Comprehensive Reconstruction and In Silico Analysis of *A e .. e* Genome-Scale Metabolic Network Model That Accounts for 1210 ORFs

Hongzhong Lu,[1] Weiqiang Cao,[1] Liming Ouyang,[1] Jianye Xia,[1] Mingzhi Huang,[1] Ju Chu,[1] Yingping Zhuang,[1] Siliang Zhang,[1] Henk Noorman[2]

[1] B , C , 200237, . . C ; : +86-021-64253522; : +86-021-64253702; - : . @ . . ; : +86-021-64253021; : +86-021-64253702; - : @ . .

[2] B C , ,

ABSTRACT: *Aspergillus niger* is one of the most important cell factories for industrial enzymes and organic acids production. A comprehensive genome-scale metabolic network model (GSMM) with high quality is crucial for efficient strain improvement and process optimization. The lack of accurate reaction equations and gene-protein-reaction associations (GPRs) in the current best model of *A. niger* named GSMM iMA871, however, limits its application scope. To overcome these limitations, we updated the *A. niger* GSMM by combining the latest genome annotation and literature mining technology. Compared with iMA871, the number of reactions in iHL1210 was increased from 1,380 to 1,764, and the number of unique ORFs from 871 to 1,210. With the aid of our transcriptomics analysis, the existence of 63% ORFs and 68% reactions in iHL1210 can be verified when glucose was used as the only carbon source. Physiological data from chemostat cultivations, ^{13}C-labeled and molecular experiments from the published literature were further used to check the performance of iHL1210. The average correlation coefficients between the predicted fluxes and estimated fluxes from ^{13}C-labeling data were sufficiently high (above 0.89) and the prediction of cell growth on most of the reported carbon and nitrogen sources was consistent. Using the updated genome-scale model, we evaluated gene essentiality on synthetic and yeast extract medium, as well as the effects of NADPH supply on glucoamylase production in *A. niger*. In summary, the new *A. niger* GSMM iHL1210 contains significant improvements with respect to the metabolic coverage and prediction performance, which paves the way for systematic metabolic engineering of *A. niger*.

Biotechnol. Bioeng. 2017;114: 685–695.

KEYWORDS: *Aspergillus niger*; genome-scale metabolic model; multi-omics; glucoamylase

Aspergillus niger is widely used in production of enzymes, heterologous proteins, and organic acids, which plays an important role in industrial areas of food, energy, feed addition, etc. (Knuf and Nielsen, 2012; Show et al., 2015; Ward, 2012). In 2007, the whole genome sequence of *A. niger* was published (Pel et al., 2007), which

^{13}C metabolic flux analysis. Using the new model, the secretion of by-products, the essential genes for growth on different media, and the effects of different NADPH sources were successfully predicted. Overall, the expanded version of *A. niger* GSMM could better describe the cell metabolic characteristics, which makes it more suitable to serve as a good platform for multi-omics integration and systematic metabolic engineering of *A. niger* microbial cell factories.

Materials and Methods

Procedure for Model Reconstructions

The procedure for GSMM reconstructions could be divided into three phases. Firstly, considering that the imbalance of mass and electrical charge in most reactions of iMA871, we re-annotated the information of all metabolites in the original model referring to the databases of KEGG and PubChem, to ensure that each metabolite has a correct element structure and electrical charge at pH 7.2 (Thiele and Palsson, 2010). Then, each reaction in the model was manually checked and corrected, thus making the published GSMM model compliant with mass and charge balance constraints.

Subsequently, to expand the *A. niger* GSMM in a systematic way, the latest genome annotation information from mainly four databases (KEGG, UniprotKB, AspGD, IMG) was gathered and summarized. To establish the GPRs, the genes which were annotated as enzyme and transport proteins were selected from each database. With the aid of the KEGG database, the relations of proteins and reactions were further checked to safeguard the high quality of GPRs in the new *A. niger* GSMM (Thiele and Palsson, 2010). The addition of each new reaction into the original model followed the standard protocol (Thiele and Palsson, 2010) to ensure the quality of the final model. Next, the Gapfind function of the Cobra Toolbox v2.0 (Schellenberger et al., 2011, La Jolla, CA) was adopted to identify all existing dead-end metabolites, and the essential reactions from KEGG were added to minimize the number of dead-end metabolites and improve the predictability of this updated *A. niger* metabolic network model (iHL1210).

Lastly, the cell composition and the energy parameters were updated. Compared with the original model (Andersen et al., 2008), some additional cell components, like the intracellular free 20 essential amino acids and the cofactors including biotin, folate, heme, thiamin diphosphate, etc., were added to ensure that the cell composition was more comprehensive and closer to the actual values.

Constraint-Based Modeling

The constraint-based flux balanced analysis (FBA) is widely used in model reconstruction and prediction of strain phenotypes (Orth et al., 2010). Generally, FBA was used to predict the flux distribution based on the linear optimization of an objective function (typically the rate of biomass production). The constraints used in FBA (Bordbar et al., 2014) include the balance of intracellular metabolites Equation (2), the reversibility of reactions, the maximum enzyme capacity, as well as the exchange rate Equation (3). The optimized objective function could be set as the maximization of cell growth Equation (1) or optimization of energy utilization (Schuetz et al., 2007). The FBA was conducted using the Cobra Toolbox v2.0 and the Gurobi 5 linear optimization algorithm based on Matlab.

$$\text{Objective}: \ \max/\min Z = C^{\mathrm{T}} \times v \tag{1}$$

$$\text{Constrains}: \ S \times v = 0 \tag{2}$$

$$\mathrm{lb} \leq v \leq \mathrm{ub} \tag{3}$$

where S is a sparse matrix of $m \times n$, m is the number of metabolites, n is the number of reactions, v represents the rate of all reactions, lb and ub, respectively, defined the lower and upper bounds of each reaction. In Formula 1, C^{T} represents the coefficient of each metabolite in the objective function.

