., Bach, L.S., Rattleff, S., Maury, J., Brix, S., Nielsen, J., Mortensen, U.H., 2011. Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes. Appl. Environ. Microbiol. 77, 1033-40.

Fierro, F., Barredo, J.L., Diez, B., Gutierrez, S., Fernández, F.J., Martin, J.F., 1995. The penicillin gene cluster is amplified in tandem repeats linked by conserved hexanucleotide sequences. Proceedings of the National Academy of Sciences 92, 6200.

Futcher, A., Cox, B., 1984. Copy number and the stability of 2-micron circle-based artificial plasmids of *Saccharomyces cerevisiae.*. J. Bacteriol. 157, 283-90.

Futcher, A., Cox, B., 1983. Maintenance of the 2 microns circle plasmid in populations of *Saccharomyces cerevisiae*. J. Bacteriol. 154, 612-22.

Geu-Flores, F., Nour-Eldin, H.H., Nielsen, M.T., Halkier, B.A., 2007. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Res. 35, e55.

Gietz, R.D., Schiestl, R.H., 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nature protocols 2, 31-4.

Haber, J.E., Debatisse, M., 2006. Gene amplification: yeast takes a turn. Cell 125, 1237-40.

Juretzek, T., Le Dall, M.T., Mauersberger, S., Gaillardin, C., Barth, G., Nicaud, J.M., 2001. Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. Yeast 18, 97-113.

Kealey, J.T., Liu, L., Santi, D.V., Betlach, M.C., Barr, P.J., 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proceedings of the National Academy of Sciences 95, 505.

Lesage, P., Todeschini, A., 2005. Happy together: the life and times of Ty retrotransposons and their hosts. Cytogenetic and genome research 110, 70-90.

Lisby, M., Mortensen, U.H., Rothstein, R., 2003. Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. Nat. Cell Biol. 5, 572-7.

Lopes, T.S., Klootwijk, J., Veenstra, A.E., van der Aar, P.C., van Heerikhuizen, H., Raué, H.A., Planta, R.J., 1989. High-copy-number integration into the ribosomal DNA of Saccharomyces cerevisiae: a new vector for high-level expression. Gene 79, 199.

Mikkelsen, M.D., Buron, L.D., Salomonsen, B., Olsen, C.E., Hansen, B.G., Mortensen, U.H., Halkier, B.A., 2012. Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. Metab. Eng. .

Murakami, T., Burian, J., Yanai, K., Bibb, M.J., Thompson, C.J., 2011. A system for the targeted amplification of bacterial gene clusters multiplies antibiotic yield in *Streptomyces coelicolor*. Proceedings of the National Academy of Sciences 108, 16020-5.

Newbert, R., Barton, B., Greaves, P., Harper, J., Turner, G., 1997. Analysis of a commercially improved *Penicillium chrysogenum* strain series: involvement of recombinogenic regions in amplification and deletion of the penicillin biosynthesis gene cluster. J. Ind. Microbiol. Biotechnol. 19, 18-27.

Partow, S., Siewers, V., Bjørn, S., Nielsen, J., Maury, J., 2010. Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. Yeast 27, 955-64.

Sakai, A., Shimizu, Y., Hishinuma, F., 1990. Integration of heterologous genes into the chromosome of Saccharomyces cerevisiae using a delta sequence of yeast retrotransposon Ty. Appl. Microbiol. Biotechnol. 33, 302-6.

Sherman, F., Fink, G.R., Hicks, J.B., 1986. Laboratory course manual for methods in yeast genetics.

Tsai, S., Ames, B.D., 2009. Structural Enzymology of Polyketide Synthases. Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates 459, 17-47.

Wattanachaisaereekul, S., 2007. Production of Polyketides by *Saccharomyces cerevisiae*. Ph. D. Thesis Technical University of Denmark.

Wattanachaisaereekul, S., Lantz, A.E., Nielsen, M.L., Nielsen, J., 2008. Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. Metab. Eng. 10, 246-54.

Wattanachaisaereekul, S., Lantz, A.E., Nielsen, M.L., Andresson, O.S., Nielsen, J., 2007. Optimization of Heterologous Production of the Polyketide 6-MSA in *Saccharomyces cerevisiae*. Biotechnology and Bioengineering 97, 893-900.

Weissman, K.J., 2009. Introduction to Polyketide Biosynthesis. Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates 459, 3-16.

Yanai, K., Murakami, T., Bibb, M., 2006. Amplification of the entire kanamycin biosynthetic gene cluster during empirical strain improvement of *Streptomyces kanamyceticus*. Proceedings of the National Academy of Sciences 103, 9661-6.

Zhang, Z., Moo-Young, M., Chisti, Y., 1996. Plasmid stability in recombinant *Saccharomyces cerevisiae*. Biotechnol. Adv. 14, 401-35.

Supplementary S1 Plasmids and Primers

The primers used to create the fragments for the USER® vector pCSN1_LM1.

Table 4 List of primers used for amplification of the gene fragments to be inserted into the vectors.

DNA fragment	**Primers**	**Sequence 5'-3'**
Dual Promoter	FW prom CSN	**ACCATTTGU**TTTATATTTGTTGTAAAA
	RV prom CSN	**ATGCATTTGU**AATTAAAACTTAGATTAG
6-MSA	FW 6-MSA CSN	**ACAAATGCAU**TCCGCTGCAACTTCTACA
	RV 6-MSA CSN	**CACGCGAU**TTAATGGTGATGGTGATGATGTT
npgA	FW npgA RC CSN	**CGTGCGAU**TTAGGATAGGCAATTACACACC
	RV npgA RC CSN	**ACAAATGGU**GCAAGACACATCAAGCGC

In figure 7 is the plasmid map of the plasmid used to generate gene targeting substrate by *Not*I restriction.

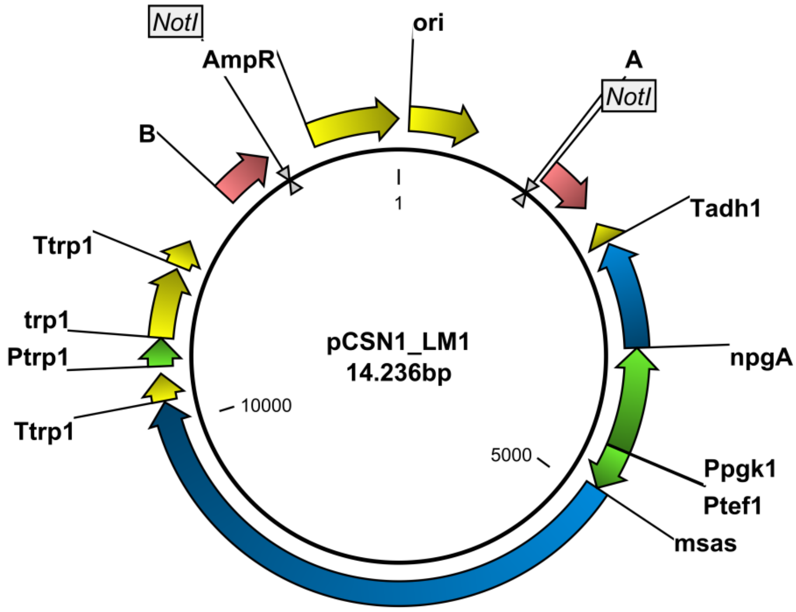


Figure 7 The figure shows a graph of the pCSN1_LM1 vector which contains a dual promoter the *npgA* gene and the *6-msas*-6xhis gene. The targeting sites for the integration into the gene amplification system are called A and B and the trp1 gene flanked by direct repeats.

This chapter is intended as an application paper that will be submitted together with a paper describing the technical aspects of the herein evaluated technology. The results intended for the first paper have thus been put in the supplementary of this chapter and was included for the sake of completeness.

Chapter 6

Application of CCD-flatbed scanning technology for quantification of microbial biomass and metabolites

Prashant Madhusudan Bapat*[1,2], Louise Mølgaard[*1], Lucas Binnenkade*[4],Claudia Matos[1], Ana Rita Brochado[1,3], Wouter Duetz[5], Anna Eliasson Lantz[1] and Kiran Raosaheb Patil[1,3§]

[1]Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, Building 223, DK-2800 Kgs. Lyngby, Denmark

[2]Present address: Novozymes A/S Hallas Alle 1, 4400 Kalundborg, Denmark.

[3]Present address: Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117, Heidelberg, Germany

[4]Present address: Department of Ecophysiology, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

[5] Enzyscreen B. B., Tingietersweg 127, 2031 ER Haarlem, The Netherlands

*These authors contributed equally to this work

[§]Corresponding author: Kiran Raosaheb Patil. Telephone: +49 6221 387- 8473.

Email: patil@embl.de

Email addresses:

LB: lucas.binnenkade@mpi-marburg.mpg.de

PMB: pmba@novozymes.com

LM: lom@bio.dtu.dk

CM: klaudiamatos@gmail.com

AEL: ael@bio.dtu.dk

ARB: rita.brochado@embl.de

WD: duetz@enzyscreen.com

KRP: patil@embl.de

Abstract

Background

Towards the goal of developing an optimized cell factory, it is essential to be able to screen a large number of conditions and strains. To achieve this goal the use of miniaturized systems such as microtiter plates and automatic pipetting systems that can handle a large number of strains in a reproducible manner has become essential. In this paper we evaluated the possibility of using a CCD-flatbed scanning based technology for high-throughput quantification of microbial biomass and product formation.

Results

The CCD-flatbed scanning platform was demonstrated as a useful platform for screening both bacteria and yeast cultures. Our results demonstrated good correlation between image intensity and biomass or product formation using standard transparent and black 24-well plates and 24-square, deepwell propylene microtiter plates. Our results show that growth of the bacterium *Pseudomonas putida* and the yeast *Saccharomyces cerevisiae* can be effectively monitored in microtiter plates using the CCD-flatbed scanning technology. Furthermore, we were able to clearly distinguish different growth behaviours in mutants of *S. cerevisiae*. In addition, we have demonstrated that the platform can be used to quantify the formation of the coloured antibiotic actinorhodin from the filamentous bacterium *Streptomyces coelicolor* and also to clearly identify the onset and rate of antibiotic production in different recombinant strains, and hence the technology has potential to be used in screening to select for the best performing strain.

Conclusions

The usefulness of the CCD-flatbed scanning technology to quantify biomass and product formation has been clearly demonstrated. The system is easily applicable to monitoring growth of bacteria and yeast. Furthermore it can aid in the advancements of discovery and production of especially coloured natural products. The technology is based on the use of commonly available and easy to handle microtiter plates and can thus be used for high throughput screening of a wide array of strains and products. The CCD-flatbed scanning technology may one day be combined with liquid handling robots and analytical equipment allowing further data generation in a very efficient manor.

Background

The detection of growth and metabolite characteristics of individual microbial strains out of large sets of strains (e.g. mutants generated in a strain improvement program) plays an important role in biotechnological and pharmaceutical research. Growth characteristics of microbial cultures such as lag phase, exponential phase, oxygen limited phase (aerobic microorganisms) and the final cell density can be used to determine strain-specific properties and evaluate growth conditions. Moreover, growth characteristics are used together with information about production characteristics to detect potential high-yield or high-productivity recombinant strains.

However, the success of a screening is often limited by the number of strains or culture conditions tested. The number of strains that are required to detect mutants with a particular function can rise up to several thousands. Consequential, the cultivation in Erlenmeyer shake flasks is often not practicable. This has resulted in the development of microtiter plate based systems. The advantages of microtiter plate based systems are that microtiter plates are cheap, robust, easy to handle, high-through-put and standardized. In addition, miniaturised growth systems and high-throughput screening methods can reduce the effort of labour, time and costs.

