Received <day> <Month>, <year>; | Revised <day> <Month>, <year>; | Accepted <day> <Month>, <year>

DOI: xxx/xxxx

ARTICLE TYPE

Error propagation in constraint-based modeling of Chinese hamster ovary cells

Diana Széliová[1,2] | Dmytro Iurashev[1,2] | David E Ruckerbauer[1,2] | Gunda Koellensperger[4] | Nicole Borth[1,2] | Michael Melcher[1,3] | Jürgen Zanghellini*[2,4]

[1]Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Vienna, Austria, EU
[2]acib – Austrian Centre of Industrial Biotechnology, Vienna, Austria, EU
[3]Institute of Statistics, University of Natural Resources and Life Sciences, Vienna, Vienna, Austria, EU
[4]Department of Analytical Chemistry, University of Vienna, Vienna, Austria, EU

Correspondence
*Jürgen Zanghellini, Corresponding address.
Email: juergen.zanghellini@univie.ac.at

Present Address
Muthgasse 11, 1190 Vienna, Austria, EU

Abstract

Chinese hamster ovary (CHO) cells are the most popular mammalian cell factories for the production of glycosylated biopharmaceuticals. To further increase titer and productivity and ensure product quality, rational systems-level engineering strategies based on constraint-based metabolic modeling, such as flux balance analysis (FBA), have gained strong interest. However, the quality of FBA predictions depends on the accuracy of the experimental input data, especially on the exchange rates of extracellular metabolites. Yet, it is not standard practice to devote sufficient attention to the accurate determination of these rates. In this work we investigated to what degree the sampling frequency during a batch culture and the measurement errors of metabolite concentrations influence the accuracy of the calculated exchange rates and further, how this error then propagates into FBA predictions of growth rates. We determined that accurate measurements of essential amino acids with low uptake rates are crucial for the accuracy of FBA predictions, followed by a sufficient number of analysed time points. Finally, we observed that the measured difference in growth rates of two cell lines can only be reliably predicted when both high measurement accuracy and sampling frequency are ensured.

KEYWORDS:
Chinese hamster ovary cells; genome-scale metabolic modeling, flux balance analysis; uptake and secretion rates; error propagation

1 | INTRODUCTION

Chinese hamster ovary (CHO) cells are the primary host for the production of biopharmaceuticals, particularly monoclonal antibodies and other complex therapeutic proteins[1]. Their main advantages include the ability to perform human-like post-translational modifications, in particular glycosylation, suspension growth in serum-free chemically defined media and a low risk of viral infections, which makes them a safe host for the production of human therapeutic proteins[2,3]. However, the development of new producer cell lines is a time-consuming, costly and laborious process, based mostly on laboratory evolution and high-throughput screening of thousands of clones. It often takes several months to obtain a high producer and this trial-and-error process must be repeated for each new product[4]. Furthermore, it is not clear which factors limit product formation[5].

[0]**Abbreviations:** AAs, amino acids; CHO, chinese hamster ovary; FBA, flux balance analysis; mAb, monoclonal antibody; RSD, relative standard deviation; RSE, relative standard error; SE, standard error

As a result of the success of metabolic modeling in designing microbial cell factories[6], interest has grown in applying these methods for the analysis and optimisation of CHO as well[7]. However, applications of modeling to CHO cells remain limited as key (bioinformatic) resources became available only recently. The publication of CHO's genome sequence[8] enabled the *in silico* reconstruction of its metabolism[9]. Along with recent updates[10,11] these genome-scale metabolic models sit at the heart of constraint-based modeling approaches that allow to computationally connect genotype and phenotype[12]. In fact, such a genome-scale metabolic model coupled with a model of the secretory pathway[13] successfully predicted gene knock-outs that led to increased productivity and growth[14]. Hence, genome-scale modeling has great potential to improve CHO cell line development by identifying bottlenecks in productivity and designing engineering strategies to resolve these.

One of the most common metabolic modeling approaches is constraint-based modeling, such as flux balance analysis (FBA)[15,16], especially for analysis of genome-scale metabolic models[17,18]. The accuracy of FBA predictions depends on the correct reconstruction of the metabolic pathways, the biomass composition and the uptake and secretion rates of extracellular metabolites which are used as constraints[16,19]. Previously we showed that the quality of growth rate predictions by FBA in CHO depends mostly on the accuracy of the measured uptake and secretion rates[20]. While simple organisms such as bacteria or yeast can grow on a single carbon source (thereby simplifying the analysis), mammalian cells grow in complex media containing numerous essential metabolites and secrete several byproducts. Typically, more than 20 exchange rates have to be measured in order to perform FBA, which makes the analysis much more challenging than for simpler organisms on minimal medium. However, despite the importance and the difficulty of determining accurate exchange rates for mammalian cells, little attention has been paid to this topic so far.

To determine exchange rates accurately, it is necessary to measure the metabolite concentration throughout the cultivation at a sufficient sampling frequency. In literature, the concentrations of extracellular metabolites are commonly measured once per day, typically resulting in 4-6 time points in total[21,22,23,24,25,26,27,28]. However, this might not be enough to obtain accurate uptake and secretion rates.

Here we investigated the impact of sampling frequency and the error of metabolite concentration measurements on the calculation of exchange rates and subsequently on FBA predictions of growth rate. We determined which exchange rates have the biggest impact on FBA and what accuracy and sampling frequency is required to detect a given difference between two cell lines.

2 | METHODS

The simulations of experimental data, statistical analysis and visualisation were done in R version 3.4.3[29]. FBA[15,16] was performed in python 3.7.4 using package COBRApy[30]. All data and scripts were deposited in Mendeley Data at https://data.mendeley.com/datasets/5vn5m33wpr/draft?a=3e7afba4-42e0-45b4-9e16-7adb419251f9.

2.1 | Reference FBA

To generate a reference state, FBA was performed with the experimental data and genome-scale metabolic models taken from Széliová et al.[20] for two cell lines (`K1par-8mMAP` and `HYher-8mMCD`). The experimentally measured specific uptake and secretion rates of glucose, amino acids, lactate and ammonium (Table A1) were used as constraints and biomass production was set as the objective function (reaction "`R_biomass_specific`"). Cysteine and tryptophan uptakes were left unconstrained due to the lack of quantitative experimental data. The predicted growth rates were used as reference states of the two cell lines in the subsequent analyses.

2.2 | Simulation of concentration profiles

The experimental exchange rates from Széliová et al.[20], the predicted growth rate and the uptake rates of cysteine and tryptophan from the reference FBA (see Section 2.1 and Table A1) were used to simulate concentration profiles using Equation (1),

$$[\hat{i}] = [\hat{i}]_0 + \frac{\hat{q}_i \widehat{\mathrm{BM}}_0}{\hat{\mu}} \left(e^{\hat{\mu} t} - 1 \right), \tag{1}$$

where $[\hat{i}]$ is an ideal concentration of metabolite i during exponential phase, $[\hat{i}]_0$ is the initial concentration of metabolite i in the medium, $\hat{q}_i$ is the reference exchange rate, $\hat{\mu}$ is the growth rate, t is time of cultivation and $\widehat{\mathrm{BM}}_0$ is the initial amount of biomass, calculated from an initial cell concentration (1.6×10^5 viable cells/mL) and the experimentally measured dry mass per cell

(`K1par-8mMAP`: 252.3 pg/cell, `HYher-8mMCD`: 279 pg/cell). The values for $[\hat{i}]_0$, $\hat{q}_i$, cell dry mass and initial cell concentration were taken from[20] (see Table A1). $\hat{\mu}$ is the growth rate predicted by the reference FBA. The time of cultivation corresponds to the length of exponential phase, which was set to 90 hours (the mean length of exponential phases of the two cell lines). Samples were generated at regular intervals of 6, 12 or 24 hours or – to resemble typical working shifts – at irregular intervals with 4 samples per day - every 4 h with a 12 h gap or every 2.5 h with a 16.5 h gap (the gap was positioned either in the beginning of each day or after the first four "dense" sampling time points, e.g. hours 0, 12, 16, 20, 24, 36 etc. or 0, 4, 8, 12, 24, 30 etc.).

To simulate experimental data, normally distributed noise was added to the ideal concentrations

$$[i] = [\hat{i}] + \varepsilon_c \quad \text{with} \quad \varepsilon_c \sim \mathcal{N}(0, \sigma_c^2) \tag{2}$$

where ε_c is a standard normal random variable with relative standard deviation (RSD) $\sigma_c/[\hat{i}]$ chosen between 0.02 and 0.20. Note that the noise increases with the concentration level. The range of the RSDs was chosen based on the available data from Széliová et al.[20], where median RSDs of the concentration measurements from 3-6 replicates were 1.7-7.1% (mean RSDs 5.3-19.8%). Thousand concentration profiles were generated for each metabolite. Afterwards, an equation of form (1) was fitted to each profile with R function `nls`, using q_i and $[i]_0$ as the fitting parameters, thereby obtaining 1000 sets of exchange rates. Standard errors (SEs) of the fitted rates were used as a measure of accuracy. For a graphical representation of the workflow see Figure A1, Sim 1.