In this study, the matrix S of the constraint-based model iHL1210 Equation (2) is composed of 1,254 rows and 1,764 columns. The number 1,254 represents the total number of metabolites in four compartments including the extracellular, cytoplasmic, mitochondrial, and peroxisomal spaces. The number 1,764 represents the total number of reactions including the exchange, transport, and metabolic reactions. During the simulation with iHL1210, the non-growth-associated maintenance (NGAM) and growth-associated maintenance (GAM) were set as 3.73 mmol ATP/gDCW.h and 61 mmolATP/gDCW (Lu et al., 2015), respectively. The lower bound of exchange reactions except for glucose, oxygen, ammonia, sulfur, phosphorus was set as 0. To predict the specific growth rate (μ), the exchange rate of glucose was set at -0.52, -0.93, -1.36, and -1.93 mmol/gDCW.h, respectively (Lameiras et al., 2015). For flux prediction, the parsimonious FBA (pFBA) (Schellenberger et al., 2011) was adopted, which optimizes the objective function at first and then minimizes the flux through the model. For in silico strain design, the minimization of metabolic adjustment (MOMA) method was exploited to simulate the effects of gene deletion or overexpression on cell growth of *A. niger* or citric acid production using *A. niger*. The detailed procedure could be found in the literatures (Boghigian et al., 2012; Segre et al., 2002).

Sensitivity Analysis

During the sensitivity analysis with iHL1210, the qs was changed between 0 and 1.8 mmol glucose/gDCW.h. According to batch cultivations, the maximum q_{O2} was set as 3.5 mmol O2/gDCW.h. For each simulation, only one of the following model parameters was changed, including the protein composition (26–40%), the RNA composition (1.8–5%), the lipid composition (10.9–20%), GAM (61 mmol ATP/gDCW $\pm$50%), NGAM (3.73 mmol ATP/gDCW.h $\pm$50%), and P/O (2.64 for mitochondrial NADH, $\pm$50%). The specific growth rate and oxygen uptake rate were calculated to evaluate the effects of each parameter change on the model prediction.

Prediction of Possible Growth-Supporting Carbon and Nitrogen Sources

To investigate the prediction capability of iHL1210, the reported 69 carbon and 30 nitrogen sources relating to *A. niger* were collected.

Then FBA was used to analyze the growth conditions on each carbon or nitrogen source. For prediction of carbon utilization, only $NH_3 \cdot H_2O$ was set as the nitrogen source while the phosphorus source and sulfur source were maintained as sulfate and phosphate, respectively. At the same time, the flux of the other exchange reactions containing carbon source were set zero except for the aimed carbon source. For prediction of nitrogen utilization, only the glucose was taken as the carbon source. The target substrate was considered growth-supporting if the predicted growth rate was above zero.

Prediction of Essential Genes

The prediction of essential genes was conducted using the singleGeneDeletion function based on the Cobra Toolbox v2.0 (Schellenberger et al., 2011). Based on the calculated specific growth rate, the genes were grouped into the essential genes (the predicted specific growth rate was equal to 0), the partially essential genes (the predicted specific growth rate was in the range from 0 to the maximum value), and the non-essential genes (the predicted specific growth rate is the same as that without the change). Both synthetic and complex medium were used to predict the essential genes for cell growth. The synthetic medium was made up of glucose, ammonia, oxygen, and inorganic sulfur and phosphorus while the complex medium contained yeast extract, which allows the uptake of 20 common amino acids.

Transcriptome Analysis

The strain *A. niger* DS03043 was kindly donated by DSM. The media and procedures for spore germination, seed culture, and batch cultivations were described in (Lu et al., 2015) with the glucose as the only carbon source.

During exponential growth phase of *A. niger* DS03043, 10–15 mL broth was taken from the fermentor into three 50 mL precooled centrifuge tubes. Samples were centrifuged at 12000 rpm for 5 min at 4°C and supernatant was discarded. The harvested mycelium was washed twice with 0.9% NaCl, dried with sterilized filter paper, and then quickly frozen in liquid nitrogen. The samples were stored in −80°C refrigerator until RNA extraction. About 90 mg frozen mycelium was placed in a mortar which was precooled in −80°C refrigerator. Then the mycelium was quickly ground into powder in liquid nitrogen in the mortar. Total RNA was extracted from the powder using TRIzol Reagent according to the standard protocol (Invitrogen, Thermo Fisher Scientific, Waltham, MA). The RNA sequence, read mapping, and quantification were accomplished by Beijing Genomics Institute (Shenzhen, China) according to a series of standard approaches (Jex et al., 2014). Reads were mapped to the genome sequence assembly of *A. niger* CBS513.88 (Pel et al., 2007) as *A. niger* DS03043 was most closely related to this sequenced strain (Lu et al., 2015). The FPKM (Fragments per kilobase of transcript per million fragments mapped) value was calculated to characterize the expression level of each gene (Mortazavi et al., 2008).

Results and Discussion

Reconstruction of *A. e* GSMM-iHL1210

The update of the *A. niger* GSMM was based on iMA871 published in 2008 (Andersen et al., 2008), which contained 871 ORF and total 2,240 reactions including isozymes. Merging the reactions with the same isozymes, the total number of reactions was reduced to 1,380, which is the base for subsequent comparison between different models. The process for iterative upgrade is shown in Figure 1. First of all, based on the metabolites annotation information, the elements and electrical charge in all reactions were manually checked and balanced. The gap analysis was also conducted and 25 dead-end metabolites were found (Supplementary file 1), which were reduced by adding essential reactions from KEGG. To update iMA871 systematically, the *A. niger* gene annotation information