Nevertheless, down-scaling of cultivation vessels does not come without consequences. One may risk changing the growth conditions and oxygen limitation may be a problem. To determine the effects of miniaturization, several studies have been conducted. These studies have lead to more insights into the influences of well dimensions, culture volumes, orbital shaking conditions and surface tension on oxygen-transfer rates and the degrees of mixing (Duetz et al., 2000; Hermann et al., 2003; Duetz and Witholt, 2004; Kensy et al., 2005a; Kensy et al., 2005b; Funke et al., 2009). It has been shown that e.g. shaking amplitude of 50 mm and a shaking velocity of 300 rpm is sufficient to reach oxygen transfer rates of 39 mmol O_2/l/h in standard round 96-lowwell plates or 40mmol O_2/L/h in standard round 24-lowwell plates, respectively. At such conditions the aerobic bacterial strain *Pseudomonas putida* CA-3 reached cell densities up to 9 g dry weight/ l during growth on a glucose mineral medium (Duetz et al., 2000). To further enhance oxygen transfer rates, square wells (Duetz and Witholt, 2004) and flower plates (Funke et al., 2009; m2p-labs, 2012a) have been applied. Further, the development of adequate well-closure systems, that assure the prevention of (cross)-contamination during vigorous shaking, defined exchange rates of headspace air and limited evaporation, has improved the appliance of microtiter plates as mini bioreactors crucially (m2p-labs, 2012b). Moreover, for *Str. coelicolor* cultivations it has been shown that the data obtained from microtiter plates is more accurate and with reduced variation than data from shake flasks (Siebenberg et al., 2010; Sohoni et al., 2012).

Most industrial biological production processes can benefit from a miniaturized testing system for the production strain. Examples of relevant organisms include mammalian cell cultures, bacteria and fungi. Single cell growing organisms like *Escherichia coli* and *S. cerevisiae* used for a large range of industrial products like recombinant proteins and commodity chemicals can easily be cultured in microtiter plates (Papini et al., 2010; Mattanovich et al., 2012). Filamentous organisms are more challenging to cultivate in small scale. However, for the streptomycetes (a group of filamentous bacteria), that are well known for their ability to produce antibiotics such as kanamycin, neomycin and hygromycin B (Wehmeier and Piepersberg, 2009), successful cultivations in microtiter based set-ups have been reported (Sohoni et al., 2012). Many natural products are coloured and having a colour screen for product formation can be very useful. This strategy has been efficiently applied both for polyketide production by *Str. coelicolor*, using actinorhodin as a model compound (Sohoni et al., 2012) and for isoprenoid production by *E. coli* using lycopene as a model compound (Alper et al., 2005). Often natural product clusters must be induced in order to produce a secondary metabolite and thus being able to screen many media conditions at the same time is a valuable asset (Bode et al., 2002; Frisvad and Samson, 2004). Compared to reading the colour of the individual compounds on a petri dish, an online monitoring system could potentially also identify the onset of antibiotic production and kinetics for production and not just final concentrations. The biggest challenge for microtiter plates are filamentous fungi that tend to form pellets and grow on the sides of the vessels and thus studies of filamentous fungi in microtiter plates are not as widespread.

Standardized microtiter plate formats allow researchers to use multipipettes, replicators and automated systems for incubation, sampling and storage. However, growth-control implicates various manual and highly repetitive steps. In many applications, biomass measurements are performed with spectrophotometers. Withdrawing a sample is hereby indispensable, which results in four main disadvantages for miniaturised growth systems. Firstly, even small sampling volumes decrease culture volumes (e.g. a total volume of 100µl is often used in standard 96-lowwell plates) significantly which leads to less reproducible results. Secondly, repetitive sampling increases the risk for (cross-) contamination. Thirdly, long stopping periods between orbital shaking can result in stress reactions due to reduced oxygen transfer rates. Lastly, optical density measurements with common spectrophotometer are time-consuming.

There are a number of commercially available systems for cultivation in small-scale bioreactors or microtiter plates that avoid the sampling problems stated above. The systems use different approaches for monitoring growth in microtiter plates. One system is the BioLector® that can monitor biomass and protein formation by GFP tagging and has been demonstrated as a tool for screening bacterial and yeast strains in 96 well plates (Kensy et al., 2009) (m2p-labs, Germany). Furthermore pH and DOT can be monitored without interrupting the shaking movement. A drawback of the BioLector® is that only one microtiter plate can be handled at a time reducing the number of strains to be tested. A further development of BioLector® is the

RoboLector, where the BioLector® has been combined with a liquid handling robot (Huber et al., 2009).

A different system is the SimCell from Bioprocessors, Inc. now a part of Seahorse Bioscience, Inc. (Morris Plains, NJ). The SimCell reactors can measure pH, dissolved oxygen tension (DOT) and optical density through the use of fluorescence detection and can handle 1260 experiments at a time. However, they are only applicable to higher eukaryotic systems such as mammalian cell cultures as the oxygen transfer is too low to support the growth of microbial cultures (Seahorse Bioscience, 2012). They also use their own custom made reactors. In addition, the Applikon Biotechnology® Micro-24 Bioreactor system is designed for mammalian as well as microbial cells (Isett et al., 2007; Chen et al., 2009; Pall Corporation, 2012). The Micro-24 Bioreactor system is capable of controlling gas supply, temperature and pH in each well. It automatically monitors and logs each reactors temperature, pH and oxygen. Cultivations run in custom made 24-well cassettes.

Samorski *et al*. described a novel method for quasi-continuous combined light scattering and fluorescence measurements of one microtiter plate that can be used to monitor growth (Samorski et al., 2005). Furthermore, ATP- consumption and fluorescence can be used to monitor microbial growth in microtiter plates.

However, all of the above mentioned methods have individual light detectors for each individual well, which necessarily limits the number of cultures to be scanned simultaneously by one apparatus. This problem is overcome by the CCD-flatbed scanning technology presented in this paper as it can measure several plates simultaneously. It also allows for higher shaking frequency and amplitude which improves the mass transfer capabilities.

In this paper we evaluated the feasibility to apply CCD-flatbed scanning technology to quantify the concentration of biomass and potentially also secondary metabolites in microbial cell suspensions in vessels with transparent bottoms (e. g. microtiter plates). The use of scanned images to follow growth allows for the continuous documentation of growth and coloured metabolite formation as all images are stored, which makes it easy to retrace the results of an experiment. This can be used either for regulatory purposes or backtracking if there are unexpected results. The CCD-flatbed scanning technology is easy to set up and uses standard microtiter plates for the cultivations and can handle several microtiter plates simultaneously as stated above. Furthermore, cultivations in standard microtiter plates allows for automatic handling by pipetting robots to facilitate efficient down stream analysis. Black 24- round low-well microtiter plates with transparent bottoms were evaluated on their applicability for the quantification of biomass in growing *P. putida* CA-3. The system was further validated as a tool for distinguishing different *S. cerevisiae* mutants based on growth characteristics. In addition, we were able to identify onset of product formation as well as distinguishing between *Str. coelicolor* recombinant strains based on volumetric productivity.

Results and discussion

The detailed set up and validation of the CCD-flatbed scanning technology will be presented in the following sections. Furthermore, applications of the technology for quantification of biomass and product formation in *P. putida, S. cerevisiae* and *Str.*

coelicolor are demonstrated. The system was developed in cooperation with Enzyscreen BV, Netherlands where the behaviour of light and physics related to the scanning technology was undertaken. This information is provided in the supplementary section.

CCD-flatbed scanning technology

The set-up of the CCD-flatbed scanner to record growth based on image analysis is illustrated in figure 1. The system was constructed by using a commercially available flatbed-scanner that was modified (steering software, scanning area, light bundling, housing etc.) and combined with a rack-system that allowed us to fix vessels and microtiter plates in a minimal distance to the scanning-unit. The rack was able to perform orbital shaking with an amplitude of 50 mm and up to 300 rpm. The software allowed for the shaking to be stopped at regular intervals and subsequently images were acquired by the scanner and the shaking restarted.

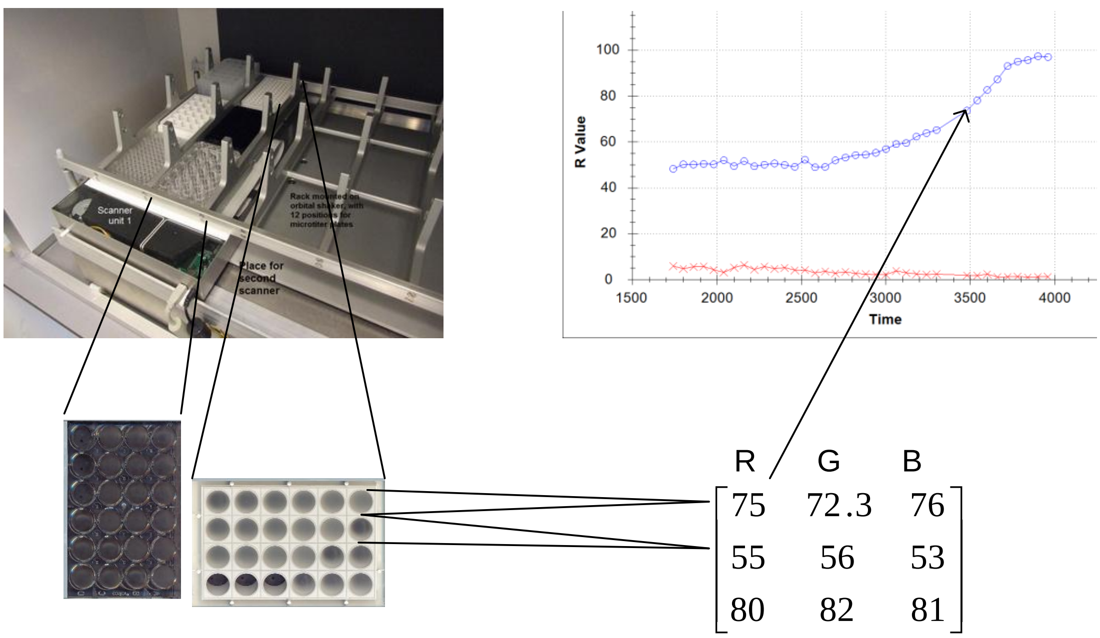


Figure 1. Illustration of the CCD-based cultivation platform and data analysis. The strains are

inoculated in microtiter plates. Then the plates are placed in the metal frame of the shaker and the shaker started. Afterwards the shaking is automatically stopped and the plates are scanned from below at regular time intervals. After the images are recorded image analysis software is used to analyse the red, green and blue colour components. Based on the colour intensity a growth curve of the microorganism can be plotted (blue curve).

In order to measure growth through analysis of scanned images, a significant response in image intensity as a result of microbial growth is required. The ability to detect colour differences by the scanner was clearly demonstrated as illustrated in supplementary figure 8. Furthermore, the behaviour of the light in the concentrated cell suspension is dependent on how the light is scattered by encounters with cells and media components. The properties of light interactions with *P. putida* cells were tested and can be seen in the supplementary figure 9. In addition, it was determined that the light intensities were also dependent on whether the image was analysed in the centre or closer to the walls of the well. This was most apparent for small well diameters such as 96 well plates and the wall effects disappeared as the diameter of the well increased. It was also shown that for standard media compositions there is only a negligible effect of peptone and glucose concentrations in the media on image intensity (figure 13, supplementary). In addition white, gray and black microtiter plates were tested for measuring growth of *P. putida* cells (figure 14, supplementary).

Application I: Monitoring of growth in *P. putida* cultivations

First, we built a correlation between biomass concentrations and scattering intensity of *P.putida* CA-3 cells suspensions in black 24-well microtiter plates (figure 2). The cell mass concentration was calculated from the scattering intensity by the first order exponential equation: biomass concentration is equal to 0.01*(scattering intensity)$^{1.67}$. The scattering intensity was calculated from the green value (in Red-Green-Blue colour space) obtained from images taken by the scanner by the correlation given in supplementary figure 8. The calculated correlation between biomass and scattering intensity has a correlation factor (R^2) close to 1 giving an excellent fit between scattering intensities and biomass concentrations for *P. putida* cells, which is also apparent from figure 2.

