Bioprocess and Biosystems Engineering (2018) 41:657–669
https://doi.org/10.1007/s00449-018-1900-9

RESEARCH PAPER

An enhanced genome-scale metabolic reconstruction of *Streptomyces clavuligerus* identifies novel strain improvement strategies

León Toro[1] · Laura Pinilla[1] · Claudio Avignone-Rossa[2] · Rigoberto Ríos-Estepa[3]

Received: 30 August 2017 / Accepted: 19 January 2018 / Published online: 5 February 2018

Abstract

In this work, we expanded and updated a genome-scale metabolic model of *Streptomyces clavuligerus*. The model includes 1021 genes and 1494 biochemical reactions; genome-reaction information was curated and new features related to clavam metabolism and to the biomass synthesis equation were incorporated. The model was validated using experimental data from the literature and simulations were performed to predict cellular growth and clavulanic acid biosynthesis. Flux balance analysis (FBA) showed that limiting concentrations of phosphate and an excess of ammonia accumulation are unfavorable for growth and clavulanic acid biosynthesis. The evaluation of different objective functions for FBA showed that maximization of ATP yields the best predictions for cellular behavior in continuous cultures, while the maximization of growth rate provides better predictions for batch cultures. Through gene essentiality analysis, 130 essential genes were found using a limited in silico media, while 100 essential genes were identified in amino acid-supplemented media. Finally, a strain design was carried out to identify candidate genes to be overexpressed or knocked out so as to maximize antibiotic biosynthesis. Interestingly, potential metabolic engineering targets, identified in this study, have not been tested experimentally.

Keywords Genome-scale metabolic reconstruction · Flux balance analysis · *Streptomyces clavuligerus* · Strain improvement · Clavulanic acid

Introduction

Streptomyces clavuligerus is a gram-positive bacterium, widely used in the production of clavulanic acid (CA), cephamycin C (CephC) and 5*S* clavam compounds with antifungal and antitumor activities. CA is a β-lactamase inhibitor clinically used in combination with β-lactam antibiotics to treat infections caused by β-lactamase-producing bacteria [1].

The biosynthetic pathway leading to CA starts with the condensation of the C3 precursor glyceraldehyde-3-phosphate and the C5 amino precursor L-arginine, to form *N*2-(2-carboxyethyl) arginine, a reaction catalyzed by carboxyethylarginine synthase (*CeaS2*). Further cyclization, hydroxylation and hydrolysis reactions produce clavaminic acid [2, 3]; these so-called early steps end by the action of clavaminate synthase (*Cas*): first, proclavaminic acid undergoes cyclization to produce dihydroclavaminic acid, which is subsequently desaturated to yield clavaminic acid [3]. In *S. clavuligerus*, the biosynthetic pathway leading to CA and 5S clavams is partially shared up to the clavaminic acid intermediate; the early steps are well characterized, but little is known about the reactions beyond clavaminic acid leading to CA or 5S clavams. In these late steps, *N*-glycyl-clavaminic acid synthetase (*gcas*) converts clavaminic acid into *N*-glycyl-clavaminic acid, but details of the reactions involved are still unknown. Finally, clavaldehyde dehydrogenase (*cad*) reduces clavaldehyde to form CA; both clavaldehyde and CA share the 3R, 5R stereochemistry associated with its

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00449-018-1900-9) contains supplementary material, which is available to authorized users.

✉ Rigoberto Ríos-Estepa
rigoberto.rios@udea.edu.co

[1] Grupo de Bioprocesos, Instituto de Biología, Universidad de Antioquia, Calle 70 No. 52-21, Medellín, Colombia

[2] Department of Microbial Sciences, School of Biosciences and Medicine, University of Surrey, Guildford, Surrey GU2 7XH, UK

[3] Grupo de Bioprocesos, Departamento de Ingeniería Química, Universidad de Antioquia, Calle 70 No. 52-21, Medellín, Antioquia, Colombia

inhibitory activity [4]. The genes involved in CA and 5S clavam biosynthesis are located in three separated locations within the *S. clavuligerus* chromosome (secondary biosynthetic clusters). The CA cluster is located next to the CephC gene cluster; the CA-CephC-supercluster and clavam cluster both lie on the chromosome, whereas the paralog gene cluster is located in the plasmid pSCL4 [5].

In view of the clinical and industrial importance of CA, the biosynthetic pathway, the associated gene cluster(s) and related regulatory mechanisms that control its production, have been topics of continuous research. *S. clavuligerus* also has a particular response system to nutrient-limiting conditions, linked to morphological differentiation and the onset of secondary metabolism [2]. These unclear aspects of *S. clavuligerus* metabolism reveal the need of using systemic methodologies that enable a holistic view of CA biosynthesis and accumulation under diverse environmental perturbations, so as to capture and eventually understand the genotype–phenotype relationship.

To improve the overall understanding of the metabolic capabilities of microorganisms, a systems biology approach, combining experimental and mathematical tools, has been established. Genome scale metabolic models (GSMMs) have been developed so as to bridge the gap between genome-derived biochemical information and metabolic phenotypes in a suitable manner, thus providing a solid interpretative framework for experimental data related to metabolic states, and enabling simple in silico experiments considering whole-cell metabolism. The basis of a GSMM lies in the functional annotation of the genes, including biosynthetic gene clusters of secondary metabolites that, when connected to the biochemical reactions catalyzed by the corresponding enzymes, provide a complete summary of the metabolic capabilities of the organism. Since parts of these biosynthetic gene clusters have not been fully characterized, GSMM-based simulations, along with bioinformatics tools and experimental data, might contribute to further identify their biological function.

Moreover, GSMMs simulate the operation of the metabolism in response to diverse perturbations, thus contributing to decipher potential interactions among metabolic, regulatory and signaling networks in a holistic manner. Investigating different in silico metabolic scenarios might facilitate the identification of better nutrient and environmental culture conditions for enhanced secondary metabolite production. Several actinomycete GSMMs are currently available for *S. coelicolor* [6], and *S. lividans* [7]; for *S. clavuligerus*, Medema et al. [8] presented a GSMM describing secondary metabolite production [8].

In this work, a reconstructed and updated GSMM was built and used to represent the *S. clavuligerus* metabolic activity. The resulting model was improved by curation of the genome-reaction information, and by adding reactions related with clavam metabolism and an updated biomass equation. By simulating different metabolic scenarios, the model allowed us to identify potential metabolic targets that might be modified by genetic engineering techniques so as to obtain overproducer strains.

Materials and methods

Genome-scale metabolic model reconstruction and refinement

The genome-scale stoichiometric model of *S. clavuligerus* [8] (kindly provided by Prof Marnix Medema, Wageningen University, The Netherlands) and a model for the phylogenetically close organism *S. coelicolor* A (3), for which genome scale reconstruction information is available [6], were utilized as base models to generate the so called *i*LT1021 model. Besides, a genome scale reconstruction was performed as a parallel model to search for functions that might not be included in the base models; for this, the genome of *S. clavuligerus* ATCC 27064 (accession number CM000913.1–CM000914.1) [9], and the Model SEED genome-scale metabolic reconstruction pipeline [10], were used. The parallel draft reconstruction had false and missing reactions due to wrong or incomplete annotations [11]; hence, it was compared step by step with genome annotations deposited in databases such as StrepDB (http://strepdb.streptomyces.org.uk), BioCyc [12], PATRIC [13], and Kyoto encyclopedia of genes and genomes (KEGG) [14]; this enabled us to build a complete list of reactions and the corresponding Gene–Protein–Reaction (GPR) associations, a very important step for further in silico gene deletion studies.