In the second version of the simulations, not only concentration profiles, but also biomass concentrations were perturbed. For this, cell concentration data BM was simulated with Equation (3),

$$\mathrm{BM} = \widehat{\mathrm{BM}}_0\, e^{\hat{\mu}t} + \varepsilon_B \quad \text{with} \quad \varepsilon_B \sim \mathcal{N}(0, \sigma_B^2) \tag{3}$$

where t is time of cultivation (0-90 h), $\widehat{\mathrm{BM}}_0$ is the initial biomass (the same as in Equation (1)) and $\hat{\mu}$ is the growth rate from the reference FBA. Normally distributed noise of 6% ($\sigma_B = 0.06 \cdot \widehat{\mathrm{BM}}_0\, e^{\hat{\mu}t}$) was added in the same way as described above for the concentration profiles. The accuracy of 6% is provided by the manufacturer of Vi-CELL XR (Beckman Coulter)[31], an automated cell counting device commonly used to quantify cell concentration. The parameters μ and BM_0 were estimated by fitting the simulated biomass concentrations with an exponential growth function using R function `nls`. Note that the estimated growth rate (μ) was used for the simulations of metabolite concentrations with Equation (1) (instead of the growth rate predicted by the reference FBA) (Figure A1, Sim 2).

2.3 | FBAs with the simulated exchange rates

The simulated exchange rates (q_i) and the SEs were used as constraints for FBA. The lower and upper bounds of the exchange reactions were set to q_i - SE_i and $q_i + \mathrm{SE}_i$, respectively. In rare cases, it was not possible to fit Equation (1) to the concentration profiles due to high noise. In those cases, the exchange rates were left unconstrained. Biomass production was used as the objective function (reaction "`R_biomass_specific`").

3 | RESULTS

First, we defined a reference state for further simulations. We used experimental data and cell line specific genome-scale metabolic models of two CHO cell lines[20] in exponential phase of a batch culture. FBA was performed based on experimental data as described in Methods. The uptake reactions of cysteine and tryptophan were left unconstrained due to the lack of quantitative experimental data. We maximized the biomass production and obtained growth rates ($\hat{\mu}$) of 0.0384 and 0.0321 h^{-1} for cell lines `K1par-8mMAP` and `HYher-8mMCD`, respectively. These were considered as reference or "true" values for the subsequent simulations. The predicted uptake rates of cysteine and tryptophan exactly corresponded to the requirements for the synthesis of biomass and were used for further simulations.

3.1 | Low exchange rates are highly inaccurate

First, 1000 sets of concentration profiles of 23 extracellular metabolites (amino acids (AAs), ammonium, glucose, and lactate) were simulated by adding normally distributed, independent random errors with a mean of zero and a RSD ranging from 2% to 20% to all reference distributions. Three replicates were simulated (representing three independent experiments, e.g. three

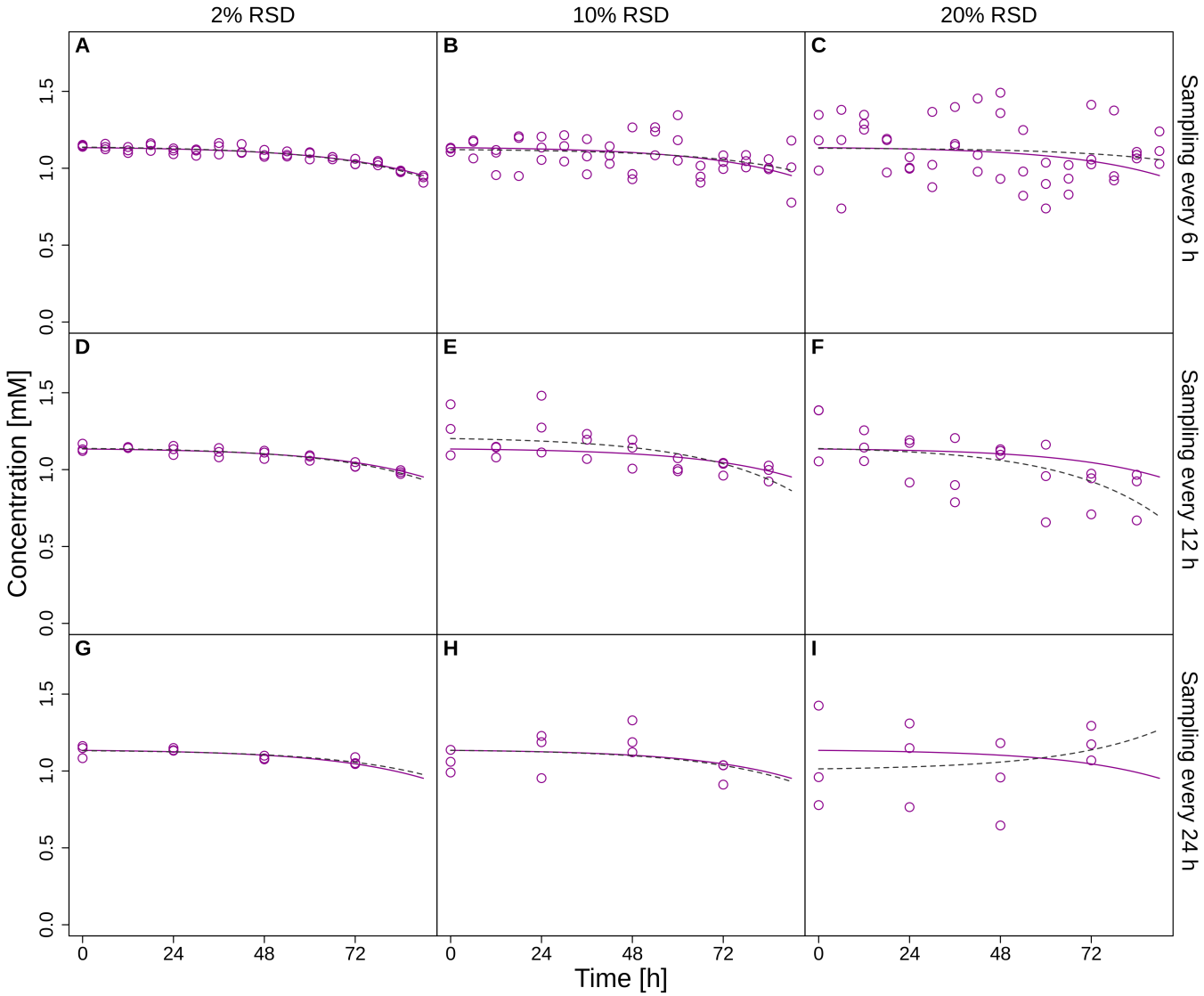


FIGURE 1 Simulated reference profiles (solid lines) and sampled perturbed concentrations (points) of histidine during the exponential phase of a batch cultivation at a sampling frequency of 6 h (panels A, B, C), 12 h (panels D, E, F) or 24 h (panels G, H, I) with 2% (panels A, D, G), 10% (panels B, E, H) or 20% (panels C, F, I) RSD of the replicate concentration measurements. Dashed lines represent fits to the data using Equation (1). Data is shown for cell line `K1par-8mMAP`.

bioreactor runs) and the distributions were sampled at three regular intervals - every 6 h, 12 h or 24 h (corresponding to 16, 8 or 4 time points throughout the batch, respectively). Figure 1 exemplarily shows simulated histidine concentration profiles for sampling intervals of 6 h, 12 h, and 24 h and RSDs of 2%, 10%, and 20%.

Next, we calculated exchange rates by fitting Equation (1) to the simulated concentration profiles and analytically evaluating its derivative. Low sampling frequency and high error can result in bad fits, that in some cases may even predict secretion instead of consumption (e.g. see Figure 1 , panel I).