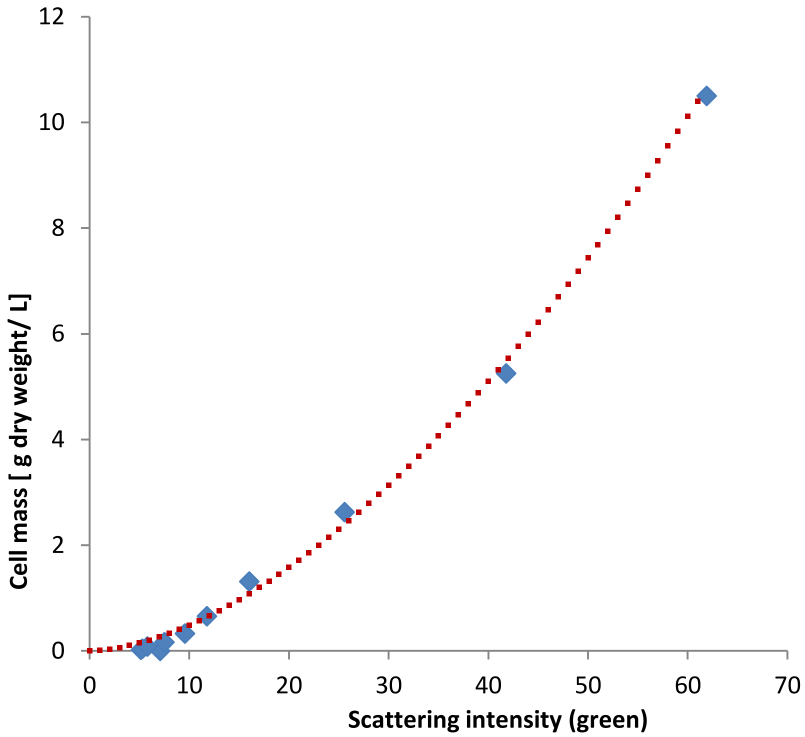


Figure 2 The correlation between biomass concentration and scattering intensity (green-value) for *P. putida* cultivations in 24-well black microtiter plates.

Growth of *P. putida* CA-3 was then monitored over time in black 24-well microtiter plates by the flatbed scanner at 1 hour intervals in triplicate cultivations. Light scattering intensity values were calculated from the green channel and converted into biomass concentrations using the correlation curve of figure 3.

We found that light scattering intensities measured with the flatbed scanner can be utilized to accurately estimate biomass concentrations of growing *P. putida* CA-3 cultures (figure 3). Our results demonstrate that digital-images produced by a commercially available flatbed scanner contain sufficient information (digital colour-values) to quantify light-scattering intensities and biomass concentrations of bacterial cell suspensions, respectively.

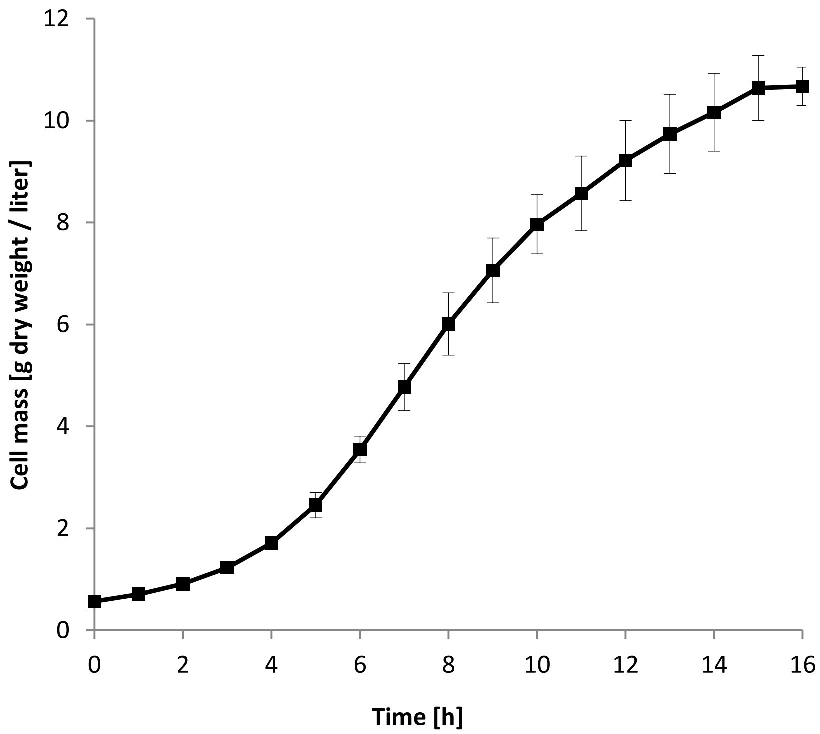


Figure 3. Accumulation of biomass during growth of liquid *P. putida* CA-3 cultures in (shaken) 24-well black microtiter plates monitored with CCD-flatbed scanning technology. The standard deviations are based on triplicate cultivations.

Application II: Monitoring of growth in *S. cerevisiae* cultivations

Next we set up an experiment to determine the ability of the CCD-flatbed scanning technology to accurately measure growth of yeast cultures. We first compared white and transparent microtiter plates. For the white microtiter plates we observed a similar behaviour with the *S. cerevisiae* strains as seen for *P. putida* in figure 14 (supplementary) with a good sensitivity at low dry cell weight, but not much response at higher dry cell weight (data not shown). We also wanted to test if transparent microtiter plates could be used and found that the intensity response to growth was similar to that observed for the black plates with *P. putida* cultures (data not shown). Transparent plates have the advantage that they are commonly available in most research labs and cheaper than the coloured plates. However, there is more reflection in the plate and the amount of growth in the neighbouring wells does cause a minor disturbance of the signal. The transparent plates gave a slightly higher inter-well variation than white plates (data not shown).

To build a more robust correlation we wanted to compare the image intensities to values obtained in a stirred tank reactor with standard off-line analysis of biomass concentration. The yeast *S. cerevisiae* was cultivated in a 2 L stirred tank reactor for 30 hours. During the cultivation dry cell weight and optical density was measured every 2-3 hours from a 5 mL sample of the fermentation broth. At each time point 750 µL of the fermentation broth was also distributed into 6 wells of a transparent 24-well microtiter plate. The microtiter plate was placed in the metal frame, shortly shaken, stopped and an image acquired by the CCD flatbed scanner. The images were analysed by the scanning software for the RGB colour values.

We built correlations for the red, green and blue colour and selected the red colour for our yeast cultures as it had the best correlation coefficient (data not shown). The data showing the biomass concentration correlation with the intensity of the red colour in the image are plotted in figure 4. Our data demonstrate that there was an excellent correlation between the biomass concentration of *S. cerevisiae* and the image intensity in the range from 1-7 g dry wt/l (R^2= 0.99). We only observed a minor variation between the image intensities measured within the different locations of the 24-well plate as indicated by the error bars in figure 4. This validated that there was no significant influence of position in the plate of an individual strain on the measurement. At low image intensities the measurements were less accurate. This is similar to what we observed in the black plates with *P. putida* (supplementary figure 14). For dry cell weights above 1 g/l a very good sensitivity was seen. The change in intensity as a response to dry cell weight was approximately 5 for *S. cerevisiae* under the tested growth conditions.

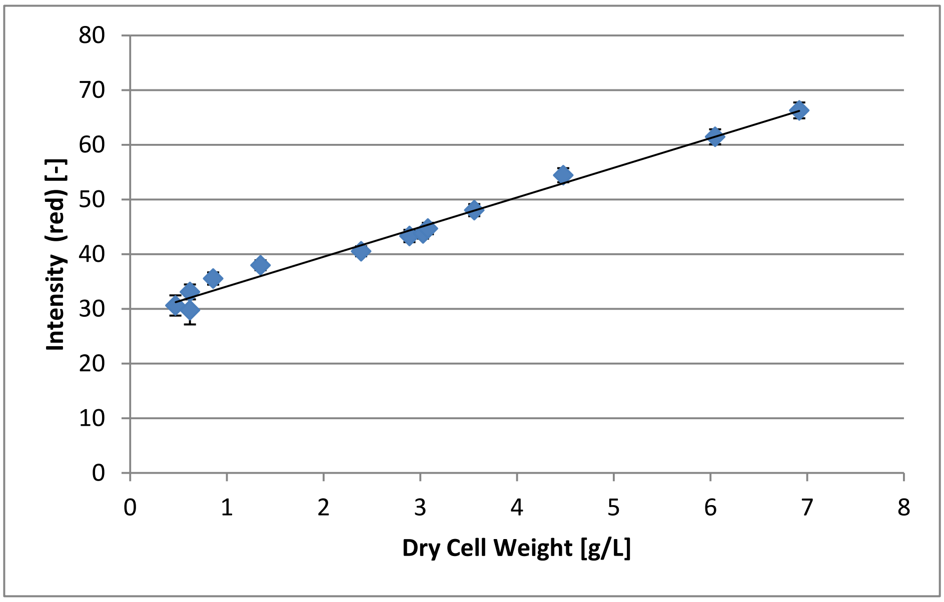


Figure 4 Comparison of image intensity and dry cell weight. *S. cerevisiae* was grown at 30 ºC in a 2 L bioreactor and samples were taken every 2-3 hours. From each sample culture broth was distributed with 750 µL in 6 wells of the 24-well transparent microtiter plate. Then the plate was shaken and scanned. At the same time dry cell weight was measured as described in materials and methods. The error bars represent the standard deviation between the 6 individual wells measured.

Application III: Screening of different *S. cerevisiae* strains for identification of growth characteristics

When screening a mutant library the ability to distinguish different growth behaviours is important. Furthermore, it is essential to be able to estimate changes in growth rate. In order to verify the potential of the CCD flatbed scanner, we grew a wildtype reference and two different *S. cerevisiae* mutants on a medium containing both glucose and galactose. The cultivations were run in white microtiter plates as these have the highest response in image intensity as function of growth. Thus they are most likely to be able to identify different growth behaviours (albeit at low biomass concentrations). The three strains were wildtype (green), mutant 1 (red) and mutant 2 (blue). Mutant 1 lacks glucose repression on galactose and can thus consume both carbon sources at the same time [5]. As can be seen on figure 5 mutant 1 only has two growth phases. The first growth phase during which the glucose and galactose are consumed (0-16 h) and the second in which the ethanol produced in the first phase is consumed (16-30 h). In contrast the wildtype (green) has three apparent growth phases. This is a result of the *S. cerevisiae* strain first consuming glucose (0-10 h). Then when all glucose is consumed repression is relieved and galactose is used (11-16 h). Finally, the primary metabolites produced in the growth phase are consumed, which for the most part is ethanol (16-30 h). Mutant 2 (blue) grows considerably slower than the other two strains which is also apparent in bioreactor cultivations (Otero, 2009).

