Received: 26 October 2017 | Revised: 3 March 2018 | Accepted: 19 March 2018
DOI: 10.1002/bit.26598

ARTICLE

WILEY BIOTECHNOLOGY & BIOENGINEERING

Analysis of metabolic networks of *Streptomyces leeuwenhoekii* C34 by means of a genome scale model: Prediction of modifications that enhance the production of specialized metabolites

Valeria Razmilic | Jean F. Castro | Barbara Andrews | Juan A. Asenjo

Department of Chemical Engineering and Biotechnology, Centre for Biotechnology and Bioengineering (CeBiB), Universidad de Chile, Santiago, Chile

Correspondence
Juan A. Asenjo, Department of Chemical Engineering and Biotechnology, Centre for Biotechnology and Bioengineering (CeBiB), Universidad de Chile, Santiago, Chile.
Email: juasenjo@ing.uchile.cl

Funding information
Comisión Nacional de Investigación Científica y Tecnológica, Grant number: FB0001; UK Newton Project for UK-Chile collaboration, Grant number: JIC CA 586; Basal Programme of CONICYT (Chile)

Abstract

The first genome scale model (GSM) for *Streptomyces leeuwenhoekii* C34 was developed to study the biosynthesis pathways of specialized metabolites and to find metabolic engineering targets for enhancing their production. The model, *i*VR1007, consists of 1,722 reactions, 1,463 metabolites, and 1,007 genes, it includes the biosynthesis pathways of chaxamycins, chaxalactins, desferrioxamines, ectoine, and other specialized metabolites. *i*VR1007 was validated using experimental information of growth on 166 different sources of carbon, nitrogen and phosphorous, showing an 83.7% accuracy. The model was used to predict metabolic engineering targets for enhancing the biosynthesis of chaxamycins and chaxalactins. Gene knockouts, such as *sle03600* (L-homoserine *O*-acetyltransferase), and *sle39090* (trehalose-phosphate synthase), that enhance the production of the specialized metabolites by increasing the pool of precursors were identified. Using the algorithm of flux scanning based on enforced objective flux (FSEOF) implemented in python, 35 and 25 over-expression targets for increasing the production of chaxamycin A and chaxalactin A, respectively, that were not directly associated with their biosynthesis routes were identified. Nineteen over-expression targets that were common to the two specialized metabolites studied, like the over-expression of the acetyl carboxylase complex (*sle47660* (*accA*) and any of the following genes: *sle44630* (*accA_1*) or *sle39830* (*accA_2*) or *sle27560* (*bccA*) or *sle59710*) were identified. The predicted knockouts and over-expression targets will be used to perform metabolic engineering of *S. leeuwenhoekii* C34 and obtain overproducer strains.

KEYWORDS
chaxalactins, chaxamycins, genome scale model, metabolic engineering, specialized metabolites, *Streptomyces leeuwenhoekii*

1 | INTRODUCTION

Microbial specialized metabolites are generally synthesized in discrete amount by the cell during the secondary metabolism using precursors produced during the primary metabolism (Hiltner, Hunter, & Hoskisson, 2015), these metabolites accomplish diverse roles within the microbial environment (Demain & Fang, 2000). Specialized metabolites have shown to have diverse chemical structures and many different applications, especially in medicine (Gross, 2007; McMurry & Begley, 2005).

Nowadays the need for novel specialized metabolites with antibiotic activity is urgent (Laxminarayan et al., 2013), and to accomplish their therapeutic use, it is necessary to enhance their production (Parekh, Vinci, & Strobel, 2000). Traditional methods for increasing the production of specialized metabolites involved changes in the expression/regulation of genes/proteins. However, the improvement due to genetic regulation would increase the production of the biosynthetic genes and resistance genes, but not the availability of precursors, which would still be a limit on the way of enhancing the titers of the metabolite (Chen, Smanski, & Shen, 2010). Understanding the connection between primary and secondary metabolism, to study the pathways devoted to precursor biosynthesis, and to exploit the genomic potential of the microorganisms is essential in order to produce them and improve their yield. Also, metabolic engineering targets for improving the production of the precursors can be identified through genome-scale models.

The majority of the specialized metabolites with antibiotic activity come from *Streptomyces* strains (Kieser, Bibb, Buttner, Chater, & Hopwood, 2000). Examples are: chloramphenicol from *S. venezuelae*, kanamycin from *S. kanamyceticus*, streptomycin from *S. griseus* and daptomycin from *S. roseosporus*. The sequenced genome of strains, like *S. coelicolor* (Bentley et al., 2002), or *S. avermitilis* (Ōmura et al., 2001), have allowed study of the biosynthesis of specialized metabolites and also identification of silent or cryptic biosynthetic gene clusters (BGCs) associated to new compounds (Challis, 2008; Lautru, Deeth, Bailey, & Challis, 2005).

A new *Streptomyces* strain was isolated from the Chaxa lagoon of the Atacama Desert of northern Chile (Okoro et al., 2009). This novel strain was named *Streptomyces leeuwenhoekii* C34 and a draft genome sequence was obtained (Busarakam et al., 2014). This strain produces novel specialized metabolites with antibiotic and anticancer activity, the chaxamycins and chaxalactins (Rateb, Houssen, Arnold et al., 2011, Rateb, Houssen, Harrison et al., 2011). The chaxamycins A-D are ansamycin-type polyketides synthesized from 3-amino-5-hydroxy-benzoic acid (AHBA) as starting molecule and malonyl-CoA and (*S*)-methylmalonyl-CoA as extensor units. The chaxamycin polyketide synthases (PKS) are similar to that of the rifamycin biosynthesis (Castro et al., 2015). Also, chaxamycin A-C have antitumor activity through inhibition of the heat shock protein 90 (Hsp90) (Rateb, Houssen, Arnold et al., 2011). The chaxalactins A-C are polyketides synthesized from malonyl-CoA and (*S*)-methylmalonyl-CoA (Castro, 2015). They have antibiotic activity against *Staphylococcus aureus* and methicillin resistant *Staph. aureus* (MRSA) (Rateb, Houssen, Arnold et al., 2011, Rateb, Houssen, Harrison et al., 2011).

Lately, a new genome sequence was achieved using next generation DNA sequencing techniques (Gomez-Escribano et al., 2015). The genome sequence allowed the identification of the biosynthetic gene cluster (BGC) that encoded for chaxamycins (Castro et al., 2015), and the chaxalactins BGC (currently being studied by Castro et al., 2015 in preparation). In addition, 32 BGCs encoded in the chromosome of *S. leeuwenhoekii* C34, plus one BGC encoded in the plasmid pSLE2 were identified. The detailed bioinformatic evaluation of the BGCs revealed that twenty-seven were probably involved in the synthesis of unknown specialized metabolites, therefore possible new compounds. The study of the metabolic pathways of *S. leeuwenhoekii* C34 would allow identification of metabolic engineering targets that would improve the production of chaxamycins, chaxalactins, or other specialized metabolites.

The study of the microorganism metabolism can be accomplished through a GSM. GSMs are mathematical models that are developed to study the behavior of the cell taking into account all the available information of the genome, omics data, and literature. These models have allowed determination of the behavior of the cell under various conditions simulating genes deletion or over-expression and their effect in the organism studied and identification of drug targets (Campodonico, Andrews, Asenjo, Palsson, & Feist, 2014; Contador, Rodŕıguez, Andrews, & Asenjo, 2015). Specifically for *Streptomyces* strains, GSM have been developed for *S. coelicolor* (Alam, Merlo, Takano, & Breitling, 2010; Borodina, Krabben, & Nielsen, 2005; Kim et al., 2014; Kim, Smith, Micklefield, & Mavituna, 2004), *S. clavuligerus* (Medema et al., 2010), *S. lividans* (D'Huys et al., 2012), and *S. tsukubaensis* (Huang, Li, Xia, Wen, & Jia, 2013). These models have been used to predict gene targets for metabolic engineering in order to improve specialized metabolite production. For example, the production of the immunosuppressant FK506 was 1.47 fold higher in *S. tsukubaensis* modified with gene knockouts and overexpression predicted by a GSM (Huang et al., 2013).

In this work, we present the first GSM for *S. leeuwenhoekii* C34 constructed using the genome sequence information obtained previously (Gomez-Escribano et al., 2015). The GSM was used to study precursor biosynthesis of specialized metabolites, such as chaxamycins, toward improving their yield. Also, identification of gene targets for overexpression or deletion that would enhance specialized metabolite biosynthesis was addressed.