We used the SEs to assess the accuracy of the calculated exchange rates (q_i). Even though the simulated concentration data has the same RSD for all metabolite concentrations, the relative standard errors (RSEs) of the calculated exchange rates (SEs divided by the absolute value of the exchange rates, q_i) markedly vary. More specifically, the closer the exchange rate is to zero, the higher is the RSE and the wider is its distribution (Figure 2 , Figure A2). Figure 2 a shows examples for a sampling frequency of 6 h (purple colours), where the simulated RSD of the concentrations is 2% and the medians of the RSEs of the rates are 1% for glutamine (high uptake), 2% for asparagine (medium uptake) and 6% for histidine (low uptake). 10% RSD of

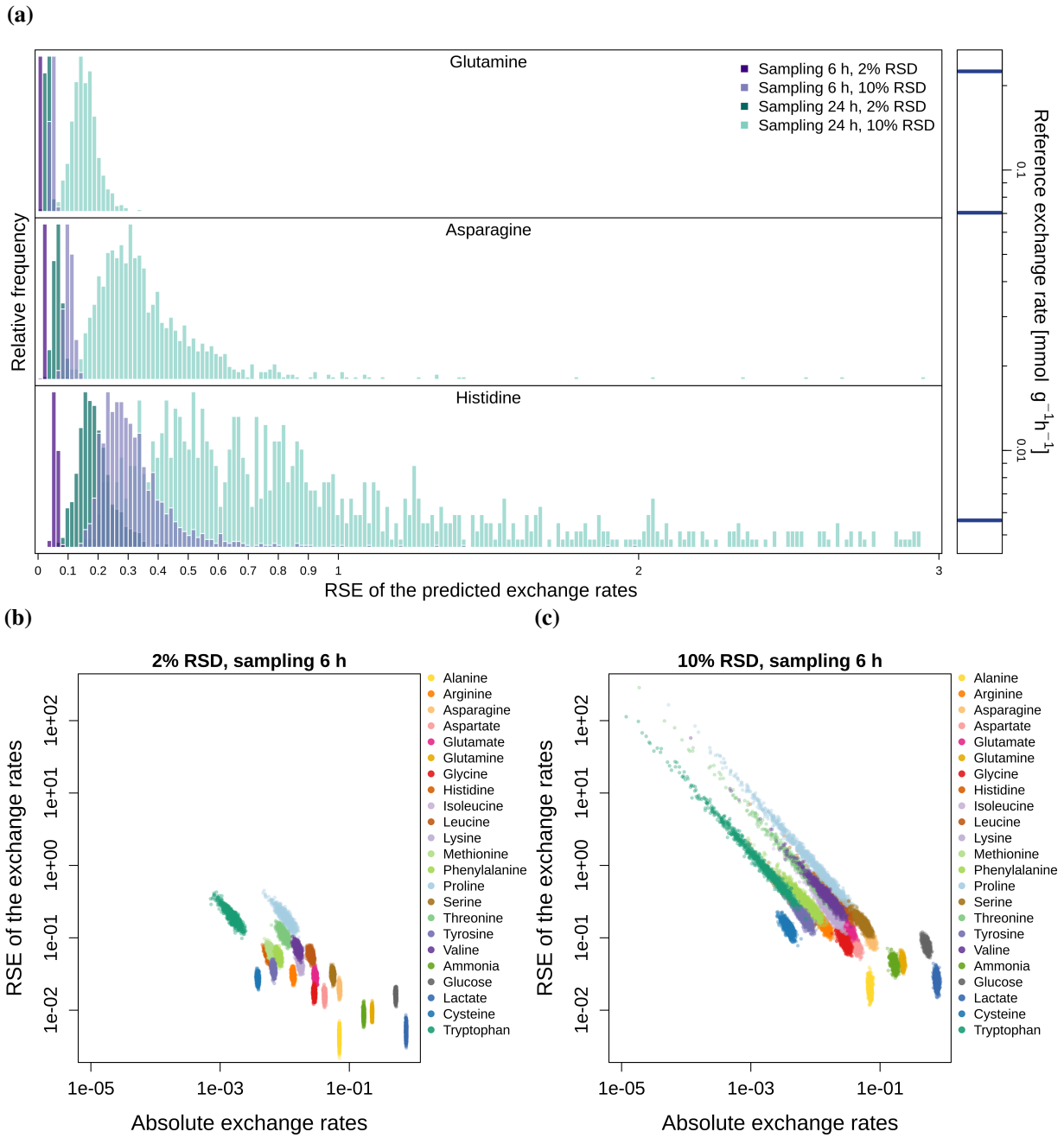


FIGURE 2 Panel a: distribution of the RSEs of three example uptake rates (for cell line `K1par-8mMAP`). The bars on the right side indicate the magnitudes of the uptake rates (top, glutamine; middle, asparagine; bottom, histidine). The lower the exchange rate, the higher the RSE and the wider the distribution. Purple colours show the distributions for sampling frequency of every 6 h, green colours for frequency of every 24 h, each for 2% and 10% RSD of the concentration measurements. Bottom: RSEs of the exchange rates as a function of the absolute values of exchange rates shown for 2% (panel b) and 10% (panel c) RSD of the concentration data.

concentration measurements leads to median RSEs of the rates of 5%, 10% and 29% for glutamine, asparagine and histidine, respectively. If the sampling interval is increased from 6 h to 24 h (green colours), the resulting RSEs of the exchange rates triplicate. For instance, with a 10% concentration error, the median error in the uptake of histidine increases up to 86% and the distribution has a very long tail. Figure A3 shows the relationship between RSDs of the metabolite concentrations and RSEs of the calculated exchange rates at different sampling frequencies for cell line `K1par-8mMCD`.

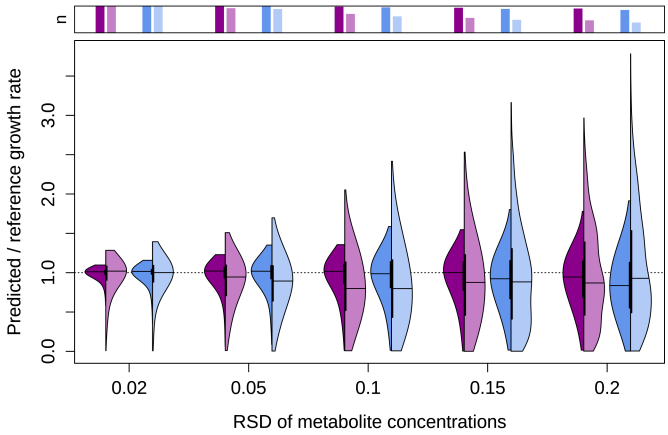


FIGURE 3 FBA predictions at different concentration RSDs for two cell lines (purple: `K1par-8mMAP`, blue: `HYher-8mMCD`). The left side of the violin plots corresponds to sampling every 6 h, the right side to every 24 h. The barplots above the plots indicate the fraction of feasible FBA solutions. The apparent cutoffs on the top at lower RSDs are artifacts due to the visualization.

3.2 | Growth predictions are strongly sensitive even to small metabolite concentration errors

We constrained the genome-scale metabolic model iCHO1766 with the computed exchange rates $[q_i - \mathrm{SE}_i, q_i + \mathrm{SE}_i]$ and predicted maximal growth rate with FBA. Figure 3 shows the ratios between predicted and reference growth rates. As expected, the distribution of growth rates gets wider with the increasing RSD in the metabolite concentrations and decreasing sampling frequency. At RSD of 20% and sampling frequency of 24 h, the predicted growth range varies between zero and three times the true value. Note that the distribution gets skewed, because the growth rates cannot be negative. Figure A4 displays the RSDs of the predicted growth rates as a function of concentration RSDs, showing that the data follows a square root law. That means that even small concentration errors lead to significant deviations of the growth rate predicted by FBA. However, growth rate RSD for concentration RSD of 20% is expected to be only twice higher than the one for concentration RSD of 5%.

Furthermore, with increasing RSD and decreasing sampling frequency we observe that there is a growing number of infeasible FBA problems (indicated by the bars on top in Figure 3). At RSD of 20% and sampling frequency of 24 h, more than half of the associated linear problems are infeasible. The predictions for the slower cell line (`HYher-8mMCD`, blue colours) are consistently worse, which can be explained by the on average lower uptake rates, which increases the RSEs of the calculated uptake rates (Figure A2).

Apart from the metabolite concentrations, the calculation of exchange rates takes the growth rate μ as an input (see Equation (1)). To check the impact of the error of the cell concentration measurements on the calculations of the exchange rates and the FBA predictions, we added noise not only to the metabolite concentrations, but also to the cell concentrations, where we introduced 6% RSD (the accuracy of ViCell XR). Figure A5 shows that the predictions are practically indistinguishable from the results in Figure 3 . Growth rates are typically high enough (0.02-0.04 h^{-1})[20] and thus can be determined with high accuracy. Hence, we concluded that measurement errors of the growth rate do not affect our simulations and need not be considered further.

Another variable in Equation (1) is the cell dry mass (contained in BM_0). However, this parameter is commonly determined only once for each particular cell line and reused for all further experiments, so the error in this parameter was not considered and the value was fixed to a previously measured value[20].

