Biotechnology Journal

Supporting Information for DOI 10.1002/biot.201300242

iOD907, the first genome-scale metabolic model for the milk yeast *Kluyveromyces lactis*

Oscar Dias, Rui Pereira, Andreas K. Gombert, Eugénio C. Ferreira, Isabel Rocha

1.1 MODEL CURATION PROTOCOL

A first curation of the model was performed, to check for reactions with wrong compartmentation information. For instance, KEGG reactions R00124 (ATP + ADP <=> ADP + ATP) and R07480 (N-(4-Aminobutylidene)-[eIF5A-precursor]-lysine + [eIF5A-precursor]-lysine <=> N-(4-Aminobutylidene)-[eIF5A-precursor]-lysine + [eIF5A-precursor]-lysine) only make sense if the reactants and products are in different compartments. Moreover, reactions promoted by enzymes assigned by WoLF PSORT to the plasma membrane were verified to determine the likelihood of those reactions taking place outside the cell.

After the first curation step, the size of the model was decreased by automatically eliminating reactions, either biochemical or transport, with dead end metabolites, from the model. That is, if in a given reaction a reactant was neither produced by another reaction nor transported from another compartment, or a product was neither being consumed by another reaction nor transported to other compartments, then such reaction was automatically excluded from the model. This step was performed recursively, because removing a reaction can render other metabolites as dead ends.

Then, the model was checked for unbalanced reactions. KEGG has several reactions unbalanced at the stoichiometric level, namely reactions involving glycans. KEGG reactions include a field labelled 'number of chains' to balance those reactions, where adding a monomer to a metabolite having *n* chains will create an oligomer with *n+1* chains. However, metabolic models only use the stoichiometry; thus, these reactions were simplified by setting *n*=0. For instance, the equation for KEGG's reaction R00887 is GDP-mannose + Mannan(n) <=> GDP + Mannan(n+1). In our model, *n* was replaced by zero, turning it into GDP-mannose <=> GDP + Mannan.

Also, several reactions were unbalanced at the protons level. Reactions such as R00137 or R02744 had more protons being produced than consumed. This step was very important, as an unbalanced model would not provide feasible predictions.

Next, the connection between reactions in distinct compartments, the duplication of various reactions in different compartments and the direction of irreversible reactions (onward or backward) was verified manually.

The next verification steps implied performing simulations and, for that purpose, the model was exported from *merlin* in the SBML format, and imported into OptFlux [1], so that simulations could be performed. When a metabolic model is imported into OptFlux, boundaries are automatically created for each reaction. Moreover, drains (exchange reactions that allow performing simulations) are also created for all external metabolites. More details on this formulation can be found in Rocha *et al*. [1]. The method used to perform the simulations was the Flux Balance Analysis. More information on this methodology is provided for example on [2].

The main purpose of these initial simulations was to guarantee biomass formation and flux over important pathways such as the oxidative phosphorylation, central carbon metabolism, the fatty acids biosynthesis or the glycolipids metabolism. The medium used in those simulations was adapted from the well–known defined medium, the synthetic Verduyn medium [3] supplemented with 5 fold the nicotinate contents [4], as shown in Table S7 of the supporting information. All boundaries for the uptake of the metabolites provided by the environmental conditions described in Table S7, of the supporting information, were unrestricted, except alpha-D-Glucose that was constrained to 1 $mmol.g^{-1}.h^{-1}$ for simulating a glucose limited chemostat. During this process, several gaps were identified and corrected by adding new reactions to the model. These gaps in the network could exist due to several reasons. The enzymes promoting those

reactions could have been assigned to the wrong compartments. If that was the case, the reactions were either relocated to the correct compartment, or duplicated to the compartments of interest if the enzyme is hypothesised to be active in several locations. Instead, those gaps could be related with the removal, in a previous step, of reactions from the model, when the metabolites involved in such reactions were unconnected from the remaining network. However, when these reactions were found to be relevant, they were re-included in the model, and the dead end metabolites connections verified. When a previously unidentified reaction for *K. lactis* was the only option to solve a gap, that reaction was added to the model and genomic evidence was sought in the discarded annotation results.

On the other hand, some reactions were incorrectly included in the network. Several enzymes catalyze reactions that accept different cofactors in different organisms, and in KEGG these reactions are usually replicated for each cofactor. Therefore, some of these reactions had to be removed from the model, and the iMM904 model was used to determine which cofactors are usually used by yeasts. Also, some enzymes were assigned to several compartments, although they were only relevant in one of them. In these cases, only one location was kept.

Subsequently, the model was again trimmed, yet now only for transport reactions with unconnected metabolites, as the removal or re-location of reactions may disconnect transport reactions from the network. Once the connectivity to the biomass precursors was also re-verified, the model was tested by simulating growth using the environmental conditions presented in Table S7, of the supporting information.

