

## 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](https://doi.org/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):*  
Vongsangnak, W., Figueiredo, L. F., Förster, J., Weber, T., Thykær, J., Stegmann, E., ... Nielsen, J. (2012). Genome-scale metabolic representation of *Amycolatopsis balhimycina*. *Biotechnology and Bioengineering* (Print), 109(7), 1798-1807. DOI: 10.1002/bit.24436

## DTU Library

Technical Information Center of Denmark

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Genome-Scale Metabolic Representation of *Amycolatopsis balhimycina*

Wanwipa Vongsangnak,<sup>1,2</sup> Luís Filipe Figueiredo,<sup>1,2</sup> Jochen Förster,<sup>2</sup> Tilmann Weber,<sup>3</sup> Jette Thykaer,<sup>1</sup> Evi Stegmann,<sup>3</sup> Wolfgang Wohlleben,<sup>3</sup> Jens Nielsen<sup>1</sup>

<sup>1</sup>Department of Systems Biology, Technical University of Denmark, Denmark; telephone: +46-0-31-772-3804; fax: +46-0-31-772-3801; e-mail: nielsenj@chalmers.se

<sup>2</sup>Fluxome A/S, Gymnasievej 5, 3660 Stenlose, Denmark

<sup>3</sup>Department of Microbiology and Biotechnology, Interfaculty Institute of Microbiology and Infection Medicine, Eberhard Karls University Tübingen, Tübingen, Germany

Received 1 September 2011; revision received 17 November 2011; accepted 3 January 2012

Published online 6 February 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bit.24436

**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.

Biotechnol. Bioeng. 2012;109: 1798–1807.

© 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

Wanwipa Vongsangnak and Luís Filipe Figueiredo contributed equally to this work. Wanwipa Vongsangnak's present address is Center for Systems Biology, Soochow University, Suzhou 215006, China.

Luís Filipe Figueiredo's present address is Department of Bioinformatics, Friedrich-Schiller-Universität Jena, Germany.

Jens Nielsen's present address is Department of Chemical and Biological Engineering, Chalmers University of Technology, Sweden.

Correspondence to: J. Nielsen

Additional supporting information may be found in the online version of this article.

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 (Baltz, 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 reconstructed 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 (Omura et al., 2001), *Saccharopolyspora erythraea* strain NRRL23338 (Oliynyk 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., 2004), TIGRFAMs (Haft et al., 2003), Gene3D (Lees et al., 2010), 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

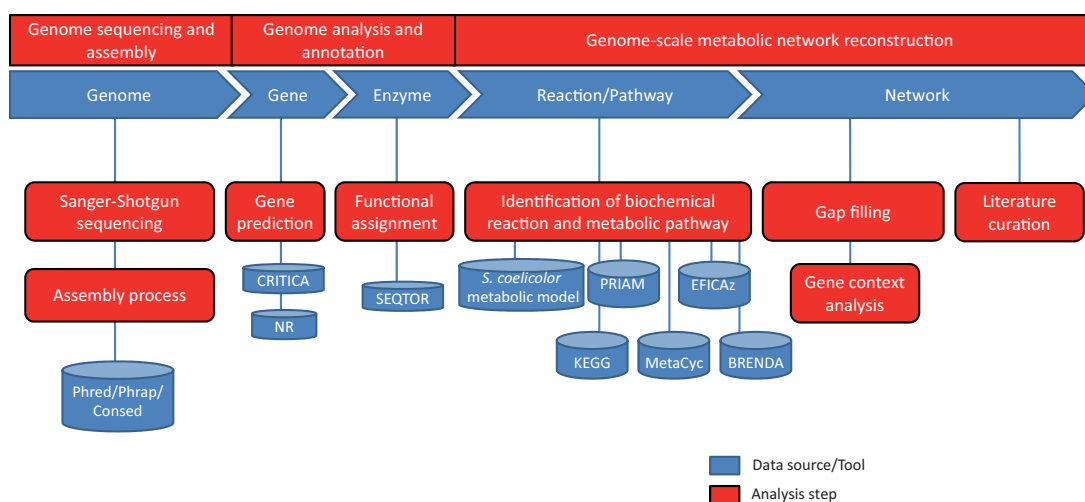
The reconstruction of a genome-scale metabolic network aimed at representing the whole metabolism of *A. balhimycina*, which consists of primary catabolism of carbohydrates, biosynthesis of nucleotides, lipids, amino acids, cofactors and production of Gibbs free energy required for biosynthesis, as well 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., EFlCAz (Tian et al., 2004), PRIAM (Claudel-Renard