3.3 | Essential amino acids determine the growth rate predictions by FBA

Next, we determined which exchange rates have the largest influence on growth rate predictions with FBA. Based on Figure A3, we hypothesized that the metabolites with largest relative error have the largest impact. Thus, if those were measured accurately it would improve the prediction accuracy. To test this, we constrained the RSD of the concentrations to 2% for the top seven metabolites with highest RSEs of the rates. This strongly improves the prediction accuracy as measured by a decrease in the interquartile ranges from 0.05, 0.13, 0.24, 0.37 and 0.46 to 0.05, 0.09, 0.12, 0.16 and 0.20 for 2-20% RSDs, respectively (see

Figure 4 a), for K1par-8mMAP. The seven most error-prone uptake rates were uptakes for proline, tryptophan, threonine, valine, methionine, leucine and histidine, the last 6 of which are essential components. For cell line HYher-8mMCD see Figure A6.

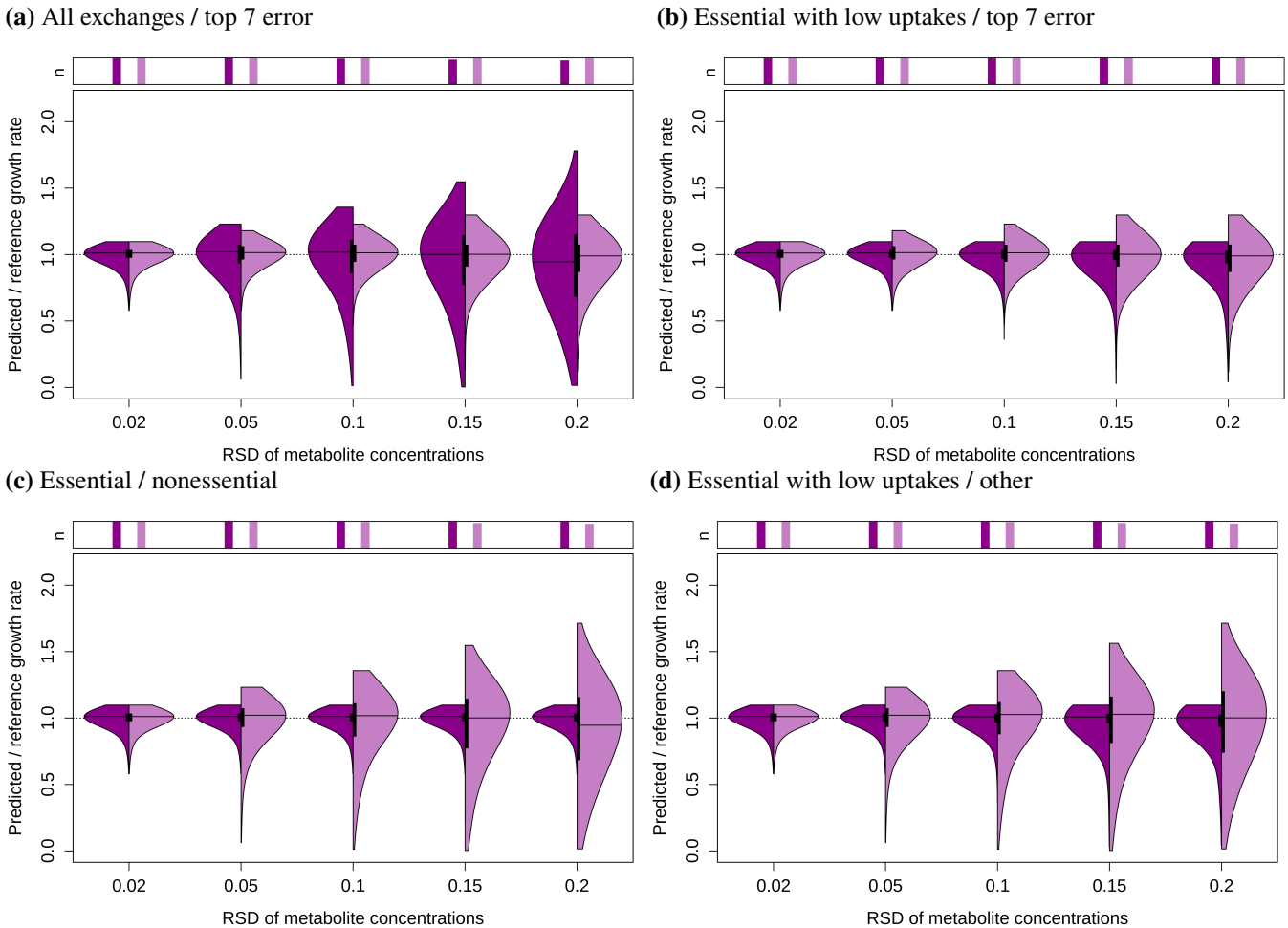


FIGURE 4 FBA predictions when different groups of metabolites are measured accurately. In both panel a and b, the right halves of the violin plots show results when the RSDs of metabolites with top 7 highest RSEs of the rates are set to 2%. In panel a, the left halves show results when RSDs of all metabolites are varied (the same as the left halves in Figure 3) and in panel b when errors of the amino acids with low normalized uptake rates are set to 2%. In panel c, in left halves, the RSDs of the essential AA concentrations are set to 2%, while all others are varied; in the right halves, the other way around. In panel d, in left halves, the RSDs of essential AAs with uptakes $< 1.5\times$ the biomass requirements are set to 2%, while all others are varied (the same as the left halves in panel b); in the right halves, the other way around. For simplicity, the data is shown only for cell line K1par-8mMAP at the sampling frequency of every 6 h. The barplots above the plots indicate the fraction of feasible FBA solutions. The apparent cutoffs on the top are artifacts due to the visualization.

As the uptake of essential AAs cannot be compensated by the uptake of other AAs, we next tested whether growth rate is limited mainly by the uptake rates of essential AAs. Therefore, we split the exchange rates into two groups - group 1: uptake rates of essential AAs (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, valine, cysteine, arginine) and group 2: all other uptake and secretion rates. We varied the RSD of the concentration data for one group, while keeping the RSD of the second group constant. When the error of the essential AA concentrations is kept constant (group 1), varying the error of the nonessential AAs has no effect on the FBA predictions (Figure 4 c and Figure A6 c, left halves). On the other hand, increasing the error of the essential AA concentrations leads to a large increase in the variability of the FBA solutions (Figure 4 c and Figure A6 c, right halves) and also has a minor effect on the number of infeasible FBA problems.

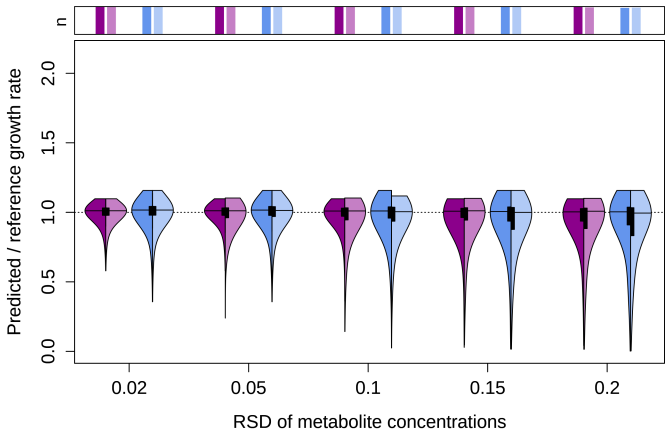


FIGURE 5 The most important exchange rates for FBA predictions are cell line specific - the left halves show the predictions when the AAs with low normalized uptakes are chosen according to the cell line specific data; in the right halves they are chosen based on the data of the other cell line. In both cases, the RSDs of the chosen AAs are set to 2%, the remaining RSDs are varied. Purple: `K1par-8mMAP`, blue: `HYher-8mMCD`. For simplicity, the data is shown only for the sampling frequency of every 6 h. The barplots above the plot indicate the fraction of feasible FBA solutions. The apparent cutoffs on the top are artifacts due to the visualization.

The uptake rates and biological functions of the essential AAs vary. Some of them are used only for the generation of cell biomass (see next paragraph), while others are also metabolized for other purposes (e.g. generation of energy), so we expected that their impact on the FBA predictions might differ.

Therefore, the next grouping was based on the biomass requirements of the essential AAs, which can be calculated as growth rate × AA coefficient in the biomass equation. Again, the exchange rates were divided into two groups: essential AAs with uptake rates smaller than 1.5× their biomass requirement (group 1; for cell line `K1par-8mMAP`: arginine, lysine, phenylalanine, threonine, tryptophan, cysteine, histidine; for cell line `HYher-8mMCD`: arginine, leucine, lysine, methionine, phenylalanine, tryptophan, cysteine); group 2: all other exchange rates. Figure 4 d and Figure A6 d show that varying the concentration errors of AAs in group 1 (right halves), which are used only for biomass formation, has much bigger impact on FBA predictions than varying the errors of metabolites in group 2 (left halves). Conversely, it makes only a tiny difference whether or not we keep the error of group 2 at 2% or leave the exchanges unconstrained (right halves of Figure 4 d and Figure A6 d vs. Figure A7). Finally, no such effects were apparent when we repeated the same procedure with random groupings (see Figure A8).