2 | METHODS

2.1 | Construction of the genome scale model

The methodology for the construction of the GSM of *S. leeuwenhoekii* C34 is summarized in Figure 1. Concisely, the information of genes/proteins/reactions of other *Streptomyces*, especially that of *S. coelicolor*, enzyme commission (EC) numbers and pathways was downloaded from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database (Kanehisa & Goto, 2000; Kanehisa et al., 2014) using the bioservices module (Cokelaer, Pultz, Harder, Serra-Musach, & Saez-Rodriguez, 2013) with a script written in python programming language

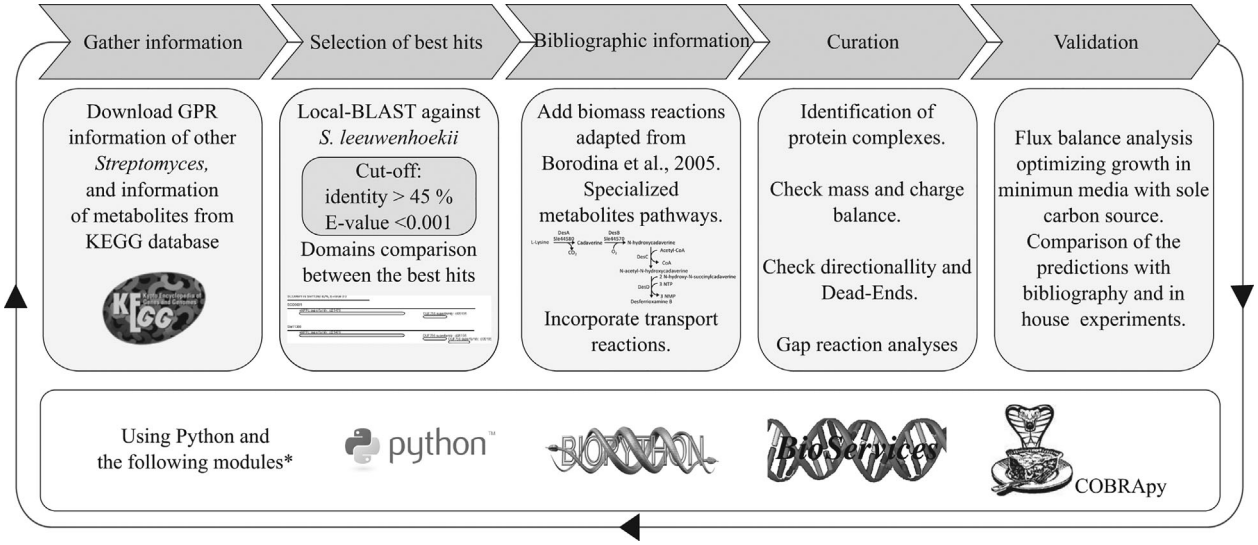

FIGURE 1 Schematic representation of the methodology used for the reconstruction of the genome scale model of *Streptomyces leeuwenhoekii* C34. *Several other modules were also used during this methodology, such as Matplotlib (Hunter, 2007) and Tkinter

version 2.7 (python programme available at https://www.python.org/download/releases/2.7/).

All the data of genes and proteins of *S. coelicolor* associated to a metabolic pathway were retrieved and using the orthology numbers information the reactions were designated. If needed, the information was complemented with data obtained from other *Streptomyces*. For transport reactions, information of the associated genes and proteins was obtained from either KEGG or transportDB (Ren, Chen, & Paulsen, 2007).

Local blast of those genes/proteins of *S. coelicolor* or another *Streptomyces* was done against a local gene database of *S. leeuwenhoekii* C34. A threshold of at least 45% identity and e-value lower than 0.001 was defined as a cut-off to select for the best hits. These lower cut-off settings were selected in order to include all possible homologous proteins. Later, selection of true homologous proteins was performed through analysis of conserved domains (CD). CD-blast (Marchler-Bauer & Bryant, 2004; Marchler-Bauer et al., 2009, 2011, 2014) was performed separately for *S. leeuwenhoekii* C34 and for the other *Streptomyces*. The CD-blast information was used to compare the domains of homologous proteins to confirm their similarity in functionality, specially for the proteins with low identity.

The model was refined with bibliographic information. Specialized metabolite biosynthesis pathways were incorporated (for more details see section 2.2). Artificial reactions for the biosynthesis of biomass were included in order to simulate cell growth (for more details see section 2.3).

Manual curation of the list of proteins, genes, and associated reactions was done for all the metabolic pathways of *S. leeuwenhoekii* C34 to determine the proper Gene-Protein-Reaction (GPR) relationship. Also, enzyme complexes were manually identified and properly annotated. It was checked that the production of biomass was possible by simulating the precursor supply of each building block of the biomass reaction. Gap-filling reactions were incorporated if needed, especially when all the genes of the pathway down the gap were present in *S. leeuwenhoekii* C34. Charge and mass balances were automatically checked. Mass balances that required water or protons were automatically fixed, and otherwise, were manually curated.

The predictability of the model was assessed by comparison of in silico results with experimental data (for more details see section 2.6).

2.2 | Incorporation of specialized metabolite pathways

Biosynthesis pathways to produce chaxamycins (Castro et al., 2015), chaxalactins (Castro, 2015), desferrioxamines (Barona-Gómez, Wong, Giannakopulos, Derrick, & Challis, 2004) amongst other specialized metabolites pathways, were manually included in the model. All the reactions needed for the biosynthesis of the precursors of the specialized metabolites were incorporated into the model. It was checked that the production of each specialized metabolite included was possible by setting the biomass production rate to a 10% of the growth in default conditions, therefore ensuring the viability of the cells, and optimizing the specialized metabolite biosynthesis.

2.3 | Biomass composition

The biomass elementary composition was adapted from the one described for *S. coelicolor* (Borodina et al., 2005), using information of *S. leeuwenhoekii* C34 when it was available and complementing with information of other *Streptomyces* (for a detailed description please see Supplementary Information: Formulation of the biomass equation and their components, Supplementary Tables S3–12). Briefly, the DNA composition was obtained from the deoxyribonucleoside composition of the chromosome of *S. leeuwenhoekii* C34. RNA composition was derived from the deoxyribonucleotide composition of rRNA, tRNA and mRNA, the later one inferred from the DNA sequence of the chromosome. The energy requirement for polymerisation of triphosphates was obtained from Ingraham, Maaløe, and Neidhardt (1983).

Composition of proteins and small molecules was assumed as for *S. coelicolor*. Triacylglycerol (TAGs) and lipid proportion of DW was taken from *S. coelicolor*, but TAG and lipid composition was taken from *S. leeuwenhoekii* C34 (Busarakam, 2014).

2.4 | Curation of the model

The model obtained was manually curated and the production of biomass was evaluated with each precursor at a time. The directionality of the reactions was checked according to ΔG energies (Flamholz, Noor, Bar-Even, & Milo, 2011) and with bibliographic information of reversibility available from the closest phylo-genetically related microorganisms.

Furthermore, the predicted domains through the conserved domain database (CDD) batch analysis for all the proteins related to reactions within the model reconstruction, were compared to *Streptomyces* homologous proteins assigned to each reaction in KEGG. To do so, a python script that compares the domains of the *S. leeuwenhoekii* C34 proteins and other *Streptomyces* was developed. Within this script it was possible to obtain a graphic representation of the proteins that did not match in order to facilitate further analysis.

Also, it was checked that the domains of the proteins assigned to each reaction had a coherent functionality. In addition, genes encoding for protein complexes or isoenzymes were identified, and properly annotated.

2.5 | Simulations of the metabolism

Simulations were done by flux balance analysis (FBA) or Minimal of metabolic adjustment (MOMA) using the Constrain Based Reconstruction Analysis for python (COBRApy) toolbox (Ebrahim, Lerman, Palsson, & Hyduke, 2013) in python programming language version 2.7 (https://www.python.org/download/releases/2.7/). The GSM was represented as a stoichiometric matrix S of size $m \times n$, where m represents the number of metabolites and n the number of reactions. Under steady state assumption the system of linear equations of mass balance is defined by:

$$Sv = 0 \tag{1}$$

where, S is a stoichiometric matrix and v is a vector of length n that represents the flux through all the reactions.

The solution space was constrained by the boundaries of each reaction Equation (2), and for the definition of an objective function Z Equation (3). For reversible reactions the lower (LB) and upper bounds (UB) were set as −1,000 and 1,000 respectively, leaving the reaction practically unconstrained, while for irreversible reactions the lower bound was set to 0. For simulation of specialized metabolite production, a percentage of the maximum biomass production rate calculated was added as a constraint, and a specific specialized metabolite was set as a new objective function Z.

$$\upsilon LB \leq \upsilon \leq \upsilon UB \tag{2}$$

$$Z = C^T \upsilon \tag{3}$$

The default carbon, nitrogen, phosphorous and sulphur sources were glucose, ammonium, phosphate and sulphate, respectively. Default conditions for simulating growth were set at −10 mmol gDW^{-1} hr^{-1} uptake rates (lower bound), where a minus sign represents that the metabolite is being consumed. To simulate aerobic growth the oxygen uptake rate was set at −10 mmol gDW^{-1} hr^{-1}. Small inorganic ions needed for biomass biosynthesis were allowed to freely enter or exit the system by setting the lower and upper bound of each exchange reaction to −1,000 and 1,000, respectively. The lower bound of the rest of the exchange reactions was constrained to zero when simulating growth in MM. When simulating growth in MM with other carbon, nitrogen or phosphorous sources, the lower bound of the default source was set to 0, and the uptake rate of the evaluated source was set to −10 mmol gDW^{-1} hr^{-1}.

The gurobi optimizer version 6.5.1 (http://www.gurobi.com/), under a free academic license, was used as the linear programming solver.

2.6 | Validation of the model

Experimental results of growth or no-growth under sole carbon, nitrogen or phosphorous sources obtained using Biolog (http://www.biolog.com/), was used to study the predictability of the model. The uptake rates of the exchange reaction of each carbon, nitrogen or phosphorous source was set to −10 mmol gDW^{-1} hr^{-1} one at a time. A viability threshold of 10% of growth of default conditions was considered.

2.7 | Gene knockout analysis and experimental studies

FBA and MOMA were used to simulate gene knockouts and find which gene deletion could enhance synthesis of precursors toward specialized metabolite biosynthesis.

In order to simulate gene deletions, simulation of growth in complex media was used as the default condition. The optimization was carried out with the biomass production rate as objective function, and the initial values of production of chaxamycin A and chaxalactin A were obtained. Then the gene(s) associated to each reaction was knocked-out and it was evaluated if the deletion of the gene(s) increased the initial production rate of each specialized metabolite evaluated. The gene(s) knockout that generated the higher increase in production was used as the genetic background for the search for double or triple gene knockouts that would further improve the yield of chaxamycin A and chaxalactin A. To ensure the viability of the cell, each knockout was selected considering a viability threshold of at least 10% of the maximum growth in default conditions.