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

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

| Biomass building blocks | Refs.  | Mass (g/gDW) |
|-------------------------|--|--------------|
| Protein                 | Nielsen (1995)   | 0.412000     |
| DNA                     | Genome sequencing  | 0.036000     |
| RNA                     | Neidhardt et al. (1990)  | 0.167000     |
| Phospholipids           | Wink et al. (2003)   | 0.027300     |
| Triacylglycerols        | Wink et al. (2003)   | 0.018180     |
| Peptidoglycan           | Borodina et al. (2005)   | 0.124770     |
| Carbohydrates           | Borodina et al. (2005)   | 0.074860     |
| Teichoic Acid           | Borodina et al. (2005)   | 0.049910     |
| Total                   | Shahab et al. (1996)<br>Olukoshi and Packter (1994);<br>Borodina et al. (2005) | 0.910020     |

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](http://www.kegg.com)) about the biosynthesis of the vancomycin group antibiotics and on additional literature available for balhimycin and vancomycin biosynthesis.



**Figure 1.** Illustration of overall pipeline for the reconstruction process of the genome-scale metabolic network of *A. balhimycina*. Red boxes correspond to the analysis steps—genome sequencing and assembly, genome analysis and annotation, and genome-scale metabolic network reconstruction. Blue boxes represent the data sources and tools used in the reconstruction process. [Color figure can be seen in the online version of this article, available at <http://wileyonlinelibrary.com/bit>]

## 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<sup>®</sup> 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 ( $K_{ATP}$ ) 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  $0.05\text{ h}^{-1}$  (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  $K_{ATP}$  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  $\beta$ -hydroxytyrosine and essential amino acids. All the simulations were made using the commercial optimization package LINDO<sup>®</sup>.

## 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  $0.1\text{ h}^{-1}$  without antibiotic production. In order to predict the theoretical productivity (i.e., balhimycin production × biomass production), FBA was used to compute the maximal balhimycin yield for specific growth rates in a range between 0.01 and  $0.09\text{ h}^{-1}$ . OptGene was set to use a population of 100 mutants subjected to up to 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<sup>®</sup> 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 analysis shows a very high percentage of homologous protein sequences between *A. balhimycina* and *S. coelicolor* A3 (2), and about 34.07% of the proteins are orthologues between these two organisms. This suggests that *S. coelicolor* A3 (2) is probably evolutionarily closer to *A. balhimycina* than the other organisms considered in this investigation. 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%, 14.54%, 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 SEquence 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

