Genome-scale metabolic representation of Amycolatopsis balhimycina

Vongsangnak, Wanwipa; Figueiredo, L. F.; Förster, Jochen; Weber, Tilmann; Thykær, Jette; Stegmann, E.; Wohlleben, W.; Nielsen, Jens

Published in:
Biotechnology and Bioengineering (Print)

Link to article, DOI:
10.1002/bit.24436

Publication date:
2012

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

Link back to DTU Orbit

Citation (APA):
Genome-Scale Metabolic Representation of Amycolatopsis balhimycina

Wanwipa Vongsangnak,1,2 Luis Filipe Figueiredo,1,2 Jochen Förster,2 Tilmann Weber,3 Jette Thykaer,1 Evi Stegmann,3 Wolfgang Wohlleben,3 Jens Nielsen1

1Department of Systems Biology, Technical University of Denmark, Denmark; telephone: +46-31-772-3804; fax: +46-31-772-3801; e-mail: nielsenj@chalmers.se
2Fluxome A/S, Gymnasievej 5, 3660 Stenlose, Denmark
3Department of Microbiology and Biotechnology, Interfaculty Institute of Microbiology and Infection Medicine, Eberhard Karls University Tübingen, Tübingen, Germany

ABSTRACT: Infection caused by methicillin-resistant Staphylococcus aureus (MRSA) is an increasing societal problem. Typically, glycopeptide antibiotics are used in the treatment of these infections. The most comprehensively studied glycopeptide antibiotic biosynthetic pathway is that of balhimycin biosynthesis in Amycolatopsis balhimycina. The balhimycin yield obtained by A. balhimycina is, however, low and there is therefore a need to improve balhimycin production. In this study, we performed genome sequencing, assembly and annotation analysis of A. balhimycina and further used these annotated data to reconstruct a genome-scale metabolic model for the organism. Here we generated an almost complete A. balhimycina genome sequence comprising 10,562,587 base pairs assembled into 2,153 contigs. The high GC-genome (69%) includes 8,585 open reading frames (ORFs). We used our integrative toolbox called SEQTOR for functional annotation and then integrated annotated data with biochemical and physiological information available for this organism to reconstruct a genome-scale metabolic model of A. balhimycina. The resulting metabolic model contains 583 ORFs as protein encoding genes (7% of the predicted 8,585 ORFs), 407 EC numbers, 647 metabolites and 1,363 metabolic reactions. During the analysis of the metabolic model, linear, quadratic and evolutionary programming algorithms using flux balance analysis (FBA), minimization of metabolic adjustment (MOMA), and OptGene, respectively were applied as well as phenotypic behavior and improved balhimycin production were simulated. The A. balhimycina model shows a good agreement between in silico data and experimental data and also identifies key reactions associated with increased balhimycin production. The reconstruction of the genome-scale metabolic model of A. balhimycina serves as a basis for physiological characterization. The model allows a rational design of engineering strategies for increasing balhimycin production in A. balhimycina and glycopeptide production in general.

© 2012 Wiley Periodicals, Inc.

KEYWORDS: Amycolatopsis balhimycina; balhimycin; genome annotation; metabolic reconstruction; modeling

Introduction

Amycolatopsis balhimycina belongs to the actinomycetes group of bacteria that encompass a wide range of antibiotics producers, of which several are used for commercial production of antibiotics. Actinomycetes are a group of Gram-positive bacteria with a high GC content in their genomic DNA and in recent years the genomes of several actinomycetes have been sequenced (Bentley et al., 2002; Oliynyk et al., 2007; Omura et al., 2001; Zhao et al., 2010). With the many years of antibiotics use, the development of resistant strains has evolved. Especially, methicillin-resistant Staphylococcus aureus (MRSA) shows resistance to a wide range of antibiotics which cause serious clinical problems (Chamber and Deleo, 2009; Köck et al., 2010). Nowadays, the antibiotics of last resort against MRSA infection are the glycopeptide vancomycin and the lipoglycopeptide teicoplanin, which are both in clinical use. However, there have so far been reported around ten isolated vancomycin-resistant S. aureus (VRSA) strains (Chang et al., 2003; Chamber and Deleo, 2009; Gunnarsson, 2004a; Pearson, 2002) and there is hence a need for novel antibiotics. Despite more than a century of efforts to eradicate or control bacterial pathogen resistances, infection remains a major problem and a growing threat to the public health (Chamber...
was assembled using the Phred/Phrap/Consed software and Deleo, 2009; Köck et al., 2010; Malabarba et al., 1997; Süßmuth and Wohlleben, 2004). Thus, in Europe alone, MRSA infection is estimated to affect more than 150,000 patients annually (Köck et al., 2010), and there is therefore much interest in alternative glycopeptide antibiotics for treatment of MRSA infections. The glycopeptide balhimycin produced by *A. balhimycina*, which is very similar in its structure and activity to vancomycin (Chen et al., 2001), appears to be an attractive antibiotic (Pelzer et al., 1999) since balhimycin shows higher activity towards anaerobic bacteria (Nadkarni et al., 1994). Furthermore, the producing bacterium (*A. balhimycina*) is amenable to genetic manipulation, and novel optimized derivatives of the complex molecule can therefore be generated by genetic engineering (Pelzer et al., 1999; Wohlleben et al., 2009). Balhimycin is consequently proposed as a target for future glycopeptide antibiotic production.