2.8 | Identification of gene overexpression targets

The algorithm Flux Scanning based on Enforced Objective Flux (FSEOF) was used to select gene targets for overexpression (Choi, Lee,

TABLE 1 Statistics of *i*VR1007

	Number
Total reactions	1,722
Metabolic conversions	1,483
Transport reactions	239
Exchange reactions	186
Reactions with ORF assigned	1,297
Reactions without ORF assigned (GAPs)	425
Percentage of reactions with ORF	75.2%
Transport reactions with ORF	108
Percentage of transport reactions with ORF	45.2%
Metabolic conversions with ORF	1,189
Percentage of metabolic conversions with ORF	80%

Kim, & Woo, 2010). Through FSEOF it is possible to identify the fluxes that are increased at the same time the flux through a objective reaction is incremented, and the production of biomass is used as an objective function. The algorithm to perform FSEOF (Choi et al., 2010) was implemented in python and used in conjunction with the COBRApy toolbox (to see details of the script used, go to Supplementary Information). As the FBA result is not unique, flux variability analysis (FVA) was applied to determine if the predicted targets for overexpression were true positives.

2.9 | Implementation and software usage

The genome scale model was constructed using biopython and implemented using COBRApy (Ebrahim et al., 2013) installed in a computer with 64-bit Windows 10 system, Intel® Core™ i5-2430M CPU @ 2.40 GHz with 8 GB RAM. General drawings and figures were done using Inkscape (https://inkscape.org) or python. Drawings of fluxes were done with Escher (King et al., 2015) and edited in Inkscape (https://inkscape.org).

3 | RESULTS

3.1 | The model

A consistent and validated model was obtained and named *i*VR1007. The model *i*VR1007 has 1,722 reactions, 1,463 metabolites, and 1,007 genes. There are 425 reactions that do not have an associated gene but are needed for connectivity of the model or for biomass production (Table 1). The distribution of *i*VR1007 reactions is shown in Figure 2.

Growth associated maintenance (GAM) requirements were assumed to be 40 mmol ATP $gDW^{-1}\,hr^{-1}$ as observed for *S. coelicolor* (Borodina et al., 2005). This term was added as ATP hydrolysis to the artificial biomass reaction. Non-growth associated maintenance (NGAM) requirements are imposed by setting the lower bound of the ATPM reaction to 3 mmol ATP $gDW^{-1}\,hr^{-1}$ as stated for *S. coelicolor* (Borodina et al., 2005).

The pathways needed for the use of different carbon/nitrogen and phosphorous sources were included in the model, specially if there was experimental evidence of growth on those sources. Sixty-five percent of the gap-filling reactions were related to transport reactions or the fatty acid metabolism, and they were necessary for completeness of the model. The remaining 35% gap-filling reactions were included for completion of different metabolisms (Figure 3). They were added to complete the metabolic pathways when the metabolic pathway was almost complete or there was evidence of the consumption or production of the compounds participating in the pathway. For example, orphan reactions included to allow the degradation of D-galacturonate and L-rhamnose.

3.1.1 | Fatty acid biosynthesis

The fatty acid composition of *S. leeuwenhoekii* C34 (Busarakam, 2014), includes branched-chain fatty acids (BCFA) as well as straight-chain fatty acids (SCFA) which were incorporated in the model. BCFA are biosynthesized from branched-chain amino acids, 2-methylpropanoyl-CoA, 3-methylbutanoyl-CoA and (*S*)-2-methylbutanoyl-CoA, and use malonyl-CoA as extender units, to produce *iso*- (odd numbered chain), *iso*- (even numbered chain), and *anteiso*- (even numbered chain) fatty acids, respectively. Also,

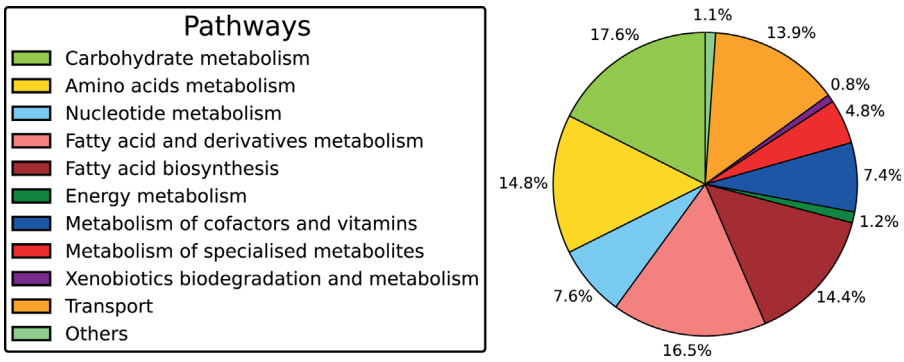

FIGURE 2 Distribution of reactions of *i*VR1007 in each metabolism category

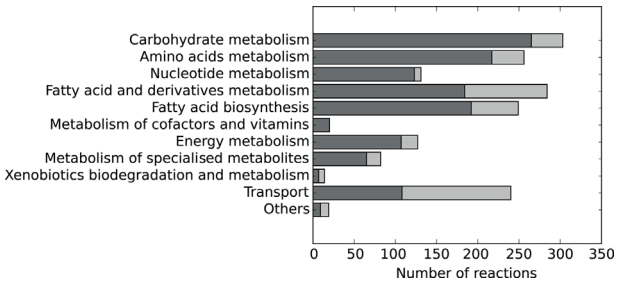

FIGURE 3 Number of reactions of *i*VR1007 in each metabolism. Reactions with ORF assigned are shown in gray and gap-filling reactions in light gray

*i*VR1007 includes the biosynthesis of unsaturated fatty acids. Malonyl-CoA is produced from acetyl-CoA in the reactions rx0404 and rx0405 (in KEGG: r04386 and r04385, respectively).

All the reactions of fatty acid biosynthesis were added stepwise instead of being lumped in one reaction. Because of this, 37% of the gap-filling reactions were from the fatty acid biosynthesis and metabolism (Figure 3).

3.1.2 | Specialized metabolite pathways

Specialized metabolites like polyketides or non-ribosomal peptides are assembled with precursors synthesized during the primary metabolism. polyketide synthase (PKS)s use acyl-CoA molecules as building blocks such as acetyl-CoA, malonyl-CoA, or (*S*)-methylmalonyl-CoA. Non-ribosomal peptide synthetase (NRPS)s use amino acids and derivatives as precursors and extender units. The Shikimate pathway is important for generation of amino acids and derivatives that are used during the primary metabolism and also in the biosynthesis of specialized metabolites. Reactions needed for the biosynthesis of chaxamycins, chaxalactins, desferrioxamines, and ectoine were included in the model.

Chaxamycins

The proposed metabolic pathway of chaxamycin biosynthesis (Castro et al., 2015) is based on rifamycin biosynthesis. Biosynthesis reactions of chaxamycins were lumped according to the PKS genes (Figure 4a). A total of 21 reactions that were exclusive, were included in the model in order to allow production of chaxamycin A and B. The key precursors

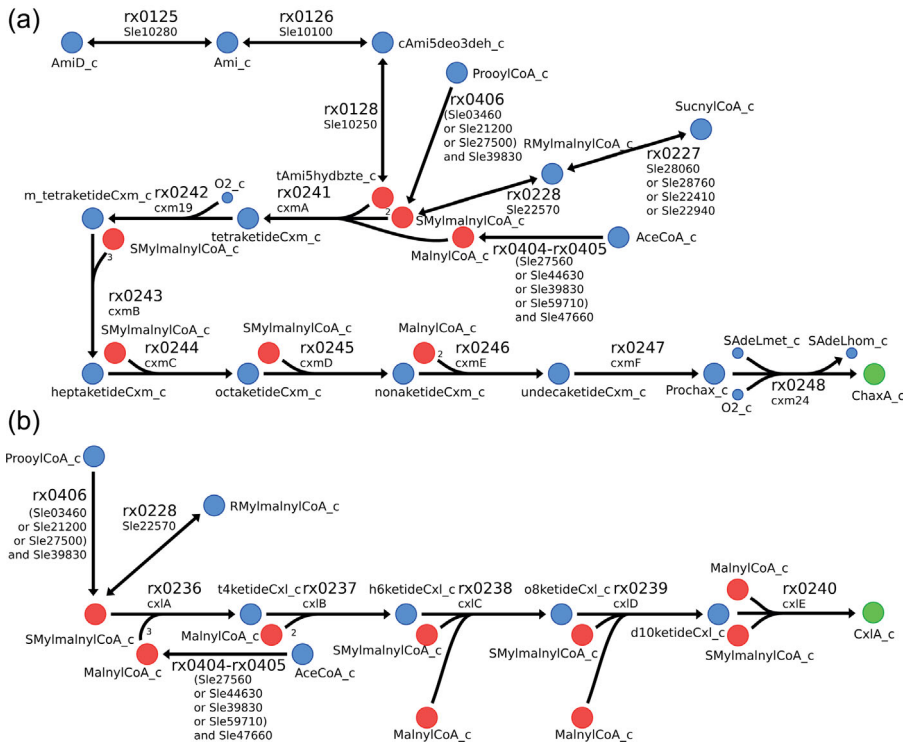

FIGURE 4 Specialized metabolites biosynthesis pathways in *S. leeuwenhoekii*. Chaxamycin A biosynthesis (a) and Chaxalactin A biosynthesis pathway (b). Green: specialized metabolites product, red: main used precursors, blue: intermediates metabolites. Abbreviations: SucnylCoA_c (succinyl-CoA), RMylmalnylCoA_c (R-methylmalonyl-CoA), ProoylCoA_c (propanoyl-CoA), SMylmalnylCoA_c (*S*-methylmalonyl-CoA), MalnylCoA_c (malonyl-CoA), AceCoA_c (acetyl-CoA), AmiD_c (aminoDAHP), Ami_c (aminoDHQ), cAmi5deo3deh_c (5-Amino-5-deoxy-3-dehydroshikimate), tAmi5hydbzte_c (3-Amino-5-hydroxybenzoate), tetraketideCxm_c (tetraketide intermediate of chaxamycin), m_tetraketideCxm_c (modified tetraketide intermediate of chaxamycin), heptaketideCxm_c (heptaketide intermediate chaxamycin), octaketideCxm_c (octaketide intermediate chaxamycin), nonaketideCxm_c (nonaketide intermediate chaxamycin), undecaketideCxm_c (undecaketide intermediate chaxamycin), Prochax_c (prochaxamycin), ChaxA_c (chaxamycin A), SAdeLmet_c (*S*-adenosyl-L-methionine), SAdeLhom_c (*S*-adenosyl-L-homocysteine), t4ketideCxl_c tetraketide intermediate chaxalactin), h6ketideCxl_c (hexaketide intermediate chaxalactin), o8ketideCxl_c (octaketide intermediate chaxalactin), d10ketideCxl_c (decaketide intermediate chaxalactins), CxlA_c (chaxalactin A)

needed for the biosynthesis of chaxamycins are: 3-Amino-5-hydroxybenzoate (AHBA), malonyl-CoA and (*S*)-methylmalonyl-CoA.