| <i>C. glutamicum</i><br>ATCC13032<br>2,993 ORFs | <i>M. tuberculosis</i><br>H37RV<br>3,991 ORFs | <i>M. tuberculosis</i><br>CDC1551<br>4,189 ORFs | <i>S. avermitilis</i><br>MA-4680<br>7,577 ORFs | <i>S. coelicolor</i><br>A3 (2)<br>7,769 ORFs | <i>S. erythraea</i><br>NRRL23338<br>7,198 ORFs | <i>A. mediterranei</i><br>U32<br>9,228 ORFs | <i>A. balhimycina</i><br>8,585 ORFs |   |
|---|---|---|--|--|--|---|-------------------------------------|---|
| 897/2,993<br>(29.97%)                           | 1,382/3,991<br>(34.63%)                       | 1,344/4,189<br>(32.08%)                         | 2,595/7,577<br>(34.25%)                        | 2,637/7,769<br>(33.94%)                      | 3,075/7,198<br>(42.72%)                        | 6,349/9,228<br>(68.80%)                     | 2,319/8,585<br>(27.01%)             | <i>A. balhimycina</i><br>10,562,587 bp<br>(69.9 %GC)          |
| 1,036/2,993<br>(34.61%)                         | 1,617/3,991<br>(40.52%)                       | 1,584/4,189<br>(37.81%)                         | 3,086/7,577<br>(40.73%)                        | 3,164/7,769<br>(40.73%)                      | 3,472/7,198<br>(48.24%)                        | 2,333/9,228<br>(25.28%)                     | 6,662/8,585<br>(77.60%)             | <i>A. mediterranei</i> U32<br>10,236,715 bp<br>(71.3 %GC)     |
| 818/2,993<br>(27.33%)                           | 1,465/3,991<br>(36.71%)                       | 1,418/4,189<br>(33.85%)                         | 2,663/7,577<br>(35.15%)                        | 2,788/7,769<br>(35.89%)                      | 1,732/7,198<br>(24.06%)                        | 3,744/9,228<br>(40.57%)                     | 2,563/8,585<br>(29.85%)             | <i>S. erythraea</i> NRRL23338<br>8,212,805bp<br>(71.15 %GC)   |
| 879/2,993<br>(29.37%)                           | 1,197/3,991<br>(29.99%)                       | 1,175/4,189<br>(28.05%)                         | 4,814/7,577<br>(63.53%)                        | 1,879/7,769<br>(24.19%)                      | 2,777/7,198<br>(38.58%)                        | 3,402/9,228<br>(36.87%)                     | 2,925/8,585<br>(34.07%)             | <i>S. coelicolor</i> A3 (2)<br>8,667,507 bp<br>(72.10 %GC)    |
| 858/2,993<br>(28.67%)                           | 1,247/3,991<br>(31.25%)                       | 1,226/4,189<br>(29.27%)                         | 1,644/7,577<br>(21.70%)                        | 4,806/7,769<br>(61.86%)                      | 2,723/7,198<br>(37.83%)                        | 3,359/9,228<br>(36.40%)                     | 2,924/8,585<br>(34.06%)             | <i>S. avermitilis</i> MA-4680<br>9,025,608 bp<br>(70.70 %GC)  |
| 903/2,993<br>(30.17%)                           | 3,751/3,991<br>(93.99%)                       | 609/4,189<br>(14.54%)                           | 1,394/7,577<br>(18.40%)                        | 1,368/7,769<br>(17.61%)                      | 1,616/7,198<br>(22.45%)                        | 1,857/9,228<br>(20.12%)                     | 1,657/8,585<br>(19.30%)             | <i>M. tuberculosis</i> CDC1551<br>4,403,837 bp<br>(65.60 %GC) |
| 911/2,993<br>(30.44%)                           | 649/3,991<br>(16.26%)                         | 3,701/4,189<br>(88.35%)                         | 1,412/7,577<br>(18.64%)                        | 1,389/7,769<br>(17.88%)                      | 1,640/7,198<br>(22.78%)                        | 1,876/9,228<br>(20.33%)                     | 1,679/8,585<br>(19.56%)             | <i>M. tuberculosis</i> H37RV<br>4,411,532 bp<br>(65.60 %GC)   |
| 281/2,993<br>(9.39%)                            | 912/3,991<br>(22.85%)                         | 905/4,189<br>(21.60%)                           | 991/7,577<br>(13.08%)                          | 1,046/7,769<br>(13.46%)                      | 1,179/7,198<br>(16.38%)                        | 1,215/9,228<br>(13.17%)                     | 1,096/8,585<br>(12.77%)             | <i>C. glutamicum</i> 13032<br>3,282,708 bp<br>(53.80 %GC)     |

Homology within organism Homology between organisms

**Figure 2.** Pairwise comparative BLAST matrix of protein sequences of *A. balhimycina* and closely related organisms. The figure shows how orthologues and paralogues existed in the genome of *A. balhimycina* as well as in the genomes of other related organisms. BLASTP cut-off: *E*-value (<1 E-30); % Coverage (>50); % Identity (>40). [Color figure can be seen in the online version of this article, available at <http://wileyonlinelibrary.com/bit>]



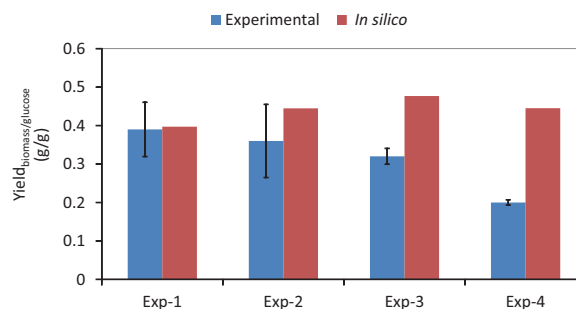
**Table II.** Main characteristics of *A. balhimycina* network.

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

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. balhimycina* and *S. coelicolor* models shows that they share 363 reactions, which represent 57% of the unique reactions present in the *A. balhimycina* model (see Table II).