Usually secondary metabolites, in particular antibiotics, are produced at very low yields and often random mutagenesis is applied to increase the yield towards the target metabolite (Balz, 2000). To improve the yield of the balhimycin antibiotic production, we seek to use rational and directed engineering of metabolic pathways using genomics as a tool (Ayar-Kayali and Tarhan, 2010; Gunnarsson et al., 2004b; Thykaer et al., 2010). In this study, a low coverage draft sequence of the genome of *A. balhimycina* with an approximately three-fold coverage was generated by Sanger-Shotgun sequencing. This draft sequence allowed us to perform sequence assembly and annotation of *A. balhimycina*. A genome-scale metabolic model was constructed based on the annotation data which was integrated into pathway databases and complemented by available literature on the physiology and biochemistry of this organism. To comprehensively understand the physiological characteristics of *A. balhimycina*, we further used the developed model to simulate its phenotypic behavior. Linear programming applied by flux balance analysis (FBA) (Orth et al., 2010), quadratic programming applied by minimization of metabolic adjustment (MOMA), (Gertz and Wright, 2003) and evolutionary programming applied by OptGene (Patil et al., 2005) were used to simulate and predict the phenotypic behavior of *A. balhimycina* as well as identify metabolic engineering targets with the goal to increase the yield of balhimycin production.

### Materials and Methods

#### Genome Sequencing and Assembly

The genome sequence of *A. balhimycina* was generated by Sanger-Shotgun sequencing (Fleischmann et al., 1995) of a small insert library and end-sequencing of a large insert cosmid library. Sequencing was performed by MWG-Biotech AG, Ebersberg, Germany. The raw paired-end data was assembled using the Phred/Phrap/Consed software package (Ewing and Green, 1998a; Ewing et al., 1998b; Gordon et al., 1998).

#### Genome Analysis and Annotation

The strategy of the genome analysis and annotation for the *A. balhimycina* can be described as follows: initially, coding regions were identified using the gene finder CRITICA (Badger and Olsen, 1999) with standard parameters and using three iterations. Homologues of the gene products in each coding sequence were predicted with BLASTP (Altschul et al., 1990) against the non-redundant protein database NR. The predicted gene products of *A. balhimycina* were then analyzed and annotated function by comparative sequence alignment analysis against protein sequences from related organisms, i.e., *Amycolatopsis mediterranei* strain U32 (Zhao et al., 2010), *Streptomyces coelicolor* strain A3 (2) (Bentley et al., 2002), *Streptomyces avermitilis* strain MA-4680 (Oamura et al., 2001), *Saccharopolyspora erythraea* strain NRRL23338 (Oligny et al., 2007), *Mycobacterium tuberculosis* strains H37RV (Cole et al., 1998) and CDC1551 (Fleischmann et al., 2002), and *Corynebacterium glutamicum* strain ATCC 13032 (Kalinowski et al., 2003). Towards functional assignment, we integrated multiple protein databases and sequence analysis tools in the form of a toolbox called SEQuence annotaTOR (SEQTOR). The toolbox includes Interproscan (Quevillon et al., 2005), Gene Ontology in 2010 (The Gene Ontology Consortium, 2010), Pfam (Finn et al., 2008), KEGG (Kanehisa, 2002; Kanehisa et al., 2010), TIGR FAMs (Haft et al., 2003), Gene3D (Lees et al., 2003), SUPERFAMILY (Gough and Chothia, 2002), PROSITE (Sigrist et al., 2010), and PANTHER (Thomas et al., 2003). Once our toolbox was used for functional annotation, a prioritized list of potential candidate genes encoding for enzyme functions was generated and then used to reconstruct the metabolic network of *A. balhimycina*.