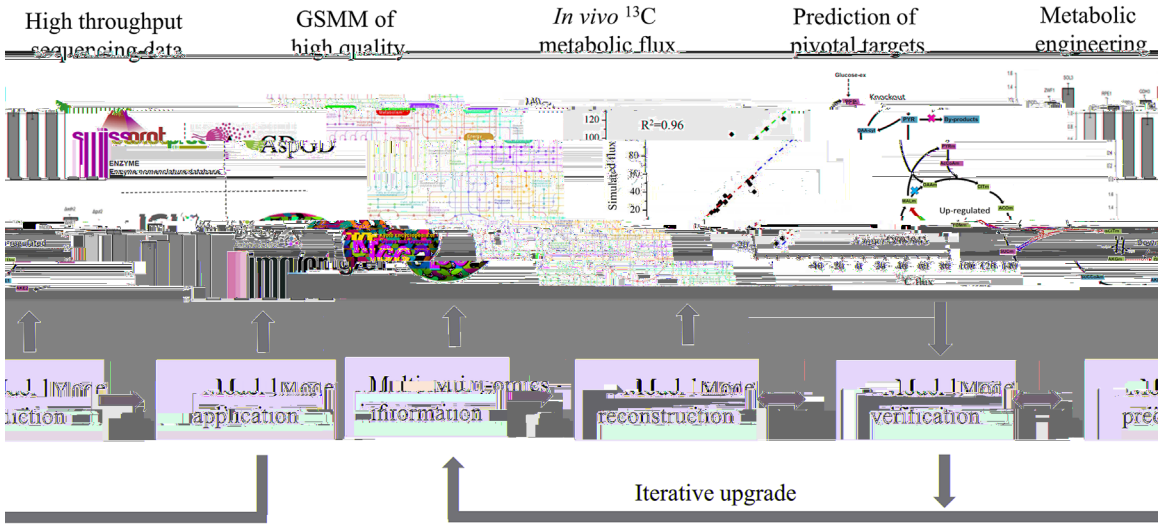

Figure 1. Pipeline for reconstruction of *A. niger* GSMM in this study.

from four common databases (AspGD, KEGG, UniProtKB, and IMG) were firstly sorted and compared (Supplementary file 2, Fig. S1). By comparison, the gene annotation from AspGD was the most comprehensive among four databases as the number of genes encoding enzymes was largest (Arnaud et al., 2012).

Compared to iMA871, the new iHL1210 harbors obvious improvements. The total number of reactions was increased from 1,380 in iMA871 to 1,764 in iHL1210, with the number of unique metabolites increasing from 775 to 902. Accordingly, the number of unique ORFs increased from 871 to 1,210 (Table I) and the detailed model information of iHL1210 could be found in Supplementary file 1. Further, the peroxisome was added, next to the previous three compartments covering the extracellular, cytoplasmic and mitochondrial spaces. In total, there are about 86 reactions in the peroxisome (Table I). It is known that an accurate cell composition is important for GSMM improvement (Feist and Palsson, 2010). Referring to the latest GSMMs of *E.coli* and yeast, the essential cell trace constituents were added, including the essential cofactors, like biotin, folate, heme, thiamin, etc.

As shown in Table I, there are now 1,748 unique reactions and 902 unique metabolites in iHL1210, which is more than that in the model of *Aspergillus terreus* NIH 2624 (Liu et al., 2013), indicating that this update enlarged the scale of the *A. niger* GSMM. The ORF number in iHL1210 is 1,210, which is lower than that of *A. terreus* NIH 2,624,

Table I. Comparison among GSMMs of *Aspergillus*.

Features	*A. terreus* NIH 2624	*A. niger* CBS 513.8	
Genome size(Mb)	29.3	34.9	34.9
Metabolic model	iJL1454	iMA871	iHL1210
Genes	1,454	871	1,210
ORF coverage (%)	14	7.8	10.8
Unique functional proteins	774	426	667
Multigene complexes	34	18	33
Genes involved in complexes	235	109	160
Instances of isozymes	350	144	232
Reactions	1,451	1,380	1,764
Unique reactions	1,357	1,190	1,748
Metabolic reactions	1,105	1,022	1,289
Cytoplasmic	811	786	998
Mitochondria	244	218	182
Extracellular	50	18	23
Peroxisome	–	–	86
Transport reactions	206	189	285
Cytoplasm, mitochondria	71	37	61
Cytoplasm, extracellular	135	152	189
Cytoplasm, peroxisome	–	–	35
Exchange reactions	140	169	189
Metabolites	1,155	1,041	1,254
Unique metabolites	833	775	902
Cytoplasmic	721	663	785
Mitochondria	260	204	179
Extracellular	174	175	195
Peroxisome	–	–	94
Gene—protein-reaction associations	1,233	969	1,266
Gene associated (metabolic/transport)	1,090/143	849/3	1,121/145
No gene association (metabolic/transport)	15/63	173/186	168/140

The data of *A. terreus* NIH 2624 was from the reference Liu et al. (2013), a part of the data relating to iMA871 was recalculated using the same method.

showing that the gene information in *A. niger* can be further increased. Based on the latest gene annotations, iHL1210 contains more genes compared to the previous model, such as the genes relating to the catabolism of polysaccharides, the uptake of extracellular substances, the membrane transport reactions, etc. In summary, the *A. niger* GSMM was systematically and significantly improved based on the integration of gene annotation from different databases.

Validation of GSMM

Model Verification by RNA-seq

The ORF in the newly updated *A. niger* GSMM iHL1210 was firstly verified with the latest RNA-seq data due to the fact that it has huge advantages in identifying the genes with low-expressions compared with the microarray data (Wang et al., 2009). The RNA-seq result showed that the expression of about 6,000 genes could be determined in the sampling condition and the KEGG categories of the measured genes are shown in Figure 2E. The iHL1210 contains 1,210 genes and the expression of most genes (63%) could be verified according to the results of RNA-seq (Fig. 2A, Supplementary file 1) when glucose was used as the only carbon source. Deducting the exchange reactions, among the remaining 1,575 reactions, the expression of genes from 12% of these reactions was not measured (Fig. 2B). After further deduction of 308 reactions without corresponding annotated genes, there are 587 and 680 reactions are associated with single and multiple genes, respectively. Based on the transcriptome analysis, about 73% of the single-gene reactions and 96% of the multi-gene reactions were verified respectively (Fig. 2C and D). More importantly, the results of RNA-seq could be used to optimize the GPRs in the new model. For example, both of the two genes (An12g03260 and An19g00140) encoding 3-hydroxyacyl-CoA dehydrogenase, relating to oxidation of fatty acid metabolism, were unmeasurable, implying there exist other isogenes in *A. niger* genome. After detailed search in AspGD database, we found that An08g09450, An11g08440 An01g09740, An04g02170, and An11g08060 also have the gene annotation of encoding 3-hydroxyacyl-CoA dehydrogenase and the expression of the latter three genes could be measured using RNA-seq. The above results indicated that the existence of most reactions was reliable in the new iHL1210, which lays a good foundation for the subsequent model evaluation.