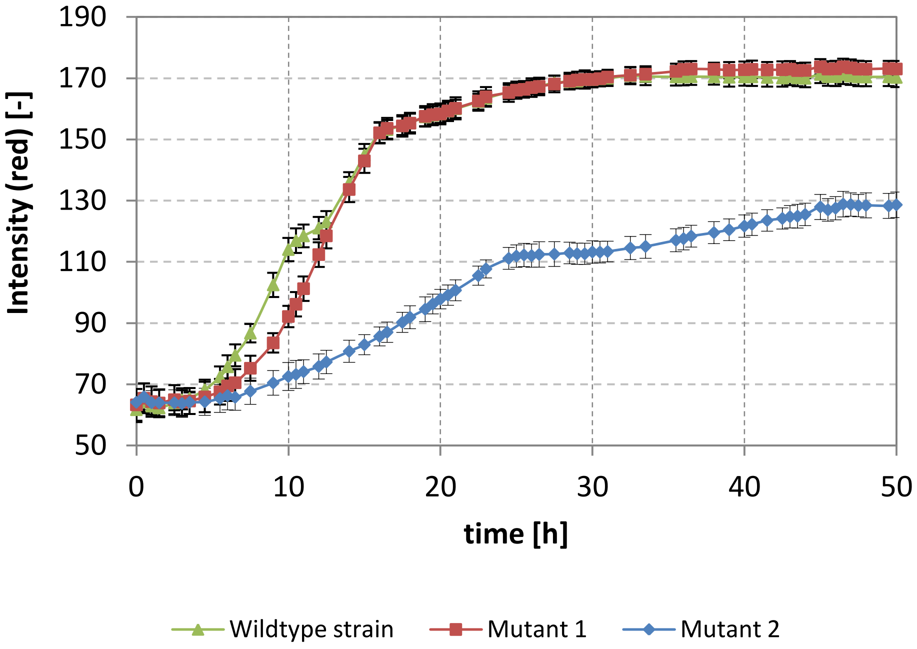


Figure 5 Application III *S. cerevisiae* mutants. For the three different *S. cerevisiae* strains the diauxic shifts between different growth phases are easily observed by using the image intensity. Wildtype strain: green triangles, Mutant 1: red squares and Mutant 2: blue diamonds. The strains were grown at 30 ºC in white walled round low well microtiter plates for 50 hours (24 wells). The scanner was set to obtain images every 30 minutes. The error bars are based on the standard deviation between the 8 wells measured for each strain.

Application IV: Product formation in *Str. coelicolor*.

We wanted to also evaluate the potential for using the described image analysis set-up for monitoring of production of secondary metabolites. Several secondary metabolites are coloured (Alper et al., 2005), hence, enabling detection by image analysis. If the product of interest would not be coloured, a coloured model compound may be used during strain optimisation. Polyketide production by the filamentous bacteria *Streptomyces coelicolor* was chosen as a case study. *Str. coelicolor* naturally produces a blue-coloured polyketide, actinorhodin. Three strains known to differ in actinorhodin production were selected. Two of the strains have been modified in the promoter of actIIorf4, a positive regulator of the actinorhodin biosynthetic pathway, and the third strain included was the corresponding wildtype strain. For the *Str. coelicolor* cultivations, square, 24-deepwell propylene microtiter plates with a working volume of 3 ml were used, as they earlier have been shown to provide better mixing and hence improved oxygen availability than standard round 96- or 24-well plates (Duetz and Witholt, 2004). This is important in order to obtain reproducible growth and production of *Str. coelicolor*. This type of plates are challenging when it comes to the image analysis as they are opaque, which resulted in that growth could not be followed in these experiments (data not shown), whereas actinorhodin production could be accurately monitored as described below.

Firstly, we wanted to investigate whether the time for onset of actinorhodin production could be determined. The light intensity of the blue channel was plotted against time for the three strains (figure 6). The light intensity started to increase around 24 h for recombinant strain 1, whereas an increase in light intensity first was seen around 48 h for recombinant 2 and the wildtype. These time points were in good agreement with onset of actinorhodin production detected by off-line analysis of actinorhodin (data not shown).

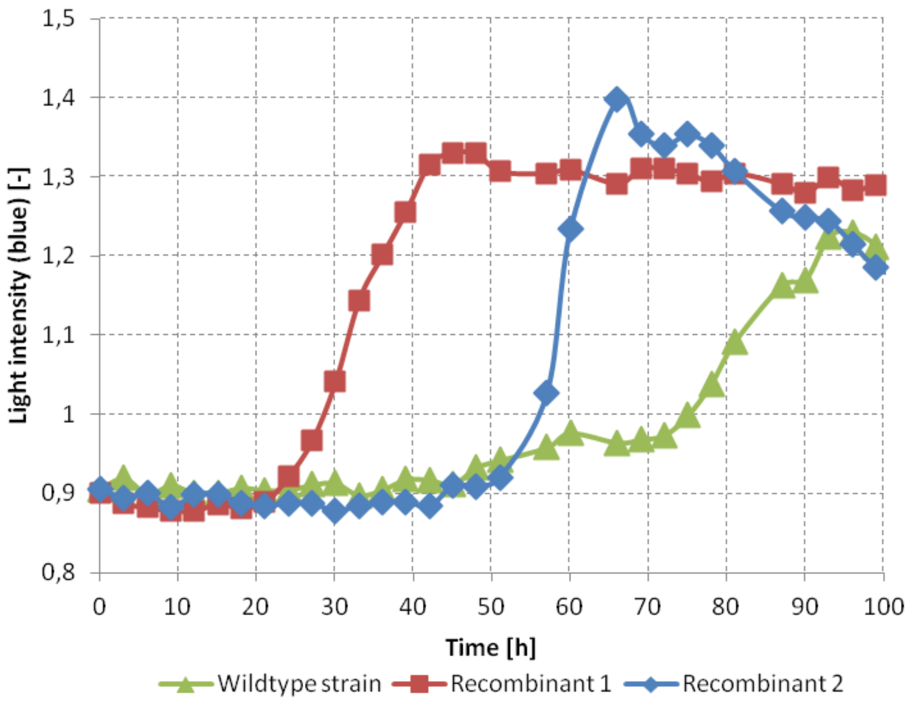


Figure 6. Light intensity of the blue channel over time for the three *Str. coelicolor* strains investigated. Wildtype strain: green triangles, recombinant1: red squares, recombinant 2: blue diamonds. The strains were grown at 28 ºC in 24-square, deepwell propylene microtiter plates for 100 hours. The scanner was set to take images every 3 hours.

To further verify the potential of the image analysis set-up, we investigated whether information about productivity could be extracted from the light intensity data. Values for the slopes of the linear range of the light intensity curves, i.e. in the range of 24-36 h for recombinant strain 1, 51-66 h for recombinant strain 2 and 72-87 h for the wildtype, were estimated (table 1). Volumetric productivities based on off-line actinorhodin concentration measurements were also calculated (table 1). When comparing the values extracted from the image analysis to the volumetric productivities, one can see that the strains follow the same order. The sensitivity of the image analysis seemed to be somewhat lower than for volumetric productivities determined from off-line data. A 1.7 and 2.5 increase for recombinant strains 1 and 2 compared to the wildtype strain was observed using the image analysis data whereas a 1.4 and 4.8 increase was seen in off-line determined volumetric productivities (table 1). This might be of less importance for screening applications as long as strains can be ranked in the correct order and may only pose a limitation for strains that have very small differences in productivity.

Table 1: Productivity data for the three *Str. coelicolor* strains

	Volumetric productivity	
Strain	Off-line analysis[1] (mg L^{-1} h^{-1})	Image[2] (intensity h^{-1})
Wildtype	11.0	0.131
Recombinant 1	15.7	0.217
Recombinant 2	52.2	0.330

[1] Values were calculated from off-line determined actinorhodin concentration

[2] Values were calculated from the linear range of the slope of the blue channel light intensity curve

Growth does not seem to disturb the signal. For recombinant 1, production was growth associated, whereas no or very limited growth took place simultaneous with production for the other two strains (data not shown). However, the extracted information matched off-line data equally well for all three strains.

Earlier start of production of a secondary metabolite and increased volumetric productivity are both wanted properties during strain improvement. The described image analysis set-up provides a valuable screening tool during strain improvement, when often characterisation of large numbers of strains is wanted.

Conclusions

We have demonstrated the applicability of an image analysis based CCD-flatbed scanning technology as a method of monitoring both growth and product formation in a variety of microorganisms. The system has been demonstrated to accurately characterize growth of *P. putida* and *S. cerevisiae*. Furthermore, we were able to distinguish between the growth rates of different *S. cerevisiae* mutants grown in the microtiter plates. Lastly, the ability to determine the volumetric productivity of actinorhodin as well as onset time of production in *Str. coelicolor* makes the CCD-based system ideal for identifying efficient antibiotic producers. In conclusion the CCD-based scanning technology has proven itself useful as a screening platform for many types of products and microorganisms.

Methods

The experimental work is described in detail in the following sections.

Image Analysis

All images were acquired by using a Plustek flat-bed scanner (Plustek Technology GmbH, Germany). Images were analyzed by using custom MATLAB© scripts that made use of MATLAB© image processing toolbox. Images of individual wells from the micro-titer plates were extracted and RGB (Red Green Blue) channel values were averaged for each well. Pixel values that deviated by more than two standard deviations from the average value were discarded and the average recalculated. The effects of discarding deviating pixels were checked manually to insure that only noise and not significant data was discarded.

Cultivations

The methods used in connection with the cultivation of the different microorganisms used are described in the following sections.

Experimental work with *P. putida*

The work carried with *P. putida* was performed as described below.

Strains

The *P.putida* CA-3 strain was a kind gift of Keur Olomá (Dublin).

Media and Cultivation Conditions

P. putida CA-3 was grown overnight in a 300-ml Erlenmeyer flask (without baffles) filled with 35 ml of mineral medium (Evans et al., 1970) supplied with nitrilotriacetic acid (4 mM) as the complexing agent, glucose (150 mM) as the sole carbon and energy source, and a K_2HPO_4-KH_2PO_4 buffer (200 mM) at pH 7.2. The flask was incubated at 25°C on an orbital shaker (300 rpm; shaking diameter, 5 cm). This preculture was used to inoculate a second-batch culture with the same properties. When this second culture had reached an optical density at 540 nm (OD540) of approximately 1.0 (corresponding to 0.40 g [dry weight] liter^{-1}, as determined by drying the washed pellet of a 20-ml cell suspension [OD540 = 24] at 110°C for 18 h and subsequent weighing on an analytical balance), aliquots of 1ml were transferred to the wells of several 24-lowwell microtiter plates (Greiner Bio-one, Netherlands). The microtiter plates were inoculated at 30°C on the scanner-shaker apparatus

(300rpm, 50 mm shaking diameter). Orbital shaking was stopped every hour for 1 min for scanning of the microtiter plates (300 dpi scanning resolution).

Experimental work with *S. cerevisiae*

The work carried out with *S. cerevisiae* was performed as described below.

Strains

The *S. cerevisiae* wildtype strain CEN.PK-113-7D was used for establishing a correlation between the bioreactor and the microtiter plates. It was also used in the study of the different colour components. For the detection of different growth phases the strains used can be seen in table 2.

Table 2 The yeast strains used for cultivation in microtiter plates. The strains CEN.PK 113-5D were used for screening of different *S. cerevisiae* strains.

Strain	Genotype	Origin
Wildtype	*MAT**a*** *MAL*2-8^c *SUC*2	P. Kötter (Frankfurt, Germany)
Mutant 1	*MAT**a*** *mig*1Δ::*MEL*1 *mig*2Δ::*URA3* *MEL*1 *MAL*2-8^c*SUC*2	(Klein et al., 1999)
Mutant 2	*MAT**a*** *URA3-52* *MAL*2-8^c *SUC*2 *ser*3Δ *ser*33Δ *sdh*3Δ	(Otero, 2009)

Media

The media used for the 2 L batch reactor with *S. cerevisiae* was a Delft medium with the following composition glucose 20 g/L, $(NH_4)_2SO_4$ 10 g/L, KH_2PO_4 3 g/L, $MgSO_4 \cdot 7H_2O$ 1 g/L, 0.05 mL/L Antifoam 298 (Sigma-Aldrich, St. Louis, MO, USA) and 2 mL/L trace metal solution ($FeSO_4 \cdot 7H_2O$ 3 g/L, $ZnSO_4 \cdot 7H_2O$ 4.5 g/L, $CaCl_2 \cdot 6H_2O$

4.5 g/L, $MnCl_2 \cdot 2H_2O$ 0.84 g/L, $CoCl_2 \cdot 6H_2O$ 0.3 g/L, $CuSO_4 \cdot 5H_2O$ 0.3 g/L, $Na_2MoO_4 \cdot 2H_2O$ 0.4 g/L, H_3BO_3 1 g/L, KI 0.1 g/L, and $Na_2EDTA \cdot 2H_2O$ 15 g/L), and 2 mL/L vitaminsolution (d-biotin 50 mg/L, Ca-pantothenate 1 g/L, thiamin–HCl 1 g/L, pyridoxin–HCl 1 g/L, nicotinic acid 1 g/L, p-aminobenzoic acid 0.2 g/L, and m-inositol 12.5 g/L).