#### Reconstruction of a Genome-Scale Metabolic Network

The reconstruction of a genome-scale metabolic network aimed at representing the whole metabolism of *A. balhimycina*, which consists of primary catabolism of carbohydrates, biosynthesis of nucleotides, lipids, amino acids, cofactors and production of Gibbs free energy required for biosynthesis, as well as of secondary metabolism. Integration of different types of data is indispensable to achieve a high-quality reconstruction (Henry et al., 2010). In the reconstruction process, we first collected genes and enzymes obtained from the genome analysis and annotation step of *A. balhimycina*. Then, we linked genes and enzymes to corresponding reactions via EC numbers. To do this, we used the earlier related genome-scale metabolic model of *S. coelicolor* as a template (Borodina et al., 2005). To further extend reaction addition in the network reconstruction, we used protein databases and biochemical pathway databases, e.g., EFICAz (Tian et al., 2004), PRIAM (Claudel-Renard et al., 2006), and KEGG (Kanehisa et al., 2010).
et al., 2003), BRENDA (Scheer et al., 2011), KEGG (Kanehisa, 2002; Kanehisa et al., 2004), and MetaCyc (Caspi et al., 2010). The resulting metabolic network contained several gaps, which meant that there were a number of metabolic reactions without corresponding genes or enzymes. To reduce the number of gaps, we performed gap filling by gene context analysis. This analysis was done using the concept of gene order conservation (Osterman and Overbeek, 2003). If a gene could still not be assigned function based on this, the metabolic network reconstruction was carried out by manual curation based on the literature. Here physiological data from the literature were used to infer the presence of reactions, e.g., when there was information of presence of a specific enzyme activity involved in consumption of a given substrate or formation of a given metabolic product, then the underlying reaction was added to the network. Figure 1 shows an overall pipeline for the reconstruction process of the genome-scale metabolic network of A. balhimycina. During the reconstruction process, the selection of reaction directionality, transport reaction and metabolite name were also considered. Concerning directionality, we specified this for few reactions where there was evidence from the biochemical literature, e.g., key kinases in the glycolysis. We mainly used KEGG (Kanehisa, 2002; Kanehisa et al., 2004) and BRENDA (Scheer et al., 2011) as reference reaction direction databases. For metabolite name, we used MetaCyc (Caspi et al., 2010) in addition to KEGG and BRENDA (Kanehisa, 2002; Kanehisa et al., 2004; Scheer et al., 2011). The biomass formation reaction is important for having a high-quality metabolic model (Rocha et al., 2008), and for A. balhimycina the so-called biomass equation was formulated based on the macromolecular cell dry weight composition of Streptomyces species because there was no information available on the cell dry weight composition of A. balhimycina. The biomass equation expresses the drain of building blocks into biomass and takes into account the amount of growth associated ATP. The references used to specify these equations are summarized in Table I and further details are given in Supplementary File 1. The balhimycin synthesis reactions are another important part of the network. The biosynthesis of balhimycin includes the synthesis of specific precursors and many of the reactions to produce them are not associated with an EC number. As vancomycin only differs from balhimycin with respect to the glycosylation pattern, the biosynthetic reactions were written based on KEGG information (www.kegg.com) about the biosynthesis of the vancomycin group antibiotics and on additional literature available for balhimycin and vancomycin biosynthesis.

### Table I. Biomass compositions of A. balhimycina. Note that the measurements compiled from literature account for 91% of the cell dry weight.