Model Verification by Availability of Different Carbon and Nitrogen Sources

To comprehensively assess, the prediction ability of iHL1210, reported *A. niger* phenotype data from 2008 to 2016 were collected, together with accumulated physiological data from our own lab. *A. niger* could grow on a wide range of carbon sources and nitrogen sources. The growth-relating 69 carbon and 30 nitrogen sources can be found in Tables II and III. The in silico growth capabilities of *A. niger* on 57 carbon and 22 nitrogen sources could be predicted using iHL1210 and the accuracy rate was 83% and 73%, respectively, which were higher than that using iMA871. Overall, the remaining inconsistencies have limited impact on the main metabolic function of the network, and they provide new space for a next round of upgrading.

Model Verification by Physiological Parameters Related to Cell Growth

Table II. Prediction of growth capabilities of *A. niger* on different carbon sources (+ for growth and − for non-growth) using iMA871 and iHL1210, respectively.

Carbon sources	Experiment	iMA871	iHL1210
Acetate	+	+	+
Alanine	+	+	+
Amylose	+	−	+
Anthranilic acid	+	−	−
Arabinan	+	−	+
Arabinose	+	+	+
Arabitol	+	+	+
Aspartic acid	+	+	+
Benzoate	+	−	−
Benzyl-vanillic acid	+	−	−
Citrate	+	+	+
Coumaric acid	+	+	+
Dethiobiotin	+	−	+
D-Galactose	+	−	+
D-galacturonic acid	+	−	+
Dihydroxyacetone	+	+	+
Dimethylterephthalate	+	+	+
DL-mandelic acid	+	−	−
D-xylose	+	+	+
Erythritol	+	+	+
Ferulic acid	+	+	+
Fructose	+	+	+
Fumarate	+	+	+
Galactan	+	−	−
Galactose	+	+	+
Galacturonic acid	+	+	+
Gluconate	+	+	+
Gluconic acid lactone	+	−	+
Glucosamine	+	−	−
Glucose	+	+	+
D-glucuronic acid	+	−	−
Glutamate	+	+	+
Glycerol	+	+	+
Indole	+	+	+
Inulin	+	−	−
Lactose	+	+	+
L-arabinose	+	−	−
L-rhamnose	+	−	+
Maltodextrin	+	−	+
Maltose	+	+	+
Maltotriose	+	−	−
Mannitol	+	+	+
Mannose	+	+	+
Methanol	+	+	+
m-hydroxybenzoic acid	+	−	+
Myo-inostiol	+	−	−
Oxalate	+	+	+
Pectin	+	−	+
Phenylacetic acid	+	−	+
Phenylalanine	+	+	+
Propionate	+	+	+
Pyruvate	+	+	+
Resorcinol	+	+	+
Rhamnose	+	+	+
Ribose	+	+	+
Salicylate	+	+	+
Sorbitol	+	+	+
Starch	+	−	+
Succinate	+	+	+
Sucrose	+	+	+
Tannic acid	+	+	+
Tartrate	+	+	+
Trehalose	+	+	+
Tryptophan	+	+	+
Vanillic acid	+	+	+
Xylan	+	−	+
Xylitol	+	+	+
Xylose	+	+	+
α-ketoglutarate	+	+	+

TABLE II. (*Continued*)

2015) mmol ATP/gDCW.h for *A. niger* DS03043. As this energy parameter has an obvious effect on cell growth, the NGAM in the iHL1210 was verified using *A. niger* NW185 chemostat cultivation data without by-product formation (Lameiras et al., 2015). As shown in Figure 3, when the NGAM was changed from 1.3 (Andersen et al., 2008) to 3.73 mmol ATP/gDCW.h, iHL1210 allows a more reasonable prediction of μ, q_{O2}, q_{CO2}, and RQ, initially indicating that the adjustment in energy parameters could improve the performance of new model. The better fit is obtained with a NGAM value of 3.73 mmolATP/gDCW.h, indicating that the NGAM may still vary between different strains.

Table III. Prediction of growth conditions of *A. niger* on different nitrogen sources (+ for growth and − for non-growth) using iMA871 and iHL1210, respectively.

Nitrogen sources	Experiment	iMA871	iHL1210
Alanine	+	+	+
Ammonia	+	+	+
Anthranilate	+	+	+
Arginine	+	+	+
Asparagine	+	+	+
Aspartate	+	+	+
Butylamine	+	+	+
Citrulline	+	+	+
4-aminobutanoate	+	−	+
Glutamate	+	+	+
Glutamine	+	+	+
Glycine	+	+	+
Nitrate	+	+	+
Ornithine	+	−	+
Pentylamine	+	−	−
Phenylalanine	+	+	+
Proline	+	−	+
Serine	+	+	+
Threonine	+	+	+
Tryptophan	+	+	+
Tyrosine	+	−	+
Urea	+	+	+
Xanthin	+	+	+
Acetamide	+	−	−
Agmatine	+	−	−
4-guanidinobutyrate	+	−	−
3-guanidinopropionate	−	−	+
L-homoarginine	−	−	+
Guanidinoacetate	−	−	+
Putrescine	+	−	−

Sensitivity Analysis of iHL1210

During the sensitivity analysis, the influence of different cell compositions (protein, RNA, DNA) and the energy parameters (GAM, NGAM, and P/O) on the prediction accuracy of the strain phenotype (the specific growth rate-μ and specific oxygen uptake rate-q_{O2}) were systematically investigated. As shown in Figure 4, μ and q_{O2} were hardly affected by the changes of the protein and RNA content. However, as the ratio of lipid increased, μ was slightly decreased. Different from the cell compositions, the predicted μ and q_{O2} were more seriously affected by energy parameter changes. As GAM and NGAM increased, the predicted μ decreased greatly while the q_{O2} increased accordingly. The simulated strain phenotypes are also sensitive to the changes in P/O as shown in Figure 4F, which is consistent with the simulated results using *E. coli* GSMM (Feist et al., 2007).

