RESEARCH ARTICLE

Development and validation of an updated computational model of *Streptomyces coelicolor* primary and secondary metabolism

Adam Amara, Eriko Takano and Rainer Breitling[*]

Abstract

Background: *Streptomyces* species produce a vast diversity of secondary metabolites of clinical and biotechnological importance, in particular antibiotics. Recent developments in metabolic engineering, synthetic and systems biology have opened new opportunities to exploit *Streptomyces* secondary metabolism, but achieving industry-level production without time-consuming optimization has remained challenging. Genome-scale metabolic modelling has been shown to be a powerful tool to guide metabolic engineering strategies for accelerated strain optimization, and several generations of models of *Streptomyces* metabolism have been developed for this purpose.

Results: Here, we present the most recent update of a genome-scale stoichiometric constraint-based model of the metabolism of *Streptomyces coelicolor*, the major model organism for the production of antibiotics in the genus. We show that the updated model enables better metabolic flux and biomass predictions and facilitates the integrative analysis of multi-omics data such as transcriptomics, proteomics and metabolomics.

Conclusions: The updated model presented here provides an enhanced basis for the next generation of metabolic engineering attempts in *Streptomyces*.

Keywords: Genome-scale metabolic modelling, *Streptomyces*, Secondary metabolism, Omics, Synthetic biology, Natural products

Background

Streptomyces species are usually soil-dwelling bacteria, which have adapted to their competitive ecological niches by developing a notably diverse secondary metabolism (e.g., antimicrobials). Currently, more than two thirds of the antibiotics used have been derived from natural products discovered in *Streptomyces* and related species [1]; however, the antibiotic discovery pipeline is drying up, while the antimicrobial resistance threat is growing. *Streptomyces coelicolor* A3(2) is a well-studied model organism for the production of antibiotics in this genus. The genome of this soil-dwelling bacterium encodes more than 20 secondary metabolite biosynthetic

gene clusters (BGCs) [2], and the species is known to produce multiple antibiotics such as Actinorhodin (Act), Undecylprodigiosin (Red), Calcium-Dependant Antibiotic (CDA) and the yellow Coelicolor Polyketide, Coelimycin P1 (yCPK) [3]. Recent developments in metabolic engineering, synthetic and systems biology have opened new opportunities to exploit *Streptomyces'* secondary metabolism diversity to discover novel antibiotics and natural product-derived drugs [4, 5]. However, expensive and time-consuming strain optimization is usually required to achieve industrially competitive production levels. A major issue faced in strain design is the ability to integrate test data (e.g. metabolomics) to improve the design [6], and many of the issues encountered are related to metabolic optimization, such as metabolic bottlenecks to increase production [7], heterologous biosynthetic pathway

* Correspondence: rainer.breitling@manchester.ac.uk
Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, School of Chemistry, Faculty of Science and Engineering, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK

Amara *et al. BMC Genomics* (2018) 19:519

Page 2 of 9

precursors production [8], or accurate predictions for metabolic engineering [9].

Genome-scale metabolic models (GSMM) have been shown to be a powerful tool to guide metabolic engineering strategies for accelerated strain optimization [10–12], and several generations of models of *Streptomyces* metabolism have been developed for this purpose [13–17]. The use of constraint-based modelling, in particular with flux balance analysis (FBA), enables the reconstruction and analysis of large metabolic networks from the genome sequence as well as predictions of growth associated phenotypes (metabolic fluxes, growth rates, metabolic gene essentiality) [18]. Informative models for this purpose can be constructed even when enzyme kinetic data or metabolite concentrations are unknown in the target organism, making this approach particularly attractive for less well-studied organisms like *Streptomyces* strains. In 2005, the first generation GSMM of *S. coelicolor*, iIB711, was published [19], which was used to identify metabolic gene knock-outs to drive the enhanced production of antibiotics in the strain [20]. In 2010, an updated model, iMA789, was published [21], which introduced more detailed antibiotics metabolic pathways and was used to interpret time-course gene expression data, which was then used to improve the model and update the genome annotation of the organism in the area of secondary metabolism. The most recent model update, iMK1208, was published by Kim et al. [22]; this model significantly expanded the number of reactions and genes, as well as updating the biomass reaction. This model was then used in a transcriptomics-based optimization for actinorhodin overproduction in *S. coelicolor* [15].

Furthermore, several genome-scale metabolic models for other biotechnologically relevant *Streptomyces* strains have been reconstructed since the first *S. coelicolor* model, iIB711. A model of the *Streptomyces tenebrarius* metabolic network, which was derived from the iIB711 model of Borodina, Krabben & Nielsen has been used to identify targets to optimize production of tobramycin [23]. A model of *Saccharopolyspora erythraea* has been reconstructed based on the iMA789 model of Alam et al. to improve the production of erythromycin [24]. One of the most recent model reconstructions derived from Kim et al.'s *S. coelicolor* iMK1208 model was used for model-guided engineering of ethylmalonyl-CoA pathways in *Streptomyces hygroscopicus* to increase production of ascomycin [14]. A large collection of minimally curated metabolic models of different *Streptomyces* strains and other actinomycetes was used to evaluate potential host strains for overproducing different chemical classes of secondary metabolites using comparative multi-objective modelling [25].

Based on recent advances in our understanding of *Streptomyces* metabolism and technical progress in the concepts of computational model building, we constructed and validated an updated GSMM of *S. coelicolor*, iAA1259, to provide a more precise metabolic flux and biomass predictions and to facilitate the integration of metabolomics, proteomics, and transcriptomics information with the model predictions.