<table>
<thead>
<tr>
<th>Biomass building blocks</th>
<th>Refs.</th>
<th>Mass (g/gDW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>Nielsen (1995)</td>
<td>0.412000</td>
</tr>
<tr>
<td>DNA</td>
<td>Genome sequencing</td>
<td>0.036000</td>
</tr>
<tr>
<td>RNA</td>
<td>Neidhardt et al. (1990)</td>
<td>0.167000</td>
</tr>
<tr>
<td>Phospholipids</td>
<td>Wink et al. (2003)</td>
<td>0.027300</td>
</tr>
<tr>
<td>Triacylglycerols</td>
<td>Wink et al. (2003)</td>
<td>0.018180</td>
</tr>
<tr>
<td>Peptidoglycan</td>
<td>Borodina et al. (2005)</td>
<td>0.124770</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Borodina et al. (2005)</td>
<td>0.074860</td>
</tr>
<tr>
<td>Teichoic Acid</td>
<td>Borodina et al. (2005)</td>
<td>0.049910</td>
</tr>
<tr>
<td>Total</td>
<td>Shahab et al. (1996)</td>
<td>0.910020</td>
</tr>
</tbody>
</table>

Shahab et al. (1996); Olukoshi and Packter (1994); Borodina et al. (2005)
Model Construction and Analysis

Once the metabolic network was reconstructed, it was converted to a genome-scale metabolic model by imposing flux balancing of all the metabolites. The resulting model was analyzed using LINDO® for flux balance analysis (FBA) and object-oriented quadratic programming package (Gertz and Wright, 2003) for minimization of metabolic adjustment (MOMA) calculations. For reaction deletion analysis, OptGene (Patil et al., 2005) was used aiming at the optimization of balhimycin productivity (i.e., balhimycin production × biomass production). FBA and MOMA were used to compute the fitness of the in silico strains as reported by Patil et al. (2005).

The A. balhimycina model is available in three formats (Excel, BioOpt, and SBML) with metabolite name list in Supplementary File 2. Besides, the A. balhimycina model is provided in the BioMet Toolbox (www.sysbio.se/BioMet) (Cvijovic et al., 2010), which it is available for use by the systems biology and modeling research community. The model syntax, e.g., BioOpt, and the format instructions are described in the BioMet Toolbox.

Model Validation

For validation of the model, in silico A. balhimycina model simulations were compared with experimental data and with simulations of the S. coelicolor model (Borodina et al., 2005). The first analysis step was to compare the model behavior with experimental results of batch fermentation of A. balhimycina. The model was evaluated by simulating A. balhimycina cell growth on glucose and comparison of the simulated data to experimentally determined growth rate and biomass yield from experimental data (Poli, 2005; Thykaer et al., 2010). Based on different glucose uptake rates according to experimental values (Poli, 2005; Thykaer et al., 2010) and free uptake of inorganic phosphorous, ammonia, oxygen and ions as input data to the model, the flux distributions corresponding to optimal growth were calculated by maximizing the flux of reaction leading to biomass.

Concerning the energetic parameters of the A. balhimycina model, i.e., ATP requirement for non-growth associated purposes (mATP), ATP requirement for synthesis of biomass from macromolecules (KaTP) and the operational ATP formation to oxygen consumption ratio (P/O ratio), the model simulation was fitted with experimental data obtained at a specific growth rate of 0.05 h⁻¹ (Poli, 2005) with glucose as the sole carbon source. The operational P/O ratio was assumed to be 1.5 (Nielsen, 1995), and the energetic parameters mATP and KaTP was hereby estimated to be 3.685 mmol/gDW and 40 mmol/gDW, respectively.

Besides A. balhimycina model simulations, we performed a comparison with the S. coelicolor model under the same condition of minimal medium and glucose uptake rate. The comparison between these two models was carried out for production of important precursor metabolites such as, phosphoenolpyruvate, d-erythrose-4-phosphate, acetate, malonyl-CoA, leucine, asparagine, tyrosine, prephenate and β-hydroxytyrosine and essential amino acids. All the simulations were made using the commercial optimization package LINDO®.

Improvement of Balhimycin Biosynthesis

Before starting to do reaction deletion for improvement of balhimycin production, the simulations were carried out to calculate the theoretical optimal balhimycin production. The simulations were carried out assuming minimal medium in which the model could only take up inorganic phosphorous, ammonia, oxygen, and ions. The main substrate was glucose and the uptake rate selected enabled a specific growth rate of 0.1 h⁻¹ without antibiotic production. In order to predict the theoretical productivity (i.e., balhimycin production x biomass production), FBA was used to compute the maximal balhimycin yield for specific growth rates in a range between 0.01 and 0.09 h⁻¹. OptGene was set to use a population of 100 mutants subjected to up 10 deletions during 500 generations, which were estimated as suitable parameters (Patil et al., 2005), to define a set of knockouts likely to result in an increased antibiotic productivity. To reduce the time consumed on reaction deletion analysis, the model was reduced to only contain unique reactions (corresponding to reaction deletion). Moreover, a set of essential reactions was defined by identifying reactions that resulted in no growth due to a single reaction deletion. After computing a set of possible knockout reactions with OptGene, all combinations of knockouts were tested using LINDO® and the object-oriented programming package.