Bioreactor Cultivations

The bioreactor (Biostat Braun Biotech International GmbH, Mulsungen, Germany) for the batch had a working volume of 2 L. The temperature of the cultivation was kept at 30 °C, and the pH was automatically controlled at 5.0 by addition of 2M KOH. The agitation was set to 500 rpm and the bioreactor was aerated with 2 L air/min. The concentrations of oxygen and carbon dioxide were measured in the off-gas by an Innova 1311 acoustic gas analyzer (Innova Airtech Instruments, Denmark). Dry cell weight was determined by filtering 5 mL of fermentation broth through a dried, pre-weighed nitrocellulose filter (Satorius AG, Geottingen, Germany) with a pore size of 0.45 mm. The residue was washed twice with the same volume saline water (0.9 % NaCl). The filter was then dried in a microwave oven at 150 W for 10 minutes to get a constant weight. The filter was cooled down in a desiccator for at least 2 hours and the cell weight was determined.

Samples from the bioreactor fermentation broth were transferred to 6 wells of the transparent 24-well microtiter plate (4 wells in the centre and 2 wells in the corners of the plate) (Costar® 24 Well Clear TC-Treated Multiple Well Plates (Product #3524)) and the plate was covered with black sandwich covers obtained from Enzyscreen B. V. (Haarlem, Netherlands). The plate was shaken for 15 seconds at 175 rpm and 50 mm shaker diameter and then scanned to capture the image for measuring image intensity.

Microtiter plate cultivations

For the microtiter plates the media was the same as for the bioreactor cultivations. In application III the carbon source was changed to be 10 g/L of glucose and 10 g/L galactose. Mutant 2 was supplemented with uracil (300 mg/L). OD was measured as previously described.

Each well of the white 24-well clear bottom microtiter plate (Visiplate TC, Wallac Oy, Turku, Finland) had a culture volume of 750 µL and the shaker with the amplitude of 50 mm. Each strain was grown in eight of the 24 wells of a microtiter plate. The strains were distributed as shown in figure 7 below (green is wildtype, red is mutant 1 and blue is mutant 2)

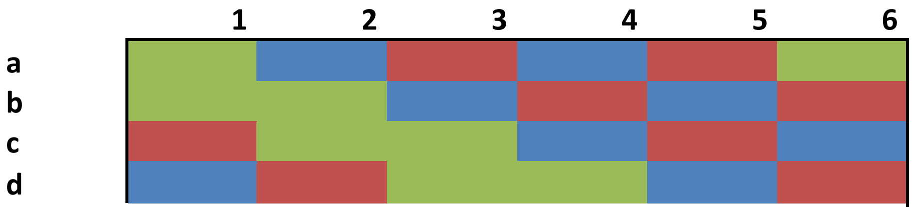


Figure 7 The distribution of the *S. cerevisiae* mutants in the 24-well microtiter plate. Wildtype (green), mutant 1 (red) and mutant 2 (blue)

The shaker was running at 150 rpm and temperature was set to 30 ºC. The plate was covered with black sandwich covers that contained special filters to minimize water evaporation during cultivations and were obtained from Enzyscreen B. V. (Haarlem, Netherlands). The shaker was stopped and an image acquired every 30 minutes during the 50 hour cultivation (150 dpi scanning resolution).

Experimental work with *Str. coelicolor*

The work carried out with *Str. coelicolor* was performed as described below.

Strains

The wildtype strain applied in this study, *Streptomyces coelicolor* A3(2), was a kind gift from Mervyn Bibb, John Innes Centre, Norwich, UK. The two recombinant strains modified in regulation of actinorhodin production were constructed in-house by replacing the native promoter of actIIorf4 with either the strong constitutive promoter ermE* or with a synthetic promoter from a promoter library. The cultures of *Str. coelicolor* were cultivated from frozen mycelia (FM) as described in (Sohoni et al., 2012).

Medium

The defined minimal medium used for bench-scale cultivations was a modification to Evans medium (1970), and was limited in phosphate. The medium was prepared as described by Borodina *et al.* (2008) and contained 3 mM NaH_2PO_4, 100 mM NH_4Cl, 10 mM KCl, 2 mM Na_2SO_4, 2 mM citric acid as chelating agent, 1.25 mM $MgCl_2$, 0.25mM $CaCl_2$, as well as the following per liter: 30 g glucose, 5 ml trace elements solution (20mM $FeCl_3$, 10mM $CuCl_2$, 50 mM $ZnCl_2$, 10mM $MnCl_2$, 0.02mM Na_2MoO_4, 20mM $CoCl_2$, 10 mM H_3BO_4), 1 ml vitamins solution (0.05 g of biotin, 1 g of calcium pantothenate, 1 g of nicotinic acid, 25 g of myo-inositol, 1 g of thiamine-HCl, 1 g of pyridoxine-HCl, 0.2 g of para-aminobenzoic acid/liter) and 100mM 3-(N-morpholino) propanesulfonic acid (MOPS) as buffer. pH was adjusted in the range of 6.80 to 6.90 using sterile NaOH solution (4M) before distribution of sterile medium into the plates.

Cultivation conditions

24-square, deepwell propylene microtiter plates and sandwich covers that contained special filters to minimize water evaporation during cultivations were obtained from Enzyscreen B. V. (Haarlem, Netherlands) and used for all cultivations. Each well was filled with 6 glass beads (3 mm in diameter) before autoclavation. 80 µl of frozen mycelia stock was mixed with 80 ml of sterile re-constituted medium. 3.15 ml of this mixture was dispensed into each well of a sterile microtiterplate. The plates were incubated at 28°C and 150 rpm. The shaker had an amplitude of 50 mm.

Actinorhodin quantification

To extract actinorhodin (ACT), 1.8 ml of 2M NaOH was added to 200µl sample. The mixture was vortexed and centrifuged at 10000 X g for 10 minutes at 4°C and the absorbance of the supernatant measured at 640nm. Actinorhodin concentrations were calculated as described in (Borodina et al., 2008). All the measurements were done in duplicate.

Authors' contributions

PMB conceptualized image analysis using flat bed scanning technology for microbial cultivations. The experimental work and data analysis was carried out by LB, LM, PMB and CM. The manuscript was drafted by LB and LM. KRP and AEL initiated the study and contributed to the experimental design, interpretation of the data and writing of the manuscript. ARB contributed to the discussion of results and manuscript draft. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the Technical University of Denmark, the Danish Council for Independent Research, Technology and Production Sciences (09-064240) and European commission (IP005224 ActinoGEN) for support during this project. Sujata Sohoni is acknowledged for providing the recombinant *Str. coelicolor* strains.

References

Alper, H., Miyaoku, K., Stephanopoulos, G., 2005. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat. Biotechnol. 23, 612-6.

Bateman, J.B., Wagman, J., Carstensen, E.L., 1966. Refraction and absorption of light in bacterial suspensions. Colloid & Polymer Science 208, 44-58.

Bode, H.B., Bethe, B., Höfs, R., Zeeck, A., 2002. Big effects from small changes: possible ways to explore nature's chemical diversity. ChemBioChem 3, 619-27.

Borodina, I., Siebring, J., Zhang, J., Smith, C.P., van Keulen, G., Dijkhuizen, L., Nielsen, J., 2008. Antibiotic overproduction in *Streptomyces coelicolor* A3 (2) mediated by phosphofructokinase deletion. J. Biol. Chem. 283, 25186.

Brown, M., 2004. Advanced Digital Photography. Media Publishing.

Chantrapornchai, W., Clydesdale, F., McClements, D., 2001. Influence of flocculation on optical properties of emulsions. J. Food Sci. 66, 464-9.

Chen, A., Chitta, R., Chang, D., Amanullah, A., 2009. Twenty-four well plate miniature bioreactor system as a scale-down model for cell culture process development. Biotechnol. Bioeng. 102, 148-60.

Cross, D.A., Latimer, P., 1972. Angular Dependence of Scaftering from *Escherichia Coli* Cells. Appl. Opt. 11, 1225-8.

Duetz, W.A., Witholt, B., 2004. Oxygen transfer by orbital shaking of square vessels and deepwell microtiter plates of various dimensions. Biochem. Eng. J. 17, 181-5.

Duetz, W.A., Ruedi, L., Hermann, R., O'Connor, K., Buchs, J., Witholt, B., 2000. Methods for Intense Aeration, Growth, Storage, and Replication of Bacterial Strains in Microtiter Plates. Appl. Environ. Microbiol. 66, 2641-6.

Evans, C.G.T., Herbert, D., Tempest, D.W., 1970. The continuous cultivation of microorganisms Part 2, Construction of a chemosstat Methods in Microbiology 2, 277-327.

Frisvad, J.C., Samson, R.A., 2004. Polyphasic taxonomy of *Penicillium* subgenus *Penicillium*. A guide to identification of food and air-borne terverticillate Penicillia and their mycotoxins. Stud. Mycol. 49.

Funke, M., Diederichs, S., Kensy, F., Müller, C., Büchs, J., 2009. The baffled microtiter plate: increased oxygen transfer and improved online monitoring in small scale fermentations. Biotechnol. Bioeng. 103, 1118-28.

Hermann, R., Lehmann, M., Büchs, J., 2003. Characterization of gas–liquid mass transfer phenomena in microtiter plates. Biotechnol. Bioeng. 81, 178-86.

Huber, R., Ritter, D., Hering, T., Hillmer, A., Kensy, F., Muller, C., Wang, L., Buchs, J., 2009. Robo-Lector - a novel platform for automated high-throughput cultivations in microtiter plates with high information content. Microbial Cell Factories 8, 42-.

Isett, K., George, H., Herber, W., Amanullah, A., 2007. Twenty-four-well plate miniature bioreactor high-throughput system: Assessment for microbial cultivations. Biotechnol. Bioeng. 98, 1017-28.

Kensy, F., John, G., Hofmann, B., Büchs, J., 2005a. Characterisation of operation conditions and online monitoring of physiological culture parameters in shaken 24-well microtiter plates. Bioprocess and biosystems engineering 28, 75-81.

Kensy, F., Zimmermann, H.F., Knabben, I., Anderlei, T., Trauthwein, H., Dingerdissen, U., Büchs, J., 2005b. Oxygen transfer phenomena in 48-well microtiter plates: Determination by optical monitoring of sulfite oxidation and verification by real-time measurement during microbial growth. Biotechnol. Bioeng. 89, 698-708.

Kensy, F., Engelbrecht, C., Buchs, J., 2009. Scale-up from microtiter plate to laboratory fermenter: evaluation by online monitoring techniques of growth and protein expression in *Escherichia coli* and *Hansenula polymorpha* fermentations. Microbial Cell Factories 8, 68-.

Klein, C.J.L., Rasmussen, J.J., Rønnow, B., Olsson, L., Nielsen, J., 1999. Investigation of the impact of *MIG1* and *MIG2* on the physiology of *Saccharomyces cerevisiae*. J. Biotechnol. 68, 197-212.

m2p-labs, 2012a. http://www.m2p-labs.com/microtiter-plates/flowerplates. Accessed May, 29th 2012.

m2p-labs, 2012b. http://www.m2p-labs.com/plate-covers. Accessed 29th of May 2012.