Results & Discussion

Genome-scale reconstruction and characteristics updated

The construction of the updated GSMM of *S. coelicolor* A3(2), iAA1259, was based on all three previously published iterative reconstructions of *S. coelicolor* metabolic models [19, 21, 22], by updating and adding data in the model based on new genetic (e.g., gene–protein–reaction relationships) or biochemical knowledge. A summary of the main updates and new features added is available in the Additional file 1: Tables S3, S4, S5 and S6.

Multiple pathways were added or updated. 1) Polysaccharide degradation pathways (e.g., for xylan, cellulose) were introduced to enable simulated growth in complex media containing these carbon sources. 2) The biosynthetic pathway for the secondary metabolite yCPK [26–28] was added to the model. This cryptic BGC is awakened under phosphate-limited condition, in nitrogen and carbon rich media [26, 29], such as in the minimal media used for systems biology studies of *S. coelicolor* [30]. 3) The biosynthetic pathways for the signalling molecules gamma-butyrolactones (SCB1, 2 and 3) were added [31, 32]; secondary metabolite production in *Streptomyces* (e.g., yCPK) can be activated through these small diffusible molecules, and they are an interesting target for synthetic biology engineering [33, 34]. 4) The futalosine pathway, an alternative menaquinone biosynthesis pathway, which was highlighted as incomplete in the previous model [22], has now been updated following recently published studies [35, 36]. 5) The oxidative phosphorylation associated reactions have been manually curated. 6) Following the above modifications, the biomass reaction has also been updated to reflect more detailed knowledge on biomass composition such as the presence of 2-demethylmenaquinol in *S. coelicolor* (MetaCyc) [37], and organic polyphosphate storage [38], as well as an update in the stoichiometry of menaquinol based on *Mycobacterium tuberculosis* data [39] (see details in Additional file 2: Table S1).

In order to facilitate metabolomics data analysis, all metabolites in the model have now been annotated with standard identifiers for a variety of relevant databases (PubChem and ChEBI) [40, 41], and chemical and structural information about each metabolite has been added (InChi and SMILES strings) to ensure unambiguous metabolite identification [42, 43]. The model capacity to facilitate metabolomics data analysis has been tested by mapping metabolites annotated with mzMatch

Amara *et al. BMC Genomics* (2018) 19:519

Page 3 of 9

[44] from an untargeted metabolomics dataset of *S. coelicolor* [45]; the metabolites were mapped automatically onto the iAA1259 metabolic network (see details in Additional file 1: Figure S1). In addition, to facilitate transcriptomics data analysis and comparative modelling, gene annotation has been expanded to include identifiers for multiple standard databases (Gene Ontology, Ensembl, and RefSeq) [46–48]. Finally, to integrate proteomics data analysis, standard database identifiers (UniProt, Pfam, and Panther) [49–51] and key reference data, such as protein sequence, length, and mass, have been added. The final model, iAA1259, is fully compliant with the current standards for high-quality GSMMs [52–54], iAA1259 is available as a SBML file in Additional file 3 and as an excel file in Additional file 4.

### Validations of the metabolic model predictions

As the first step in model validation, chemostat data collected by Melzoch et al. *for S. coelicolor* in a glucose-limited minimal defined media [55] were used to compare biomass predictions by the four generations of model: iIB711, iMA789, iMK1208, and iAA1259. Specific growth rates for each model were predicted in silico, using the known glucose and O_2 uptake rates as constraints on the model, along with the production rates of CO_2 and γ-actinorhodin, the extracellular lactone form of actinorhodin [56]. Biomass production was maximized to estimate the optimal

predicted growth rate. Then, the growth rate predicted in silico was compared to the dilution rate that corresponds to the observed growth rate at steady state (Fig. 1). Since the first published model iIB711, there have been some significant improvements in biomass predictions; iAA1259 shows a slight improvement in predictions compared to the previous model update, iMK1208 (8.2% average error for iMK1208 predictions versus 7.0% with iAA1259). This first validation confirms that the predictive performances of the updated model iAA1259 are at least as good as the previous models generations. However, the next validation step requires more complex and quantitative datasets. The data used as constraints and the predicted growth rates data for the different models are available in Additional file 1: Table S7.