The *i*LT1021 model was converted to SBML format to be fully compatible with COBRA Toolbox [11] and other FBA tools like SurreyFBA [15]. The COBRA Toolbox was used to run and refine the model; gaps in the model were then identified using the steady-state model and constraint consistency checker (MC^3) for biochemical networks to detect connectivity and topological issues in the stoichiometric matrix and flagging issues that might arise during constraint-based optimization [16].

To perform the in silico simulations and to predict the metabolic flux distribution under different conditions in *S. clavuligerus* cultures, flux balance analysis (FBA) was carried out, as described elsewhere [17].

Biomass composition

The biomass composition of *S. clavuligerus* consisted of protein, DNA, RNA, lipids (phospholipids and triacylglycerol), cell wall constituents (peptidoglycan, carbohydrates,

teichoic acid and trehalose), and small molecules. Defining the biomass equation is a crucial step to obtain high-quality models [18]; yet, there is no detailed information about biomass constituents of *S. clavuligerus*. The first GSMM for *streptomycetes* species was the iLB711 model for *S. coelicolor* [6]; since then, biomass composition and equations for this species have been used for the related emerging models as for *S. clavuligerus*.

Robustness analysis (RA)

The effect of different substrate concentrations (C-, N-, P- and O sources) on biomass and antibiotic production (CA) was studied; the maximum consumption rate for each substrate was set to 1 mmol gDW^{-1} h^{-1} and, for the prediction of different effects on the antibiotic biosynthesis rate, the growth rate was constrained to 90% of the maximum value so as to avoid the unrealistic case of CA production without biomass growth. The flux towards CA biosynthesis was maximized as the objective function. RA was performed to evaluate the phenotypic response (growth rate and CA production) to perturbations on the different substrates generally used in *S. clavuligerus* cultures. For this, an in silico minimal medium was used, GAPI, containing glycerol, ammonia and phosphate; all other exchange reactions were constrained to zero, except for oxygen, ions, water and hydrogen, which were left unconstrained.

Validation of the *i*LT1021 model and evaluation of alternative objective functions

In this work, we used experimental data reported by Bushell et al. [19] for a continuous culture, as constraints to investigate the validity of the *i*LT1021 model. Different studies [20–23] consider growth rate maximization as the most appropriate objective function; however, this may not be the case for experimental instances such as the production of secondary metabolites, e.g., antibiotics [24]. Therefore, additional experimental datasets, including batch cultures, were proposed as scenarios for the evaluation of different linear and non-linear objective functions (obfuncs). The obfuncs used in this study were: max biomass, max ATP, max ATP/flux and max biomass/flux (maximization of ATP and biomass yield per unit flux, respectively); these obfuncs have been reported for other microbial growth simulations [22]. Using all the experimental constraints, the prediction of every objective function was observed by the correlation of scatter plots of separated flux-by-flux comparisons, to see how each experimental flux distribution matches the corresponding predictions. To quantify the overall agreement, we used the similarity between the multiple computational and experimental results by means of the Euclidean distance criteria, as described elsewhere [21].

Essential gene analysis

We analysed the essentiality of individual genes of *S. clavuligerus* under minimal (GAPI) and complex medium condition (CPX); the amino acid exchange reactions were set to 0.1 mmol gDW^{-1} h^{-1} and genes were sequentially deleted one at a time. The genes were categorized in three classes: (1) essential—genes required for cellular growth; (2) conditionally essential—considerable reduction of growth rate; and (3) non-essential—genes not required for cellular growth. Simulations using FBA as well as the *'grRatio'* function (relative growth rate ratio) computed the growth rate ratio between the one obtained from the model with a deleted gene and the original model without deletions. '*grRateWT*' represents the cell growth rate calculated using the original model (1/h) and the cell growth considering each knockout strain '*grRateKO*' [cell growth rate calculated from the model with a reaction deletion/gene (1/h)]. The essential gene analysis was conducted using the "*singleGeneDeletion*" function of the COBRA toolbox [11]: if the *grRatio* was lower than 10^{-6}, the gene was defined as essential. The biomass growth rate was set as the objective function for the different simulations.

Metabolic targets for increasing clavulanic acid production (strain design)

The use of genome scale models has been widely adopted, not only to simulate cell growth but also to predict beneficial gene knockout/overexpression to improve the production of metabolites of interest [22]. One of the most recent approach for the in silico strain design, RoBoKoD (Robust, Overexpression, Knockout and Dampening), is able to optimize the target product considering simultaneously the cellular growth using a multi-objective problem [23]. The RoBoKoD method uses flux variability analysis (FVA) to profile each reaction under differing production percentages of target compound and biomass synthesis [23]. Using these profiles, reactions were identified as potential knockouts, overexpression or dampening targets; the last targets correspond to those less intuitive alternatives that were not considered, according to the RoBoKoD method [23]. RoBoKoD was implemented for strain design using the *i*LT1021 model to identify the group of candidate genes to be altered (overexpressed/knocked-out), keeping those essential for cell growth and maintenance; as selection criteria, the rank of the reactions, overexpression (OER) or Knockout Ranking (KOR) >0 were used.

All calculations were made in MATLAB R2014a 8.3.0.5.32 (Mathworks Inc.; Natick, MA, USA) utilizing the SMBL Toolbox [25], and the COBRA Toolbox [11]. LP and MILP problems were solved using the Gurobi Solver [26].

Results and discussion

Reconstruction and properties of the model *i*LT1021

The most remarkable feature of modelling *S. clavuligerus* is the large size of its genome, 8,556,892 bps (6.760 Mb chromosome and 1.796 Mb plasmid), compared to typical bacterial genomes [8]. Regarding the annotation, 7281 protein-coding genes were predicted as open-reading frames (ORFs), including at least 1838 hypothetical proteins, 1367 putative proteins, 6 rRNAs operons, 73 tRNAs and 48 sary metabolite clusters. Only 1175 ORFs have an enzyme commission (E.C.) number assigned.

For the purposes of metabolic model reconstruction, we considered all the metabolic pathways that have been reported as active/operative in *S. clavuligerus*, and also for which most of the pathway-specific enzymes were present. The model included the biosynthesis of the β-lactam antibiotics CA and CephC. The biosynthetic route of clavams 5S and the last steps of the CA pathway are not yet completely elucidated [27]. Figure 1 presents the overall metabolic pathway content in terms of the percentage of total genes assigned to each pathway.

The updated genome scale model for *S. clavuligerus* was named *i*LT1021; it contains 1494 metabolic reactions (1295 metabolic conversions and 199 transport reactions), 1360 metabolites and 1021 genes that account for 14% of the total ORFs (see Table 1). Using the model SEED pipeline [10], a reconstructed/parallel draft model was obtained; it comprised 1358 reactions, 1470 metabolites and 1039 genes to obtain reaction–protein association automatically and include missing information in the final reconstruction project. The correct annotation was guided by checking correspondence to the PATRIC annotation [13] for the *S. clavuligerus* strain (prj:47867) and StrepDB database (http://strepdb.streptomyces.org.uk/). Finally, it was used as template to model GPR association.