Model Verification by Production of By-Products

The GSMM could be used to predict the secretion of by-products (Feist et al., 2007). To evaluate the performance of iHL1210, the secretion of by-products was simulated. Generally, organic acids and polyols were the two main by-products produced by *A. niger* during industrial production (Pedersen et al., 2011). For example, oxygen limitation could lead to accumulation and secretion of organic acids from the TCA cycle according to the previous research (Diano et al., 2009). In the simulation by iHL1210, the citric acid was secreted under lower q_{O2} values, which matched the experimental data well (Supplementary file 2, Fig. S2), and the overflow of citric acid is believed to be beneficial for the reduction of excess NADH under hypoxia condition. In addition, no direct relation was found between q_{O2} and oxalic acid formation in the model which coincided with the actual experimental data, thus showing the good performances of iHL1210 in qualitative prediction of organic acid secretion under oxygen-limited conditions.

Model Verification by In Vivo ^{13}C Metabolic Flux

Although, the GSMM could be used to predict important physiological parameters, it is still difficult to ensure a full accuracy as there are obvious discrepancies between the predicted flux distribution and the in vivo estimated flux (Damiani et al., 2015). To further evaluate the prediction performance of iHL1210, the predicted flux using FBA was compared with the estimated metabolic flux using the ^{13}C labeled technique (Lu et al., 2015). It should be noted that the FBA calculates a solution range for the flux network and not a single value for all fluxes (Orth et al., 2010). Usually, there are two strategies to solve such problem. Firstly, the average flux values could be calculated by a specific sampling algorithm (Megchelenbrink et al., 2014). Secondly, new constraints can be added according to a cell metabolic objective function, such as the optimization of energy utilization, minimization of metabolic adjustment (MOMA), etc. Because it was reported that there exists a direct relation ($R^2 = 0.92$) (Song et al., 2014) between the predicted fl

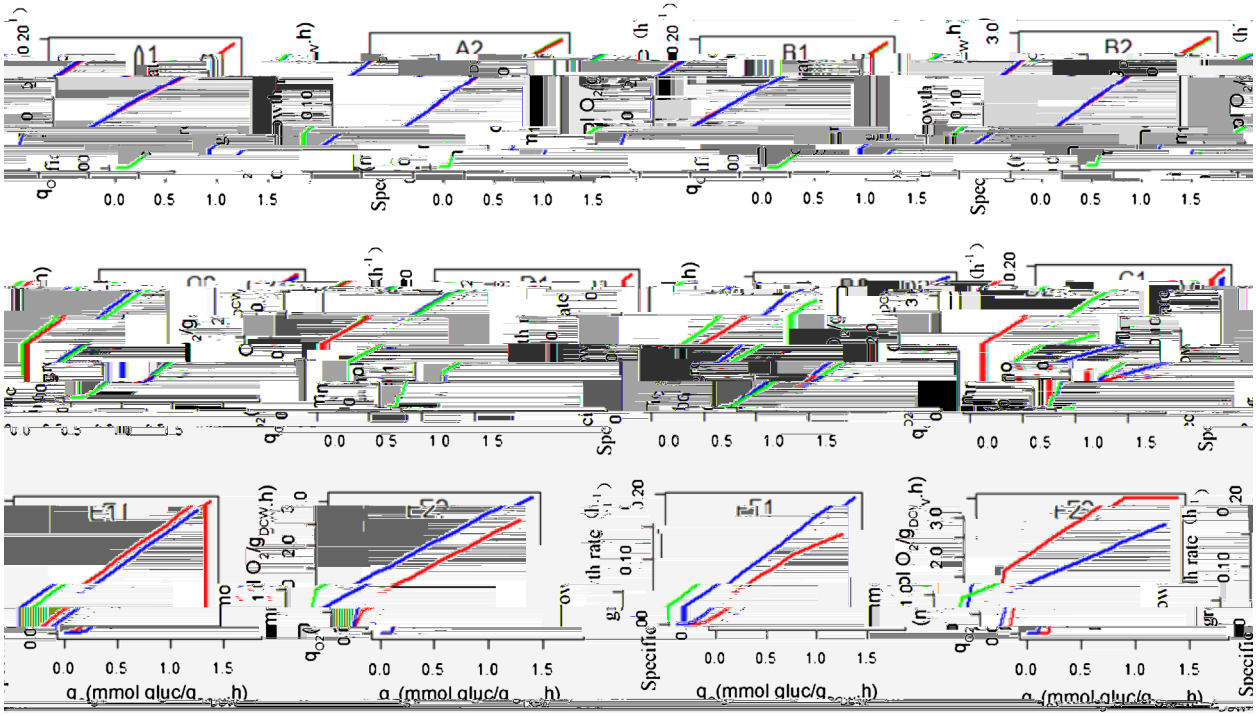

Figure 4. Effects of model parameters on specific growth rate (**A1**, **B1**, **C1**, **D1**, **E1**, **F1**) and specific oxygen uptake rate (**A2**, **B2**, **C2**, **D2**, **E2**, **F2**) using sensitivity analysis with iHL1210. The simulations were performed for aerobic glucose-limited cultivations by varying the protein content (26–40%) (**A**), the RNA content (1.8–5%) (**B**), the lipid content (10.9–20%) (**C**), the GAM (61 mmol ATP/gDCW ±50%) (**D**), NGAM (3.73 mmol ATP/g_{DCW}.h ±50%) (**E**), and P/O (2.64 for mitochondrial NADH and 1.64 for succinate and cytosolic NADH ±50%) (**F**). Blue represents the simulated results for the high value of the input parameter, and red the lower in all cases. The space between the blue and red lines represents the response of the specific growth and oxygen uptake rates depending on the varying input model parameters.

the correlation coefficient between the predicted fluxes and the ^{13}C fluxes for CBS513.8 (wild strain) and *A. niger* DS0304 (mutant strain) were 0.83 and 0.96, respectively, initially indicating the good performance of iHL1210. However, there still are obvious gaps at some points, which might be caused by the following two reasons. Firstly, the model size varies greatly between the GSMM and the ^{13}C flux model as the latter only contains less than 100 reactions involved in the central carbon metabolism. Secondly, the computational algorithms were different between pFBA and ^{13}C metabolic flux analysis. Generally, with more constraints from labeling information of amino acids and intermediate metabolites, the ^{13}C fluxes were more accurate (Zamboni et al., 2009). On the whole, the high consistency between the predicted flux by pFBA and measured ^{13}C metabolic flux confirms the good prediction performance of iHL1210.