A more substantial improvement in prediction quality is observed when comparing the dynamic growth predictions of the metabolic models iAA1259, iMK1208, and iMA789, to published experimental growth data (Fig. 2) [57]. The dynamic growth was predicted by applying dynamic constraints from fermenter data (see Methods for details). The comparison of the predicted and experimental dynamic cell growth shows a significant improvement in quantitative and qualitative biomass prediction using the updated model iAA1259 (moving from an average absolute error of 37.6% with iMK1208 predictions to 5.3% with iAA1259; Fig. 2,
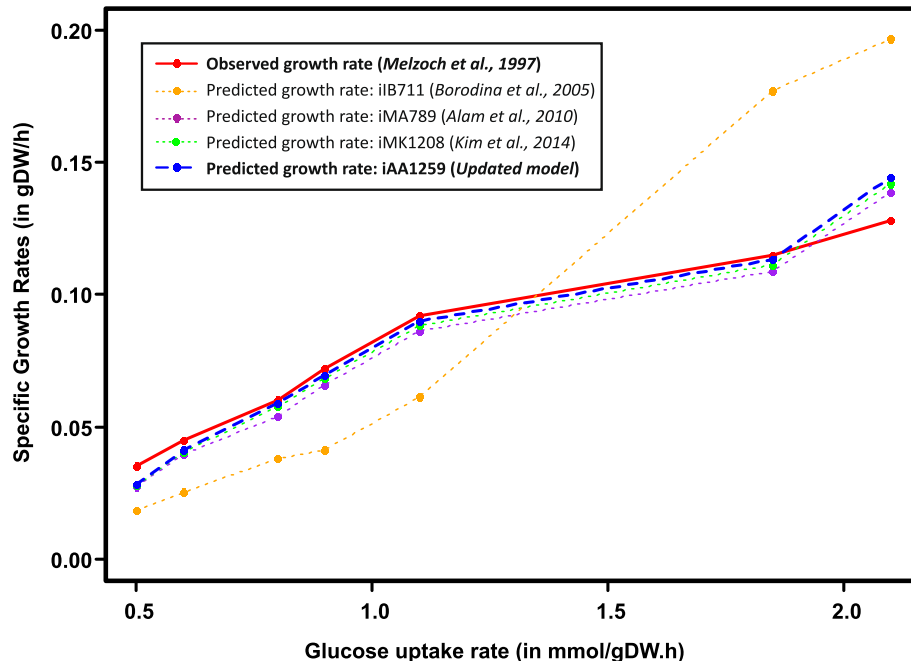
Fig. 1 Initial model growth predictions validation. Comparison of the specific growth rate predicted *in silico* with different models to the measured growth rates in chemostat data [55] with a glucose-limited minimal defined media. The published data on the rate of glucose uptake, oxygen consumption, CO_2 production and γ-actinorhodin production for seven different conditions were used as metabolic constraints in the different models. Growth prediction by iAA1259 shows a slight improvement compared to its immediate predecessor, iMK1208 [22]
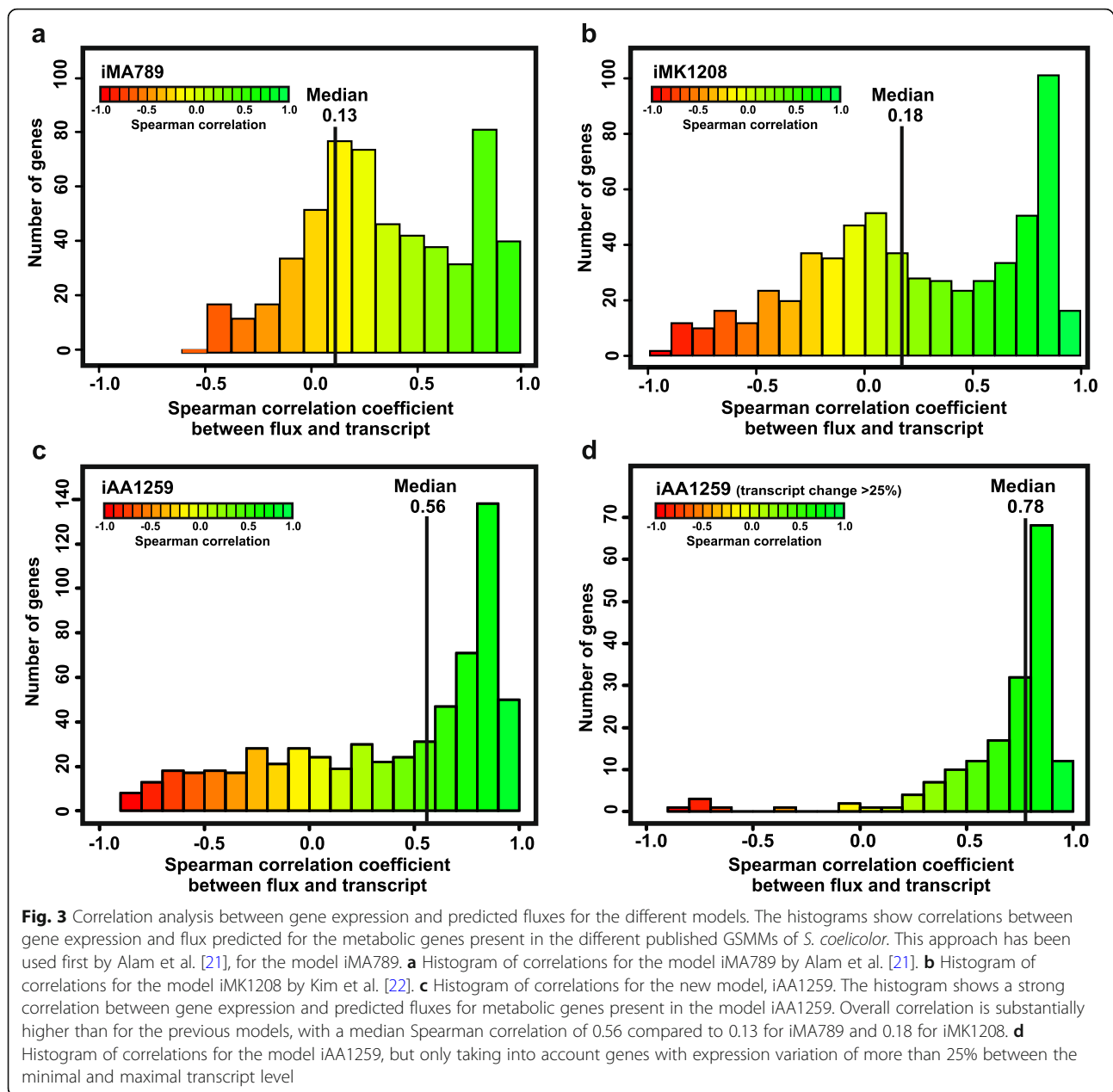
Amara *et al. BMC Genomics* (2018) 19:519