Chaxalactins

There are three types of chaxalactins (A-C) but as all of them use the same precursors it has been proposed that chaxalactin A is synthesized first; only chaxalactin A biosynthesis reactions were added to the model. The reactions needed for chaxalactin A biosynthesis was inferred from the putative biosynthetic pathway proposed (Castro, 2015). There are five PKS biosynthesis genes and each of them was annotated as a reaction (Figure 4b).

Desferrioxamines

Desferrioxamines are siderophores with high affinity to iron. They are produced when the concentration of iron is low. Desferrioxamines are released by the cells, they form an iron complex and then the complex is transported into the cell to recover the iron. The BGC of desferrioxamines has been studied for *S. coelicolor* (Barona-Ǵomez et al., 2004). In the same work the biosynthesis pathway of desferrioxamine E was described, complementing the previous information (Günter, Toupet, & Schupp, 1993; Schupp, Toupet, & Divers, 1988).

Ectoine

Ectoine is a compatible solute that has a protective role within the cell. It is produced in higher amounts when the cell is in conditions of salinity or heat stress. The biosynthesis pathway of ectoine has been previously described (Bursy et al., 2008; Peters, Galinski, & Trüper, 1990). Ectoine is synthesized by enzymes encoded in a highly conserved BGC:*ectABC*. Homologous genes of the ectoine BGC have been identified in the genome of *S. leeuwenhoekii* C34.

3.2 | Validation of the model

The default conditions for simulation of growth were set as specified in the methodology (section 2.5 and 2.6).

Simulations were carried out to challenge the model to correctly predicted growth in aerobic conditions and no growth in anaerobic conditions, by analyzing the growth capabilities when the oxygen exchange reaction was blocked. The model predicted that there was no growth when the oxygen exchange reaction was blocked, and that growth depends on the uptake rate of oxygen.

The predictability of the model was evaluated and validated using data of growth/no-growth in sole carbon, nitrogen or phosphorous sources inferred from Biolog data. Comparison between the prediction and experimental results are shown in the supplementary information (Supplementary Figures S1 and S2). The model correctly predicted growth/no-growth with 72 out of 89 carbon sources, 45 out of 54 nitrogen sources and 22 out of 23 phosphorous sources, corresponding to an 83.7% accuracy. There were five cases for which the model was not capable of predicting growth contradicting the experimental information. These discrepancies could be due to missing content or inconsistencies. On the other hand there were 16 cases where the model predicts growth and the experimental evidence showed the opposite. The differences could be due to errors, inconsistencies or missing constrains, such as transcriptional regulatory constraints.

To allow the model to use several carbon/nitrogen sources, it was necessary to include gap-filling reactions. For example, the experimental evidence showed that *S. leeuwenhoekii* C34 was capable of growing using L-lysine as carbon/nitrogen source, however the reactions required to allow L-lysine consumption and connection to the metabolic network did not have the *S. leeuwenhoekii* C34 genes associated. Further studies are required in order to identify the genes involved.

The metabolism of D-melibiose, D-raffinose and stachyose goes through the production of D-galactose. The model predicts that *S. leeuwenhoekii* C34 is able to grow using either of the mentioned carbon sources, however the experimental information indicated that there is no growth with D-melibiose, D-raffinose, and stachyose, while there is growth with D-galactose. The conversion of D-melibiose, D-raffinose, or stachyose to D-galactose is catalyzed by the genes *sle10440*, *sle08850*, or *sle63600*. Additional analyses are required in order to see if these reactions are subject to transcriptional constraints.

Similar cases occur for D-alanine, formic acid, D-glucose 6-phosphate, sucrose, uridine, glycolate, L-homoserine, L-leucine, L-ornithine, D-glutamate, cytidine, and cytosine, all the genes needed to use these carbon/nitrogen sources are present in *i*VR1007 and consequently the model predicts growth, however the experimental evidence showed that *S. leeuwenhoekii* was not able to use those sole carbon/nitrogen sources. This suggests that there could be regulatory constrains over those genes.

The model predicts growth with L-lactic acid as carbon source in contrast with experimental evidence. The metabolism of this compound in the model includes gap-filling reactions that were added to complete the fructose and mannose metabolism. If these reactions were eliminated from the model, *i*VR1007 predicts not-growth with L-rhamnose contradicting the experimental information. Similarly in the case of formamide, where if the gap-filling reaction is deleted the model wrongly predicts growth with other sources.

The experimental information showed that D-glucosamine could be used as sole carbon source but not as nitrogen source. In contrast, the model predicts that it is possible to use it as either carbon or nitrogen source. Further studies are required in order to understand why *S. leeuwenhoekii* is not capable of using this substrate as a nitrogen source.

The capacity to grow in complex media was also evaluated. The model predicted a higher growth rate in complex media compared to growth on sole carbon sources.

3.3 | Identification of metabolic engineering targets

The GSM was used to identify non-intuitive gene knockouts or overexpression that would improve specialized metabolite production like chaxamycins and chaxalactins. A simplification of the core metabolism of *S. leeuwenhoekii* C34 showing the biosynthesis of precursors for specialized metabolites biosynthesis, as well as the metabolic engineering targets for increasing their production is shown in Figure 5.