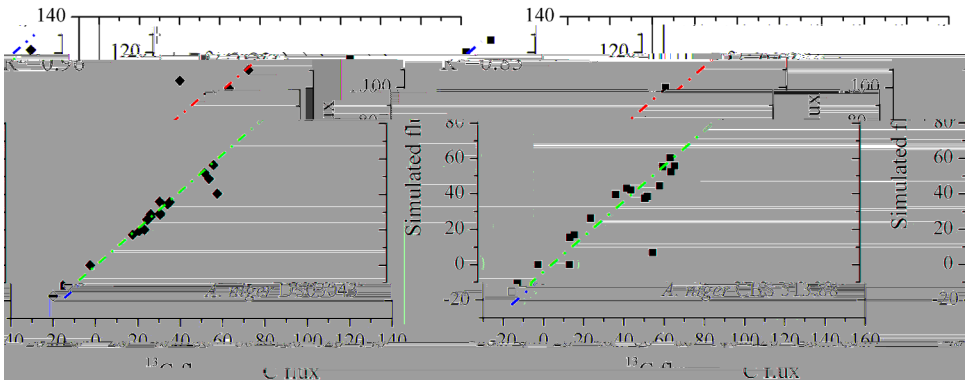

Figure 5. Consistent changes in flux values can be predicted with reasonable accuracy using iHL1210. The growth kinetic parameters used for pFBA simulation, as well as the ^{13}C flux distribution of of *A. niger* CBS513.88 and *A .niger* DS03043 were described in previous research (Lu et al., 2015).

Gene Target Prediction In Silico for Strain Design

As the GPRs were established in the GSMM, the relations between the genotype and phenotype could be predicted efficiently using GSMM (Thiele and Palsson, 2010). With the singleGeneDeletion function of the Cobra Toolbox v2.0, the essential genes, the partially essential genes, and the non-essential genes were predicted (Schellenberger et al., 2011). Using cultivations on synthetic medium, 216 essential genes were obtained, which were related to energy metabolism, TCA cycle, amino acids metabolism, etc. (Fig. 6). When complex medium was adopted, the number of essential genes decreased from 216 to 166 for the main reason that the number of genes responsible for the essential amino acids synthesis decreased (Fig. 6). Furthermore, on synthetic medium, 96 partially essential genes were identified by model simulation. These partially essential genes may be pivotal for subsequent in silico strain design (Pan and Hua, 2012) because the yield of enzyme production may increase together with a slowdown of cell growth.

To further evaluate the potential of the latest *A. niger* GSMM in rational strain design, experimental data of some engineered *A. niger* (the rates of cell growth and citric acid production using *A. niger*) were gathered from published studies and iHL1210 was used to simulate the effects of the corresponding gene deletion or overexpression. The experimental and simulated data showed a high consistence (Supplementary file 3), indicating that iHL1210 can facilitate the process of engineering more efficient *A. niger* microbial cell factories for the production of proteins and organic acids.

Evaluation of Different NADPH Sources on Enzyme Production

As reported before (Driouch et al., 2012; Lu et al., 2015), the metabolic pathways relating to NADPH formation, like the PP pathway and the oxidation of malate to pyruvate, were up-regulated when *A. niger* produced high level of enzyme. Therefore, the effects of different metabolic pathways providing NADPH on glucoamylase production were initially evaluated in this study. In iHL1210, there are total 173 reactions involving NADPH and 49 reactions among them have the possibility to form NADPH. After detailed calculation, we found that each of the remaining 19 NADPH sources from 22 reactions allowed the cell growth and enzyme production (Supplementary file 1) separately when the other NADPH sources were simply set to zero. These 22 reactions were grouped into four categories based on multi-omics integrative analysis (Fig. 7). Among them, the existence of eight reactions can be directly verified from studies of transcriptomics (this work) and proteomics (De Ferreira Oliveira et al., 2010), for example the PP pathway ($G6P + NADP \geq D6PG + NADPH$, and $D6PC + NADP \geq R15P + NADPH + CO2$),