Page 4 of 9
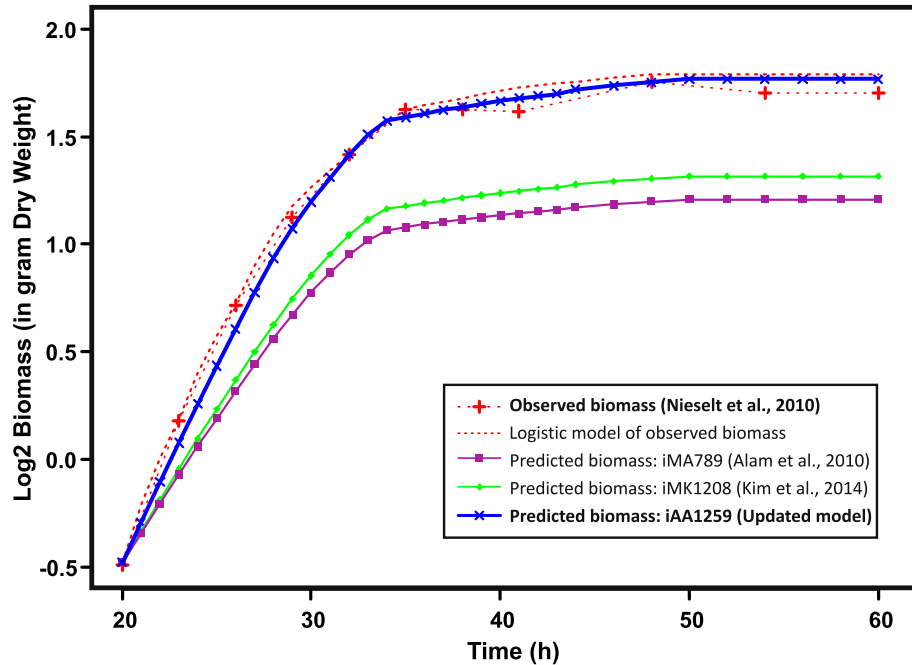
Fig. 2 Comparison of dynamic cell growth predictions. The quantitative in silico growth predictions are compared to measured biomass and predictions with previously published models across time points. The models were constrained using phosphate, glucose, and glutamate consumption, as well as production of the antibiotics actinorhodin and undecylprodigiosin measured in a fermenter experiment [57]. The updated model's predictions are closer to experimental observations than those of previous models, showing a significant improvement in growth prediction with the iAA1259 model

and Additional file 1: Figure S2). This improvement in biomass predictions is most likely due to the update of the biomass reaction and in the oxidative phosphorylation-related reactions updates (i.e., cytochrome oxidases and/or menaquinone pathway), as these are the main adjustments affecting biomass-related reactions directly.

The next validation step involved individual metabolic flux predictions across the model. For this purpose, the models were constrained with time series fermenter data for glucose and O_2 uptake rates and the production rates of γ-actinorhodin and CO_2 for 32 time points from Nieselt et al. using the method introduced by Alam et al. [21]. The metabolic flux predictions were compared to time series of gene-and protein expression reported by Lahtvee and colleagues [58], as proxies for the relative metabolic flux across the time course. For the majority of genes, the gene expression changes over time are strongly correlated to the predicted metabolic fluxes through the associated reactions (Fig. 3), and the correlation is substantially improved in the updated model presented here (median Spearman correlation coefficient 0.56, compared to 0.18 in the most recent predecessor, iMK1208). When focusing only on the correlation for genes that change at least 25% in expression across the time course (Fig. 3d), i.e. those genes that should show correlation, the quality of the correlation is even more

pronounced (the Pearson correlation coefficient increases by 39% from 0.56 to 0.78), and it becomes clear that only a very small number of genes show anti-correlated behaviour, i.e. a strong disagreement between gene expression and predicted fluxes. A similar trend is observed when applied to the fluxes predicted with iMA789 and iMK1208 models, both showing an increase of overall Pearson correlation from 0.13 to 0.38, and from 0.18 to 0.56, respectively (Additional file 1: Figure S1). The trend of a progressive increase in predictive power is still observed from iMA789 to iMK1208 (47% increase in correlation), and from iMK1208 to iAA1259 (42% increase in correlation).

In the updated model, there are two major genetic features showing anti-correlation (Fig. 4, row a), the genes associated to Calcium Dependant Antibiotics (CDA) biosynthesis, and the *nuo* operon genes associated to an NADH dehydrogenase (complex I). The CDA genes (Fig. 4, row e) are anti-correlated because their gene expression unexpectedly increases during the transition phase (Fig. 4, row c), whereas the model does not produce CDA. The metabolite production was not switched on in the model, as there is no calcium in the media conditions used by Nieselt et al. [57]. It has been shown previously that CDA could not be detected at significant levels if there was no calcium in the media [59]. Furthermore, production of the associated proteins is not

Amara *et al. BMC Genomics* (2018) 19:519

Page 5 of 9

Fig. 3 Correlation analysis between gene expression and predicted fluxes for the different models. The histograms show correlations between gene expression and flux predicted for the metabolic genes present in the different published GSMMs of *S. coelicolor*. This approach has been used first by Alam et al. [21], for the model iMA789. **a** Histogram of correlations for the model iMA789 by Alam et al. [21]. **b** Histogram of correlations for the model iMK1208 by Kim et al. [22]. **c** Histogram of correlations for the new model, iAA1259. The histogram shows a strong correlation between gene expression and predicted fluxes for metabolic genes present in the model iAA1259. Overall correlation is substantially higher than for the previous models, with a median Spearman correlation of 0.56 compared to 0.13 for iMA789 and 0.18 for iMK1208. **d** Histogram of correlations for the model iAA1259, but only taking into account genes with expression variation of more than 25% between the minimal and maximal transcript level

confirmed by the proteomics data [60] (Fig. 4, row b). Thus, in this case, the model prediction (no flux increase during the transition phase, Fig. 4 row d) appears to be correct, and gene expression in this exceptional case might not be correlating with metabolic flux. The Fig. 4 is available in high-definition as Additional file 5.