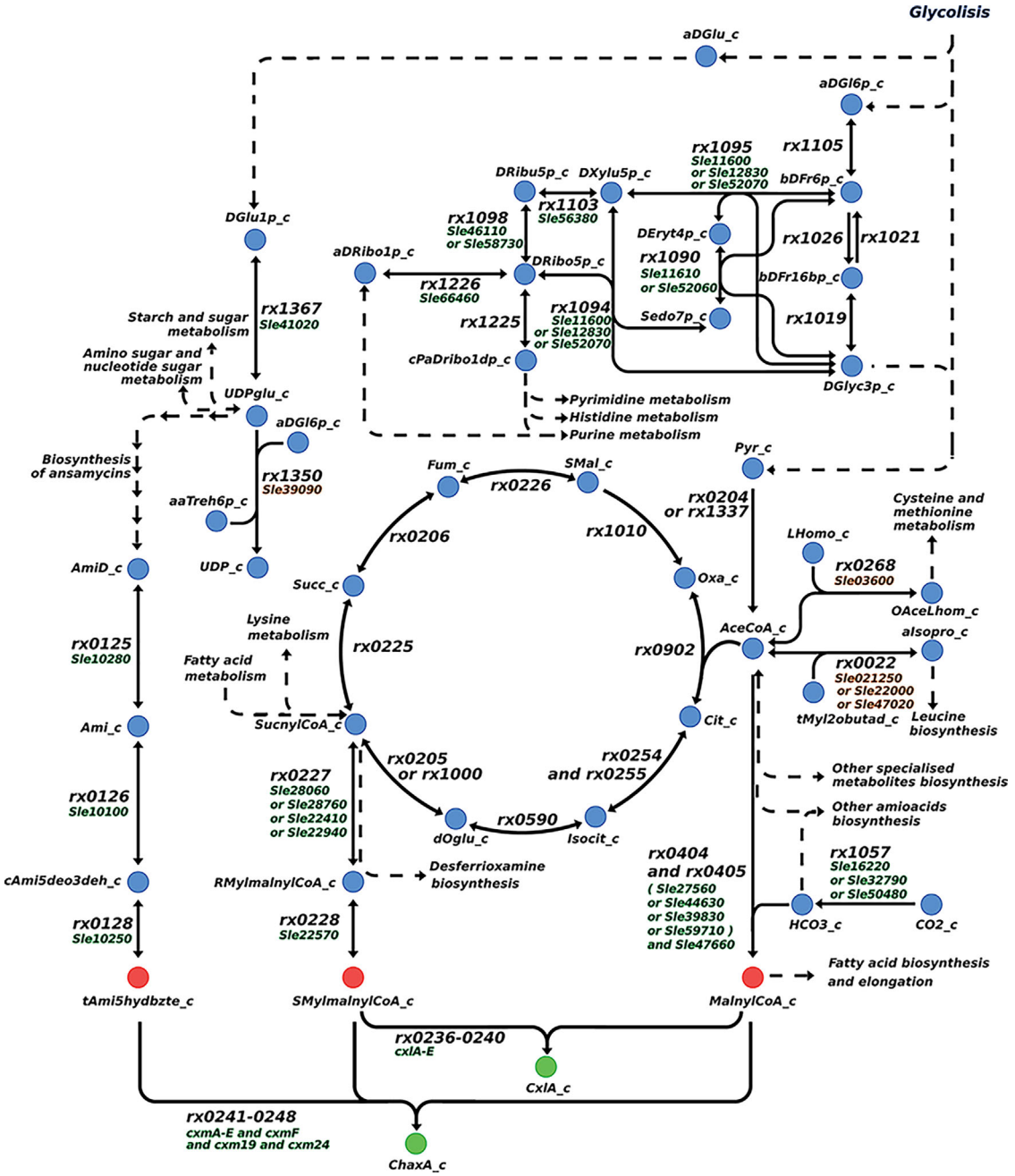

FIGURE 5 Simplified core metabolism of *S. leeuwenhoekii* C34 showing gene targets for over-expression (highlighted in green) and knockouts (highlighted in orange) for increasing specialized metabolites biosynthesis. The genes involved in other reactions had been omitted for simplicity. Abbreviations: ChaxA_c (chaxamycin A), CxlA_c (chaxalactin A), aDGlu_c (α-D-Glucose), DGlu1p_c (D-Glucose 1-phosphate), UDPglu_c (UDP-glucose), aDGl6p_c (α-D-Glucose 6-phosphate), aaTreh6P_c (α,α-Trehalose 6-phosphate), AmiD_c (aminoDAHP), Ami_c (aminoDHQ), cAmi5deo3deh_c (5-Amino-5-deoxy-3-dehydroshikimate), tAmi5hydbzte_c (3-Amino-5-hydroxybenzoate), bDFr6p_c (β-D-Fructose 6-phosphate), bDFr16bp_c (β-D-Fructose 1,6-bisphosphate), DGlyc3p_c (D-Glyceraldehyde 3-phosphate), DEryt4p_c (D-Erythrose 4-phosphate), Sedo7p_c (Sedoheptulose 7-phosphate), Dxylu5p_c (D-Xylulose 5-phosphate), DRibu5p_c (D-Ribulose 5-phosphate), Dribo5p_c (D-Ribose 5-phosphate), aDRibo1p_c (α-D-Ribose 1-phosphate), cPaDribo1dp_c (5-Phospho-α-D-ribose 1-diphosphate), Pyr_c (Pyruvate), AceCoA_c (acetyl-CoA), MalnylCoA_c (malonyl-CoA), LHomo_c (L-Homoserine), OAceLhom_c (O-Acetyl-L-homoserine), tMyl2obutad_c (3-Methyl-2-oxobutanoic acid), alsopro_c (α-Isopropylmalate), Cit_c (Citrate), Isocit_c (Isocitrate), dOglu_c (2-Oxoglutarate), SucnylCoA_c (succinyl-CoA), Succ_c (Succinate), Fum_c (Fumarate), SMal_c (S-Malate), Oxa_c (Oxaloacetate), RMylmalnylCoA_c (R-methylmalonyl-CoA), SMylmalnylCoA_c (*S* methylmalonyl-CoA)

3.3.1 | Gene knockout analysis

The procedure described in section 2.7, allowed obtention of sets of gene deletions that increment the production of chaxalactin A. For increasing chaxamycin A production no single gene knockout was found.

The gene *sle03600* (*metX*), encoding for an acetyl-CoA:L-homoserine *O*-acetyltransferase is associated to the reaction rx0268 (KEGG No. r01776) in the cysteine and methionine metabolism. Deletion of this gene produced more availability of acetyl-CoA that in principle, could be converted to malonyl-CoA and increase chaxalactin A production to a 60% of the theoretical maximum yield. The deletion of *sle03600* was used as a starting point to predict further improvement of chaxalactin A production. It was found that an additional deletion of the genes *sle21250* (*leuA_3*), *sle22000* (*leuA_2*), and *sle47020* (*leuA_1*) (encoding for 2-isopropylmalate synthase), associated to the reaction rx0022 (KEGG No. r01213) predicted an increase in chaxalactin A production to 67% of the maximum yield (Figure 6a).

The chaxamycin A production was predicted to be increased in the presence of at least double gene deletions. The deletion of *sle03600* and *sle39090* (otsA, encoding for a trehalose 6-phosphate synthase) produced an increase to 8% of the theoretical maximum. Further deletion of *sle29000*, *sle29010* (gauA_2) and *sle56410* (*gauA_1*) (encoding for IMP dehydrogenase), associated to the reaction rx1231 (KEGG No. r02661, r01130), predicted an improvement of chaxamycin A production to 44 % of the theoretical maximum (Figure 6b).

A summary of the predicted gene knockouts that enhance the production of specialized metabolites and their associated reactions is shown in Table 2.

3.3.2 | Identification of targets for overexpression

FSEOF was applied to identify reactions that had an incremented flux when a specific specialized metabolite production was enforced and the production of biomass was optimized in growth in complex media, this procedure allows obtention of overexpression gene targets that would enhance the production of chaxamycins and chaxalactins (Supplementary Tables S1 and S2).

Fifty-four reactions that had increasing fluxes while the production of chaxamycin A was enforced were identified. Nineteen of them are directly involved in the production of the precursor AHBA and in the production and transport of chaxamycin A, while 35 reactions are not. When FSEOF was applied to study targets of overexpression for chaxalactins enhancement, a total of 31 reactions were detected as possible targets.

Of the not directly related reactions several of them were found for more than one of the specialized metabolites. The identification of common overexpression targets is probably due to the use of the same extensor unit (malonyl-CoA).

4 | DISCUSSION

4.1 | Discrepancies between model predictions and the experimental data

The reported accuracy of other models developed and validated using growth phenotype data was around 74–90% (Durot et al., 2008; Feist et al., 2007; Oberhardt, Puchałka, Fryer, Dos Santos, & Papin, 2008;

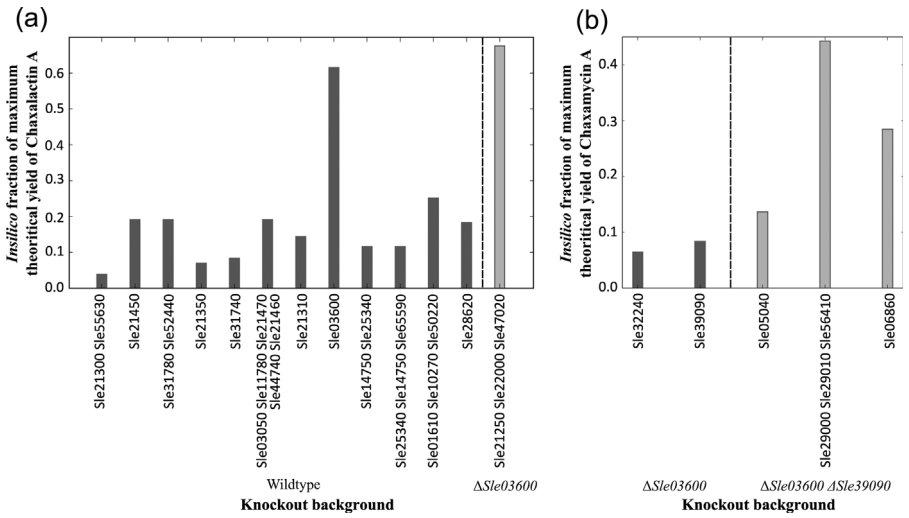

FIGURE 6 Gene knockout search for increased chaxalactin A production in the wildtype strain (gray) and Δ*sle03600* (light gray) (a); and gene knockout search for increasing chaxamycin A production in Δ*sle03600* (gray) and Δ*sle39090* (light gray) background (b). The proteins are: Sle21300 and Sle55630 (branched-chain amino acid aminotransferase); Sle21450 (ketol-acid reductoisomerase); Sle31780 and Sle52440 (dihydroxy-acid dehydratase); Sle21350 (1-pyrroline-5-carboxylate dehydrogenase); Sle31740 (pyrroline-5-carboxylate reductase); Sle03050, Sle11780, Sle21470, Sle44740 and Sle21460 (acetolactate synthase complex); Sle21310 (3-isopropylmalate dehydrogenase); Sle03600 (homoserine *O*-acetyltransferase); Sle14750 and Sle25340 (catalase); Sle25340, Sle14750 and Sle65590 (catalase); Sle01610, Sle10270 and Sle50220 (3-deoxy-7-phosphoheptulonate synthase); Sle28620 (phosphoribosylaminoimidazolecarboxamide formyltransferase / IMP cyclohydrolase); Sle21250, Sle22000 and Sle47020 (2-isopropylmalate synthase); Sle32240 (2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase); Sle39090 (trehalose 6-phosphate synthase); Sle05040 (cystathione beta-lyase); Sle29000, Sle29010 and Sle56410 (IMP dehydrogenase); Sle06860 (L-histidine N-α-methyltransferase)

TABLE 2 Predicted genes knockout targets for increasing chaxamycin A and chaxalactin A production

Predicted effect of knockout	Gene(s)	No.	Reaction
Increase chaxalactin A production	*sle03600*	rx0268	AceCoA_c + LHomo_c ↔ CoA_c + OAceLhom_c
Increase chaxalactin A production in a Δsle03600 background	*sle21250*, *sle22000*, *sle47020*	rx0022	AceCoA_c + tMyl2obutad_c + H2O_c → alsopro_c + CoA_c + H_c
Increase chaxamycin A production in a Δsle03600 background	*sle39090*	rx1350	UDPglu_c + aDGl6p_c → UDP_c + aaTreh6p_c
Increase chaxamycin A production in a Δsle03600 and Δsle39090 background	*sle29000*, *sle29010*, *sle56410*	rx1231	IMP_c + NAD_c + H2O_c → Xant5p_c + NADH_c + H_c

AceCoA_c (acetyl-CoA), LHomo_c (L-homoserine), CoA_c (coenzyme A), OAceLhom_c (*O*-acetyl-L- homoserine), tMyl2obutad_c (3-methyl-2-oxobutanoic acid), alsopro_c (α-isopropylmalate), UDPglu_c (UDP-glucose), aDGl6p_c (α-D-glucose 6-phosphate), aaTreh6p_c (α,α-trehalose 6-phosphate), Xant5p_c (xanthosine 5-phosphate).