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., Porro, D., 2012. Recombinant protein production in yeasts. Methods in molecular biology (Clifton, NJ) 824, 329.

Otero, J.M., 2009. Industrial Systems Biology and Metabolic Engineering of *Saccharomyces cerevisiae* A case study in succinic acid production.

Pall Corporation, 2012. http://www.pall.com/main/Biopharmaceuticals/Product.page?id=52961. Accessed 20th of April 2012.

Papini, M., Salazar, M., Nielsen, J., 2010. Systems biology of industrial microorganisms. Biosystems Engineering I , 51-99.

Robertson, T.B., 1909. On The Refractive Indices Of Solutions Of The Caseinates And The Acid-And Alkali-Equivalents Of Casein. J. Phys. Chem. 13, 469-89.

Samorski, M., Müller-Newen, G., Büchs, J., 2005. Quasi-continuous combined scattered light and fluorescence measurements: A novel measurement technique for shaken microtiter plates. Biotechnol. Bioeng. 92, 61-8.

Seahorse Bioscience, 2012. http://www.seahorsebio.com/bioprocessing/overview.php. Accessed 15th of May 2012.

Siebenberg, S., Bapat, P.M., Lantz, A.E., Gust, B., Heide, L., 2010. Reducing the variability of antibiotic production in Streptomyces by cultivation in 24-square deepwell plates. Journal of Bioscience and Bioengineering 109, 230-4.

Sohoni, S.V., Bapat, P.M., Lantz, A.E., 2012. Robust, small-scale cultivation platform for *Streptomyces coelicolor*. Microbial Cell Factories 11, 9.

Wehmeier, U.F., Piepersberg, W., 2009. Enzymology of Aminoglycoside Biosynthesis—Deduction from Gene Clusters. Meth. Enzymol. 459, 459-91.

Supplementary

The technical aspects in terms of response and light scattering angles that are needed for the scanner to work as a cell quantification platform are described in the following sections. The goal of the results presented in the supplementary was to gain insight in the behaviour of visible light after entering the transparent bottom of a vessel containing microbial cells. We varied the vessel diameter, vessel-wall colour, filling height, cell concentration, and the presence or absence of a light absorbing layer on top of the culture, and measured the amount of light exiting the vessel perpendicularly through the same transparent bottom, using a flatbed CCD scanner.

Dose-response of the CCD-scanner

Image capturing of CCD scanners is based on a non-linear conversion from light intensities to digital RGB-values (normally 256 levels / 8-bit). To convert the scanner response into absolute light intensities the transmittance of light through light-absorbing dilutions with known extinction coefficients was plotted against the scanner response (Figure 8). Exclusively the transmittance of green-light (green-value) was measured for direct comparison to photometric measurements of microbial suspensions at 540nm. Figure 8 shows that a linear correlation exists between digital green-values higher than 20 and the intensity of light transmitted. In the range of low light intensities the curve showed non-linear pattern. This is because in the low light range digital capture systems often reduce the bit depth obtained from the analogue-to-digital converter by choosing fewer levels, but levels that represent more visually equal increments for the human eye (Brown, 2004). A correlation fitting the entire curve can be given as scattering intensities equals 0.118 multiplied by the green value to the power of 1.3. This equation gives the best fit of the data to the curve over all due to the exponential shape in the beginning of the curve ($R^2 = 1.0$). The conversion into absolute values enabled us to quantify and to analyse light scattering intensities from microbial cell suspensions in the experiments.

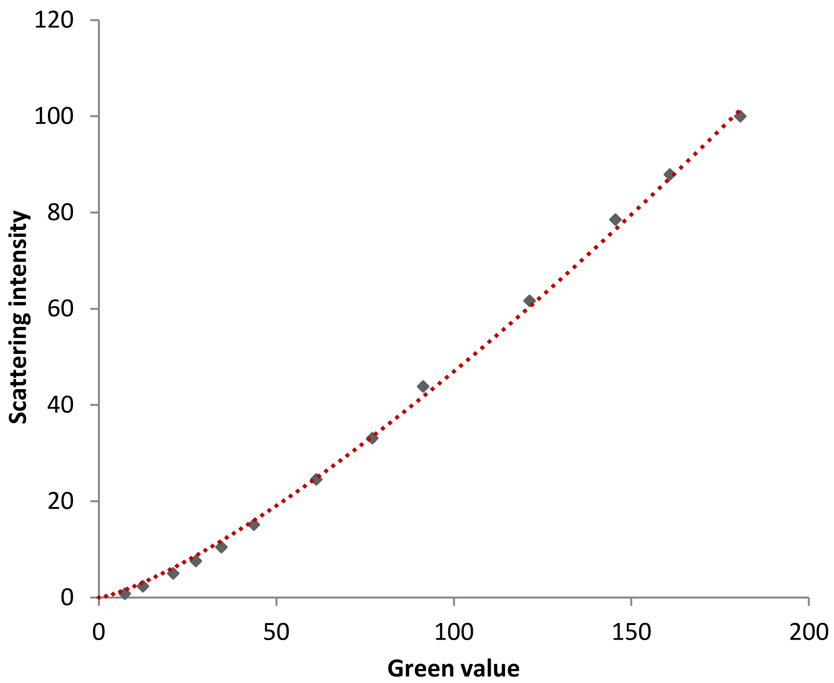


Figure 8. Dose response-curve. Correlation of relative intensities of light transmitted through ink-dilutions of known extinction coefficients to digital arbitrary green-values obtained from scanning images of the same dilutions.

Distribution of scattering angles after interaction with single *P. putida* cells

To understand the complex behaviour of light in concentrated cell suspensions, it is a prerequisite to determine the behaviour of light striking a single cell. For that reason a petri dish filled with a dilute *P. putida* cell suspension was illuminated from below with a green laser beam. The laser beam was set perpendicular to the surface. The light intensities at various angles were measured using a lux meter. It appeared that more than 95% of the light that interacts with a single *P. putida* cell is forwardly scattered (by means of refraction or reflection) at angles lower than 40 degrees (Fig. 9). The maximal intensity of scattered light was detected at an angle of 11 degrees in the forward direction. Less than 1% of the light was reflected in the backward direction. These results are in accordance with experiments on the angular dependence of scattering from *E. coli* cells which demonstrated that the intensity of refracted light is maximal at scattering angles lower than 30° (Cross and Latimer, 1972). The fact that the major fraction of light striking a single cell is scattered in the forward directions indicates that light returning from concentrated cell suspensions is not the cause of single reflections at the cell surface but of many light scattering events that have randomly redirected the light beam.

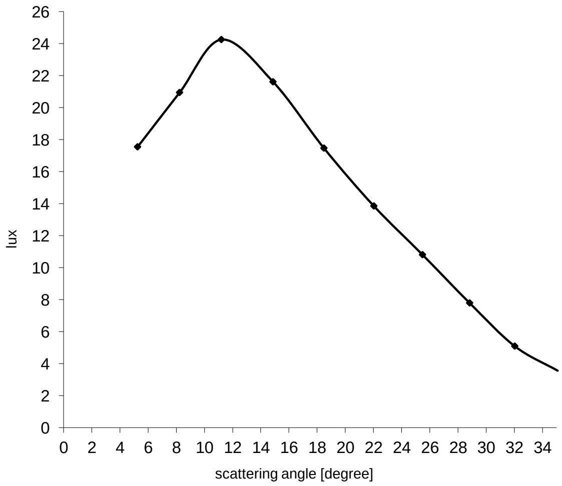


Figure 9. Intensity of light scattering dependent on the refraction angle. A green laser beam was directed perpendicular to the bottom of a Petri dish filled with a dilute *P. putida* CA-3 cell suspension (0.12 g dry wt l^{-1}). Light intensities were measured 15 cm above the suspension at angles between 5 and 35 degrees.

Light scattering by cells in black-walled vessels

To verify that the major fraction of light returning from a cell suspension has been scattered via multiple cells we determined the impact of light absorbing and light reflecting boundaries of the cell suspension on light scattering intensities measured with the CCD-scanner system. If dilute cultures in a black-walled 16 mm vessel are covered with a light absorbing layer on top of the culture, almost no light was found to exit through the bottom of the vessel (situation A in Fig 10). The small amount of light that does return is the sum of i) the little light that is directly reflected by a single cell (change of direction by an angle of approx 150 degrees), and ii) the light changed in direction by 150 degrees as a result of multiple small-angle scatterings by the subsequent interaction with multiple cells (e.g. 5 encounters, with and average change in direction of 30 degrees). The chance of the latter happening is apparently quite small as well.

If the light absorbing layer on top of the culture is removed, the amount of light returning to the CCD is 3-5 fold higher (Situation B in Fig. 10). The extra light returning to the CCD is a consequence of light reflecting to the culture-air interface before or after being scattered (by a small angle) by one or more cells. The apparently relatively high frequencies at which this is occurring may be explained by the circumstance that only a single encounter with a cell (at a scattering angle of 30 degrees) is sufficient (instead of the at least 5 encounters as described above). The amount of light returning from vessels filled with water was found to be negligible. Accordingly, direct reflections from the suspension surface to the CCD do not have an impact on light scattering intensities measured by the scanner system.

Light scattering by cells in white-walled vessels

The amount of light returning from white-walled 16 mm vessels was found to be 10-50 fold higher compared to black-walled vessels for *P. putida* cultures in the range of

0.1-3 g dry wt l^{-1} (Situation D in Fig. 10). The reason behind this large difference is probably two-fold. First of all (most relevant at low cell densities) light may reflect (unhindered until this time) to the white walls (change in direction between 150 and 180 degrees), and subsequently be scattered by a single cell (change in direction of e.g. 30 degrees), so that the light can exit the vessel from the bottom and arrive at the CCD. This sequence of events is similar to what is happening in the black-walled vessel when light is allowed to reflect to the culture-air interface, but with the difference that on a white surface the reflection is highly diffuse and almost 100%, while at a culture-air interface the amount of light reflected can be as low as 5% (dependent on the angle). Secondly, light that gets trapped in the cell suspension from hundreds of scattering events occurring at higher cell densities does not get absorbed when interacting with a vessel wall. Instead it gets reflected, and so obtains another chance to eventually exit the vessel through the bottom and arrive at the CCD. Assuming that the second situation is responsible for most of the light detected at moderate cell densities (e.g. 2 g dry wt l^{-1}), it seems reasonable to conclude that most light that does reach the CCD has reflected to the white wall 10-50 times since it entered the vessel. That the availability of a light-reflecting culture-air interface is also important in this 'trapping' of the light into the well, is proven by the much lower light intensities measured in white-walled vessels where the cultures are covered with a light absorbing layer (situation C in Fig 10).