Results and Discussion

Genome Sequencing

A low coverage draft sequence of the genome of A. balhimycina with an approximately three-fold coverage was generated based on Sanger-Shotgun sequencing (see Materials and Methods). In total, 34,156,807 bases were obtained in 53,213 shotgun reads and further sequences resulted in an assembly into 2,153 contigs with a total length of 10,562,587 nucleotides. The genome sequence of A. balhimycina is registered at the European Nucleotide sequence Archive (ENA) under the project ID 72785.

Analysis of A. balhimycina Genome

The A. balhimycina genome consists of a circular chromosome of 10,562,587 currently estimated base pairs (bp) with an average GC content of 69.9% and 8,585 predicted open reading frames (ORFs). These ORFs have an average length of 308 bp (ranging from 15 to 3,075 bp). To analyze the...
A. balhimycina genome in terms of sequence similarity with other related organisms, comparative sequence alignment analysis was performed. All predicted gene products of A. balhimycina were searched against protein sequences from related organisms (see Materials and Methods). As expected, we observed that A. balhimycina and A. mediterranei has the largest number of orthologues (i.e., 77.60% for one direction) (Fig. 2). Interestingly, the A. mediterranei has the largest number of orthologues from related organisms (see Materials and Methods). A. balhimycina analysis was performed. All predicted gene products of other related organisms, comparative sequence alignment A. balhimycina existed in the genome of M. tuberculosis NRRL23338 other related organisms, A. mediterranei contains the largest number of paralogues (27.01%). For the was performed. As also shown in Figure 2, Besides, an analysis of homology within each the genome was performed. As also shown in Figure 2, A. balhimycina contains the largest number of paralogues (27.01%). For the other related organisms, A. mediterranei U32, S. erythraea NRRL23338, S. coelicolor A3 (2) and S. avermitilis MA-4680, M. tuberculosis strains H37RV and CDC1551, C. glutamicum ATCC 13032, the number of paralogues are found to be 25.28%, 24.06%, 24.19%, 21.70%, 16.26%, and 9.39%, respectively. 

Linking Genes to Enzymes Towards Reconstruction of Metabolic Network

Using the pipeline for reconstruction of a genome-scale metabolic network from genome sequence data (see Fig. 1), we generated a network for A. balhimycina. In the analysis step of the pipeline, the protein sequences of A. balhimycina were imported into a SÉquence annotaTOR (SEQTOR) as our integrative toolbox (see Materials and Methods). Once the protein sequences were annotated by SEQTOR (see Supplementary File 3), we then extracted enzyme functions involved in cellular metabolism, and passed on to the next analysis step. This step involved a reconstruction of the metabolic network, where enzymes were linked to reactions via EC numbers. To perform this, we used the S. coelicolor genome-scale metabolic model as a template network (Borodina et al., 2005) because it is evolutionarily closer to A. balhimycina than the other organisms (see Fig. 2). In addition, we also used pathway databases such as, KEGG (Kanehisa, 2002; Kanehisa et al., 2004), BRENDA (Scheer et al., 2011), MetaCyc (Caspi et al., 2010), EFICAz (Tian et al., 2004), and PRIAM (Claudel-Renard et al., 2003) for formulation of biochemical reactions. The network generating in this way contained several gaps, and we therefore performed gene context analysis and manual curation based on the literature for filling the gaps.

Metabolic Network Characteristics of A. balhimycina

The reconstruction process resulted in a metabolic network containing 583 ORFs as seen in Table II (see Supplementary File 4 in the form of protein sequences). These ORFs linked enzymes to metabolic reactions and catalyzed 1,200 reactions of which 634 were unique reactions, i.e., there