Regarding the second major anti-correlation, which is seen for the genes encoding the 14 subunits of NADH dehydrogenase I (NDH-I), the nuo complex, the disagreement between gene expression and predicted flux is due to a regulatory phenomenon, which in general is difficult to capture in a constraint-based model: two isoenzyme complexes are present in *S. coelicolor*, the relative expression of which is controlled by a regulatory loop dependent on NADH/NAD$^+$ ratio [61]. Of these, the *nuo* genes are preferentially expressed during fumarate respiration (stationary phase, [62]). So, while fluxes through the reaction catalysed by the NADH-dehydrogenase are reduced after the transition phase, *nuo* gene expression increases and results in an anti-correlation of *nuo* gene expression with the flux prediction (which does not distinguish between the isoenzyme complexes). The second NADH dehydrogenase, NDH-II, encoded by three copies of the *ndh* gene, is preferentially expressed during exponential phase and switched off after the transition phase; hence, the *ndh* genes show high correlation with the

Amara *et al. BMC Genomics* (2018) 19:519

Page 6 of 9

Fig. 4 Validation by integrated transcriptomics and proteomics analysis. Gene expression and proteomics data were mapped to metabolic enzyme-coding genes and the associated metabolic fluxes predicted over time. Overall, the predicted flux trends are strongly correlated (green colour in the top bar) to the observed gene expression trend across the metabolic switch event (around between 35 and 36 h). Two highly anti-correlated gene clusters are highlighted (red colour in the top bar). **a** Correlation: Pearson (P) and Spearman (S) correlation coefficient between the experimental gene expression level and the predicted fluxes through the corresponding reaction for each individual metabolic gene (green: good correlation; yellow: no correlation; red: anti-correlation). **b** Proteome: Protein abundance observed in experimental time course data: red: high: green: low abundance, black: missing data (only a small subset of enzymes was quantified). Proteomics data from Thomas, et al. [60]. **c** Gene expression: Gene expression levels observed in the same experimental time course (red: high, blue: low expression). A much larger number of time course were studied than in the proteomics analysis. Gene expression data from Nieselt et al. [57]. **d** Predicted flux: Flux predicted during a simulated time course (green: high; red: low predicted flux). **e** Genome features: Selected genomic regions discussed in the text are annotated. The data is ordered based on the position of analysed genes in the reference genome (from left to right, from 161,237 bp to 8,468,158 bp). Genome sequence from Bentley et al. [2]

predicted flux. While regulatory phenomena like this are not considered in this type of model, the misprediction highlights an interesting phenomenon for future study, i.e. the impact of the relative role of the two sources of reducing co-factors on secondary metabolism in *S. coelicolor*.

Conclusions

Here, we have presented an updated computational model of *S. coelicolor* primary and secondary metabolism, iAA1259; this model shows improved predictive abilities compared to previous model generations for metabolic changes at different scales, from overall biomass dynamics to fluxes through individual reactions.

Another important improvement is that the model has been also updated to enable integrative multi-omics data analysis, to be used for designing and debugging of engineered *Streptomyces* strains using a synthetic biology approach [6], and is now fully compliant with current modelling standards [52, 54].

The model presented here will be a good basis for the next round of computer-aided design of metabolically enhanced *Streptomyces* strains. The principled construction of the model using standard identifiers will facilitate the transfer of information to related strains beyond *S. coelicolor* (e.g., recently emerging popular biotechnological hosts, such as *Streptomyces albus* and *Streptomyces venezuelae* [63, 64]). It will also serve as a solid starting point for the next generation of updated metabolic models, which will address the challenge of including kinetic and regulatory constraints, in a similar way as the recently published genome-scale metabolic

models for the well-studied microorganisms *Escherichia coli* [65] and *Saccharomyces cerevisiae* [66].

Methods
Metabolic model reconstruction
The model reconstruction was initiated by updating the iMK1208 *S. coelicolor* model. The standard protocol for reconstruction of high-quality constraint-based GSMMs was followed when adding new genes, reactions, and metabolites [53].

In summary, the initial stoichiometric matrix was generated by comparing and using the iMK1208 model [22], KEGG [67], ScoCyc [37], and two automated reconstructions using RAST annotations and SEED reconstructions [68, 69]. The resulting matrix was manually curated for specific pathways (e.g., secondary metabolites biosynthesis, oxidative phosphorylation), to add or correct missing reactions, metabolites, genes associated, or reversibility constraints; this was supported by extensive literature survey to identify new knowledge or gaps in the previous model. Comparative analysis of transcriptomics data with iMK1208 helped to identify gene mis-annotations to be corrected [21, 57]. The biomass reaction was updated, as multiple reactions impacting biomass have been added (e.g., demethylmenaquinone, cytochrome oxidases or NADH dehydrogenase reactions). This was followed by a recalculation of the ATP fluxes for growth-associated and non-growth-associated maintenance using chemostat data [55], following the Varma & Palsson protocol [70]; the resulting values were very similar to those used in iMK1208 (with a GAM of 75.7 ATP in iMK1208 versus 75.79 ATP in

Amara *et al. BMC Genomics* (2018) 19:519

Page 7 of 9

iAA1259, and an NGAM of 2.65 in iMK1208 versus 2.64 in iAA1259). The detailed modifications on the biomass are available in the Additional file 2: Table S1.

Finally, multiple database identifiers were added either by automatic matching or by manual curation when necessary. The metabolites were annotated with multiple database identifiers; BiGG and KEGG identifications were already present in iMK1208, and other databases relevant to metabolomics data analysis were added: ChEBI, HMDB, CAS, IUPAC, ChemSpider, Metlin and PubChem identifications, wherever available. Chemical structure-related annotations (SMILES or InChi) were also introduced for all metabolites. Furthermore, when available, all reactions were annotated with EC code and CAS registry number, in addition to the BiGG annotation used in iMK1208. Additional gene annotations have been included to facilitate transcriptomics data integration with identifiers for Gene Ontology (GO), RefSeq, EMBL-ENA and Ensembl. For integrated proteomics analysis, annotations have been expanded to include identifiers for UniProt, Pfam and Panther, as well as data on protein length, mass, and amino acid sequence to support the direct mapping of mass-spectrometry-based proteomics data in the future. The expansion of these annotations also aims at helping fast reconstruction of metabolic models for other *Streptomyces* strains using comparative reconstruction and modelling methods. The final model has been named iAA1259 and is compliant with current metabolic model standards [52, 54]. The final model is available in SBML format and Excel format in Additional files 3 and 4.