Throughout this phase, reactions were edited, manually added to, or removed from the model. The manual inspection of the fluxes associated to reactions involved in the formation of the biomass precursors exposed some gaps in the network. Whenever a gap was found in the model, reactions were sought to fill that gap. The KEGG pathways and the iMM904 model were used as standards. If the model lacked a reaction in a KEGG pathway, this gap was analyzed to search for additional GPR evidence and the reaction was added to the network. If the gap was outside a KEGG pathway, gap filling reactions identified within the iMM904 model were sought in KEGG. If the reaction was not available in KEGG it was manually created and added to the network. In either case, the model was improved with the new reaction and the gap was filled.

Other reactions were added whenever needed, such as the fructose 6-phosphate conversions reaction. This reaction is an artefact of the model, used to connect two KEGG pathways, as the product of the first pathway would be the specific beta-D-fructose 6-phosphate and the reactant od the second pathway was the generic fructose 6-phosphate.

This process was repeated several times, according to Figure 1, until the in *silico* results replicated the *in vivo* data. The model was tested by simulating growth using the environmental conditions presented in Table S7 (additional file 2 of the supporting information). This methodology was implemented using *merlin 2.0* for the reconstruction process and OptFlux 3.0 [47] for the validation of the model. All predictions were performed using the IBM® CPLEX solver.

This methodology was implemented using *merlin* for the reconstruction process and OptFlux 3.0 [1] for the validation of the model. All predictions were performed using the IBM® CPLEX solver.

1.2 CARBON SOURCES ASSESSMENT

As shown in Table S12, consensus between the 2 data sources and the *in silico* predictions (iOD907) was only attained for 9 carbon sources (establishing growth in 6 and not growing in 3) out of a total of 21 carbon sources tested in all datasets.

The iOD907 predictions agreed with CBS-KNAW in 10 carbon sources as show in Table S12, 8 of which indicated no growth on the Biolog experiments (surprisingly, this list includes one of the carbon sources in which *K. lactis* is known to thrive, unlike other yeasts like *S. cerevisiae*, D-galactose) and 2 carbon sources that were not tested by Biolog.

On the other hand, as show in Table S12, the iOD907 model matched growth with the Biolog experiments for 5 other carbon sources in which CBK-KNAW did not establish growth (2) or for which the carbon sources were not tested (3). As well, there is 1 carbon source which is reported by CBS-KNAW to be a viable carbon source for which growth could not be identified in both the *in silic*o strain and the Biolog experiments (L-lysine).

Surprisingly, the *in silico* strain was able to use D-gluconate as sole carbon source for growth, despite the fact that both Biolog and CBS suggest no growth in *in vivo* experiments.

Although the *in silico* strain predictions agreed with Biolog when assessing no growth for 19 metabolites not tested in CBS, they were different for 17 other metabolites also not tested by CBS. The latter set comprises carbon sources that provide viable, *in silico*, growth predictions. Most of these metabolites are amino acids, except for D-mannose, acetate, malate, fumarate, putrescine and uridine.

Finally, there are 80 carbon sources in Table S12, in which either the Biolog experiments or the CBS-KNAW catalogue assessed growth for *K. lactis* but which were not present in the model, and 13 carbon sources that, although being present in the model, do not have transport reactions in the *in silico* strain. Moreover, Biolog and CBS tested 47 carbon sources which could not be identified in the KEGG compound database. Therefore, since the iOD907 model only has metabolites with cross references to KEGG, such carbon sources were not present in the model and their viability as carbon sources for *K. lactis* growth could not be tested *in silico*.

According to Table S12 (additional file 2 of the supporting information), there were 13 carbon sources for which a transporter was missing in the model. CBS-KNAW only tested cellobiose (positive growth). Biolog tested all 13 carbon sources but it only established growth for dihydroxyacetone, deoxyribose and oxalomalic acid. All other carbon sources did not attain growth (including cellobiose). When a transport reaction for each of these carbon sources is added to the model *in silico* growth was attained with the four abovementioned carbon sources and for glyoxylic acid, D-fructose 6-phosphate, D-glucose 6-phosphate and D-glucose 1-phosphate. These transport reactions were not added to the model due to contradictory (cellobiose) or missing experimental evidences.

It should be noticed that the ability to grow in a particular carbon source is often dependent, not only on the enzymatic and transport capabilities of the organism, but also on the presence or absence of other medium components. A possible explanation for the discrepancies found between the experiments is the different setups of each test, namely the absence or presence of other nutrients.