![Figure 2](image_url)
were 566 redundant reactions. The resulting network was analyzed in terms of its topology and hereby a total of 163 gaps were identified. In the process of gap filling, we then manually added the relevant reactions to eliminate existing gaps (Supplementary File 2) in order to enable the model to produce biomass building blocks, biomass and balhimycin, to transport metabolites and to define a biological meaningful respiratory chain. The reactions added for metabolism, transport and balhimycin synthesis correspond to 45% of the unassigned reactions and the remaining reactions contribute mainly to amino acid and cofactor synthesis. Transport reactions added were linked to the few genes annotated as metabolite transporters. Alternatively, transport reactions were added due to the necessity to enable the model to secrete by-products and amino acids and to take up amino acids or other nutrients from the medium. The remaining reactions were added afterwards in order to ensure that the cells could produce cofactors like coenzyme A, acyl-carrier protein, NAD, menaquinone, and amino acids. The amino acid synthesis was distributed into six families and the most incomplete pathways in terms of gene annotation belong to the families of aromatic amino acids and histidine. From the total of 1,363 reactions presented in the model, 1,267 were biochemical transformations and 96 were transport reactions generating a network of 647 metabolites in which 45 were external metabolites. A comparison of the A. balhimycin and S. coelicolor models shows that they share 363 reactions, which represent 57% of the unique reactions present in the A. balhimycin model (see Table II).

**Table II. Main characteristics of A. balhimycin network.**

<table>
<thead>
<tr>
<th>Network characteristics</th>
<th>Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORFs</td>
<td>583</td>
</tr>
<tr>
<td>EC numbers</td>
<td>407</td>
</tr>
<tr>
<td>Metabolic reactions</td>
<td>1,363</td>
</tr>
<tr>
<td>Biochemical transformations</td>
<td>1,267</td>
</tr>
<tr>
<td>Transport reactions</td>
<td>96</td>
</tr>
<tr>
<td>Reactions with ORFs assignments</td>
<td>1,200</td>
</tr>
<tr>
<td>Reactions without ORFs assignments</td>
<td>163</td>
</tr>
<tr>
<td>Unique metabolic reactions</td>
<td>634</td>
</tr>
<tr>
<td>Metabolites</td>
<td>647</td>
</tr>
<tr>
<td>Internal metabolites</td>
<td>602</td>
</tr>
<tr>
<td>External metabolites</td>
<td>45</td>
</tr>
<tr>
<td>Comparative network</td>
<td></td>
</tr>
<tr>
<td>Reactions shared between S. coelicolor and A. balhimycin models</td>
<td>363</td>
</tr>
</tbody>
</table>

**Experimental and In Silico Model Comparison**

The validation of the model is a central step of genome-scale modeling, and if model simulations agree with experimental data, then the model can be applied for further prediction of phenotypic behavior. Here we used experimental data on biomass yield on glucose for comparison with model simulations (see Fig. 3). The A. balhimycin model shows a good agreement with experimental data (Poli, 2005; Thykaer et al., 2010), but the model required lower glucose uptake than the experimental data, in particular in comparison with the data of Thykaer et al. (2010) (see Fig. 3). This deviation can be explained by the fact that equations for biomass and macromolecular compounds synthesis were based on the macromolecular composition of other organisms (see Table I).

**Comparison of Model Simulations for A. balhimycin and S. coelicolor**

A comparison between the model simulations of A. balhimycin and S. coelicolor was carried out under minimal medium conditions (see Materials and Methods) and glucose uptake rate of 0.711 mmol/gDW/h (Poli, 2005). Objective function in the FBA was maximizing biosynthesis of important precursor metabolites, and the fluxes towards these key metabolites were identified (directly related to the yield on glucose as the glucose uptake rate was fixed in all simulations). The key metabolites evaluated are all involved in glycopeptide production: phosphoenolpyruvate, D-erythrose-4-phosphate, acetate, malonyl-CoA, leucine, asparagine, tyrosine, prephenate, and β-hydroxytyrosine. This analysis allowed for comparison of the stoichiometric efficiency in production of these different metabolites in the two organisms. From Figure 4A, it is evident that there are no big differences between A. balhimycin and S. coelicolor models. One of the main reasons is that both models shared the reactions and the central carbon metabolism and the biosynthesis of amino acids seems to be conserved between the two organisms. It is, however, not surprising that there are some differences between the two organisms. For example, we found that the S. coelicolor model is not able to produce β-hydroxytyrosine since it is a non-proteinogenic amino acid which is one of the precursors for balhimycin production (Puk et al., 2004; Recktenwald et al., 2002).
S. coelicolor, which is due to a more efficient energy metabolism of this organism. The models were also active considering the high degree of conservation with other organisms. Some metabolites like tyrosine and prephenate have the same yields on glucose because their synthesis is coupled. For example, prephenate is a tyrosine precursor and the pathway therefore leads to tyrosine, and the pathway is conserved between the two organisms and consequently, the flux towards both metabolites are the same. The fluxes towards phosphoenolpyruvate and D-erythrose-4-phosphate have higher values than towards prephenate synthesis because the synthesis of those metabolites is associated with glucose catabolism. Therefore, the pathways leading to these metabolites are shorter than the ones for prephenate synthesis. Also, the production of amino acids is predicted to be very similar with the two models (Fig. 4, panel B). The good agreement is especially found for amino acids involving shorter pathways than for amino acids produced by long pathways. It is interesting that the stoichiometric efficiency is higher for S. coelicolor, which is due to a more efficient energy metabolism of this organism. The models were also compared in terms of antibiotic yield. For this study, the objective function was the production of actinorhodin or calcium-dependent antibiotic in S. coelicolor and balhimycin in A. balhimycina using the same glucose uptake rate (Supplementary File 5). A. balhimycina has almost the same theoretically maximum yield of balhimycin as S. coelicolor when producing calcium-dependent antibiotic, but S. coelicolor has a higher yield of actinorhodin.