We revised and expanded the published model [8] to generate the *i*LT1021 model, integrating new features that represent the metabolic capabilities of *S. clavuligerus*. Initially, 16 exchange reactions without associated metabolites in the core (cytoplasm) were identified. The model rendered ten dead ends; two genes not registered in databases were missing in the original model and corrected. Notwithstanding the clavam biosynthetic pathway is not completely elucidated, hypothetical reactions and the associated genes (SCLAV_2922, SCLAV_2923, SCLAV_2928 and SCLAV_2803) were added to the model to enable simulation of 5S clavam biosynthesis [27]. The lack of information about cofactors and precursors and the absence of experimental data reported for the clavam pathway undoubtedly restricts the work on simulation for this part of secondary metabolism. However, the integration of available information constitutes an important input for CA biosynthesis studies; hence it is possible to infer carbon flux at the CA

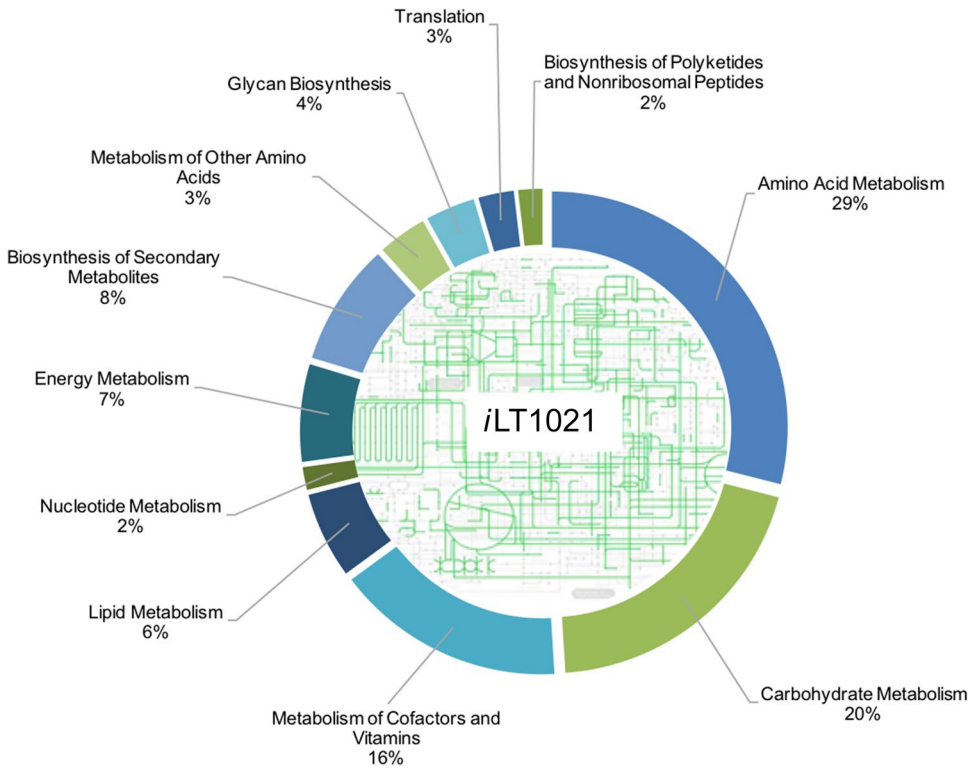

Fig. 1 Distribution of genes in the different metabolic pathways in the *i*LT1021 model

Table 1 Comparison of *Streptomyces*-species genome and model characteristics

	S. clavuligerus		*S. coelicolor A3*(2)	
Genome features				
Sequence length	8.556 Mb		8.667 Mb	
G + C content	72%		72%	
Total open-reading frames	7281		7825	
	*i*LT1021	Medema's model [8]		*i*MK1208 [24]
In silico model features				
Number of reactions	1494	1492		1643
Number of genes	1021	864		1208
Transport reactions	199	202		216
ORF associated to reactions (% total reactions)	66	58		86
ORF coverage (%)	14	11		15
Metabolites	1360	1375		1246
Compartments	2 (c, e)	2 (c, e)		2(c, e)

bifurcation point, wherein flux splits to yield either CA or 5S clavams, as well as to establish stoichiometric ratios that might explain the metabolic fate of carbon.

Further modifications included reactions for trehalose metabolism (SCLAV_0623, SCLAV_4349, SCLAV_1446 and SCLAV_3253) and most importantly, an updated biomass equation [28]. These changes in reactions/genes will become constraints for simulating *S. clavuligerus* metabolism, especially when integrating *omics* data e.g., transcriptomics or proteomics, to the model. The set of reactions, metabolites, and genes included in the model are provided in Supplementary material 1 and 2.

A summary of the features of the constructed model is shown in Table 1, along with a comparison with the most recent high-quality GSMMs for *S. clavuligerus* [8], and for *S. coelicolor* [24]. The addition of 157 genes to the *i*LT1021 (most of them located in the chromosome) results in a considerable increase of ORF coverage, an important factor in the analysis of gene essentiality and strain design. The identified missing links were filled either by the introduction of sink reactions to allow for material exchange between the cell and its surrounding environment, or by including reactions from other phylogenetically related organisms. The MC^3 software [16] was used to detect gaps and dead-end metabolites. MC^3 statistics did render a final model with 334 dead-end metabolites and 476 zero flux reactions (Supplementary material 1), showing that 70% of the reactions were active under the tested conditions; this might be explained by the inactivity of those transport reactions with no exchange of products or reactants, thus causing dead ends. Dead ends detected were involved neither in the biomass synthesis nor in the cellular metabolism.

Biomass equation

The biomass equation for *S. clavuligerus* was updated in the *i*LT1021 model; we used the macromolecular and monomeric composition of *S clavuligerus* [28]. Trehalose was included in the macromolecular composition in the updated *i*LT1021 model, as trehalose and glycogen are the two major forms of carbohydrate storage in actinomycetes [5]. The monomeric composition of biomass was calculated using *S. avermitilis* as a reference; this strain is genetically closer to *S. clavuligerus* than *S. coelicolor* (see Fig. 2). The amino acid content was calculated from codon usage. The content of glycine (Gly), leucine (Leu), aspartate (Asp), threonine (Thr) and proline (Pro) was very different compared with the protein composition generally used for the *S. coelicolor* models. Finally, while the composition of most macromolecules was identical to that of model *i*IB711, protein, RNA, DNA and carbohydrate composition were quite different (see Supplementary material 1).