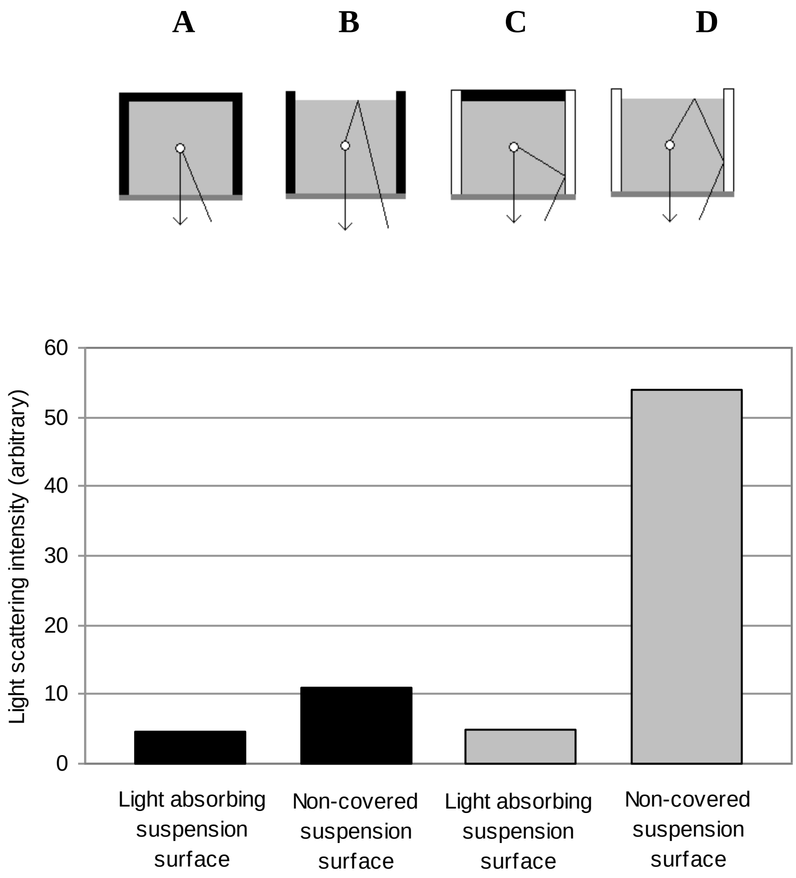


Figure 10. Influence of a light absorbing (black) and light reflecting suspension boudaries on light scattering intensities measured with the CCD-scanner system. Light scattering intensities were measured in the center of transparent bottoms of 16mm-vessels filled with *P. putida* CA-3 cell suspensions at a biomass concentration of 0.88g dry wt. l^{-1}. **Black bars**: vessels with black walls. **Grey bars**: vessels with white walls. Figure A - D: Schemata (simplified) of light paths that contribute most significantly to the intensity of light scattering detected by the CCD under each condition.

Travelling distance of light before exiting the vessels

The strong effects of the wall colour indicated that a large proportion of the light exiting from the centre of the vessel must have reflected to the wall (8 mm away) previously. This - in its turn - indicated that even at relatively high cell densities light exiting from the centre of the wells has actually travelled multiple mm's or even cm's after entering the cell suspension. We therefore quantified the "light travelling distance distribution" with the experimental setup as depicted in Fig 11: part of a petri dish was covered with a light-impermeable black layer (round, with various diameters, and a small hole in the middle), and the amount of light exiting from the small hole in the middle was quantified using the CCD flatbed scanner. It appeared that 50% of the exiting light has entered the vessel at a distance of 4 mm or more. For a cell density of 14 g dry wt l^{-1}, it can be calculated that statistically a photon must have interacted with at least several hundreds cells during this travel. Assuming that cells are randomly distributed in the suspension and each cell has an average volume of 1×10^{-15} l^3 (or 1×10^{-18} m^3) and an average dry weight of 3×10^{-13} g, approximately 5% of the suspension volume is occupied by cells and accordingly a photon would strike a cell on the average every 20µm. Hence, for a distance of 4 mm a photon would strike at least 200 cells. However, the chance that a photon strikes 200 cells without being scattered in various directions is minimal. Accordingly, the number of cells that a photon has stroked before exiting the vessel in a distance of 4mm or more is certainly in a much larger magnitude. We therefore conclude that a photon must have been scattered at least thousands or even million times before exiting the vessel bottom in such a distance.

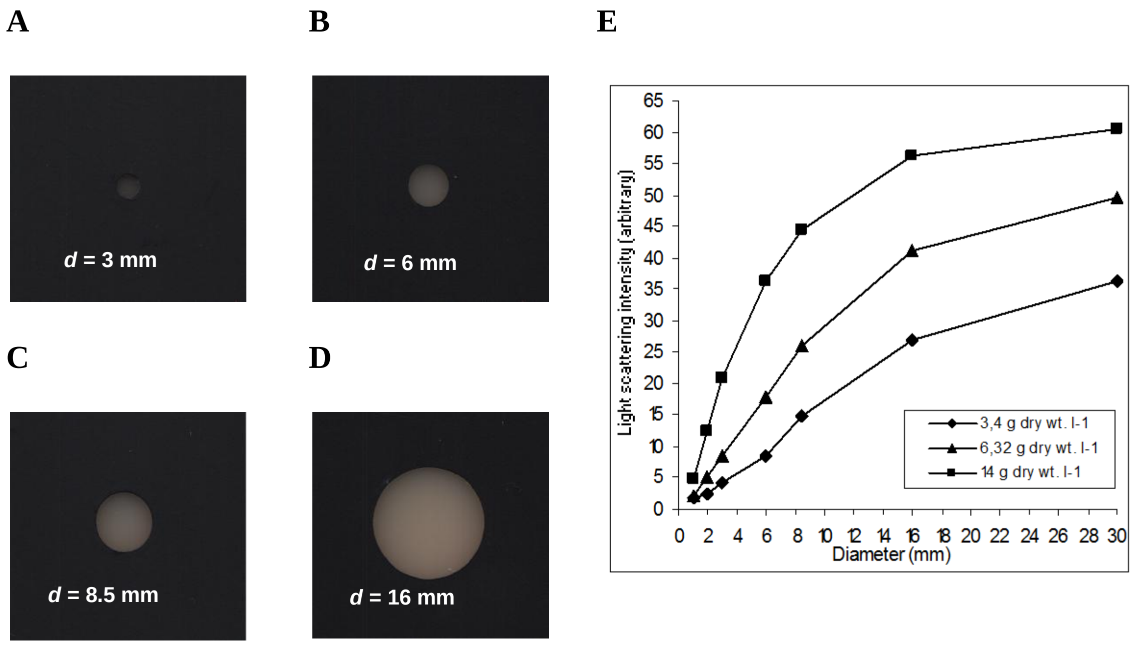


Figure 11. Light scattering intensities in *P. putida* CA-3 cell-suspensions of different concentrations measured in aperture holes of various diameters surrounded by infinite non-illuminated (non-transparent) areas. **A - D** Scanning images of the *P. putida* CA-3 cell suspensions in aperture holes of different diameters at a concentration of 14 g dry wt. l^{-1}. **E** Correlation between diameter of the illuminated area (aperture hole) and light scattering intensities

Spatial distribution of light exiting black and white walled vessels of various diameters

The intensity of light exiting from white-walled vessels (containing an identical cell suspension) appeared to be negatively correlated with the vessel diameter, but positively correlated for black-walled vessels (see figures 12A-12F). Furthermore, light intensities appeared to be highest in the centre of black-walled vessels, while the opposite is true for white-walled vessels. This stands in good agreement with the general picture arising from the results given above: The major fraction of light travels large distances even in highly concentrated suspensions before exiting the suspension at the transparent bottom. Most light that returns from the suspension has been scattered many times instead of being reflected by specular reflections near to the suspension boundary. For that reason, light absorbance by black walls and diffuse reflections by white walls have strong impacts on light intensities up to considerable distances to the walls. The distance up to which the light intensity is dependent on the wall colour is negatively correlated with the biomass concentration of the suspension (compare Fig. 12E and 12F). As assumed above, most light is scattered by cells in the forward direction which implicates that a photon needs to strike several cells for a 150 degree change in direction. At higher biomass concentrations such a change in direction can occur in a more narrow space. Accordingly, the average travel distance of a photon in a suspension is also negatively correlated with the concentration of cells. Additional experiments (not shown) demonstrated that light-scattering intensities measured in the centre of vessels with diameter of 86mm or bigger are not influenced by the wall colour. We therefore conclude that behind a certain distance aberrations of light scattering due to reflection and absorption at the vessel wall become negligible and light scatters similar to an infinite cell suspension.

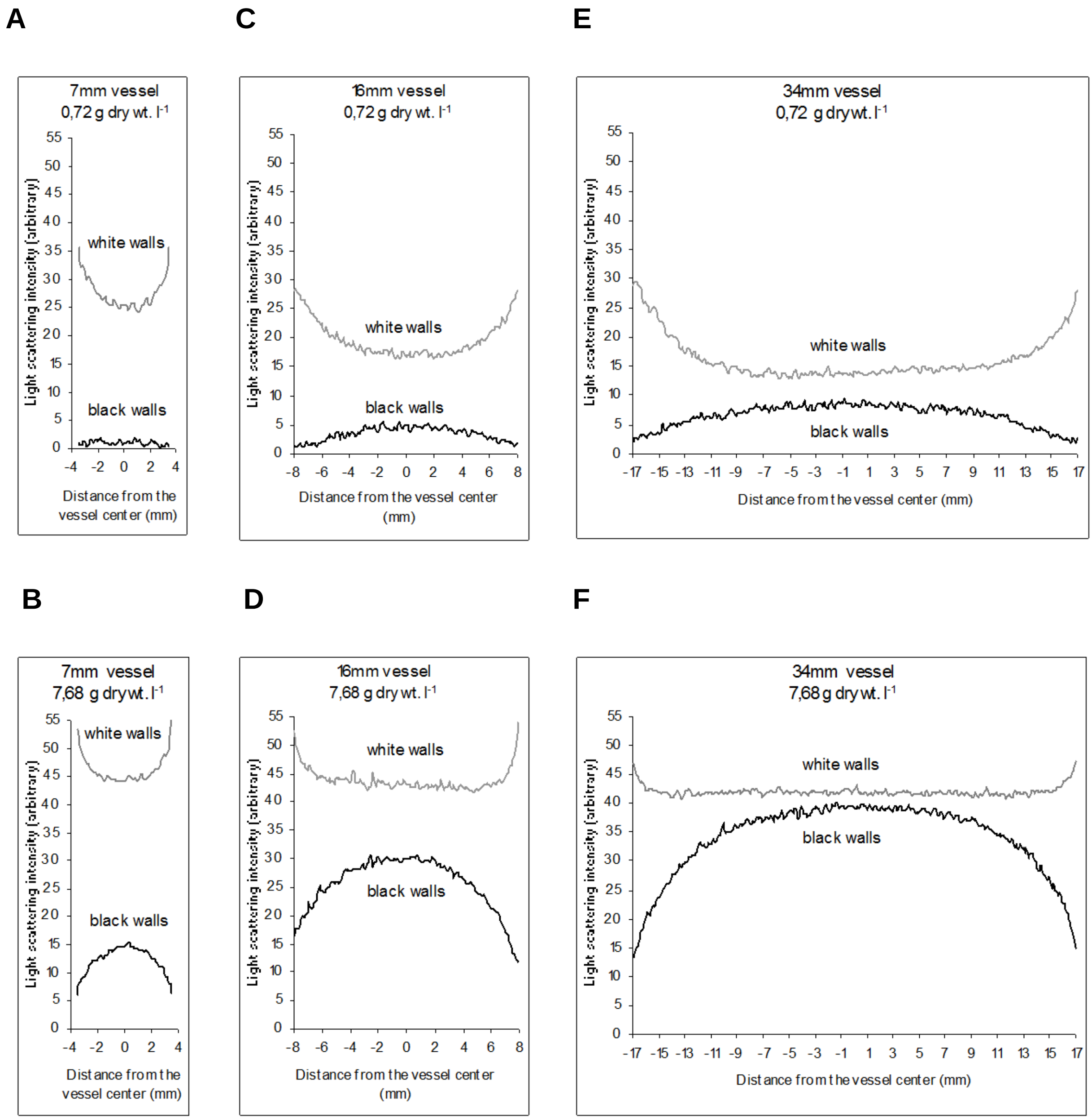


Figure 12. Impact of vessel-diameter and black and white walls on light scattering intensities in *P. putida* CA-3 cell suspensions. Light scattering intensities were measured with a CCD flatbed scanner below the transparent bottom (over the horizontal middle axis) of vessels with black and white walls filled with *P. putida* CA-3 cell suspensions at biomass concentrations of 0.72g dry wt. l^{-1} and 7,68g dry wt. l^{-1}, respectively.