Constraint-based modelling

The model was analysed by using Flux Balance Analysis (FBA) and parsimonious FBA (pFBA) to predict optimal in silico growth and metabolic flux using the COBRA toolbox in Matlab and Python [71, 72] and further evaluated using OptFlux and Sybil [73, 74]. To apply condition-specific constraints corresponding to the media composition, the uptake fluxes for exometabolites not available in the medium were set to zero, while all metabolic by-products were always allowed to leave the metabolic system. The measured nutrient uptake rates from the fermenter datasets are used to define constraints of the nutrient uptake for the model. The objective function maximized in the modelling was the growth rate (steady-state flux towards biomass).

Despite the fact that FBA is not a dynamic modelling approach (its basic assumption being a steady-state flux distribution), using dynamic constraints on CO_2, O_2, glucose, phosphate, and glutamate uptake based on fermenter time-course data [57] enabled simulation of the growth and metabolic dynamics across time. In order to simulate the production of the main antibiotics, the biomass composition was varied dynamically depending on the observed concentration of γ-Act and Red secondary metabolites in the cultures [21].

Transcriptomics and proteomics data analysis

Multiple omics data types have been used to validate the model; the proteomics data [60] have been acquired from the same time-series experiment samples as the flux constraints data and the transcriptomics data [57]. The transcriptomics and proteomics data were matched to corresponding metabolic genes associated with reactions by matching the StrepDB gene annotations. The matching procedure was similar to the one used for iMA789 [21]. Gene expression levels and predicted fluxes were compared using Pearson and Spearman correlations. The data used are available in the Additional file 6: Table S2.

Additional files

Additional file 1: Containing the **Table S3.** Summary table of the updates and new features added to the iAA1259 model compared to the previous generations. **Table S4.** Table of the new reactions added to the iAA1259 model. **Table S5.** Table of the new metabolites added to the iAA1259 model. **Table S6.** Table of the new genes added to the iAA1259 model. **Figure S1.** Correlation analysis between gene expression and predicted fluxes for iMA789 and iMK1208 (gene expression showing a variation superior to 25%). **Figure S2**. Mapping of observed metabolites in an untargeted metabolomics dataset onto the metabolic network**. Table S7.** Constraints used and predicted growth rates of the different models from the Fig. 1**. Figure S3.** Comparison of the normalized growth prediction of the metabolic models to the experimental data. (DOC 4121 kb)

Additional file 2: Table S1. Biomass modifications and recalculation of ATP consumption. (XLS 81 kb)

Additional file 3: iAA1259 metabolic model in SBML format. (XML 4249 kb)

Additional file 4: Excel file specifying metabolites, reactions, genes contained, and databases IDs present in the iAA1259 metabolic model. (XLS 2531 kb)

Additional file 5 High-resolution version of Fig. 4. Validation by integrated transcriptomics and proteomics analysis. (PDF 1881 kb)

Additional file 6 Table S2. Full detailed data used for the Fig. 4. (XLS 1690 kb)

Amara *et al. BMC Genomics* (2018) 19:519

Page 8 of 9

Availability of data and materials
The iAA1259 genome-scale metabolic model is available in SBML and Excel format: iAA1259.xml and iAA1259.xls.

Authors' contributions
AA reconstructed and validated the model, and analysed the omics data. AA, ET, and RB designed the experiments, interpreted the results, and wrote the manuscript. All authors have read, edited and approved the manuscript.