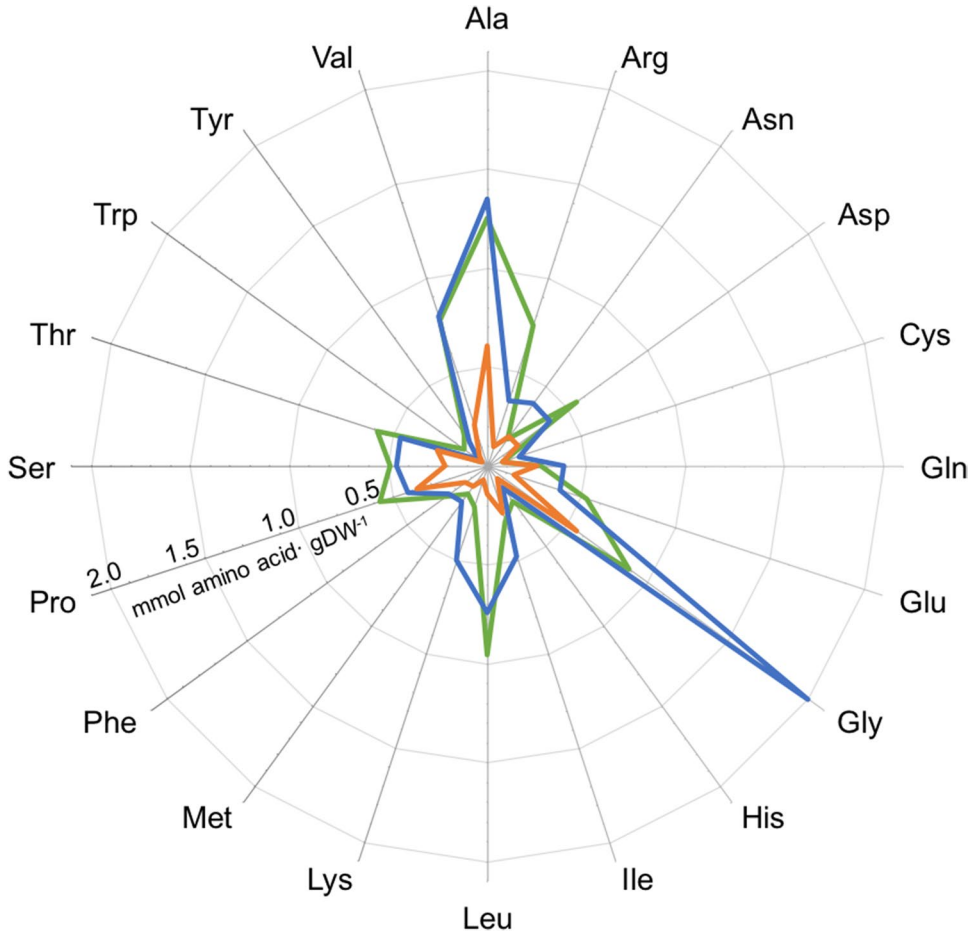

Fig. 2 Proposed protein composition for *Streptomyces clavuligerus* *i*LT1021 (green) and comparison with related *Streptomyces coelicolor* models: A(3) *i*B711 model [8], (blue) and *i*MK1208 model [24], (orange). (Color figure online)

Determination of energetic parameters

For the calculation of energetic parameters for *S. clavuligerus* (i.e., the growth and non-growth-associated ATP maintenance values, GAM and NGAM, respectively), we used available chemostat data obtained at four dilution rates [19]. As for microbial systems, the *P*/*O* ratio (the number of ATP molecules formed per oxygen consumed during respiration) does not exceed 2.0 [24], the *P*/*O* ratio in our simulations was constrained to 1.75. Lastly, the NADH dehydrogenase complex flux ratio was constrained to 1:3 (reactions Rxn0631 and Rxn0633), as it was suggested for the *i*IB711 model.

Using chemostat culture data from dilution rates in the range 0.02–0.07 h^{-1}, we calculated the intercept and slope by linear regression (from a plot of the specific growth rate as a function of the specific glycerol uptake rate). The intercept for the specific glycerol uptake rate (*x*-axis) was 0.025 mmol glycerol gDW^{-1} h^{-1}. Using this value, a NGAM value of 3.03 mmol ATP gDW^{-1} h^{-1} was obtained by FBA, while maximizing ATP consumption rate. A GAM value of 48.5 mmol ATP gDW^{-1} was obtained by FBA with maximization of ATP as the objective function. These GAM and NGAM values are in good agreement to those reported for the *S. coelicolor* A3(2) models. For *i*MK1208, NGAM and GAM were 2.65 mmol ATP $gDW^{-1}{\cdot}h^{-1}$ and 75.7 mmol ATP gDW^{-1}, respectively, while for *i*IB711, these values were 3.8 and 47.0, respectively [6, 24]. The values calculated for GAM and NGAM were used to adjust the biomass and the ATP maintenance requirement reactions, respectively.

In silico prediction of carbon and nitrogen utilization

The *S. clavuligerus* capability of utilizing different carbon and nitrogen sources for growth was predicted by FBA. Each substrate carbon source was used as a sole carbon source in the in silico minimal medium, assuming uptake rates of 1.0 mmol gDW^{-1} h^{-1}. The simulation results were compared to experimental data, as shown in Table 2. A 93.5% match was obtained, which indicates that the *S. clavuligerus* model predicts the activity of the diverse pathways for catabolism of the different carbon and nitrogen sources. These results are in good agreement with experimental reports [29–31].

Table 2 Simulation of growth for alternative carbon sources

Carbon source			Nitrogen source		
Substrate	Growth	References	Substrate	Growth	References.
Acetate	+	[29]	L-arginine	+	[29, 31]
Citrate	+	[29]	L-alanine	+	[29]
Fumarate	+	[29]	L-asparagine	+	[29]
Glycerol	+	[29]	L-aspartate	+	[29]
D-ribose	+	[29]	L-glutamate	+	[29, 31]
L-lactate	+	[29]	L-glutamine	+	[29]
L-malate	+	[29]	L-isoleucine	+	[29]
Maltose	+	[29]	L-leucine	+	[29]
Pyruvate	+	[29]	L-lysine	+	[29]
Succinate	+	[29]	L-methionine	−	[29]
Sucrose	+	[32]	L-phenylalanine	+	[29]
D-Glucose	− (*i*LT1021: +)	[29, 30]	L-proline	+	[29]
Trehalose	(*i*LT1021: +)	This study	L-valine	+	[29]
α-ketoglutarate	+	[29]	L-threonine	+	[29]
			L-ornithine	+	[31]
			Ammonia	+	[32]
			Urea	+	[32]

(+) Growth; (−) no growth

Since the model included a transport reaction for glucose as a sole carbon source, it did predict cellular growth. However, *S. clavuligerus* cannot grow on glucose due to a deficient expression of glucose permease *(glcP)* [33]. Clearly, despite the predictive power of GSMM, they possess inherent limitations related to the absence of regulatory issues in stoichiometric models. Furthermore, this linear-modelling approach does not consider the topological and flux connectivity features, which are indeed non-linear.

Robustness analysis

We performed a RA to evaluate the response of growth and CA production towards perturbations on substrate uptake and antibiotic production rates. Figure 3 shows the RA results for growth rate (Fig. 3a, b) and CA (Fig. 3c), as objective functions. The effect of phosphate uptake on cell growth shows a maximum growth rate μ of 0.07 h^{-1} for a *P* flux of 0.15 mmol gDW^{-1} h^{-1}. An increase in the flux of *P* promotes a decrease in growth rate down to zero at a *P* uptake rate of 1.0 mmol gDW^{-1} h^{-1}. Varying the oxygen uptake rate, maximum growth rate was achieved at an oxygen uptake rate of 0.01 mmol gDW^{-1} h^{-1}. The growth rate remained constant at higher uptake rates. This is consistent with the obligate aerobic metabolism of *S. clavuligerus* for CA production.

In the case of glycerol, simulations showed a linear behavior for growth rate as a function of the C-source uptake rate over the range of uptake rates tested; no inflexion point was observed. Changes in the N-source uptake rate showed that maximal growth rate is observed at an optimal uptake rate of 0.58 mmol gDW^{-1} h^{-1}. Higher uptake rates (> 0.6) caused a decrease in growth rate.

Regarding secondary metabolism (Fig. 3b), both CA and CephC biosynthesis showed similar results. Metabolic simulations allow for the study of extreme scenarios. Cell growth and no antibiotic production provide a feasible scenario, while maximum antibiotic biosynthesis rates with no cell growth becomes biologically unfeasible. Between those extreme scenarios, the phenotypic space shows all feasible biomass and product pair combinations in steady state flux vectors, where low growth rates favor antibiotic production [3].