Together this data demonstrates that the quality of FBA predicted growth rates is (primarily) determined by the uptake rates of essential AAs, which are used mostly for generation of biomass. The set of the relevant AAs is cell line-specific (according to the interquartile ranges, Figure 5). However, measuring them accurately leads to a larger improvement than accurate measurements of the high-error AAs (Figure 4 b). Note that only three out of seven AAs (threonine, tryptophan, histidine for `K1par-8mMCD` and leucine, methionine, tryptophan for `HYher-8mMCD`) from the "low uptake" group are also in the "top 7 high errors" group.

3.4 | Accurate concentrations and high sampling frequency are needed for growth comparisons

Often, the goal of FBA is to compare (predicted) growth rates between different cell lines or conditions[32,17,33]. Therefore, we wanted to determine how often we are able to correctly identify the growth rate difference between two cell lines at the different sampling frequencies or RSDs of the concentration measurements. Again the difference from the reference FBAs was regarded as a "true" difference in the growth rates (0.0384 vs. 0.0321 h^{-1} resulting in a 0.0063 h^{-1} or 16% difference). As described earlier (Section 3.2), we ran 1000 FBAs for both cell lines at fixed sampling frequencies and metabolite concentration RSDs (see Figure 3). We compared growth rates of the two cell lines for all possible pairs, which resulted in 10^6 comparisons for each sampling frequency and RSD or less if there were FBAs with no feasible solution. In Figure A9 we normalized the predicted differences by the expected difference in growth rate (0.0063 h^{-1}) and plotted them as cumulative distributions. The value of 1 represents the expected difference in growth rates. All values below zero mean that the predicted growth rate difference was opposite to what was expected - the faster-growing cell line had a smaller predicted growth rate than the slower-growing cell line.

As expected, with increasing RSDs of the concentration data and decreasing sampling frequency, there is a growing number of solutions that incorrectly identify the faster growing cell line and the distributions get broader. For example at a daily sampling frequency and 20% RSD (Figure A9, right plot, green line), there are around 20% cases where the predicted difference is 5× higher than the "true" difference and almost 50% cases where the predicted difference is in the opposite direction.

Figure 6 a shows the (cumulative) probability of correctly predicting the faster growing cell line. It can be used to determine the measurement accuracy and sampling frequency that are needed to identify the faster cell line in at least 80% of cases. If the concentration measurements are accurate enough, sampling frequencies of 6 h and 12 h lead to correct identification in at least 80% of cases. This probability quickly decreases with larger RSDs in the concentration measurements. In the worst case, less than 60% FBAs correctly predict which cell line grows faster, which is only slightly better than chance.

We also checked whether it is necessary that the time points are equally spaced, as sampling every 6 h is impractical to do in a laboratory. Therefore we simulated data with different sampling schedules where the sampling time points are close together during the (working) day (e.g. every 4 or 2.5 h), followed by a longer interval without sampling (12 or 16.5 h). In addition, we simulated sampling every 12 h, but with six rather then three replicates. The results for these alternative sampling schedules lie within the shaded area in Figure 6 a.

All previous simulations were done with three replicates at each time point (as illustrated in Figure 1). We wanted to test whether doubling the replicate number leads to a considerable improvement in the predictions of growth rate. Figure A10 a shows the probability of identifying the faster growing cell line when the number of replicates is increased to six. The improvement compared to three replicates is very small. For comparison, we increased the number of replicates to 100 (Figure A10 b), although this is usually not feasible in a laboratory. The improvement is only modest, especially for the higher RSDs of metabolite concentrations. If the RSD is bigger than 15% and sampling is done on a daily frequency, there is still a more than 20% chance to incorrectly identify the faster growing cell line even with 100 replicates.

Finally, Figure 6 b shows the probability of correctly predicting the faster growing cell line when the RSDs of essential AAs with low uptakes are kept at 2% and only the RSDs of the remaining AAs are varied (as in Figure 4 d and Figure A6 d, left halves). For the highest sampling frequency (every 6 h), the faster cell line is identified in at least 80% cases for all RSDs.

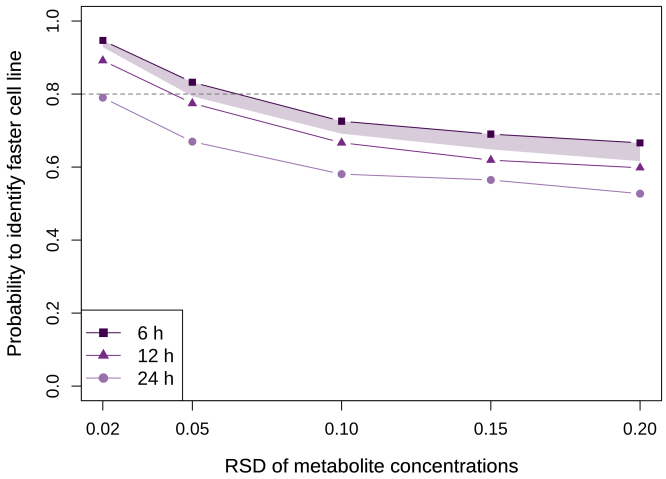
4 | DISCUSSION

Systems level analysis of CHO cells is needed in order to design rational engineering and optimization strategies to streamline cell line and process development. The genome-scale metabolic model of CHO[9] provides a basis for constraint-based metabolic modeling methods such as FBA. To successfully use these and obtain useful predictions, it is necessary to feed the model with accurate data, mainly uptake and secretion rates of extracellular metabolites. CHO cells are cultivated in complex media and consume or secrete numerous metabolites, including glucose, lactate, ammonium and all amino acids. Typically, more than 20 metabolite exchange rates have to be determined as an input for FBA.

Previously, we showed that accurate quantification of uptake and secretion rates is essential for good predictions of CHO growth rates by FBA; to achieve the required accuracy it is necessary to analyze sufficient number of time points throughout the relevant culture phases[20]. In this work we investigated how sampling frequency and the error of metabolite concentration measurements influence the accuracy of the calculated uptake and secretion rates and how the resulting error propagates into the growth rate predictions by FBA. A similar analysis was done by Goudar et al.[34] where error propagation from metabolite concentrations into rates and then into metabolic fluxes was analysed. However, only a small metabolic network was used and the experiments were done in perfusion cultures, where the procedure for determination of rates is different from batch cultivation, since they are in steady state.

In Figure 2 we showed examples of error propagation from concentration data to the calculated uptake rates. The smaller the uptake rate, the higher the relative error of the rate, which was already pointed out by Hädicke et al.[35]. The analysis also shows the importance of a sufficiently high sampling density throughout the culture. The comparisons in Figures 2 demonstrate that sampling only once per day (or 4 time points during the exponential phase, as is standard practice) leads to errors in the rates that are several times larger than the errors of the concentration data. There are also more extreme values (e.g. very high/low rates or even rates that point in the opposite direction), which are basically unusable for any meaningful analysis with FBA. This is nicely demonstrated in Figure 3 by the big variance and the high number of infeasible solutions. Consequently, insufficient measurement accuracy as well as low sampling frequency also make it almost impossible to reliably compare the predicted growth rates of two cell lines (Figure 6).

(a) RSD of all metabolites varied



(b) RSD of essential AAs with low uptakes kept at 2%

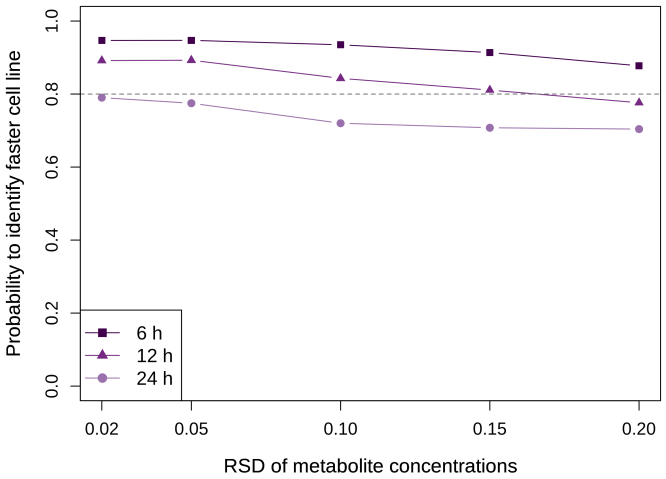


FIGURE 6 Percentage of FBAs where faster growing cell line is correctly identified at different concentration RSDs and sampling frequencies. In panel a, concentration RSDs of all metabolites are varied between 2-20%. In panel b, the RSDs of low uptake essential AAs are kept at 2% (as in Figures 4 d and A6 d), while the RSDs of the remaining metabolites are varied between 2-20%. The shaded area in panel a represents results with different sampling schedules when 4 samples per day are taken: regular sampling every 6 h; every 4 h with a 12 h gap; every 2.5 h with a 16.5 h gap; every 12 h with 6 instead of 3 replicates. For the schedules with a gap, two versions were done with different positioning of the gaps, e.g. sampling at hours 0, 4, 8, 12, 24, 30, 36 etc. or 0, 12, 16, 20, 24, 36 etc.