In conclusion, the intersection of the results of the in vivo growth tests performed by Biolog and CBS-KNAW, and the in silico simulations are fairly positive since it encompasses 9 out of 21 possible carbon sources. Within these carbon sources, there was inconsistency between the two in vivo data sources and the model predictions in 1 case, an agreement between in silico

and Biolog in 3 cases and between in silico and CBS-KNAW for 8 times. All the carbon sources for which in vivo tests gave inconsistencies should thus be double checked to confirm or refute the in silico prediction.

1.3 ADDITIONAL FIGURES

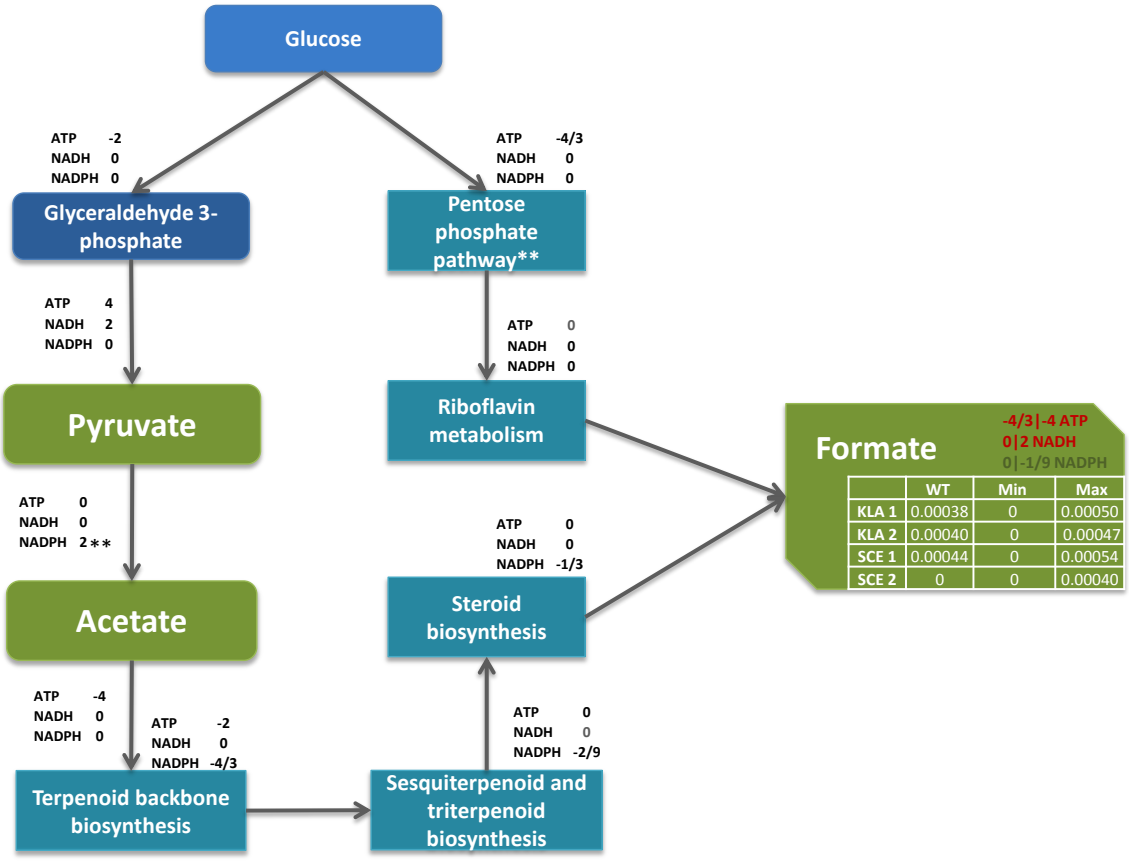

Figure S1. Analysis of the energy balances involved in the formation of several by-products (namely formate). *K. lactis* wild type simulation 1 (KLA1) and *S. cerevisiae* wild type simulation 1 (SCE1) were obtained using ENV1 environmental conditions. Likewise *K. lactis* wild type simulation 2 (KLA2) and *S. cerevisiae* wild type simulation 2 (SCE2) use ENV2. Env 1: qO2 = 1.2 mmol.g-1.h-1, qglucose = 2.49 mmol.g-1.h-1; Env 2: qO2 = 1.7 mmol.g-1.h-1, qglucose = 2.045 mmol.g-1.h-1.
The formate energetic and production yields are provided by several pathways that have origin in glycolysis. These pathways include the non-oxidative part of the pentose phosphate pathway (PPP), the riboflavin metabolism, the steroid biosynthesis and the terpenoid, sesquiterpenoid and triterpenoid biosynthesis. (* - assuming a NADPH-dependent acetaldehyde dehydrogenase, similar to *S. cerevisiae*' ALD6 [74], in *K. lactis*, ** - non-oxidative branch of the Pentose Phosphate Pathway).