With the purpose of determining the influence of the different substrates on CA biosynthesis (Fig. 3c), the value of the growth rate was restricted to 90% of the optimum value to ensure a scenario of CA synthesis along with biomass, and avoid the extreme non-feasible scenario, i.e., CA production without growth. The P-source shows a negative effect on CA production at high uptake rates and favorable at limiting conditions, coinciding with the well-known experimental data for *S. clavuligerus* where the most favorable conditions for antibiotic biosynthesis are observed at suboptimal phosphate concentrations [34].

The C-source and oxygen had the same effect on antibiotic production as that observed for growth rate as objective function. Finally, the N-source showed a negative effect on CA production at uptake rates above 0.655 mmol gDW^{-1} h^{-1}, with maximum values observed for uptake rates in the range of 0.586–0.655 mmol gDW^{-1} h^{-1}.

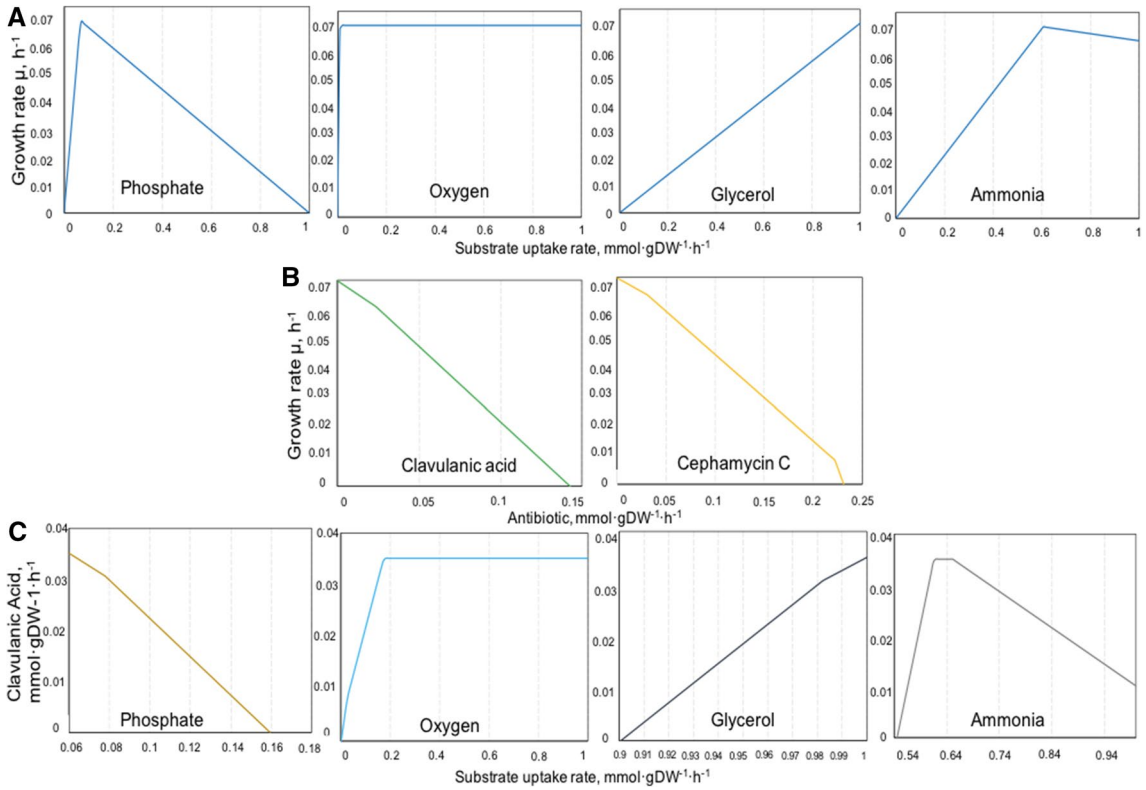

Fig. 3 Robustness analysis for *Sclav_i*LT1021 using minimal medium GAPI (glycerol, ammonia, phosphate and ions). **a** Substrate uptake effect on growth rate as objective function; **b** antibiotic biosynthesis and its dependence on growth rate; **c** substrate uptake effect on clavulanic acid biosynthesis as objective function and growth rate constrained to 90%

The negative effect of ammonia on CA production has been previously reported for in vivo studies; ammonia must be removed from the cells to avoid toxic intracellular concentrations that compromise growth [34]; the ammonia-free conditions also have been reported as favorable due to inhibitory and/or repressive effect of this nitrogen source [35].

Simulation and validation of model *i*LT1021

Available chemostat data of a phosphate-limited chemically defined media was used to run simulations [19]. In their studies, Bushell et al. [19] carried out chemostat cultures at four dilution rates; measured rates included glycerol and oxygen consumption, and CO_2 and CA excretion. The results revealed that antibiotic yields decreased with the increase of growth rate (Supplementary Figure S1). The complete dataset is shown in Table 3.

Using cell growth as objective function, the in silico growth rates were closed to experimental values at low dilution rates (differences were 4.0 and 17% from the 0.03 and 0.05 h^{-1}, respectively). However, at high growth rates (0.07 and 0.1 h^{-1}), the in silico values diverged from the in vivo values, while CO_2 rate was low (see Table 3). This discrepancy might be caused by a limited respiratory capacity rather than due to the biomass equation; a similar phenomenon was observed using the *i*MK1208 model for *S. coelicolor* [24]. In this regard, authors concluded that model prediction made by conventional FBA cannot capture a phenomenon in which carbon waste occurs due to low respiratory efficiency at high dilution rates [24].

Table 3 Comparison of in silico and in vivo growth rates of *S. clavuligerus*

Media condition (mmol gDW^{-1} h^{-1})		Growth rate (h^{-1})		CO_2 (mmol gDW^{-1} h^{-1})	
Glycerol	O_2	In vivo[a]	In silico[b]	In vivo	In silico
4.88	12.81	0.03	0.0313	6.41	8.86
6.51	15.56	0.05	0.0585	13.42	11.81
10.09	19.68	0.07	0.0836	14.86	13.63
23.67	53.43	0.1	0.188	66.52	37.04

[a]Experimental results taken from Bushell et al. [19]

[b]Simulation results

Simulations using biomass growth as objective function resulted in no CA production, since, as a secondary metabolite, the CA biosynthesis is not essential for growth and will compete for biosynthetic precursors (e.g.,

glyceraldehyde-3-phosphate) and energetic resources. This result agrees with those shown for actinorhodin (ACT) production using the *S. coelicolor* *i*IB711 model [36].

Exploring different objective functions using the *i*LT1021 model to predict growth and antibiotic production

The validation results showed that, using biomass maximization as a objective function, the model could not predict the experimentally observed CA production pattern in chemostat cultures (Supplementary Figure S1). Therefore, additional experimental datasets were used for the evaluation of different linear and non-linear objective functions (obfuncs). Besides the P-limited chemostat data by Bushell et al. [19], we used data from P-limited continuous cultures at two different dilution rates [28], which included additional data for phosphate and ammonia uptake rates. We also used data from a batch culture using a defined media supplemented with glutamate and proline [32], which included the production rate of cephamycin C. With these datasets, the suitability of each objective function (Table 4) was tested.

Objective functions were linear obfuncs (maximization of biomass and ATP yields), and non-linear maximization of ATP (max ATP/flux) and biomass yield per unit flux (max Biomass/flux) [21]. Supplementary Figure S2 shows a scatter plot for individual flux predictions contrasting predicted and experimental fluxes for all the objective functions evaluated. The dotted line represents the linear regression for correlation coefficient calculations. FBA studies have been previously used as a valuable tool for process design in antibiotic production, using maximization of cell growth as obfuncs [5–7]. Sánchez et al. [37] used CA production and ATP yield as obfuncs, as alternative functions to cell growth maximization [37]. In these simulations, using the experimental dataset from Bushell et al. [19] as constraints, the maximization of ATP yield showed to be the best predictor of cellular behavior [19].