### 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. balhimycina* model shows a

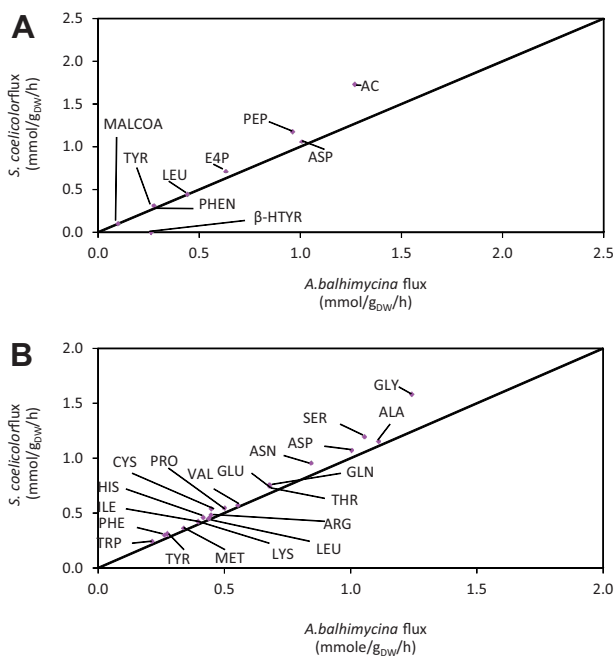


**Figure 3.** Comparison of experimental (blue) and in silico (red) yield of biomass per glucose from different experiments. These experimental data were obtained from different batch fermentations. Three batches data (Exp-1, Exp-2, and Exp-3) were taken from Poli (2005) and one batch data (Exp-4) was taken from Thykaer et al. (2010). [Color figure can be seen in the online version of this article, available at <http://wileyonlinelibrary.com/bit>]

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. balhimycina* and *S. coelicolor*

A comparison between the model simulations of *A. balhimycina* 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  $\beta$ -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. balhimycina* 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  $\beta$ -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). In



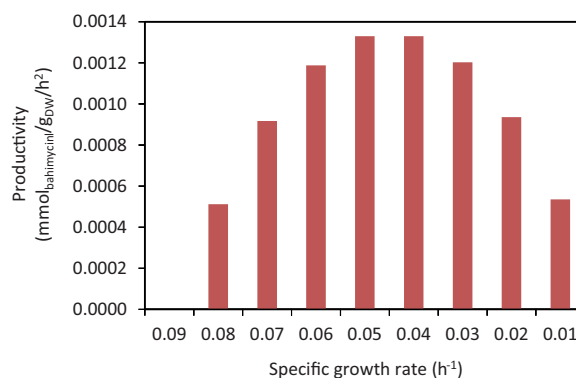
**Figure 4.** A: Comparison of precursor metabolites fluxes between *A. balhimycina* and *S. coelicolor* models. PEP, phosphoenolpyruvate; E4P, D-erythrose-4-phosphate; AC, acetate; MALCOA, malonyl-CoA; LEU, leucine; ASP, asparagine; TYR, tyrosine; PHEN, prephenate;  $\beta$ -HTYR,  $\beta$ -hydroxytyrosine. B: Comparison of amino acids fluxes between *A. balhimycina* and *S. coelicolor* models.

In addition, there are some metabolites that show some deviation, e.g., acetate and glycine. One of the reasons for this deviation has to do with the fact that a part of glycolysis is not operational due to the absence of 6-phosphofructokinase in the annotated data of *A. balhimycina*. Adding these missing reactions would probably make glycolysis fully operational as all other enzymes are present and likely to be 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 though 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  $0.05 \text{ h}^{-1}$  at a “productivity” of  $0.00133 \text{ mmol balhimycin/gDW/h}^2$ . 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



**Figure 5.** “Productivity” of balhimycin at different specific growth rates. [Color figure can be seen in the online version of this article, available at <http://wileyonlinelibrary.com/bit>]