Ethics approval and consent to participate
No humans, animals, or plants were used in this study. The data used here were publicly available.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Hopwood DA. *Streptomyces* in nature and medicine: the antibiotic makers: Oxford University Press; 2007. p. 10–6.
2. Bentley SD, Chater KF, Cerdeno-Tarraga A-M, Challis GL, Thomson NR, James KD, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature. 2002;417:141–7.
3. van Keulen G, Dyson PJ. Production of specialized metabolites by *Streptomyces coelicolor* A3 (2). Adv Appl Microbiol. 2014;89:217–66.
4. Smanski MJ, Zhou H, Claesen J, Shen B, Fischbach MA, Voigt CA. Synthetic biology to access and expand nature's chemical diversity. Nat. Rev. Microbiol. 2016;14:135–49.
5. Medema MH, Breitling R, Bovenberg R, Takano E. Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. Nat Rev Microbiol. 2011;9:131–7.
6. Carbonell P, Currin A, Jervis AJ, Rattray NJW, Swainston N, Yan C, et al. Bioinformatics for the synthetic biology of natural products: integrating across the design–build–test cycle. Nat Prod Rep. 2016;33:925–32.
7. Lu W, Ye L, Lv X, Xie W, Gu J, Chen Z, et al. Identification and elimination of metabolic bottlenecks in the quinone modification pathway for enhanced coenzyme Q10 production in *Rhodobacter sphaeroides*. Metab Eng. 2015;29:208–16.
8. Gomez-Escribano JP, Bibb MJ. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. J Ind Microbiol Biotechnol. 2014;41:425–31.
9. Fong SS. Computational approaches to metabolic engineering utilizing systems biology and synthetic biology. Comput Struct Biotechnol J. 2014;11:28–34.
10. Simeonidis E, Price ND. Genome-scale modeling for metabolic engineering. J Ind Microbiol Biotechnol. 2015;42:327–38.
11. Kim B, Kim WJ, Kim DI, Lee SY. Applications of genome-scale metabolic network model in metabolic engineering. J Ind Microbiol Biotechnol. 2015; 42:339–48.
12. Xu C, Liu L, Zhang Z, Jin D, Qiu J, Chen M. Genome-scale metabolic model in guiding metabolic engineering of microbial improvement. Appl Microbiol Biotechnol. 2013;97:519–39.
13. Huang D, Li S, Xia M, Wen J, Jia X. Genome-scale metabolic network guided engineering of *Streptomyces tsukubaensis* for FK506 production improvement. Microb Cell Factories. 2013;12:52.
14. Wang J, Wang C, Song K, Wen J. Metabolic network model guided engineering ethylmalonyl-CoA pathway to improve ascomycin production in *Streptomyces hygroscopicus* var. ascomyceticus. Microb Cell Factories. 2017;16:169.
15. Kim M, Yi JS, Lakshmanan M, Lee D-Y, Kim B-G. Transcriptomics-based strain optimization tool for designing secondary metabolite overproducing strains of *Streptomyces coelicolor*. Biotechnol Bioeng. 2016;113:651–60.
16. Toro L, Pinilla L, Avignone-Rossa C, Ríos-Estepa R. An enhanced genome-scale metabolic reconstruction of *Streptomyces clavuligerus* identifies novel strain improvement strategies. Bioprocess Biosyst Eng. 2018;41:657–69.
17. Fondi M, Pinatel E, Talà A, Damiano F, Consolandi C, Mattorre B, et al. Time-resolved Transcriptomics and constraint-based modeling identify system-level metabolic features and overexpression targets to increase Spiramycin production in *Streptomyces ambofaciens*. Front Microbiol. 2017;8:835.
18. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox v2.0. Nat Protoc. 2011;6:1290–307.
19. Borodina I, Krabben P, Nielsen J. Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. Genome Res. 2005;15:820–9.
20. Borodina I, Siebring J, Zhang J, Smith CP, van Keulen G, Dijkhuizen L, et al. Antibiotic overproduction in *Streptomyces coelicolor* A3 (2) mediated by phosphofructokinase deletion. J Biol Chem. 2008;283:25186–99.
21. Alam MT, Merlo ME, Hodgson DA, Wellington EMH, Takano E, Breitling R. Metabolic modeling and analysis of the metabolic switch in *Streptomyces coelicolor*. BMC Genomics. 2010;11:1.
22. Kim M, Yi JS, Kim J, Kim JN, Kim MW, Kim BG. 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. 2014;9:1185–94.
23. Borodina I, Schöller C, Eliasson A, Nielsen J. Metabolic network analysis of *Streptomyces tenebrarius*, a *Streptomyces* species with an active Entner-Doudoroff pathway. Appl Environ Microbiol. 2005;71:2294–302.
24. Licona-Cassani C, Marcellin E, Quek L-E, Jacob S, Nielsen LK. Reconstruction of the *Saccharopolyspora erythraea* genome-scale model and its use for enhancing erythromycin production. Antonie Van Leeuwenhoek. 2012;102:493–502.
25. Zakrzewski P, Medema MH, Gevorgyan A, Kierzek AM, Breitling R, Takano E. MultiMetEval: comparative and multi-objective analysis of genome-scale metabolic models. PLoS One. 2012;7:e51511.
26. Pawlik K, Kotowska M, Chater KF, Kuczek K, Takano E. A cryptic type I polyketide synthase (*cpk*) gene cluster in *Streptomyces coelicolor* A3(2). Arch Microbiol. 2007;187:87–99.
27. Kotowska M, Ciekot J, Pawlik K. Type II thioesterase ScoT is required for coelimycin production by the modular polyketide synthase Cpk of *Streptomyces coelicolor* A3(2). Acta Biochim Pol. 2014;61:141–7.
28. Awodi UR, Ronan JL, Masschelein J, De Los Santos ELC, Challis GL. Thioester reduction and aldehyde transamination are universal steps in actinobacterial polyketide alkaloid biosynthesis. Chem Sci. 2017;8:411–5.
29. Gottelt M, Kol S, Gomez-Escribano JP, Bibb M, Takano E. Deletion of a regulatory gene within the *cpk* gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3 (2). Microbiology. 2010;156:2343–53.
30. Wentzel A, Bruheim P, Øverby A, Jakobsen ØM, Sletta H, Omara WAM, et al. Optimized submerged batch fermentation strategy for systems scale studies of metabolic switching in *Streptomyces coelicolor* A3(2). BMC Syst Biol. 2012;6:59.
31. Kato J, Funa N, Watanabe H, Ohnishi Y, Horinouchi S. Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in *Streptomyces*. Proc Natl Acad Sci U S A. 2007;104:2378–83.
32. Takano E. γ-Butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. Curr Opin Microbiol. 2006;9:287–94.
33. Biarnes-Carrera M, Breitling R, Takano E. Butyrolactone signalling circuits for synthetic biology. Curr Opin Chem Biol. 2015;28:91–8.
34. Biarnes-Carrera M, Lee C-K, Nihira T, Breitling R, Takano E. Orthogonal regulatory circuits for *Escherichia coli* based on the γ-butyrolactone system of *Streptomyces coelicolor*. ACS Synth Biol. 2018;7(4):1043–55.
35. Ogasawara Y, Kondo K, Ikeda A, Harada R, Dairi T. Identification of tirandamycins as specific inhibitors of the futalosine pathway. J Antibiot. 2017;70:798–800.
36. Zhi X-Y, Yao J-C, Tang S-K, Huang Y, Li H-W, Li W-J. The futalosine pathway played an important role in menaquinone biosynthesis during early prokaryote evolution. Genome Biol Evol. 2014;6:149–60.
37. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 2016;44:D471–80.
38. Esnault C, Dulermo T, Smirnov A, Askora A, David M, Deniset-Besseau A, et al. Strong antibiotic production is correlated with highly active oxidative metabolism in *Streptomyces coelicolor* M145. Sci Rep. 2017;7:200.
39. Kavvas ES, Seif Y, Yurkovich JT, Norsigian C, Poudel S, Greenwald WW, et al. Updated and standardized genome-scale reconstruction of *Mycobacterium*

Amara *et al. BMC Genomics* (2018) 19:519

Page 9 of 9

tuberculosis H37Rv, iEK1011, simulates flux states indicative of physiological conditions. BMC Syst Biol. 2018;12:25.

40. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem substance and compound databases. Nucleic Acids Res. 2016; 44:D1202–13.

41. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, et al. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 2007;36:D344–50.

42. Heller S, McNaught A, Stein S, Tchekhovskoi D, Pletnev I. InChI - the worldwide chemical structure identifier standard. J Cheminform. 2013;5:7.

43. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Model. 1988;28:31–6.

44. Scheltema RA, Jankevics A, Jansen RC, Swertz MA, Breitling R. PeakML/ mzMatch: a file format, java library, R library, and tool-chain for mass spectrometry data analysis. Anal Chem. 2011;83:2786–93.

45. Jankevics A, Merlo ME, de Vries M, Vonk RJ, Takano E, Breitling R. Metabolomic analysis of a synthetic metabolic switch in *Streptomyces coelicolor* A3 (2). Proteomics. 2011;11:4622–31.

46. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

47. Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, et al. Ensembl genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Res. 2018;46:D802–8.

48. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2007;35:D61–5.

49. The UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43:D204–12.

50. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, et al. The Pfam protein families database. Nucleic Acids Res. 2004;32:138D–141.

51. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. 2012;41:D377–86.

52. Le Novere N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, et al. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat Biotechnol. 2005;23:1509–15.

53. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 2010;5:93–121.

54. King ZA, Lu J, Drager A, Miller P, Federowicz S, Lerman JA, et al. BiGG models: a platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 2016;44:D515–22.

55. Melzoch K, De Mattos MJT, Neijssel OM. Production of actinorhodin by *Streptomyces coelicolor* A3(2) grown in chemostat culture. Biotechnol Bioeng. 1997;54:577–82.

56. Bystrykh LV, Fernández-Moreno MA, Herrema JK, Malpartida F, Hopwood DA, Dijkhuizen L. Production of actinorhodin-related' blue pigments' by *Streptomyces coelicolor* A3 (2). J Bacteriol. 1996;178:2238–44.

57. Nieselt K, Battke F, Herbig A, Bruheim P, Wentzel A, Jakobsen ØM, et al. The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. BMC Genomics. 2010;11:10.

58. Lahtvee P-J, Sánchez BJ, Smialowska A, Kasvandik S, Elsemman IE, Gatto F, et al. Absolute quantification of protein and mRNA abundances demonstrate variability in gene-specific translation efficiency in yeast. Cell Syst. 2017;4:495–504.e5.

59. Diez V, Loznik M, Taylor S, Winn M, Rattray NJW, Podmore H, et al. Functional exchangeability of oxidase and dehydrogenase reactions in the biosynthesis of hydroxyphenylglycine, a nonribosomal peptide building block. ACS Synth Biol. 2015;4:796–807.

60. Thomas L, Hodgson DA, Wentzel A, Nieselt K, Ellingsen TE, Moore J, et al. Metabolic switches and adaptations deduced from the proteomes of *Streptomyces coelicolor* wild type and phoP mutant grown in batch culture. Mol Cell Proteomics. 2012;11:M111.013797.

61. Gyan S, Shiohira Y, Sato I, Takeuchi M, Sato T. Regulatory loop between redox sensing of the NADH/NAD(+) ratio by Rex (YdiH) and oxidation of NADH by NADH dehydrogenase Ndh in *Bacillus subtilis*. J Bacteriol. 2006; 188:7062–71.

62. Brekasis D, Paget MSB. A novel sensor of NADH/NAD+ redox poise in *Streptomyces coelicolor* A3(2). EMBO J. 2003;22:4856–65.

63. Zaburannyi N, Rabyk M, Ostash B, Fedorenko V, Luzhetskyy A. Insights into naturally minimised *Streptomyces albus* J1074 genome. BMC Genomics. 2014;15:97.

64. Phelan RM, Sachs D, Petkiewicz SJ, Barajas JF, Blake-Hedges JM, Thompson MG, et al. Development of next generation synthetic biology tools for use in *Streptomyces venezuelae*. ACS Synth Biol. 2017;6:159–66.

65. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. Nat Biotechnol. 2017; 35:904–8.

66. Sánchez BJ, Zhang C, Nilsson A, Lahtvee P-J, Kerkhoven EJ, Nielsen J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol Syst Biol. 2017;13:935.

67. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.

68. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol. 2010;28:977–82.

69. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. 2014;42:D206–14.

70. Varma A, Palsson BØ. Parametric sensitivity of stoichiometric flux balance models applied to wild-type *Escherichia coli* metabolism. Biotechnol Bioeng. 1995;45:69–79.

71. Ebrahim A, Lerman JA, Palsson BØ, Hyduke DR. COBRApy: COnstraints-based reconstruction and analysis for Python. BMC Syst Biol. 2013;7:74.

72. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. Nat Protoc. 2007;2:727–38.

73. Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, et al. OptFlux: an open-source software platform for in silico metabolic engineering. BMC Syst Biol. 2010;4:45.

74. Gelius-Dietrich G, Desouki A, Fritzemeier C, Lercher MJ. Sybil – efficient constraint-based modelling in R. BMC Syst Biol. 2013;7:125.