Oh, Palsson, Park, Schilling, & Mahadevan, 2007). The higher accuracy was reported when the number of considered carbon/nitrogen/phosphorous/sulfur sources was lower, such as the case for *i*MO1056 with a reported accuracy of 90% (Oberhardt et al., 2008) but just considering 30 conditions. On the other hand, accuracies of 74% and 76% were reported for the *Bacillus subtilis* model (Oh et al., 2007), and *E. coli* K-12 model (Feist et al., 2007), considering 271 and 170 conditions, respectively. In the case of *i*VR1007, an accuracy of 83.7% was obtained considering 166 conditions.

This percentage of accuracy could be improved by analysing and understanding the discrepancies between predictions and experimental data. The model *i*VR1007 did not include regulatory constraints, hence the discrepancies could be due to missing regulatory information. The model could be improved with the inclusion of regulatory constraints (Herrgård, Lee, Portnoy, & Palsson, 2006). On the other hand, the identification of genes involved in gap-filling reactions would allow polishing of the GSM. For example, the GSM of *E. coli* has been greatly improved over the years by the inclusion of experimental information (Orth et al., 2011).

GSMs are always under development, and with the future availability of more experimental data for *S. leeuwenhoekii* C34, it is going to be possible to further improve the predictability of the model.

4.2 | Identification of metabolic engineering targets for increasing chaxamycin and chaxalactin production

4.2.1 | Prediction of gene knockout targets

The GSM allowed obtention of novel knockout gene targets that could be used to improve specialized metabolite biosynthesis (Table 2). Gene knockouts were detected that predict an increment in the production of chaxamycin A and chaxalactin A.

The deletion of *sle03600* produced an increased flux through the reactions rx0404 and rx0405 due to more availability of acetyl-CoA, and consequently an enhanced flux toward chaxalactin A production is predicted. Further deletion of *sle21250*, *sle22000*, and *sle47020* (2-isopropylmalate synthases; reaction rx0022), increments the pool of acetyl-CoA that can go to the production of malonyl-CoA, a key precursor in specialized metabolite biosynthesis.

In a Δ*sle03600* background, the deletion of *sle39090* (trehalose 6-phosphate synthase) that participates in the reaction rx1350 that consumes UDP-glucose, predicts an increase in the fluxes toward the production of AHBA a key precursor for the biosynthesis of chaxamycin A.

Similarly, other gene deletions predicted to increase chaxamycin A production were found to, directly or indirectly, increment the fluxes toward the production of the building blocks of these metabolites. For example an increment in the flux through the reactions rx0404 and rx0405, generating a higher conversion of acetyl-CoA to malonyl-CoA.

The gene knockout targets differ from previously reported ones and can lead to an improvement in specialized metabolite biosynthesis. Some of the reported gene knockouts for enhancing specialized metabolites in other *Streptomyces* are: *SCO1937* and *SCO6661* in *S. coelicolor* (Ryu, Butler, Chater, & Lee, 2006). The homologous genes in *S. leeuwenhoekii* C34 are *sle52050* and *sle11620* that participate in reactions rx1099 and rx1101. These genes were not found as deletion targets with *i*VR1007, and when the deletion of those genes was simulated, an increased production of specialized metabolites was not found, nor an increased flux through rx0404 and rx0405. However, using Δ*sle52050* and Δ*sle11620* as genetic background, enables the identification of other gene deletions that enhanced specialized metabolites production (data not shown).

It has been described that gene knockouts associated with the metabolism of *N*-acetyl-glucosamine produce a higher yield of actinorhodin (Swiatek, Tenconi, Rigali, & van Wezel, 2012). However, the role of *N*-acetyl-glucosamine metabolism in enhancing specialized metabolites is associated to regulatory constraints, as those were not included in the model, it was not possible to detect if the deletion of the homologous genes, *sle39040*, *sle25040*, *sle39050*, and *sle43640* would enhance the production of specialized metabolites in *S. leeuwenhoekii* C34. Nevertheless, these and other gene knockouts that were not found could be used along with the predicted gene targets in order to boost specialized metabolite production.

As gene targets were found for improving each specialized metabolite production separately, it is possible that the deletion of the predicted genes generates an increase in the production of any of the studied metabolites. Particularly, as many of the gene deletions cause an improvement in the flux through the reactions that convert acetyl-CoA to malonyl-CoA, the main precursor used to synthesized

specialized metabolites. For this reason, it is important to consider the predicted gene deletions on modified strains that do not contain competitive BGC. For example, when improving production of chaxalactins the strain *S. leeuwenhoekii* M1614 (Δchaxamycins BGC) should be used as the starting point. Also, it would be better if the proposed modifications could be performed in a *S. leeuwenhoekii* C34 strain lacking other main active BGCs. Therefore, the metabolic profile would be easier to analyse and there would be more precursors available for the biosynthesis of the desired specialized metabolite (Gomez-Escribano & Bibb, 2011).

4.2.2 | Prediction of overexpression gene targets

Several potential gene targets for overexpression were identified using FSEOF. Nineteen reactions were identified as targets to increase the production of chaxamycins and chaxalactins, they are mainly associated with the biosynthesis of the precursor malonyl-CoA or with the pentose phosphate pathway (PPP).

An important precursor for the biosynthesis of polyketides such as chaxamycin A and chaxalactin A is malonyl-CoA. In the model, there are two essential reactions needed for malonyl-CoA biosynthesis, rx0404 and rx0405 (in KEGG:r04386 and r04385, respectively). These two reactions are usually lumped into one reaction (KEGG: r00742) that was not included in the model to avoid duplicity. The genes predicted to catalyse both reactions are *sle47660* (*accA*) and any of the following: *sle27560* (*bccA*), *sle44630* (*accA_1*), *sle39830* (*accA_2*), or *sle59710*. The first two could be better targets because they have higher homology to proteins already studied (Maharjan, Park, Yoon, Lee, & Sohng, 2010). However the protein Sle59710, despite having lower identity to the associated proteins of the reaction rx0404, has the same domains as homologue proteins and also has the domains that are present in Sle47660. Future studies would allow determination of whether this protein is capable of catalyzing the conversion of acetyl-CoA to malonyl-CoA.

The reactions rx0404 and rx0405 were found as overexpression targets for chaxamycins and chaxalactins. Overexpression of acetyl-CoA carboxylase (*acc*) has proven to be very useful to accomplish overproduction of several specialized metabolites (Maharjan et al., 2010, Maharjan, Koju, Lee, Yoo, & Sohng, 2012; Ryu et al., 2006; Zha, Rubin-Pitel, Shao, & Zhao, 2009;).

Besides the use of ATP and acetyl-CoA, for the generation of malonyl-CoA in reaction rx0404 and rx0405, bicarbonate is also needed. Bicarbonate is generated from CO_2 in the reaction rx1057 (KEGG No. r10092). This reaction was found as a target for the three specialized metabolites, and the three genes encoding for carbonic anhydrase are *sle16220* (*cynT_2*), *sle32790* (*cynT_1*), and *sle50480* (*mtcA1*). The overexpression of homologous genes of other strains had not been required when overexpressing acetyl-CoA carboxylase complex (Maharjan et al., 2010, 2012; Ryu et al., 2006; Zha, Rubin-Pitel, Shao, & Zhao, 2009), this indicates that the bicarbonate is not restricting the reaction, therefore the overexpression of the genes in *S. leeuwenhoekii* C34 could not affect specialized metabolite production.

Several targets are associated to the PPP and some of them were identified as targets for the two specialized metabolites (rx1090 and rx1095), and others for chaxalactins (rx1094, rx1098, rx1103, and rx1226). The reaction rx1090 (KEGG No. r01827) associated to the genes *sle11610* (*tal1*) or *sle52060* (*tal2*) (encoding for transaldolase), produces D-erythrose 4-phosphate, and *β*-D-fructose 6-phosphate.

The interlink between the PPP and the glycolysis pathway means that any modifications in either of the pathways directly affect the other (Olano, Lombó, Méndez, & Salas, 2008). In the case of the modifications suggested by the model, the higher flux of rx1090 could generate a higher flux of *β*-D-fructose 6-phosphate to the glycolysis pathway.

There are other reactions (rx0291, rx1711, rx0288, rx1709, rx1710, rx0283, and rx0264) that are associated with cysteine and methionine metabolism. They seem to be selected as targets for overexpression as they could increase the acetyl-CoA pool (rx0288, in KEGG r00586), however this is a gap-filling reaction in the model. There is not evidence of overexpression of homologous genes in other strains to increase specialized metabolite production. Hence, the overexpression of any of the associated genes should be addressed carefully.

Among the reactions not directly involved that were found as targets to enhance production of chaxamycins, is the reaction rx1367 (KEGG No. r00289) that uses D-glucose 1-phosphate to produce UDP-glucose, the latter compound is needed for the biosynthesis of AHBA. The gene associated to this reaction is *sle41020* (*galU*, UTP-glucose-1-phosphate uridylyltransferase). The expression of the homologous gene in *Amycolatopsis mediterranei* U32, *galU*, associated to rifamycin biosynthesis had been increased by the incorporation of nitrate to the media (Shao et al., 2015). However, the effect of the overexpression under a constitutive strong promoter over the AHBA production has not been studied.

The genes encoding for methylmalonyl-CoA mutase, associated to the reaction rx0227 (KEGG No. r00833), *sle28060* or *sle28760* or *sle22410* or *sle22940*, are also targets for overexpression as the production of (R)-methylmalonyl-CoA is predicted to be used to produce (*S*)-methylmalonyl-CoA by the gene *sle22570* in the reaction rx0228 (KEGG No. r02765). (*S*)-methylmalonyl-CoA is an important precursor in the biosynthesis of chaxamycins and chaxalactins, therefore these overexpression targets were found for both cases. Overexpression of homologous genes in other *Streptomyces* has not been performed. The overexpression of these genes should be studied in conjunction with the overexpression of the genes predicted to increase the biosynthesis of malonyl-CoA and AHBA.