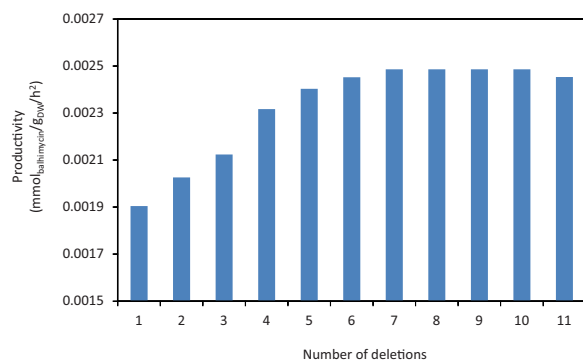


**Table III.** Set of reactions taken from optimization with OptGene which improved balhimycin synthesis using MOMA for flux calculations.

| ORF name/reaction name | EC number | Metabolic reaction   |
|------------------------|-----------|--|
| Contig1752.orf4        | 2.7.2.3   | 3-Phospho-D-glyceroyl phosphate + ADP $\leftrightarrow$ 3-Phospho-D-glycerate + ATP      |
| Contig2081.orf19       | 1.1.1.49  | Alpha-D-Glucose 6-phosphate + NAD $\rightarrow$ D-Glucono-1,5-lactone 6-phosphate + NADH |
| Contig1813.orf5        | 2.3.3.9   | Acetyl-CoA + Glyoxylate $\rightarrow$ CoA + Malate                                       |
| Contig1813.orf4        | 4.1.3.1   | Isocitrate $\rightarrow$ Glyoxylate + Succinate  |
| Contig1478.orf1        | 6.2.1.5   | ADP + Orthophosphate + Succinyl-CoA $\leftrightarrow$ ATP + Succinate + CoA              |
| Contig1502.orf4        | 1.7.1.4   | nitrite + 3 NADH $\rightarrow$ 3 NAD + NH <sub>3</sub>                                   |
| Contig1336.orf1        | 1.8.1.9   | Oxidized thioredoxin + NADPH $\rightarrow$ Reduced thioredoxin + NADP                    |
| G_0020                 | 2.3.1.30  | L-Serine + Acetyl-CoA $\leftrightarrow$ O-Acetyl-L-serine + CoA                          |
| Contig1939.orf4        | 1.3.99.3  | 3-Methylbutanoyl-CoA + menaquinone $\rightarrow$ 3-methylcrotonyl-CoA + menaquinol       |
| Contig1562.orf2        | 1.2.1.9   | D-Glyceraldehyde 3-phosphate + NADP $\rightarrow$ NADPH + 3-Phospho-D-glycerate          |
| T_0042                 | —         | L-Histidine $\rightarrow$ L-Histidine (external) + H (external)                          |

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.



**Figure 6.** “Productivity” against the number of gene deletions that result in improved productivity. [Color figure can be seen in the online version of this article, available at <http://wileyonlinelibrary.com/bit>]

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 (Contig2081.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.orf4) 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. WV 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

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.
- Ayar-Kayali H, Tarhan L. 2010. Intra- and extra-cellular flux distributions of TCA and glyoxalate cycle and vancomycin production of *Amycolatopsis orientalis* grown in different glycerol concentration. *Prikl Biokhim Mikrobiol* 46:519–526.
- Badger JH, Olsen GJ. 1999. CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol* 16:512–524.
- Baltz RH. 2000. Genetic methods and strategies for secondary metabolite yield improvement in actinomycetes. *Antonie van Leeuwenhoek* 79: 251–259.
- Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O’Neil S, Rabinowitz E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature* 417:141–147.
- Borodina I, Krabben P, Nielsen J. 2005. Genome-scale analysis of *Streptomyces coelicolor* A3 (2) metabolism. *Genome Res* 15:820–829.
- Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD. 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 38:D473–D479.
- Chamber HF, Deleo FR. 2009. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nat Rev Microbiol* 7:629–641.
- Chang S, Sievert DM, Hageman JC. 2003. Infection with vancomycin resistant *Staphylococcus aureus* containing the vanA resistance gene. *N Engl J Med* 348:1342–1347.
- Chen H, Tseng CC, Hubbard BK, Walsh CT. 2001. Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *PNAS* 98:14901–14906.
- Claudel-Renard C, Chevalet C, Faraut T, Kahn D. 2003. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res* 31:6633–6639.
- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaija F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393:537–544.
- Cvijovic M, Olivares-Hernández R, Agren R, Dahr N, Vongsangnak W, Nookaew I, Patil KR, Nielsen J. 2010. BioMet Toolbox: genome-wide analysis of metabolism. *Nucleic Acids Res* 38:W144–W149.
- Ewing B, Green P. 1998a. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186–194.
- Ewing B, Hillier L, Wendl MC, Green P. 1998b. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185.
- Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz H, Ceric G, Forslund K, Sean R, Eddy SR, Erik LL, Sonnhammer EL, Bateman A. 2008. The Pfam protein families database. *Nucleic Acids Res* 36:D281–D288.
- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton G, Fitzhugh W, Fields C, Gocyne JD, Scott J, Shirley R, Liu L, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom Eva, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496–512.
- Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D, Hickey E, Kolonay JF, Nelson WC, Umayam LA, Ermolaeva M, Salzberg SL, Delcher A, Utterback T, Weidman J, Khouri H, Gill J, Mikula A, Bishai W, Jacobs WR Jr., Venter JC, Fraser CM. 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* 184: 5479–5490.
- Gertz EM, Wright SJ. 2003. Object-oriented software for quadratic programming. *ACM Trans Math Software* 29:58–81.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202.