Table 4 Objective functions evaluated using the *i*LT1021 model

Objective function	Description	Mathematical definition
max Biomass[a]	Maximization of biomass yield	$\max \frac{v_{biomass}}{v_{glycerol}}$
max ATP[b]	Maximization of ATP yield	$\max \frac{v_{ATP}}{v_{glycerol}}$
max ATP/flux	Maximization of ATP yield per flux unit	$\max \frac{v_{ATP}}{\sum v_i^2}$
max Biomass/flux	Maximization of biomass per flux unit	$\max \frac{v_{biomass}}{\sum v_i^2}$

Most suitable obfuncs for [a]Batch and [b]Chemostat scenarios, according to simulations conducted in this study

The correlation between the experimental fluxes reported and those predicted by the *i*LT1021 model are presented in Fig S2.a. In these simulations, maximization of ATP yield renders the highest correlation coefficients (0.81–0.85), followed by maximization of biomass per flux unit (0.74–0.79) and maximization of biomass yield (0.40–0.48). The same behavior was observed using a different continuous culture dataset [28] (Fig S2.b), though higher correlation coefficients were attained, $R = 0.86$ and $R = 0.81$–0.82 for max ATP and max biomass/flux, respectively; this is due to the cellular objective being strongly dependent on the input data, in this case a larger dataset. Finally, batch culture data (Fig S2.c) showed a very good correlation by maximizing biomass yield ($R = 0.99$) and maximization of ATP yield per flux unit ($R = 0.99$). Maximization of ATP yield and biomass per flux unit also showed a high R coefficient. These results agree with previous reports [20, 21, 36, 38], which supported the maximization of biomass as the objective function for FBA in batch cultures [21]; in contrast, chemostat cultures were better predicted by the ATP yield maximization. Overall, the two non-linear objective functions seem to be reliable for the cases tested. Although some objective functions provided reasonable predictions under the conditions evaluated, there is no consensus about which one has to be used when studying secondary metabolism. Therefore, special care should be taken in the evaluation and analysis of objective functions, case by case for each particular condition.

Identification of essential genes

The gene essentiality analysis revealed that 12.73 and 10.01% of total genes in *i*LT1021 are essential for cell growth in minimal (GAPI) and complex media (CPX), respectively (see Fig. 4). A total of 100 genes were essential in both conditions and an additional group of 30 genes were essential only in minimal media (see Fig. 4). Genes participating in central metabolism SCLAV_3951 (fructose-1,6-bisphosphatase II), SCLAV_1142 (glucose-6-phosphate isomerase), SCLAV_2648 (fructose-bisphosphate aldolase), SCLAV_3252 (trehalose-6-phosphate synthase) and SCLAV_3630, a gene coding for DNA-directed RNA polymerase subunit beta essential for growth (transcription process), were essential in both the GAPI and the CPX medium conditions.

To identify genes essential for cell viability under each condition, a functional categorization was made. We identified the distribution of genes across different metabolic processes. Aminoacyl-tRNA biosynthesis shows the most lethal genes (20%), since all genes involved are used for the synthesis of biomass precursors, mainly for protein macromolecules [39], for amino acid biosynthesis (16.90%) and purine metabolism (11.53%). For the CPX

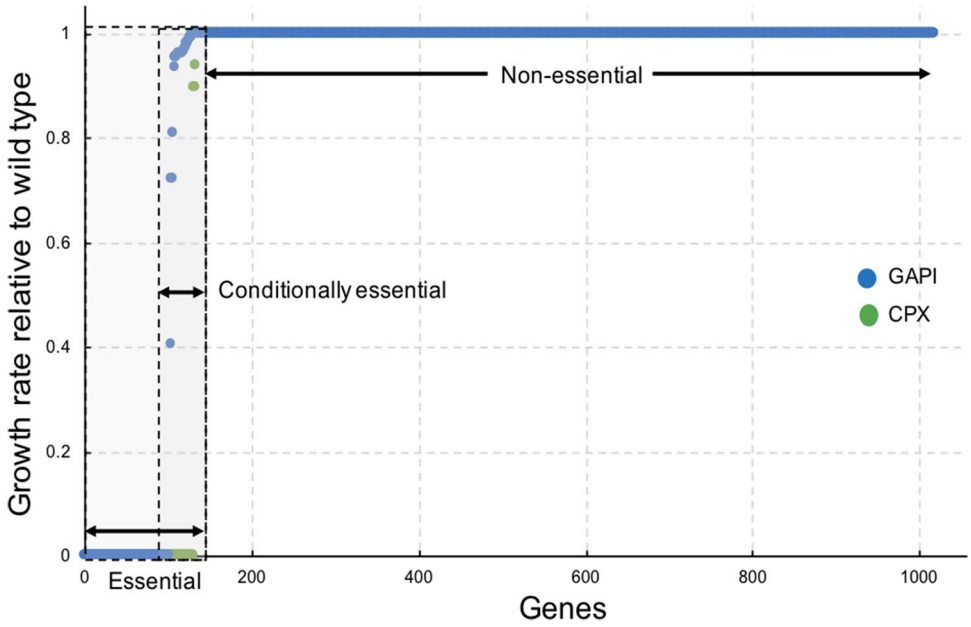

Fig. 4 Gene essentiality in the *S. clavuligerus* metabolic network. *GAPI* minimal medium (glycerol, ammonia, phosphate and ions), *CPX* complex medium

medium, supplemented with amino acids, nutrients were directly up taken and metabolized, therefore, maintaining various biosynthetic pathways inactive; genes associated with these pathways were classified as non-essential. However, certain amino acid pathways were identified as essential for cell maintenance and growth, such as the biosynthetic pathways of phenylalanine, tyrosine and tryptophan. The amino acid-associated gene argininosuccinate synthase (SCLAV_0796) is involved in the arginine biosynthetic pathway [40], essential for both growth and secondary metabolism. Argininosuccinate synthase seems to be involved in the developmental changes observed in the bacterial life cycle, as its disruption prevents the formation of aerial mycelium in several *Streptomyces* strains [41].

In a previous *S. clavuligerus* model [8], which used the same minimal media (GAPI) for validation purposes, a total of 100 essential genes were identified and the knockouts applied to the enzyme-coding genes located in the 1.8 Mb pSL4 plasmid did not show any effect on growth. This suggests that the mega-plasmid does not encode any functions essential to primary metabolism. Interestingly, gene essentiality analysis using the *i*LT1021 model reported, for the minimal media, 30 additional essential genes, which were added in the early reconstruction steps. Most of them belong to multi-copies in the chromosome except SCLAV_p0971, which is located in the pSL4 plasmid, encoding for 3-dehydroquinate synthase [EC:4.2.3.4]. This reaction is part of the shikimate pathway, involved in the biosynthesis of aromatic amino acids such as phenylalanine, tyrosine, and tryptophan.