Among the other common overexpression targets identified for chaxamycins and chaxalactins, there is a set of reactions (rx1257, rx1258, rx1262, rx1263, rx1277, and rx1628) that lead to the production of glyoxylate and urea, however the last reaction is a gap-filling reaction in the model. Therefore it is necessary to identify enzymes that could be associated to that reaction.

From the gene targets identified for overexpression other genes that could be used could be inferred. For example, it can be inferred that succinyl-CoA is required for the production of

(*S*)-methylmalonyl-CoA, a precursor molecule of chaxamycins and chaxalactins. However, the FSEOF result does not include overexpression targets for increasing succinyl-CoA because there is more than one reaction from where the model can predict succinyl-CoA production, like rx0205 or rx1000 (in KEGG: r01197 and r08549, respectively). In the case of overexpressing those reactions it will be necessary to overexpress the genes: *sle30310* (*korB*, 2-oxoglutarate ferredoxin oxidoreductase subunit *β*) with *sle30300* (*korA*, 2-oxoglutarate ferredoxin oxidoreductase subunit *α*) (for rx0205) and *sle24580* (*kgd*, 2-oxoglutarate dehydrogenase E1 component) with *sle49590* (*dlaT*, 2-oxoglutarate dehydrogenase E2 component) and with *sle49600* (*pdhD*, dihydrolipoamide dehydrogenase) (for rx1000). A similar case can be proposed for the reaction rx1551 (KEGG No. r00678), catalysed by tryptophan 2,3-dioxygenase, generating L-formylkynurenine that could be used to produce L-alanine in the reaction rx1557 (KEGG No. r03936) that could be used later to produce pyruvate in the reaction rx0049 (KEGG No. r00396). In this case, the two last reactions were not predicted as overexpression targets because there are several reactions in which the involved metabolites participate. The gene encoding for the enzyme catalysing the reaction rx1551, *sle33790*, could be used as a non-intuitive overexpression target.

Several of the reactions identified are associated with isoenzymes, so it is crucial to analyze through RT-PCR which genes are actually being expressed (see section 4.3).

All the predicted overexpression targets should be used for metabolic engineering of *S. leeuwenhoekii* C34. Also, the overexpression targets can be included as gene targets for improving PKS production identified in similar strains. On the other hand, the gene overexpression targets identified for other strains elsewhere, that were not detected in this study, could also be considered as targets. Such is the case of the overexpression of the branched-chain *α*-keto acid dehydrogenase (BCDH) complex that improves actinorhodin production in *S. coelicolor* (Kim et al., 2014).

4.3 | Addressing functionality and genetic redundancy

The genomes of *Streptomyces* strains usually contain several genes encoding for the same biochemical function (Bentley et al., 2002; Ikeda et al., 2003). This has been observed for *S. leeuwenhoekii* C34 in this work. The study of the essential genes of the central metabolism in *E. coli* (Kim & Copley, 2007) showed that a large number (80 out of 227) were nonessential under the studied conditions. The reason for this is the availability of alternative pathways, isoenzymes, multifunctional enzymes or broad-specific enzymes (Kim & Copley, 2007). The redundancy or metabolic flexibility, allows the cell to use different pathways under specific environmental conditions (Hiltner et al., 2015).

In the GSMs the isoenzymes are associated with a particular pathway. However, it is important to study their regulation to determine their biochemical relevance under different culture conditions and to evaluate their impact in the production of specialized metabolites.

4.4 | The genome scale model of *S. leeuwenhoekii* C34 would allow experimental design

The combination of the genes identified as targets for overexpression and knockouts can be used to design strains with specific phenotypes. Also, the model can be used to identify the formulation of an optimal media for the production of each specialized metabolite. Similarly, the model could be used to study gene deletions that would help in understanding and complementing the biosynthesis pathways that, currently, are incomplete.

5 | CONCLUSIONS

The first GSM of *S. leeuwenhoekii* C34, *i*VR1007, was developed. Experimental information of growth in different carbon, nitrogen, and phosphorous sources was used to validate the model. The in-silico simulations showed that the model had 83.7% accuracy. The model, *i*VR1007, was used to predict gene knockouts and overexpression that could enhance specialized metabolite production. The deletion of *sle03600* (homoserine *O*-acetyltransferase) and *sle21250*, *sle22000*, and *sle47020* (2-isopropylmalate synthases) predicted an increase in chaxalactin production as the deletion of these genes produced knockouts of reactions that consume acetyl-CoA, therefore allowing an increase in the acetyl-CoA pool that is predicted to be used to enhance the production of malonyl-CoA.

Chaxamycin production was predicted to be enhanced by the deletion of *sle03600* and *sle39090* (trehalose 6-phosphate synthase) due to the indirect increment of malonyl-CoA and AHBA production, respectively. Further improvement was predicted with the additional deletion of the genes *sle29000*, *sle29010*, and *sle56410* (IMP dehydrogenases) associated to the reaction rx1231.

Gene targets for overexpression were found for each of the analysed specialized metabolites. Specifically, 35 and 25 overexpression targets for enhancing the production of chaxamycins and chaxalactins, respectively, were predicted. These targets were not directly related to their biosynthesis. The majority of them correspond to new overexpression targets, although homologous genes to previously described overexpression targets for increasing other polyketides biosynthesis were also found, such as the genes associated with the conversion of acetyl-CoA to malonyl-CoA in the reactions rx0404 and rx0405.

ACKNOWLEDGMENTS

This paper is partly based on work carried out by V.R. at the John Innes Centre in Norwich, UK and we are very grateful to Mervyn Bibb and Juan-Pablo Gomez-Escribano who were extremely helpful in obtaining a high quality genome sequence of *S. leeuwenhoekii* C34, which was essential for the construction of the genome scale model. We are also grateful to comments on the paper by Mervyn Bibb.We are very grateful to Michael Goodfellow for the help provided in the biolog experimental design and in the facilitation of installations. We thank the help provided in the use of the biolog equipment and in the biolog data

analysis, to Imen Nouioui and to María del Carmen Montero-Calasanz, respectively. This project was partly funded by a UK Newton Project for UK-Chile collaboration (JIC CA 586). V.R. received a national Ph.D. scholarship No 21110384 from the National Commission for Scientific and Technological Research (CONICYT). This work was also financially supported by the Basal Programme of CONICYT (Chile) for funding of the Centre for Biotechnology and Bioengineering (CeBiB) (project FB0001).

ORCID

Juan A. Asenjo http://orcid.org/0000-0002-3475-7664

REFERENCES

Alam, M. T., Merlo, M. E., Takano, E., & Breitling, R. (2010). Genome-based phylogenetic analysis of *Streptomyces* and its relatives. *Molecular Phylogenetics and Evolution*, *54*, 763–772.

Barona-Gómez, F., Wong, U., Giannakopulos, A. E., Derrick, P. J., & Challis, G. L. (2004). Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. *Journal of the American Chemical Society*, *126*, 16282–16283.

Bentley, S., Chater, K., Cerdeno-Tarraga, A. M., Challis, G., Thomson, N., James, K., … Hopwood, D. A. (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 Research*, *15*, 820–829.

Bursy, J., Kuhlmann, A. U., Pittelkow, M., Hartmann, H., Jebbar, M., Pierik, A. J., & Bremer, E. (2008). Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by *Streptomyces coelicolor* A3 (2) in response to salt and heat stresses. *Applied and Environmental Microbiology*, *74*, 7286–7296.

Busarakam, K. (2014). *Novel actinobacterial diversity in arid Atacama Desert soils as a source of new drug leads* (PhD thesis). Newcastle University.

Busarakam, K, Bull, A. T, Girard, G, Labeda, D. P, van Wezel, G. P, & Goodfellow, M. (2014). *Streptomyces leeuwenhoekii* sp. nov., the producer of chaxalactins and chaxamycins, forms a distinct branch in *Streptomyces* gene trees. *Antonie van Leeuwenhoek*, *105*, 849–861.

Campodonico, M. A., Andrews, B. A., Asenjo, J. A., Palsson, B. O., & Feist, A. M. (2014). Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm GEM-Path. *Metabolic Engineering*, *25*, 140–158.

Castro, J. F. (2015). *Identification of the chaxamycin and chaxalactin biosynthesis genes through genome mining of Streptomyces leeuwenhoekii C34 and heterologous production of chaxamycins in Streptomyces coelicolor M1152* (PhD thesis). University Of Chile

Castro, J. F., Razmilic, V., Gomez-Escribano, J. P., Andrews, B., Asenjo, J. A., & Bibb, M. J. (2015). Identification and heterologous expression of the chaxamycin biosynthesis gene cluster from *Streptomyces leeuwenhoekii*. *Applied and Environmental Microbiology*, *81*, 5820–5831.

Challis, G. L. (2008). Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology*, *154*, 1555–1569.

Chen, Y., Smanski, M. J., & Shen, B. (2010). Improvement of secondary metabolite production in *Streptomyces* by manipulating pathway regulation. *Applied Microbiology and Biotechnology*, *86*, 19–25.

Choi, H. S., Lee, S. Y., Kim, T. Y., & Woo, H. M. (2010). In silico identification of gene amplification targets for improvement of lycopene production. *Applied and Environmental Microbiology*, *76*, 3097–3105.

Cokelaer, T., Pultz, D., Harder, L. M., Serra-Musach, J., & Saez-Rodriguez, J. (2013). BioServices: A common Python package to access biological Web Services programmatically. *Bioinformatics*, *29*, 3241–3242.