- Gough J, Chothia C. 2002. SUPERFAMILY: HMMs representing all proteins of known structure SCOP sequence searches, alignments and genome assignments. *Nucleic Acids Res* 30:268–272.
- Gunnarsson N. 2004a. Antibiotic production and primary metabolism in the actinomycetes *Nonomuraea* sp ATCC 39727. Ph.D. Thesis. BioCentrum, Technical University of Denmark, Denmark.
- Gunnarsson N, Mortensen UH, Sosio M, Nielsen J. 2004b. Identification of the Entner-Doudoroff pathway in an antibiotic-producing actinomycete species. *Mol Microbiol* 52:895–902.
- Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. *Nucleic Acids Res* 31:371–373.
- Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 28:977–982.
- Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, Goesmann A, Hartmann M, Huthmacher K, Krämer R, Linke B, McHardy AC, Meyer F, Möckel B, Pfefflerle W, Pühler A, Rey DA, Rückert C, Rupp O, Sahm H, Wendisch VF, Wiegäbe I, Tauch A. 2003. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol* 104:5–25.
- Köck R, Becker K, Cookson B, van Gemert-Pijnen JE, Harbarth S, Kluytmans J, Mielke M, Peters G, Skov RL, Struelens MJ, Tacconelli E, Navarro Torné A, Witte W, Friedrich AW. 2010. Methicillin-resistant *Staphylococcus aureus* MRSA: burden of disease and control challenges in Europe. *Euro Surveill* 15:19688.
- Kanehisa M. 2002. The KEGG database. *Novartis Found Symp* 247:91–9101.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277–D280.
- Lees J, Yeats C, Redfern O, Clegg A, Orengo C. 2010. Gene3D: merging structure and function for a thousand genomes. *Nucleic Acids Res* 38:D296–D300.
- Malabarba A, Nicas T, Ciabatti R. 1997. Glycopeptide resistance in multiple antibiotic-resistant Gram-positive bacteria: a current challenge for novel semi-synthetic glycopeptide derivatives. *Eur J Med Chem* 32:459–478.
- Nadkarni SR, Patel MV, Chatterjee S, Vijayakumar EK, Desikan KR, Blumbach J, Ganguli BN, Limbert M. 1994. Balhimycin, a new glycopeptides antibiotic produced by *Amycolatopsis* sp Y-86 21022 Taxonomy, production, isolation and biological activity. *J Antibiot Tokyo* 47:334–341.
- Neidhardt FC, Ingraham JL, Schaechter M. 1990. Physiology of the bacterial cell – A molecular approach. Sunderland, MA: Sinauer Associates. p. 520.
- Nielsen J. 1995. Physiological Engineering Aspects of *Penicillium chrysogenum*. Ph.D. thesis. Polyteknisk Forlag, Technical University of Denmark, Denmark.
- Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, Haydock SF, Leadlay PF. 2007. Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat Biotechnol* 25:447–453.
- Olukoshi ER, Packter NM. 1994. Importance of stored triacylglycerols in *Streptomyces*: possible carbon source for antibiotics. *Microbiology* 140:931–943.
- Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T, Kikuchi H, Shiba T, Sakaki Y, Hattori M. 2001. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci USA* 98:12215–12220.
- Orth JD, Thiele I, Palsson B. 2010. What is flux balance analysis? *Nat Biotechnol* 28:245–248.
- Osterman A, Overbeek R. 2003. Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol* 7:238–251.
- Patil KR, Rocha I, Förster J, Nielsen J. 2005. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics* 6:308.
- Pearson H. 2002. Superbug hurdles key drug barrier. *Nature* 418:469–470.