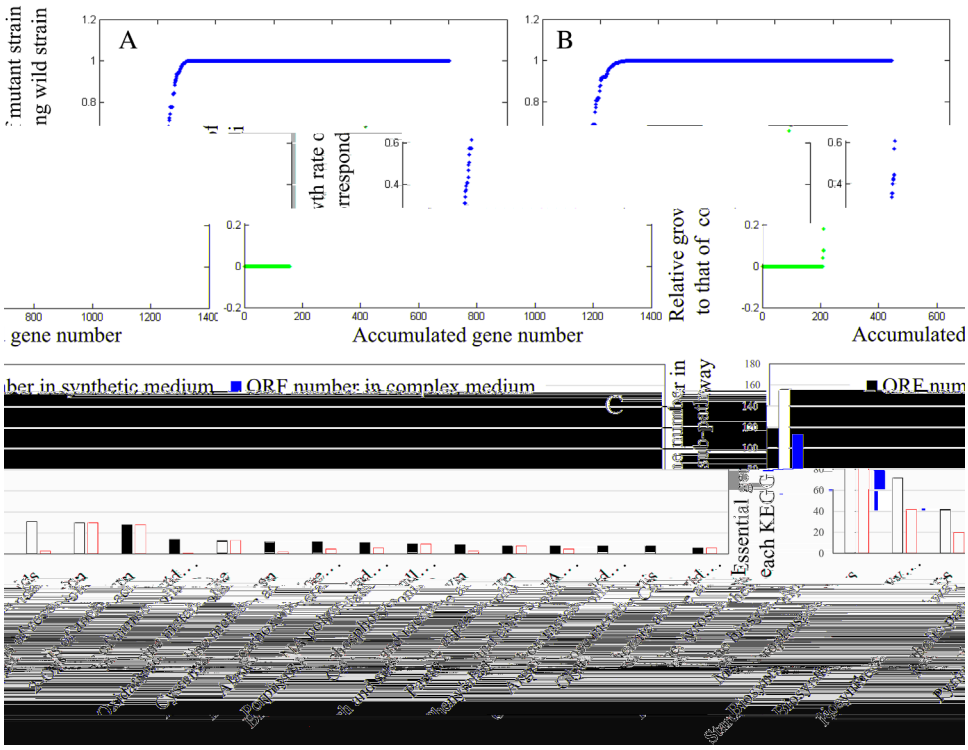

Figure 6. Results of single gene deletion study with iHL1210. Changes in relative growth rate of mutant strain to that of wild strain in synthetic medium (**A**) and complex medium (**B**). KEGG categories of essential genes in synthetic medium and complex medium (**C**).

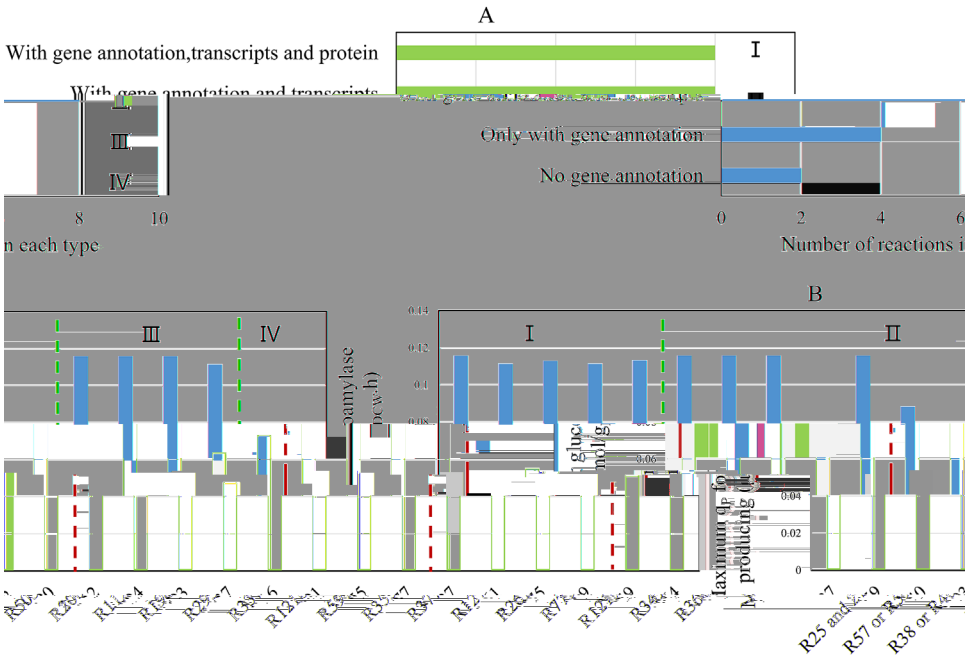

Figure 7. Classification of reactions producing NADPH (**A**) and the maximum specific glucoamylase production (**B**) under each single NADPH source from the corresponding reactions. The parameters used in the calculation: qs = 1.7 mmol/g_{DCW}.h, $\mu = 0.05h^{-1}$, NGAM = 3.73 mmol ATP/g_{DCW}.h.

the TCA cycle (ICIT + NADP ≥ AKG + NADPH + CO_2), and oxidation of malate to pyruvate (MAL + NADP ≥ PYR + NADPH + CO_2), initially indicating that these are key sources of NADPH for *A. niger* metabolic activities. The results of the simulation also indicate that the yield of glucoamylase would be the highest when the PP pathway was set as the only source for NADPH supply. Further, simulations using MOMA indicated that the deletion of genes in the PP pathway, encoding glucose 6-phosphate 1-dehydrogenase and phosphogluconate dehydrogenase, could lead to the reduction of the maximum yield of glucoamylase by 16.9%, larger than that in genes deletions from the other NADPH sources, strengthening the hypothesis that the PP pathway was the most important NADPH source for the enzyme production in *A. niger*.

Although, the NADPH supply from the PP pathway could promote enzyme production, the effects of the PP pathway flux on cell growth should not be ignored, as the PP pathway provides not only NADPH, but also precursors for cell growth, like nucleotide, cofactors, etc. (Driouch et al., 2012). As reported, reinforcement of the PP pathway may accelerate the cell growth (Park et al., 2014), which is unfavorable for the productivity. To avoid this negative effect, other candidate reactions could be selected to improve enzyme production, such as the oxidation of malate to pyruvate (MAL + NADP ≥ PYR + NADPH) and the transhydrogenase reaction (NADP + NADH ≥ NADPH + NAD). Meanwhile, simulation based on iHL1210 also revealed that the NADPH from reactions of R112, R127, R194, and so on, could also support the highest yield of enzyme formation. However, there is still uncertainty as the existence of these reactions lack definite evidences from transcriptomics and proteomics studies. The integrated analysis of multi-omics effects and model simulation indicates that the NADPH from the core carbon metabolism is the optimal choice for strain growth and enzyme production, while the actual optimal NADPH supply pathway still needs to be confirmed from molecular biology experiments.

Conclusion

Based on the latest gene annotation information and physiological data, the current *A. niger* GSMM (iMA871) was systematically updated and evaluated. The new *A. niger* GSMM iHL1210 contains 1,764 reactions and 1,210 ORFs, thus making it comparable with the latest GSMM of other *Aspergillus* species. With iHL1210, the growth capability on most of the reported carbon and nitrogen sources, as well as the accurate pathway flux distribution could be predicted well. In addition, the essential and partially essential genes for cell growth were predicted and the effects of different NADPH sources on glucoamylase enzyme production were evaluated, which provides potential gene targets for future metabolic engineering. We believe, with molecular biology experiments to further confirm the existence of certain reactions, as well as integration of more omics information, the quality of the *A. niger* GSMM could be remarkably elevated in the future.