Strain design for clavulanic acid overproduction

A total of 316 reactions were identified as potential targets for overexpression, with an overexpression ranking (OER) greater than zero, suggesting that 35.4% of the total genes would be worthy of being explored so as to acquire clavulanic acid overproduction. The higher OER value was 1.182 for the *N*-glycyl-clavaminic acid synthetase (*gcas*) and clavaldehyde dehydrogenase (*cad*); these reactions are involved in the CA biosynthetic pathway (see Table 5 and Supplementary material 1). The gene *gcas* represents the first identified biosynthetic intermediate of the pathway that is specific only for clavulanic acid and not for 5S clavams; meanwhile, *cad* reduces the clavaldehyde intermediate to clavulanic acid (last step) [42]. A second group of the CA biosynthetic genes, involved in the early steps, *ceas1/2, bls1/2* and *pah1/2* had an OER of 0.9733. The overexpression of the structural biosynthetic genes *ceas2, bls2, cas2* and *pah2* resulted in an 8.7-fold increase in CA titers [43].

Moreover, pyruvate dehydrogenase and malonate-semialdehyde dehydrogenase (OER: 0.8621), isocitrate dehydrogenase (IDH) NADP+ (OER: 0.8414) and histidinol-phosphate aminotransferase (OER: 0.8492) seem to be targets for further genetic engineering studies, since its overexpression results in higher pools of acetyl-coA and/or alpha-ketoglutarate, the latter a co-substrate in clavaminic acid biosynthesis. IDH (NADP+) is an important target of the TCA pathway as it increases NADPH availability, a cofactor in the conversion of clavaminic acid to CA. In addition, the selected reactions favor the biosynthesis of ornithine from glutamate, an essential amino acid precursor of CA via arginine [5].

Table 5 Reactions selected as potential targets for overexpression/knockout so as to increase CA biosynthesis

Reaction ID	Reaction name	Gene association	Ranking	References
Overexpression				
Rxn0967	*N*-glycyl-clavaminic acid synthetase (*gcas*)	SCLAV_4181	1.1825	This study
Rxn0968	Clavaldehyde dehydrogenase (*cad*) [EC: 1.1.1.100]	SCLAV_4190		
Rxn0961	*N*2-(2-carboxyethyl) arginine synthase (*ceas*) [EC:2.5.1.66]	SCLAV_4197 SCLAV_p1074	0.9733	[43]
Rxn0962	Carboxyethyl-arginine beta-lactam-synthase (*bls*) [EC:6.3.3.4]	SCLAV_4196		
Rxn0963 Rxn0965 Rxn0966	Clavaminate synthase 1 or 2 (*cas*) [EC:1.14.11.21]	SCLAV_2925 SCLAV_4194		
Rxn0964	Proclavaminate amidinohydrolase (*pah*) [EC:3.5.3.11]	SCLAV_4195		
Rxn0675	Pyruvate dehydrogenase [EC: 1.2.4.1]	SCLAV_1613	0.8621	[5]
Rxn0097	Isocitrate dehydrogenase (NADP+) [EC: 1.1.1.42]	SCLAV_0808	0.8414	This study
Rxn0849	Malonate-semialdehyde dehydrogenase [EC: 1.2.1.18]	SCLAV_1899 SCLAV_p0939	0.8621	
Rxn0207	Histidinol-phosphate aminotransferase [EC: 2.6.1.9]	SCLAV_1261	0.8492	
Knock-out				
Rxn0138	Glycolate oxidase [EC:1.1.3.15]	SCLAV_2040	1.058824	This study
Rxn0324	NAD + synthase (glutamine-hydrolysing) [EC:6.3.5.1]	SCLAV_1480	0.857143	
Rxn0320	L-aspartate oxidase [EC:1.4.3.16]	SCLAV_2394	0.857143	
Rxn0001	Glycerate kinase [EC:2.7.1.31]	SCLAV_0632 SCLAV_0878	0.70010	
Rxn0135	Isocitrate lyase [EC:4.1.3.1]	SCLAV_p0927	0.693038	

Glycerol can be metabolized through glycerol dehydrogenase, aldehyde dehydrogenase and glycerate kinase to form 2-phosphoglycerate and subsequently pyruvate [28]. The knock-out analysis suggested that glycerate-kinase, with a knockout ranking (KOR) of 0.7001, might be modified so as to favor the anaplerotic pathway towards the glyceraldehyde-3-phosphate pool, thus reducing amino acid biosynthesis via oxaloacetate and pyruvate, allowing glycerol dehydrogenase and aldehyde dehydrogenase to increase the C3 precursor pool for clavulanic acid biosynthesis. Concerning amino acid metabolism, SCLAV_2394 gene codes for L-aspartate oxidase; its knockout might avoid the formation of iminoaspartate thus allowing aspartate to be used as precursor in the synthesis of arginine, the C5 CA precursor. Likewise, by knocking out glycolate oxidase (KOR: 1.058) and isocitrate lyase (KOR: 0.6930), we could infer that blocking the respective reactions, the flux through the glyoxylate shunt decreases, allowing the synthesis of aspartate and alpha-ketoglutarate, key precursors in CA biosynthesis, since malate and isocitrate are not compromised in alternative routes.

Finally, although the reactions concerning the synthesis of clavams were included in the model, the lack of information about cofactors and even reactants within the clavam pathway prevented the use of RoBoKoD [23]; accordingly, the analysis did not render such reactions as potential targets. Even though RoBoKoD has not been validated beyond *E. coli* for butanol [23] and tryptophan [44] production, it proved to be a useful tool for strain design through accurate prediction of favorable genetic interventions; therefore, it would be of great interest to validate the predictions presented here for *S. clavuligerus*. To our knowledge, in silico tools for the metabolic design of *S. clavuligerus* strains have not been explored in depth, beyond the approach presented here.

Conclusions

In this work, an expanded genome-scale metabolic model of *S. clavuligerus*, *i*LT1021, was reconstructed and updated with an amended biomass equation and the introduction of clavam biosynthesis and trehalose metabolism. The developed model was validated by growth capabilities and robustness analysis; model predictions were in agreement with reported experimental data.

In constraint-based analysis of secondary metabolism in *Streptomyces*, the choice of an appropriate objective function is a key step to describe the metabolic shift from exponential to stationary growth phase. This study also reports on the analysis of different objective functions to predict the biomass growth along with CA biosynthesis using the *i*LT1021 model. Our results revealed that ATP yield is the best objective function to maximize to obtain a reliable model.

RoBoKoD was used along with the *i*LT1021 model to perform a strain design to overproduce clavulanic acid; this analysis identified a group of genes that have not been

explored experimentally in overexpression/knockout analysis, especially genes located in the CA gene cluster. Moreover, this study also advises about new potential targets for metabolic engineering; among others, the *IDH, glyK* and NAD^+ synthase (glutamine-hydrolysing) genes were identified. These results encourage the continuous development of GSMM as potential tools for bioprocess optimization.