- Pelzer S, Sübmuth R, Heckmann D, Recktenwald J, Huber P, Jung G, Wohlleben W. 1999. Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908. *Antimicrob Agents Chemother* 43:1565–1573.
- Poli E. 2005. BT Antibiotic Production by *Amycolatopsis balhimycina*. Dissertation. BioCentrum, Technical University of Denmark, Denmark.
- Puk O, Bischoff D, Kittel C, Pelzer S, Weist S, Stegmann E, Roderich D, Süßmuth RD, Wohlleben W. 2004. Biosynthesis of Chloro-β-Hydroxytyrosine, a nonproteinogenic amino acid of the peptidic backbone of glycopeptide antibiotics. *J Bacteriol* 186:6093–6100.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res* 33:W116–W120.
- Recktenwald J, Shawky R, Puk O, Pfennig F, Keller U, Wohlleben W, Pelzer S. 2002. Nonribosomal biosynthesis of vancomycin-type antibiotics: a heptapeptide backbone and eight peptide synthetase modules. *Microbiology* 148:1105–1118.
- Rocha I, Förster J, Nielsen J. 2008. Design and application of genome-scale reconstructed metabolic models. *Methods Mol Biol* 416:409–431.
- Shahab N, Flett F, Oiver SG, Butler PR. 1996. Growth rate control of protein and nucleic acid content in *Streptomyces coelicolor* A3 (2) and *Escherichia coli* B/r. *Microbiology* 142:1927–1935.
- Scheer M, Grote A, Chang A, Schomburg I, Munnaretto C, Rother M, Söhngen C, Stelzer M, Thiele J, Schomburg D. 2011. BRENDA, the enzyme information system in 2011. *Nucleic Acids Res* 39:D670–D676.
- Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N. 2010. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* 38:D161–D166.
- Süßmuth RD, Wohlleben W. 2004. The biosynthesis of glycopeptides antibiotics—a model for complex, non-ribosomally synthesized, peptidic secondary metabolites. *Appl Microbiol Biotechnol* 63:344–350.
- Tian W, Arakaki AK, Skolnick J. 2004. EFICAZ: a comprehensive approach for accurate genome-scale enzyme function inference. *Nucleic Acids Res* 32:6226–6239.
- The Gene Ontology consortium. 2010. The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res* 38:D331–D335.
- Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, Ladunga I, Ulitsky-Lazareva B, Muruganujan A, Rabkin S, Vandergriff JA, Dormeieux O. 2003. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res* 31:334–341.
- Thykaer J, Nielsen J, Wohlleben W, Weber T, Gutknecht M, Lantz AE, Stegmann E. 2010. Increased glycopeptide production after overexpression of shikimate pathway genes being part of the balhimycin biosynthetic gene cluster. *Metab Eng* 12:455–461.
- Wink JM, Kroppenstedt RM, Ganguli BN, Nadkarni SR, Schumann P, Seibert G, Stackebrandt E. 2003. Three new antibiotic producing species of the genus *Amycolatopsis*, *Amycolatopsis balhimycina* sp. nov., *A. tolypomycina* sp. nov., *A. vancoresmycina* sp. nov., and description of *Amycolatopsis keratiniphila* subsp. *keratiniphila* subsp. nov. and *A. keratiniphila* subsp. *nogabecina* subsp. nov. *System Appl Microbiol* 26:38–46.
- Wohlleben W, Stegmann E, Sübmuth RD. 2009. Molecular genetic approaches to analyze glycopeptide biosynthesis. *Methods Enzymol* 458:459–486.
- Zhao W, Zhong Y, Yuan H, Wang J, Zheng H, Wang Y, Cen X, Xu F, Bai J, Han X, Lu G, Zhu Y, Shao Z, Yan H, Li C, Peng N, Zhang Z, Zhang Y, Lin W, Fan Y, Qin Z, Hu Y, Zhu B, Wang S, Ding X, Zhao GP. 2010. Complete genome sequence of the rifamycin SV-producing *Amycolatopsis mediterranei* U32 revealed its genetic characteristics in phylogeny and metabolism. *Cell Res* 20:1096–1108.