There are more than 20 uptake and secretion rates that are used as inputs for FBA. However, their impacts on the predictions of growth rates are vastly different. We expected that the prediction accuracy might improve if we measure those metabolites accurately that have the biggest relative errors of the rates. Even though this approach led to a big improvement in the predictions, further improvement was observed when we considered the biological roles of the metabolites. Even though all AAs are used as building blocks for the biomass, there are big differences in their exchange rates and metabolism. Some AAs can be synthesized in CHO, while others are essential and need to be taken up from the medium. Most of the essential AAs are consumed at very low rates and are predominantly used for the synthesis of the biomass, but some can also be partially catabolized. We expected that the amino acids which are solely used for the biomass synthesis would have the biggest impact on the growth rate predictions, which was confirmed by our analysis in Figure 4 . Interestingly, the uptake rates of glucose and glutamine, which are the main energy sources for CHO[36], have no impact on the growth rate predictions, suggesting that energy provision is not the limiting factor for the predicted growth rates. Consequently, our analyses show that to obtain accurate FBA predictions of growth rates, the essential amino acids with low uptake rates should be quantified with the highest possible accuracy and at a sufficient sampling frequency throughout the culture.

These aspects should be taken into consideration when planning experiments. The RSDs simulated in this paper represent the overall variation among replicates. This can be divided into biological and method variation. Hädicke et al.[35] provided a guide for how the contributions of these errors to the total variance can be analysed. Generally, the biological variance is bigger than the method variance and cannot be influenced, but it can be estimated from historical data, if available. On the other hand, method variance can be reduced by choosing an appropriate method that can accurately quantify the AAs that have the biggest impact on predictions and considering all other aspects that could impact the final data quality (e.g. storage of the samples at -80 °C immediately after sampling, avoiding freeze-thawing etc.). The measurement accuracy of 6 example amino acids (important for `K1par-8mMAP`) is discussed in the supplementary section A1.

When planning experiments, simulations similar to the ones presented in this paper can be performed to determine whether the accuracy is sufficient to answer the required scientific questions.

Our analyses demonstrate the difficulty of obtaining accurate uptake rates of the AAs with low uptakes, because very high measurement accuracy and sampling frequency are needed. If this is not possible, another option might be to switch to a different objective function than the maximization of growth. For example Chen et al.[37] propose using minimization of the uptake of

a nonessential metabolite (e.g. glucose or glutamine) as the objective function and fixing the growth rate to the experimental value. The minimal uptakes of the essential amino acids are estimated such that they can sustain the experimentally determined growth rate. Hence, for this approach the uptake and secretion rates of nonessential metabolites and the biomass composition might be the most important measurements.

5 | ACKNOWLEDGEMENTS

This work has been supported by the COMET center "acib: Next Generation Bioproduction", which is funded by BMK, BMDW, SFG, Standortagentur Tirol, Government of Lower Austria and Vienna Business Agency in the framework of COMET - Competence Centres for Excellent Technologies. The COMET-Funding Program is managed by the Austrian Research Promotion Agency FFG. Additional funding came from the PhD program BioToP (Biomolecular Technology of Proteins) of the Austrian Science Fund (FWF Project W1224).

6 | CONFLICT OF INTEREST

The authors declare no commercial or financial conflict of interest.

References

1. Walsh G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol* 2018; 36: 1136–1145.
2. Berting A, Farcet MR, Kreil TR. Virus susceptibility of Chinese hamster ovary (CHO) cells and detection of viral contaminations by adventitious agent testing. *Biotechnology and bioengineering* 2010; 106(4): 598–607.
3. Kim JY, Kim YG, Lee GM. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Applied microbiology and biotechnology* 2012; 93(3): 917–930.
4. Lai T, Yang Y, Ng SK. Advances in mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals* 2013; 6(5): 579–603.
5. Mathias S, Fischer S, Handrick R, et al. Visualisation of intracellular production bottlenecks in suspension-adapted CHO cells producing complex biopharmaceuticals using fluorescence microscopy. *Journal of biotechnology* 2018; 271: 47–55.
6. Maia P, Rocha M, Rocha I. In Silico Constraint-Based Strain Optimization Methods: the Quest for Optimal Cell Factories. *Microbiology and Molecular Biology Reviews* 2016; 80(1): 45–67. doi: 10.1128/MMBR.00014-15
7. Hamdi A, Széliová D, Ruckerbauer DE, Rocha I, Borth N, Zanghellini J. Key Challenges in Designing CHO Chassis Platforms. *Processes* 2020; 8(6): 643.
8. Xu X, Nagarajan H, Lewis NE, et al. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nature biotechnology* 2011; 29(8): 735.
9. Hefzi H, Ang KS, Hanscho M, et al. A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Systems* 2016; 3: 434–443.e8. doi: 10.1016/j.cels.2016.10.020
10. Yeo HC, Hong J, Lakshmanan M, Lee DY. Enzyme capacity-based genome scale modelling of CHO cells. *bioRxiv* 2019: 2019.12.19.883652. doi: 10.1101/2019.12.19.883652
11. Fouladiha H, Marashi SA, Li S, et al. Systematically gap-filling the genome-scale model of CHO cells. *bioRxiv* 2020: 2020.01.27.921296. doi: 10.1101/2020.01.27.921296
12. Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature Reviews. Microbiology* 2012; 10(4): 291–305. doi: 10.1038/nrmicro2737

13. Gutierrez JM, Feizi A, Li S, et al. Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nature Communications* 2020; 11(1): 1–10.

14. Kol S, Ley D, Wulff T, et al. Multiplex secretome engineering enhances recombinant protein production and purity. *bioRxiv* 2020: 647214.

15. Varma A, Palsson BO. Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/technology* 1994; 12(10): 994–998.

16. Varma A, Palsson BO. Metabolic capabilities of escherichia coli: I. Synthesis of biosynthetic precursors and cofactors. *Journal of theoretical biology* 1993; 165(4): 477-502.

17. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis?. *Nature biotechnology* 2010; 28(3): 245–248.

18. O'Brien EJ, Monk JM, Palsson BO. Using genome-scale models to predict biological capabilities. *Cell* 2015; 161(5): 971–987.

19. Varma A, Palsson BO. Metabolic capabilities of escherichia coli. II. Optimal growth patterns. *Journal of theoretical biology* 1993; 165(4): 503-522.

20. Széliová D, E Ruckerbauer D, N Galleguillos S, et al. What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. *Metabolic engineering* 2020; Accepted manuscript. doi: 10.1016/j.ymben.2020.06.002

21. Provost A, Bastin G, Agathos SN, Schneider YJ. Metabolic design of macroscopic models: application to CHO cells. In: IEEE. ; 2005: 2982–2989.

22. Dahodwala H, Nowey M, Mitina T, Sharfstein ST. Effects of clonal variation on growth, metabolism, and productivity in response to trophic factor stimulation: a study of Chinese hamster ovary cells producing a recombinant monoclonal antibody. *Cytotechnology* 2012; 64(1): 27–41.

23. Selvarasu S, Ho YS, Chong WP, et al. Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnology and bioengineering* 2012; 109(6): 1415–1429.

24. Sheikholeslami Z, Jolicoeur M, Henry O. Elucidating the effects of postinduction glutamine feeding on the growth and productivity of CHO cells. *Biotechnology progress* 2014; 30(3): 535-546.

25. McAtee Pereira AG, Walther JL, Hollenbach M, Young JD. 13C Flux Analysis Reveals that Rebalancing Medium Amino Acid Composition can Reduce Ammonia Production while Preserving Central Carbon Metabolism of CHO Cell Cultures. *Biotechnology Journal* 2018; 13(10): 1700518. doi: 10.1002/biot.201700518

26. Nicolae A, Wahrheit J, Bahnemann J, Zeng AP, Heinzle E. Non-stationary 13C metabolic flux analysis of Chinese hamster ovary cells in batch culture using extracellular labeling highlights metabolic reversibility and compartmentation. *BMC Systems Biology* 2014; 8(1): 50. doi: 10.1186/1752-0509-8-50

27. Calmels C, McCann A, Malphettes L, Andersen MR. Application of a curated genome-scale metabolic model of CHO DG44 to an industrial fed-batch process. *Metabolic engineering* 2019; 51: 9-19.