Influence of glucose and protein on light scattering measurements with CCD flatbed scanning technology

Intensities of light scattered from *P. putida* CA-3 cell suspensions containing 200mM and 1M glucose and 40 g l^{-1} and 80 g l^{-1} peptone were measured in a range of different biomass concentrations (Fig. 13). It turned out that both glucose and peptone reduce light scattering intensities of cell suspensions detected with the CCD flatbed scanner. Especially at biomass concentrations higher than 1 g dry wt. l^{-1} scattering intensities were significantly lower in the presence of both substances. At the highest biomass concentration tested (14 g dry wt. l^{-1}) light scattering intensities were 6.7% and 11.8% lower in the presence of 200mM and 1M glucose, respectively. In the presence of 40 g l^{-1} and 80 g l^{-1} peptone the intensity of scattered light was 7.4% and 11.4% lower, respectively.

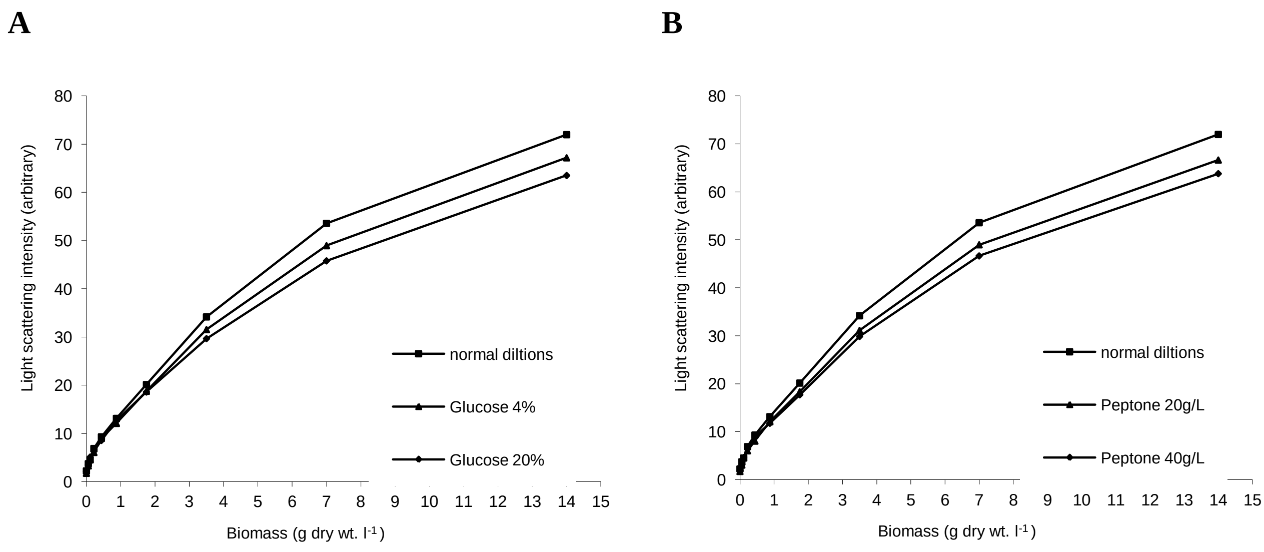


Figure 13. Influence of glucose (A) and peptone (B) on light scattering intensities in *P. putida* CA-3 cell-suspensions at different biomass concentrations.

This is because glucose and protein in the suspension approach the refractive index of the solvent to that of the cells, which reduces light scattering and the suspension appears less turbid (Robertson, 1909; Bateman et al., 1966; Chantrapornchai et al., 2001). Accordingly, components of liquid growth media such as glucose and peptone can alter the optical properties of a cell suspension and thus have to be taken into consideration for light-scattering measurements using CCD flatbed scanning technology. For example, experiments with strongly varying protein concentrations in heterogeneous culture collections would require an initial calibration of the scanner system. Glucose concentrations in standard liquid growth media, however, are usually below 200mM and cause only negligible alterations (data not shown).

Application of various types of commercially available microtiter plates

Light scattering measurements were performed with dilution-series of *P. putida* CA-3 cell-suspensions in black, grey and white low well microtiter plates in the 24-well and 96-well format and transparent bottoms (Fig. 14). It turned out that at biomass concentrations below 3.5 g dry wt. l^{-1} the resolution of the system is highest if white microtiter plates are applied, while at higher biomass concentrations the application of black microtiter plates results in a much higher resolution and therefore more accurate results. This implies that low cell densities can be most reliably determined in small white-walled vessels, and the least accurate in small black-walled vessels. For large-sized vessels, the wall colour has a smaller effect. For measurements in the full range of biomass concentrations grey microtiter plates appeared to be most reliable. In conclusion, our results demonstrate that for *P. putida* biomass concentrations in a range of 0-14 g dry wt. l^{-1} can be accurately determined using CCD-flatbed scanning technology in conjunction with grey microtiter plates with transparent bottoms.

In black 96-well microtiter plates and at biomass concentrations below 2 g dry wt. l^{-1} light scattering intensities were below the detection limit and reflections from the suspension surface caused inaccurate values. At biomass concentrations higher than 2 g dry wt l^{-1}, however, light scattering intensities increased almost linearly. In contrast, black 24-well microtiter plates behaved similar to grey ones and exhibit a relatively high resolution in the whole range of biomass concentrations.

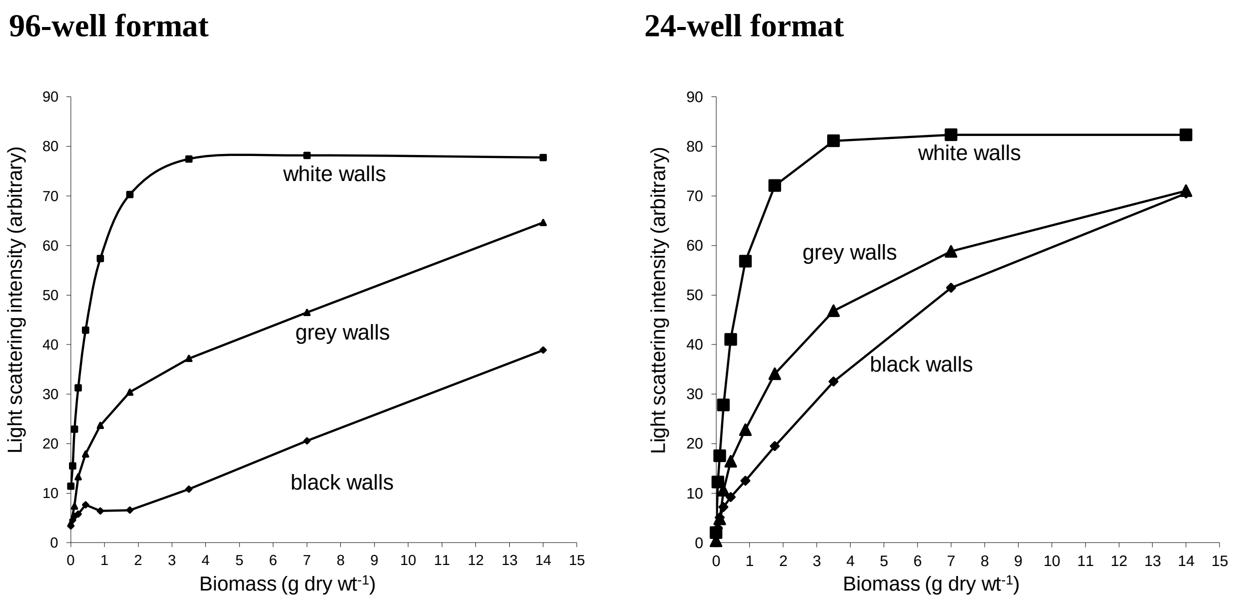


Figure 14. Correlation between light scattering intensities and biomass concentrations of *P. putida* CA-3 cell suspensions in black, white and grey 24- and 96-lowwell microtiter plates with transparent bottoms. Light scattering intensities were quantified using a flatbed scanning technology.

Conclusions and Perspectives

The production of medicine and chemicals through the use of biological cell factories has long been the goal of the biotech industry. Polyketides in particular constitute a large and diverse group of medically significant compounds. The discovery of which is still ongoing and having an efficient cell factory ready for heterologous expression is clearly beneficial. Thus this thesis set out to engineer a model polyketide cell factory into two commonly used host systems for heterologous expression namely *Aspergillus nidulans* and *Saccharomyces cerevisiae*. The model polyketide 6-MSA was used to demonstrate the applicability of these systems for polyketide production.

A stable expression system for the production of 6-MSA in *A. nidulans* has been set up. Furthermore the production of 6-MSA in *A. nidulans* was optimized using the algorithm OptGene to predict beneficial genetic engineering targets. The predicted target *gdhA* resulted in a severe growth rate reduction that was partially rescued by insertion of an extra copy of *gdhB*. This study provides the first attempts to use OptGene to guide metabolic engineering in *A. nidulans*. It indicates that the metabolism of *A. nidulans* may well be more complex than the model *i*HD666 describes. Thus future efforts should set out to provide more fluxome and transcriptome data as well as mutants targeting central carbon metabolism that can shed lights on the metabolic fluxes and connections in *A. nidulans*. Another proposal would be to tailor OptGene to deal with more regulated complex biological systems. As models of *A. niger* and *A. oryzae* have been published after the *A. nidulans* model there may be some lessons learned from these fungi that could be useful for the *A. nidulans* model.

Another well-established polyketide host is the yeast *S. cerevisiae*. In this respect a novel gene amplification system was developed that would allow for high expression and thus likely high productivity of 6-MSA in *S. cerevisiae*. Importantly, the system was constructed to have stable integration of the PKS and the PPTase in the yeast genome as opposed to the traditionally used 2µ plasmids that have been shown to be unstable. The system was successfully used to demonstrate that amplification was possible in up to four copies of the 10.5 kb gene fragments. There was a good correlation between gene copy number and 6-MSA titers in Erlenmeyer flask experiments. In future the system will be developed to harbor up to ten copies of the gene amplification cassettes aiming at an even higher productivity. This may even provide higher titers than the commonly used 2µ based plasmid systems. Furthermore the system allows for the identification of pathway bottlenecks by allowing for the stable increase in gene copy number of one gene to be combined with one copy of another gene in the pathway. In terms of polyketide production, it would be possible to investigate if the PPTase or PKS expression level are limiting for polyketide production. Sufficiently high levels of expression may also lead to limitations in acyl-CoA or NADPH substrates under which circumstances metabolic engineering would be a well chosen approach to increase productivity.

Engineering an efficient cell factory also involves the need to test a variety of production conditions such as media. In this respect the validation of a CCD-flatbed scanning platform using image analysis for measuring biomass and metabolite formation in a variety of microorganisms was undertaken. The results clearly demonstrate the usefulness of the system for monitoring growth of both bacteria and yeast. This includes screening of mutant libraries with diverse growth profiles. Furthermore productivity and onset of production of the colored antibiotic actinorhodin in *S. coelicolor* was identified through the use of the CCD-flatbed scanning platform combined with image analysis. The system could also be used for inducing novel polyketide expression by testing a variety of media and inducers in parallel. The system is still ongoing development and can monitor growth continuously during cultivations. It has the potential for being combined with liquid handling robots. Furthermore plate optimization could one day lead to a screening system than can handle more challenging filamentous fungi. This would decrease the time, labor and cost of industrial screening programs even further.

Polyketides constitute a diverse group of compounds. Even after identification of the polyketide gene clusters through an array of cultivation conditions, their heterologous expression and production present a challenge. This is in many cases due to the large size of the gene clusters. Thus the development within molecular biology and screening of polyketide cell factories is still a challenging road to follow. Modeling cellular metabolism has come a long way in only a short time and will one day be able to direct metabolic engineering even in *A. nidulans* as it has been shown for *S. cerevisiae* and *E. coli*. Increasing the time an effort for advancing our knowledge of metabolic fluxes in *A. nidulans* is essential for the successful modeling of future chemical production processes.