Contador, C., Rodŕıguez, V., Andrews, B., & Asenjo, J. (2015). Genome-scale reconstruction of *Salinispora tropica* CNB-440 metabolism to study strain-specific adaptation. *Antonie van Leeuwenhoek*, *108*, 1075–1090.

Demain, A. L., & Fang, A. (2000). The natural functions of secondary metabolites. *History of modern biotechnology I* (pp. 1–39). Heidelberg, Germany: Springer.

D'Huys, P.-J., Lule, I., Vercammen, D., Anné, J., Van Impe, J. F., & Bernaerts, K. (2012). Genome-scale metabolic flux analysis of *Streptomyces lividans* growing on a complex medium. *Journal of Biotechnology*, *161*, 1–13.

Durot, M., Le Fèvre, F., de Berardinis, V., Kreimeyer, A., Vallenet, D., Combe, C., … Schachter, V. (2008). Iterative reconstruction of a global metabolic model of *Acinetobacter baylyi* ADP1 using high-throughput growth phenotype and gene essentiality data. *BMC Systems Biology*, *2*, 85.

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: Constraints-based reconstruction and analysis for python. *BMC Systems Biology*, *7*, 74.

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., … Palsson, B. Ø. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, *3*, 121.

Flamholz, A., Noor, E., Bar-Even, A., & Milo, R. (2011). EQuilibrator—The biochemical thermodynamics calculator. *Nucleic Acids Research*, *40*, 770–775.

Gomez-Escribano, J. P., & Bibb, M. J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology*, *4*, 207–215.

Gomez-Escribano, J. P., Castro, J. F., Razmilic, V., Chandra, G., Andrews, B., Asenjo, J. A., & Bibb, M. J. (2015). The *Streptomyces leeuwenhoekii* genome: De novo sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2. *BMC Genomics*, *16*, 485.

Gross, H. (2007). Strategies to unravel the function of orphan biosynthesis pathways: Recent examples and future prospects. *Applied Microbiology and Biotechnology*, *75*, 267–277.

Günter, K., Toupet, C., & Schupp, T. (1993). Characterization of an iron-regulated promoter involved in desferrioxamine B synthesis in *Streptomyces pilosus*: Repressor-binding site and homology to the diphtheria toxin gene promoter. *Journal of Bacteriology*, *175*, 3295–3302.

Herrgård, M. J., Lee, B.-S., Portnoy, V., & Palsson, B. Ø. (2006). Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Research*, *16*, 627–635.

Hiltner, J. K., Hunter, I. S., & Hoskisson, P. A. (2015). Chapter Four-Tailoring specialized metabolite production in *Streptomyces*. *Advances in Applied Microbiology*, *91*, 237–255.

Huang, D., Li, S., Xia, M., Wen, J., & Jia, X. (2013). Genome-scale metabolic network guided engineering of *Streptomyces tsukubaensis* for FK506 production improvement. *Microbial Cell Factories*, *12*, 1.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing In Science and Engineering*, *9*, 90–95.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., … Ōmura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature Biotechnology*, *21*, 526–531.

Ingraham, J. L., Maaløe, O., & Neidhardt, F. C. (1983). *Growth of the bacterial cell*. Sunderland, MA: Sinauer.

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*, 27–30.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge, and principle: Back to metabolism in KEGG. *Nucleic Acids Research*, *42*, 199–205.

Kieser, T., Bibb, M., Buttner, M., Chater, K., & Hopwood, D. (2000). *Practical streptomyces manual* (2nd ed.). Norwich, UK: The John Innes Foundation.

Kim, H. B., Smith, C. P., Micklefield, J., & Mavituna, F. (2004). Metabolic flux analysis for calcium dependent antibiotic (CDA) production in *Streptomyces coelicolor*. *Metabolic Engineering*, *6*, 313–325.

Kim, J., & Copley, S. D. (2007). Why metabolic enzymes are essential or nonessential for growth of *Escherichia coli* K12 on glucose. *Biochemistry*, *46*, 12501–12511.

Kim, M., Sang Yi, J., Kim, J., Kim, J.-N., Kim, M. W., & Kim, B.-G. (2014). Reconstruction of a high-quality metabolic model enables the identification of gene overexpression targets for enhanced antibiotic production in *Streptomyces coelicolor* A3 (2). *Biotechnology Journal*, *9*, 1185–1194.

King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., & Palsson, B. O. (2015). Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Computational Biology*, *11*, e1004321.

Lautru, S., Deeth, R. J., Bailey, L. M., & Challis, G. L. (2005). Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nature Chemical Biology*, *1*, 265–269.

Laxminarayan, R., Duse, A., Wattal, C., Zaidi, A. K., Wertheim, H. F., Sumpradit, N., … Cars, O. (2013). Antibiotic resistance—The need for global solutions. *The Lancet Infectious Diseases*, *13*, 1057–1098.

Maharjan, S., Koju, D., Lee, H. C., Yoo, J. C., & Sohng, J. K. (2012). Metabolic engineering of *Nocardia* sp. CS682 for enhanced production of nargenicin A1. *Applied Biochemistry and Biotechnology*, *166*, 805–817.

Maharjan, S., Park, J. W., Yoon, Y. J., Lee, H. C., & Sohng, J. K. (2010). Metabolic engineering of *Streptomyces venezuelae* for malonyl-CoA biosynthesis to enhance heterologous production of polyketides. *Biotechnology Letters*, *32*, 277–282.

Marchler-Bauer, A., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., … Bryant, S. H. (2009). CDD: Specific functional annotation with the conserved domain database. *Nucleic Acids Research*, *37*, D205– D210.

Marchler-Bauer, A., & Bryant, S. H. (2004). CD-Search: Protein domain annotations on the fly. *Nucleic Acids Research*, *32*, W327–W331.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., … Bryant, S. H. (2014). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, *43*, D222–D226.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., … Bryant, S. H. (2011). CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, *39*, D225–D229.

McMurry, J., & Begley, T. P. (2005). *The organic chemistry of biological pathways*. Colorado: Roberts and Company Publishers.

Medema, M. H., Trefzer, A., Kovalchuk, A., van den Berg, M., Müller, U., Heijne, W., … Takano, E. (2010). The sequence of a 1.8 Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biology and Evolution*, *2*, 212–224.

Oberhardt, M. A, Puchaɬka, J., Fryer, K. E., Dos Santos, V. A. M, & Papin, J. A (2008). Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *Journal of Bacteriology*, *190*, 2790–2803.

Oh, Y. K., Palsson, B. O., Park, S. M., Schilling, C. H., & Mahadevan, R. (2007). Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *Journal of Biological Chemistry*, *282*, 28791–28799.

Okoro, C. K., Brown, R., Jones, A. L., Andrews, B. A., Asenjo, J. A., Goodfellow, M., & Bull, A. T. (2009). Diversity of culturable actinomycetes in hyper-arid soils of the Atacama Desert, Chile. *Antonie van Leeuwenhoek*, *95*, 121–133.

Olano, C., Lombó, F., Méndez, C., & Salas, J. A. (2008). Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metabolic Engineering*, *10*, 281–292.

Ōmura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., … Hattori, M. (2001). Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. *Proceedings of the National Academy of Sciences USA*, *98*, 12215–12220.

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., & Palsson, B. Ø. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular Systems Biology*, *7*, 535.

Parekh, S., Vinci, V., & Strobel, R. (2000). Improvement of microbial strains and fermentation processes. *Applied Microbiology and Biotechnology*, *54*, 287–301.

Peters, P., Galinski, E., & Trüper, H. (1990). The biosynthesis of ectoine. *FEMS Microbiology Letters*, *71*, 157–162.

Rateb, M. E., Houssen, W. E., Arnold, M., Abdelrahman, M. H., Deng, H., Harrison, W. T., … Jaspars, M. (2011). Chaxamycins A-D, bioactive ansamycins from a hyper-arid desert *Streptomyces* sp. *Journal of Natural Products*, *74*, 1491–1499.

Rateb, M. E., Houssen, W. E., Harrison, W. T., Deng, H., Okoro, C. K., Asenjo, J. A., … Jaspars, M. (2011). Diverse metabolic profiles of a *Streptomyces* strain isolated from a hyper-arid environment. *Journal of Natural Products*, *74*, 1965–1971.

Ren, Q., Chen, K., & Paulsen, I. T. (2007). TransportDB: A comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Research*, *35*, D274–D279.

Ryu, Y. G., Butler, M. J., Chater, K. F., & Lee, K. J. (2006). Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*. *Applied and Environmental Microbiology*, *72*, 7132–7139.

Schupp, T., Toupet, C., & Divers, M. (1988). Cloning and expression of two genes of *Streptomyces pilosus* involved in the biosynthesis of the siderophore desferrioxamine B. *Gene*, *64*, 179–188.

Shao, Z. H., Ren, S. X., Liu, X. Q., Xu, J., Yan, H., Zhao, G. P., & Wang, J. (2015). A preliminary study of the mechanism of nitrate-stimulated remarkable increase of rifamycin production in *Amycolatopsis mediterranei* U32 by RNA-s eq. *Microbial Cell Factories*, *14*, 1.

Swiatek, M. A., Tenconi, E., Rigali, S., & van Wezel, G. P. (2012). Functional analysis of the *N*-acetylglucosamine metabolic genes of *Streptomyces coelicolor* and role in control of development and antibiotic production. *Journal of Bacteriology*, *194*, 1136–1144.

Zha, W., Rubin-Pitel, S. B., Shao, Z., & Zhao, H. (2009). Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering. *Metabolic Engineering*, *11*, 192–198.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Razmilic V, Castro JF, Andrews B, Asenjo JA. Analysis of metabolic networks of *Streptomyces leeuwenhoekii* C34 by means of a genome scale model: Prediction of modifications that enhance the production of specialized metabolites. *Biotechnology and Bioengineering*. 2018;115:1815–1828. https://doi.org/10.1002/bit.26598