28. Calmels C, Arnoult S, Ben Yahia B, Malphettes L, Andersen MR. Application of a genome-scale model in tandem with enzyme assays for identification of metabolic signatures of high and low CHO cell producers. *Metabolic Engineering Communications* 2019; 9.

29. R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: 2018.

30. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology* 2013; 7(1): 74. doi: 10.1186/1752-0509-7-74

31. Beckman Coulter: Easy ... Fast ... Reliable Vi-CELL XR Cell Viability Analyzer Product Brochure. https://www.beckman.com/gated-media?mediaId={E4BE9210-2F92-4DA5-A867-A870D4C93AF2}; . Accessed: 2020-04-01.

32. Edwards JS, Covert M, Palsson B. Metabolic modelling of microbes: the flux-balance approach. *Environmental microbiology* 2002; 4(3): 133–140.

33. Jouhten P. Metabolic modelling in the development of cell factories by synthetic biology. *Computational and structural biotechnology journal* 2012; 3(4): e201210009.

34. Goudar CT, Biener R, Konstantinov KB, Piret JM. Error propagation from prime variables into specific rates and metabolic fluxes for mammalian cells in perfusion culture. *Biotechnology progress* 2009; 25(4): 986–998.

35. Hädicke O, Lohr V, Genzel Y, Reichl U, Klamt S. Evaluating differences of metabolic performances: Statistical methods and their application to animal cell cultivations. *Biotechnology and bioengineering* 2013; 110(10): 2633–2642.

36. Neermann J, Wagner R. Comparative analysis of glucose and glutamine metabolism in transformed mammalian cell lines, insect and primary liver cells. *Journal of cellular physiology* 1996; 166(1): 152–169.

37. Chen Y, McConnell BO, Dhara VG, et al. An unconventional uptake rate objective function approach enhances applicability of genome-scale models for mammalian cells. *NPJ systems biology and applications* 2019; 5(1): 1–11.

38. Hermann G, Schwaiger M, Volejnik P, Koellensperger G. 13C-labelled yeast as internal standard for LC–MS/MS and LC high resolution MS based amino acid quantification in human plasma. *Journal of Pharmaceutical and Biomedical Analysis* 2018; 155: 329–334.

39. Schwaiger M, Schoeny H, El Abiead Y, Hermann G, Rampler E, Koellensperger G. Merging metabolomics and lipidomics into one analytical run. *Analyst* 2019; 144(1): 220–229.

40. Ortmayr K, Schwaiger M, Hann S, Koellensperger G. An integrated metabolomics workflow for the quantification of sulfur pathway intermediates employing thiol protection with N-ethyl maleimide and hydrophilic interaction liquid chromatography tandem mass spectrometry. *Analyst* 2015; 140(22): 7687–7695.

How to cite this article: Diana Széliová, David E Ruckerbauer, Gunda Koellensperger Nicole Borth, Michael Melcher, and Jürgen Zanghellini (2020), Error propagation in constraint-based modeling of Chinese hamster ovary cells, , *2020;00:1–6*.

APPENDIX

TABLE A1 Initial concentration $[\hat{i}]_0$ of the medium components (mM) / biomass $\widehat{\mathrm{BM}}_0$ (g L^{-1}) and exchange rates $\hat{q}_i$ (mmol g^{-1} h^{-1}) for cell lines `K1par-8mMAP` (top half) and `HYher-8mMCD` (bottom half). For biomass, the column $\hat{q}_i$ corresponds to growth rate $\hat{\mu}$ (h^{-1}) of the reference FBA.

i	$[\hat{i}]_0$	$\hat{q}_i$	i	$[\hat{i}]_0$	$\hat{q}_i$	i	$[\hat{i}]_0$	$\hat{q}_i$	i	$[\hat{i}]_0$	$\hat{q}_i$	i	$[\hat{i}]_0$	$\hat{q}_i$
Ala	0.165	0.07	Glc	32.262	-0.523	Ile	3.292	-0.016	NH_4	3.206	0.167	Trp*	1.019	-0.002
Arg	1.488	-0.013	Gln	8.591	-0.225	Lac	4.313	0.756	Phe	1.571	-0.008	Tyr	0.891	-0.007
Asn	5.314	-0.07	Glu	2.668	0.03	Leu	5.164	-0.026	Pro	6.632	-0.011	Val	3.926	-0.016
Asp	1.819	0.041	Gly	1.384	0.029	Lys	2.707	-0.017	Ser	6.236	-0.056			
Cys*	0.379	-0.004	His	1.134	-0.006	Met	1.366	-0.006	Thr	3.646	-0.01	BM	0.040	0.038
Ala	0.151	0.036	Glc	34.831	-0.441	Ile	2.752	-0.011	NH4	0.797	0.114	Trp*	1.019	-0.001
Arg	2.081	-0.012	Gln	8.906	-0.101	Lac	7.047	0.477	Phe	1.330	-0.008	Tyr	0.985	-0.007
Asn	5.999	-0.087	Glu	1.826	0.013	Leu	4.048	-0.020	Pro	4.349	-0.020	Val	3.150	-0.017
Asp	1.381	0.020	Gly	0.128	0.019	Lys	3.016	-0.016	Ser	5.138	-0.046			
Cys*	0.379	-0.004	His	1.198	-0.006	Met	0.875	-0.004	Thr	3.016	-0.013	BM	0.045	0.032

Source: Széliová et al.[20]
* rates of Cys and Trp were predicted from the reference FBA. The initial concentrations are from the CD-CHO medium patent.

A1 REMARKS ON QUANTIFICATION OF AMINO ACIDS BY MASS SPECTROMETRY

Accurate absolute quantification of amino acids by mass spectrometry (MS) involves isotope dilution, regardless whether GC-MS (gas chromatography-mass spectrometry) or LC-MS (liquid chromatography–mass spectrometry) methods are implemented. Internal standardization is accomplished either by ^{13}C enriched or deuterated standards. Isotopic labelling of biomass proved to offer cost-effective pipelines for standard production[38]. In fermentation media, protein precipitation and efficient amino acid extraction is ensured by using polar solvents, both a prerequisite for accurate quantification. The internal standard is added as early as possible upon extraction, in order to compensate for losses during sample preparation (to ensure ideally 100% recovery). In this way, experimental precision in the low percent range can be achieved on a routine basis.

The accuracy of amino acid analysis in fermentation media can be validated by using the certified reference material SRM 1950, human plasma, as the same protocols for sample preparation and MS methods can be applied. While threonine and lysine can be accurately quantified by both platforms, LC-MS and GC-MS, respectively, accurate quantification of arginine, cysteine, phenylalanine and tryptophan demands LC-MS based analysis.

Using orthogonal GC-MS and LC-MS, threonine and lysine concentrations could be certified in the certified reference material for human plasma, SRM 1950, with expanded uncertainties of 5 and 10% respectively. The expanded uncertainty covers the measurand with approximately 95% confidence. Thus it expresses both the observed difference between the results from the LC-MS and GC-MS methods and their respective uncertainties, in addition to uncertainty components related to purity of the standards used for calibration. Using LC-MS methods alone, excellent experimental precisions ranging at 2-3% are technically feasible. The repeatabilities given in table A2 were observed by analysing independently prepared plasma samples[39]

by hydrophilic interaction chromatography (based on a zwitterionic stationary phase at pH 9) providing isomer separation of threonine and homoserine.

Decomposition upon derivatization in GC-MS hampers the analysis of phenylalanine, arginine and cysteine. Despite internal standardization and robotic just in-time derivatization, otherwise ensuring accuracy and precision, the detrimental decomposition processes result in poor recoveries and high uncertainties for these amino acids. Accordingly, for these amino acids only reference values based on LC-MS analysis were reported in SRM 1950. The expanded uncertainties ranged between 3 and 15%. Phenylalanine revealed an expanded uncertainty of 15% based on 3 independent LC-MS protocols. Using only one method as in the case of arginine, an expanded uncertainty of 3% was obtained, which corresponds to typical LC-MS precisions observed in other laboratories[39] (see table A2). The poor precision as obtained for cysteine in SRM 1950 can be improved, as in the certification campaign, no derivatization was implemented. In fact, accurate analysis of cysteine, containing a highly reactive thiol group, requires chemical protection during extraction. A protocol using N-ethylmaleimid protection[40] of the reactive thiol improved the precision achievable by LC-MS (5-10%).

Tryptophan is known to be a highly unstable compound, making proper standardization a challenge as such. This is exemplified through the fact that neither the certified reference standard for amino acids (NIST SRM 2389a) nor the human plasma SRM 1950 (see Table A2) gives concentrations for tryptophan. Special care has to be taken for the preparation of tryptophan standards. In GC-MS, accuracy is additionally jeopardized by the derivatization step. For calibration, tryptophan standards have to be prepared separately to the otherwise used SRM 2389a and freshly. Time between sampling, sample preparation and measurement has to be minimized. Taking all these measures, experimental repeatbilities of around 3% are achievable when using HILIC-MS and a minimum sample preparation (without preconcentration step by vacuum evaporation) [Mate Rusz, unpublished data].