Optimizing Balhimycin Production

To optimize balhimycin production, the maximum balhimycin productivity was simulated with the metabolic model. The antibiotic production was calculated for different specific growth rates as illustrated in Figure 5, and it is found that the maximum balhimycin production is obtained for a specific growth rate of 0.05 h\(^{-1}\) at a “productivity” of 0.00133 mmol balhimycin/gDW/h\(^2\). Reaction deletion analysis was carried out using OptGene (Patil et al., 2005) within a population of 100 mutants subjected to up to 10 deletions generated over 500 generations to identify possible gene deletions that resulted in improved balhimycin synthesis. When FBA was used for flux calculations, the simulations resulted in no balhimycin production. This result is more likely because the concept of FBA generally searches for an optimal solution that maximizes the primary objective function, in this case biomass growth. Synthesis of balhimycin requires consumption of ATP and NADPH and maximization of growth therefore leads to no production of balhimycin. Balhimycin production would only be advantageous for cell growth if there can be established redox coupling between catabolism and the biosynthetic route during synthesis of the intermediary metabolites with less energy requirement. Alternatively, if the biomass production is coupled with the antibiotic synthesis, for example, if a given byproduct can be displaced through the antibiotic synthesis. To our knowledge, there is no modeling method that allows finding...
the minimal set of gene knockout that changes the flux coupling landscape towards the coupling of biomass production with balhimycin production.

In contrast, when MOMA was used for flux calculations, the solution space is searched for a solution that minimizes the flux changes between the knockout strain and the wild type. Since MOMA does not optimize for biomass production, contrary to FBA, it allows the identification of solutions producing secondary metabolites which are a possible way to displace intermediary metabolites from blocked routes. When MOMA was used for calculations, OptGene could find gene deletion targets correspond to 11 reactions resulting in improved balhimycin production. As seen in Table III, a set of 11 reactions was selected for further calculation of the best reaction deletion combination. An object-oriented quadratic programming package was used to delete combinations of these 11 reactions starting from single reaction deletions until the all 11 reactions had been deleted. The highest productivity of each deletion is plotted against the number of reactions deleted as shown in Figure 6. These results show that deleting more than five reactions does not lead to further improvement in the productivity. We notice that the deletion of five reactions shows the highest productivity (see Fig. 6) and two combinations are identified.

These combinations differ in only one reaction. Thus, in the first combination, the reactions deleted are glucose-6-phosphate dehydrogenase with EC: 1.1.1.49 (Contig15081.orf19), malate synthase with EC: 2.3.3.9 (Contig1813.orf5), succinate-CoA ligase (ADP-forming) with EC: 6.2.1.5 (Contig1478.orf1), thioredoxin-disulfide reductase with EC: 1.8.1.9 (Contig1336.orf1) and histidine transport (T_0042). In the second combination, the reaction with EC: 4.1.3.1 (Contig1813.orf4) is deleted instead of reaction of EC: 2.3.3.9 (Contig1813.orf5). From a biochemical view point, these two combinations of gene deletion have the same phenotype representations. Indeed, these two reactions, catalyzed by isocitrate lyase (4.1.3.1) and malate synthase (2.3.3.9), are both key reactions in the glyoxylate shunt and the knockout of any of these enzymes lead to inactivation of this metabolic pathway and prevent the consumption of acetyl-CoA, which is one of the precursors of balhimycin.