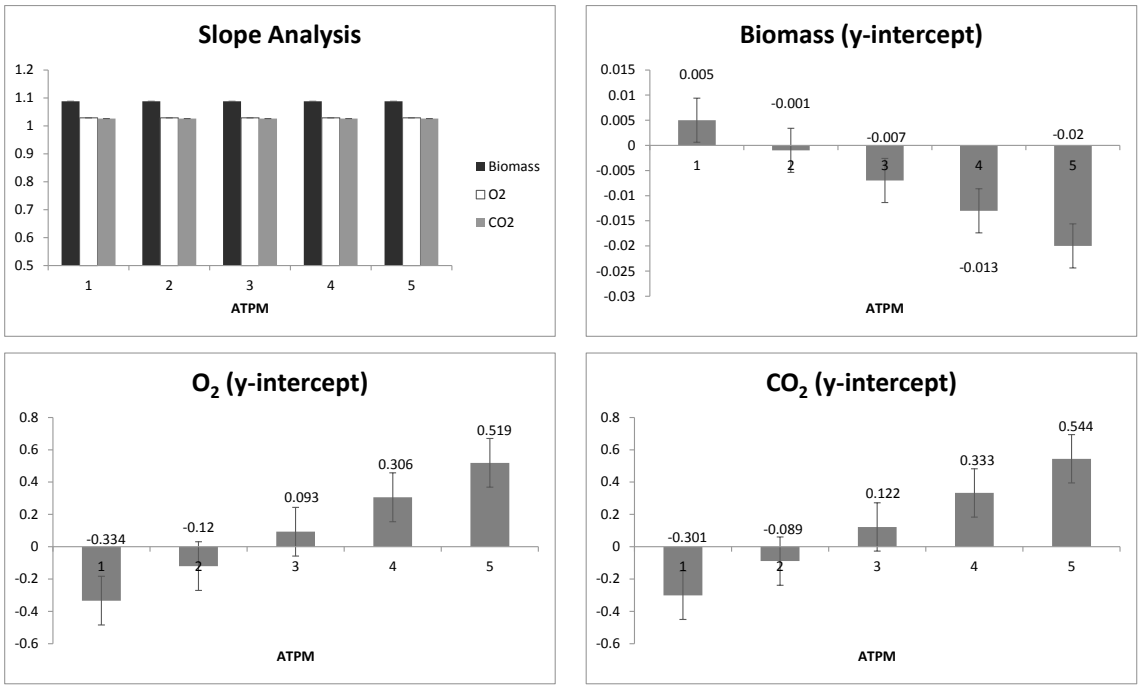

Figure S2. Analysis of the linear regression between the in vivo values (Kiers et al.[47]) and the prediction values from iOD907 in terms of growth, oxygen, and carbon dioxide yields, shown in Table S10.
The vertical bars on the columns represent the standard error, automatically calculated considering all glucose uptake rates. ATPM – Maintenance ATP flux (mmol.gDW-1.h-1). This figure shows that the slopes of the linear regressions, between the in vivo data and the predictions (biomass O2 and CO2 yields) provided by simulations under the in vivo growth environmental conditions remain constant for all ATP maintenance values tested. However, the analysis of the linear regressions' y-intercept values provides a different scenario. Setting the ATP used for maintenance to 2 mmol.gDW-1.h-1 clearly provides the best fitting for the biomass yield and the CO2 production. Moreover, the best fitting for O2 consumption is obtained when setting the ATP flux to 3 mmol.gDW-1.h-1. Therefore, the analysis of the results suggests that the ATP used for maintenance should be set to 2 mmol.gDW-1.h-1, as this value provides the best overall fitting to the in vivo data on all the predictions.

1.4 ADDITIONAL REFERENCES

[1] Rocha, I., Maia, P., Evangelista, P., Vilaça, P., et al., OptFlux: an open-source software platform for in silico metabolic engineering., *BMC Syst. Biol.*, 2010, 45, DOI: 10.1186/1752-0509-4-45.

[2] Dias, O. and Rocha, I., Systems Biology in Fungi, in: Paterson, R. (Ed.) *Molecular Biology of Food and Water Borne Mycotoxigenic and Mycotic Fungi*, 2013.

[3] Verduyn, C., Postma, E., Scheffers, W. A., and Van Dijken, J. P., Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation., *Yeast*, 1992, 501–17, DOI: 10.1002/yea.320080703.

[4] Kiers, J., Zeeman, A. M., Luttik, M., Thiele, C., et al., Regulation of alcoholic fermentation in batch and chemostat cultures of Kluyveromyces lactis CBS 2359., *Yeast*, 1998, 459–69, DOI: 10.1002/(SICI)1097-0061(19980330)14:5<459::AID-YEA248>3.0.CO;2-O.