Authors' Contributions

HL carried out the model reconstruction and validation. JC and MH participated in the design and coordination of the study. LO, JX, YZ, SZ, and NH helped to draft the manuscript. All authors read and approved the final manuscript.

This work was financially supported by Royal DSM (Delft, the Netherlands) and partially supported by National Basic Research Program (973 Program 2013CB733600), NWO-MoST Joint Program (2013DFG32630), Open Funding Project of the State Key Laboratory of Bioreactor Engineering.

References

Andersen MR, Nielsen ML, Nielsen J. 2008. Metabolic model integration of the bibliome, genome, metabolome and reactome of *Aspergillus niger*. Mol Syst Biol 4:1–13.

Arnaud MB, Cerqueira GC, Inglis DO, Skrzypek MS, Binkley J, Chibucos MC, Crabtree J, Howarth C, Orvis J, Shah P, Wymore F, Binkley G, Miyasato SR, Simison M, Sherlock G, Wortman JR. 2012. The *Aspergillus* Genome Database (AspGD): Recent developments in comprehensive multispecies curation, comparative genomics and community resources. Nucleic Acids Res 40:653–659.

Boghigian BA, Armando J, Salas D, Pfeifer BA. 2012. Computational identification of gene over-expression targets for metabolic engineering of taxadiene production. Appl Microbiol Biotechnol 93(5):2063–2073.

Bordbar A, Monk JM, King ZA, Palsson BO. 2014. Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet 15(2):107–120.

Caspeta L, Chen Y, Ghiaci P, Feizi A, Buskov S, Hallström BM, Petranovic D, Nielsen J. 2014. Altered sterol composition renders yeast thermotolerant. Science 346(6205):75–78.

Damiani AL, He QP, Jeffries TW, Wang J. 2015. Comprehensive evaluation of two genome-scale metabolic network models for *Scheffersomyces stipitis*. Biotechnol Bioeng 112(6):1250–1262.

De Ferreira Oliveira JMP, Van Passel MWJ, Schaap PJ, De Graaff LH. 2010. Shotgun proteomics of *Aspergillus niger* microsomes upon D-xylose inductions. Appl Environ Microbiol 76(13):4421–4429.

Diano A, Peeters J, Dynesen J, Nielsen J. 2009. Physiology of *Aspergillus niger* in oxygen-limited continuous cultures: Influence of aeration, carbon source concentration and dilution rate. Biotechnol Bioeng 103(5):956–965.

Driouch H, Melzer G, Wittmann C. 2012. Integration of in vivo and in silico metabolic fluxes for improvement of recombinant protein production. Metab Eng 14(1):47–58.

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3:1–18.

Feist AM, Palsson BO. 2010. The biomass objective function. Curr Opin Microbiol 13(3):344–349.

Ganzlin M, Rinas U. 2008. In-depth analysis of the *Aspergillus niger* glucoamylase (glaA) promoter performance using high-throughput screening and controlled bioreactor cultivation techniques. J Biotechnol 135(3):266–271.

Jex AR, Nejsum P, Schwarz EM, Hu L, Young ND, Hall RS, Korhonen PK, Liao S, Thamsborg S, Xia J, Xu P, Wang S, Scheerlinck JY, Hofmann A, Sternberg PW, Wang J, Gasser RB. 2014. Genome and transcriptome of the porcine whipworm *Trichuris suis*. Nat Genet 46(7):701–706.

Knuf C, Nielsen J. 2012. *Aspergilli*: Systems biology and industrial applications. Biotechnol J 7(9):1147–1155.

Lameiras F, Heijnen J, van Gulik W. 2015. Development of tools for quantitative intracellular metabolomics of *Aspergillus niger* chemostat cultures. Metabolomics 11(5):1–12.

Liu J, Gao Q, Xu N, Liu L. 2013. Genome-scale reconstruction and in silico analysis of *Aspergillus terreus* metabolism. Mol Biosyst 9(7):1939–1948.

Long MR, Ong WK, Reed JL. 2015. Computational methods in metabolic engineering for strain design. Curr Opin Biotechnol 34(0):135–141.

Lu H, Liu X, Huang M, Xia J, Chu J, Zhuang Y, Zhang S, Noorman H. 2015. Integrated isotope-assisted metabolomics and 13C metabolic flux analysis reveals metabolic flux redistribution for high glucoamylase production by *Aspergillus niger*. Microb Cell Fact 14(1):1–14.

Machado D, Herrgård M. 2014. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Comput Biol 10(4):1–12.

Megchelenbrink W, Huynen M, Marchiori E. 2014. OptGpSampler: An improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. PLoS ONE 9(2):1–8.

Meyer V, Fiedler M, Nitsche B, King R. 2015. The cell factory *Aspergillus* enters the big data era: Opportunities and challenges for optimising product formation. Berlin Heidelberg: Springer. p 1–42.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-S eq. Nat Meth 5(7):621–628.

O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BO. 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol Syst Biol 9:1–13.

Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? Nat Biotechnol 28(3):245–248.

Pan P, Hua Q. 2012. Reconstruction and in silico analysis of metabolic network for an oleaginous yeast *Yarrowia lipolytica*. PLoS ONE 7(12):1–11.

Park SH, Kim HU, Kim TY, Park JS, Kim S-S, Lee SY. 2014. Metabolic engineering of *Corynebacterium glutamicum* for L-arginine production. Nat Commun 5:1–9.

Pedersen L, Hansen K, Nielsen J, Lantz AE, Thykaer J. 2011. Industrial glucoamylase fed-batch benefits from oxygen limitation and high osmolarity. Biotechnol Bioeng 109(1):116–124.

Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, Turner G, de Vries RP, Albang R, Albermann K, Andersen MR, Bendtsen JD, Benen JA, van den