Acknowledgements The authors thank Professor Marnix H. Medema and Mohammad Tauqeer Alam for providing the template for the *Streptomyces clavuligerus* model, and Professor Andrzej Kierzek for advice on the SurreyFBA platform [15]. This work was supported by Departamento Administrativo de Ciencia, Tecnología e Innovación—COLCIENCIAS-Colombia (Grant no. 111566945929). L. Toro and L. Pinilla thank COLCIENCIAS-Colombia for scholarships. C. Avignone-Rossa was supported by Grant BB/L02683X/1 from the Biotechnology and Biological Sciences Research Council (BBSRC, United Kingdom).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Paradkar A (2013) Clavulanic acid production by *Streptomyces clavuligerus*: biogenesis, regulation and strain improvement. J Antibiot 66:411–420
2. Jensen SE (2012) Biosynthesis of clavam metabolites. J Ind Microbiol Biotechnol 39:1407–1419
3. Ozcengiz G, Demain AL (2013) Recent advances in the biosynthesis of penicillins, cephalosporins and clavams and its regulation. Biotechnol Adv 31:287–311
4. Zelyas NJ, Cai H, Kwong T, Jensen SE (2008) Alanylclavam biosynthetic genes are clustered together with one group of clavulanic acid biosynthetic genes in *Streptomyces clavuligerus*. J Bacteriol 190:7957–7965
5. Medema MH, Alam MT, Heijne WH, van den Berg MA, Muller U, Trefzer A, Bovenberg RA, Breitling R, Takano E (2011) Genome-wide gene expression changes in an industrial clavulanic acid overproduction strain of *Streptomyces clavuligerus*. Microb Biotechnol 4:300–305
6. Borodina I, Krabben P, Nielsen J (2005) Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. Genome Res 15:820–829
7. D'Huys PJ, Lule I, Vercammen D, Anné J, Van Impe JF, Bernaerts K (2012) Genome-scale metabolic flux analysis of *Streptomyces lividans* growing on a complex medium. J Biotechnol 161:1–13
8. Medema MH, Trefzer A, Kovalchuk A, van den Berg M, Müller U, Heijne W, Wu L, Alam MT, Ronning CM, Nierman WC, Bovenberg RA (2010) The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. Genome Biol Evol 2:212–224
9. Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford) 2011:bar009
10. Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA, Henry C (2013) Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. In: Systems metabolic engineering: methods and protocols, pp 17–45
11. Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc 6:1290–1307
12. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahrén D, Tsoka S, Darzentas N, Kunin V, López-Bigas N (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res 33:6083–6089
13. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D (2014) PATRIC, the bacterial bioinformatics database and analysis resource. Nucl Acids Res 42:D581-D591
14. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30
15. Gevorgyan A, Bushell ME, Avignone-Rossa C, Kierzek AM (2011) SurreyFBA: a command line tool and graphics user interface for constraint-based modeling of genome-scale metabolic reaction networks. Bioinformatics 27:433–434
16. Yousofshahi M, Ullah E, Stern R, Hassoun S (2013) MC3: a steady-state model and constraint consistency checker for biochemical networks. BMC Syst Biol 7:1–8
17. Orth JD, Thiele I, Palsson B (2010) What is flux balance analysis? Nat Biotechnol 28:245–248
18. Durot M, Bourguignon PY, Schachter V (2009) Genome-scale models of bacterial metabolism: reconstruction and applications. FEMS Microbiol Rev 33:164–190
19. Bushell ME, Kirk S, Zhao HJ, Avignone-Rossa CA (2006) Manipulation of the physiology of clavulanic acid biosynthesis with the aid of metabolic flux analysis. Enzym Microb Technol 39:149–157
20. Edwards JS, Palsson B (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci USA 97:5528–5533
21. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol 3:119
22. Blazeck J, Alper (2010) Systems metabolic engineering: genome-scale models and beyond. Biotechnol J 5:647–659
23. Stanford NJ, Millard P, Swainston N (2015) RobOKoD: microbial strain design for (over)production of target compounds. Front Cell Dev Biol 3:17
24. Kim M, Sang Yi J, Kim J, Kim JN, Kim MW, Kim BG (2014) Reconstruction of a high-quality metabolic model enables the identification of gene overexpression targets for enhanced antibiotic production in *Streptomyces coelicolor* A3(2). Biotechnol J 9:1185–1194
25. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19:524–531
26. Hutter F, Hoos HH, Leyton-Brown K (2010) Automated configuration of mixed integer programming solvers. In: International conference on integration of artificial intelligence (AI) and operations research (OR) techniques in constraint programming. Springer, Berlin, pp 186–202
27. Tahlan K, Anders C, Wong A, Mosher RH, Beatty PH, Brumlik MJ, Griffin A, Hughes C, Griffin J, Barton B, Jensen SE (2007) 5S clavam biosynthetic genes are located in both the clavam and paralog gene clusters in *Streptomyces clavuligerus*. Chem Biol 14:131–142
28. Mercier C (2006) A Genome-scale investigation of clavulanic acid biosynthesis by *Streptomyces clavuligerus* in batch and chemostat cultures using transcriptomic and fluxomic analysis. PhD Thesis. University of Surrey, UK
29. Aharonowitz Y, Demain AL (1978) Carbon catabolite of cephalosphorin production in *Streptomyces clavuligerus*. Antimicrob Agents Chemother 14:159–164

30. Müller JC, Toome V, Pruess DL, Blount JF, Weigele M (1983) Ro 22-5417, a new clavam antibiotic from *Streptomyces clavuligerus*. III. Absolute stereochemistry. J Antibiot(Tokio) 36:208–212
31. Romero J, Liras P, Martín JF (1986) Utilization of ornithine and arginine as specific precursors of clavulanic acid. Appl Environ Microbiol 52:892–897
32. Romero J, Liras P, Martín JF (1984) Dissociation of cephamycin and clavulanic acid biosynthesis in *Streptomyces clavuligerus*. Appl Microbiol Biotechnol 20:318–325
33. Pérez-Redondo R, Santamarta I, Bovenberg R, Martín JF, Liras P (2010) The enigmatic lack of glucose utilization in *Streptomyces clavuligerus* is due to inefficient expression of the glucose permease gene. Microbiology 156:1527–1537
34. Lebrihi A, Germain P, Lefebvre G (1987) Phosphate repression of cephamycin and clavulanic acid production by *Streptomyces clavuligerus*. Appl Microbiol Biotechnol 26:130–135
35. Aharonowitz Y, Demain AL (1979) Nitrogen nutrition and regulation of cephalosporin production in *Streptomyces clavuligerus*. Can J Microbiol 25:61–67
36. Khannapho C, Zhao H, Bonde BK, Kierzek AM, Avignone-Rossa CA, Bushell ME (2008) Selection of objective function in genome scale flux balance analysis for process feed development in antibiotic production. Metab Eng 10:227–233
37. Sánchez C, Quintero JC, Ochoa S (2015) Flux balance analysis in the production of clavulanic acid by *Streptomyces clavuligerus*. Biotechnol Prog 31:1226–1236
38. García Sánchez CE, Torres Sáez RG (2014) Comparison and analysis of objective functions in flux balance analysis. Biotechnol Prog 30:985–991
39. Persson BC, Gustafsson C, Berg DE, Björk GR (1992) The gene for a tRNA modifying enzyme, m5U54-methyltransferase, is essential for viability in *Escherichia coli*. Proc Natl Acad Sci USA 89:3995–3998
40. Redshaw PA, McCann PA, Pentella MA, Pogell BM (1979) Simultaneous loss of multiple differentiated functions in aerial mycelium-negative isolates of *Streptomycetes*. J Bacteriol 137:891–899
41. Wu G, Culley DE, Zhang W (2005) Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. Microbiology 151:2175–2187
42. Song JY, Jensen SE, Lee KJ (2010) Clavulanic acid biosynthesis and genetic manipulation for its overproduction. Appl Microbiol Biotechnol 88:659–669
43. Jnawali HN, Yoo JC, Sohng JK (2011) Improvement of clavulanic acid production in *Streptomyces clavuligerus* by genetic manipulation of structural biosynthesis genes. Biotechnol Lett 33:1221–1226
44. Panichkin VB, Livshits VA, Biryukova IV, Mashko SV (2016) Metabolic engineering of *Escherichia coli* for L-tryptophan production. Appl Biochem Microbiol 52:783–809