The deletion of the reaction of EC: 1.1.1.49 (Contig2081.orf19) forces the model to produce NADPH instead of NADH as the same reaction is present in the model but with NADP as cofactor. The deletion of reactions of EC: 4.1.3.1 (Contig1813.orf5) and EC: 2.3.3.9 (Contig1813.orf5) has the same impact in the model and both stop the glyoxylate cycle. Regarding deleting reaction with EC: 6.2.1.5 (Contig1478.orf1), this blocks the cyclic function of the TCA cycle by preventing the transformation of succinyl-CoA to succinate and CoA. However, the production of succinate from 2-oxoglutarate can take place through succinate semialdehyde combining the reactions with EC: 2.6.1.19 (Contig1784.orf2) and EC: 1.2.1.16 (Contig1569.orf2 or Contig 1008.orf2). This path leads to the production of NADH or NADPH, respectively, instead of the ATP (EC: 6.2.1.5). In order to improve balhimycin production, it can be concluded that the model has to redirect the electrochemical energy present as NADPH towards antibiotic production and reduce the growth rate by decreasing the ATP production.

### Conclusions

The genome sequencing and the annotation have opened possibilities for comprehensive understanding of...
A. balhimycina at the systematic level. Here a genome-scale metabolic model of A. balhimycina was reconstructed consisting of 583 ORFs as protein encoding genes (7% of the predicted 8,585 ORFs), 407 EC numbers, 647 metabolites and 1,363 metabolic reactions. The A. balhimycina model shows a good agreement between in silico data and experimental data. The model was further used to identify key reactions associated with increased balhimycin production. The reaction deletion analysis carried out by OptGene using MOMA as the framework for flux calculations in the model revealed 11 reactions that could improve antibiotic productivity when deleted. Two combinations of five reactions deletions show improved antibiotic productivity. Deletions of the candidate reactions force the model to increase the NADPH production and reduce the ATP production. The increase of NADPH can be used for synthesis of antibiotic building blocks where an increased flux through the pentose phosphate pathway in A. balhimycina may also result in a higher activity of the shikimate pathway, which has already proven to be important in optimization of the balhimycin production (Thykaer et al., 2010). Additionally, the decrease of ATP production reduces the growth rate, and hereby allow for re-directing building blocks towards antibiotic production. As described above, A. balhimycina has interesting perspectives as cell factory for production of novel glycopeptides. This organism has already been shown to be genetically assessable and it may therefore be used for production of novel antibiotics. Thus, the genome-scale metabolic model presented in this work represents a valuable resource on the metabolism of an important glycopeptide producing bacterium and this may allow for identification of novel metabolic engineering strategies for improving the production of this important class of antibiotics in the future.

We would like to thank the European Union Framework VI ”COM-BIG-TOP” (http://www.combigtop.uni-tuebingen.de), the Lundbeck Foundation, Denmark, and the Chalmers Foundation for financial support. WW would like to thank King Mongkut’s University of Technology Thonburi (KMUTT), Thailand and Soochow University, China for financial support. TW and WW were funded by the German Ministry of Education and Research (BMBF) with the grant “GenBioCom” FKZ 0315585A.

References


Ayar-Kayali H, Tarhan L. 2010. Intra- and extra-cellular flux distributions in A. balhimycina described above, directing building blocks towards antibiotic production. As production reduces the growth rate, and hereby allow for re-directing building blocks towards antibiotic production. As described above, A. balhimycina has interesting perspectives as cell factory for production of novel glycopeptides. This organism has already been shown to be genetically assessable and it may therefore be used for production of novel antibiotics. Thus, the genome-scale metabolic model presented in this work represents a valuable resource on the metabolism of an important glycopeptide producing bacterium and this may allow for identification of novel metabolic engineering strategies for improving the production of this important class of antibiotics in the future.

We would like to thank the European Union Framework VI ”COM-BIG-TOP” (http://www.combigtop.uni-tuebingen.de), the Lundbeck Foundation, Denmark, and the Chalmers Foundation for financial support. WW would like to thank King Mongkut’s University of Technology Thonburi (KMUTT), Thailand and Soochow University, China for financial support. TW and WW were funded by the German Ministry of Education and Research (BMBF) with the grant “GenBioCom” FKZ 0315585A.

References