TABLE A2 Uncertainties of six amino acids important for cell line `K1par-8mMAP`

Amino acid	Standard uncertainty % LC-MS (HILIC/RP) standards: SRM 2389a, tryptophan, N=5; 24 h isotope dilution using ^{13}C yeast as standard	Standard uncertainty % LC-MS (HILIC), plasma, protein precipitation, μM concentrations, N=5 independently prepared samples isotope dilution using ^{13}C yeast as standard	Expanded uncertainty % Certificate SRM 1950, NIST Human plasma reference material* isotope dilution using ^{13}C yeast as standard
Arginine	2	3	3
Cysteine**	4	5-10	15
Lysine	2	3	10
Phenylalanine	1-2	4	13
Threonine***	2-3	3	5
Tryptophan	2	3	No certified/no indicative value given

* SRM 1950: for lysine and threonine, certified concentrations are given in SRM 1950 (orthogonal GC-MS, LC-MS measurements available using isotope dilution); for arginine, cysteine and phenylalanine only reference values are given since only LC-MS data (using isotope dilution) were available; no data for tryptophan. The uncertainty provided with each value is an expanded uncertainty about the mean to cover the measurand with approximately 95% confidence; it expresses both the observed difference between the results from the methods and their respective uncertainties, in addition to Type B components related to purity of the standards used, consistent with the ISO/JCGM Guide.

** accurate quantification of cysteine requires derivatization upon extraction. NEM derivatization can be accomplished during the extraction/protein precipitation step[40].

*** The isomers threonine/homoserine are not always separated, only by zicHILIC.

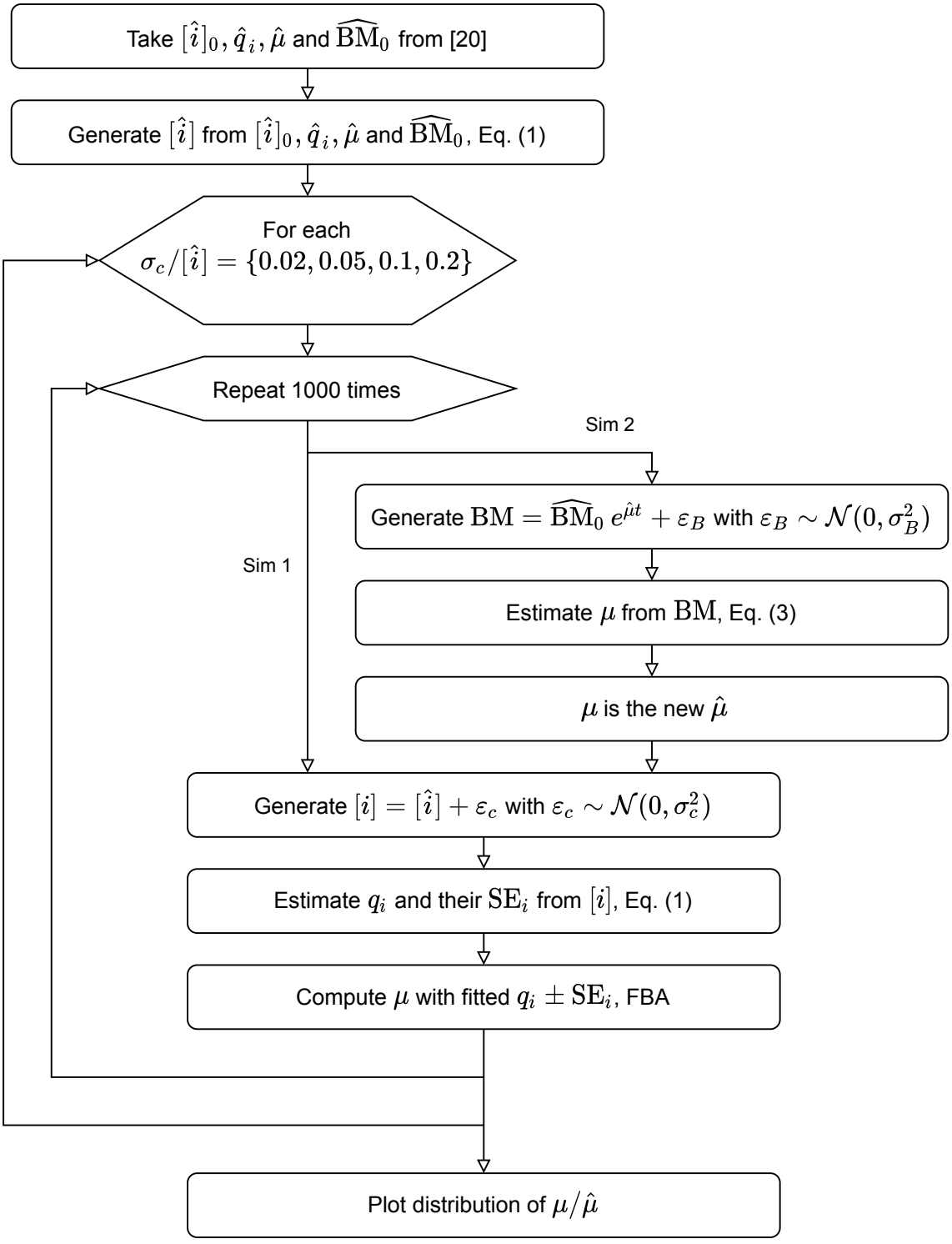


FIGURE A1 Workflow of the simulations.

(a) K1par-8mMAP **(b)** HYher-8mMCD

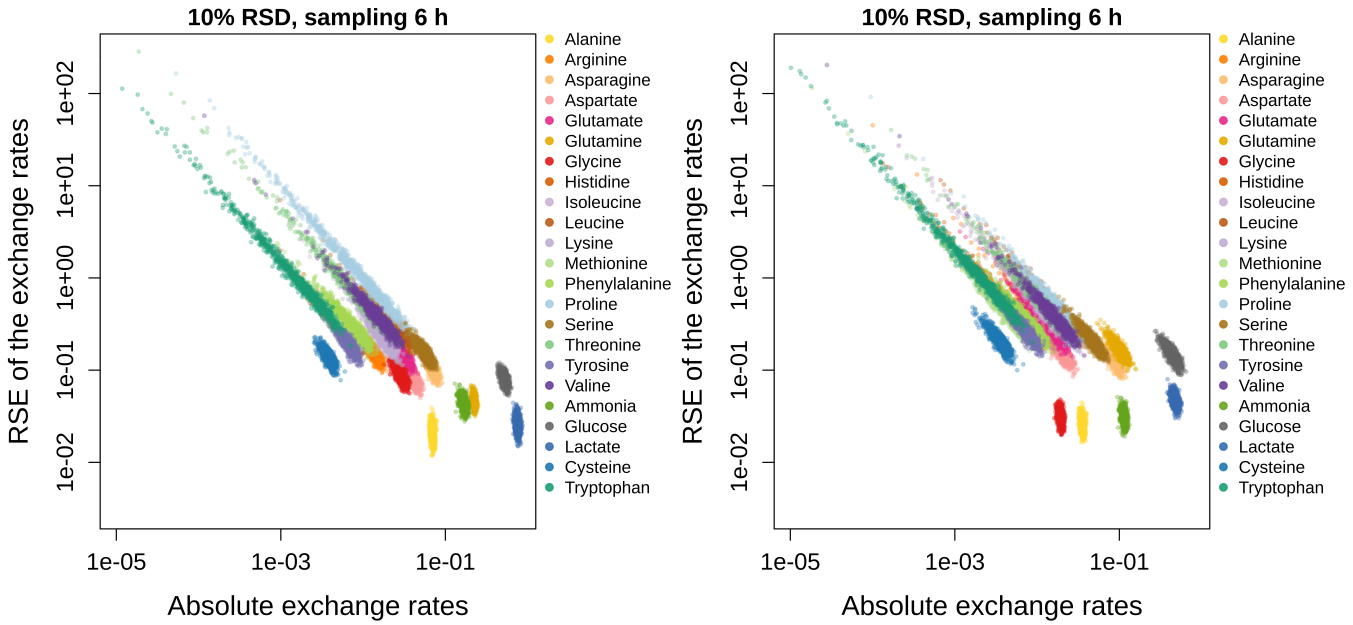


FIGURE A2 Related to Figure 2 . RSEs of the exchange rates as a function of the absolute values of exchange rates for 10% RSD of the concentration data for cell lines `K1par-8mMAP` (panel a) and `HYher-8mMCD` (panel b).

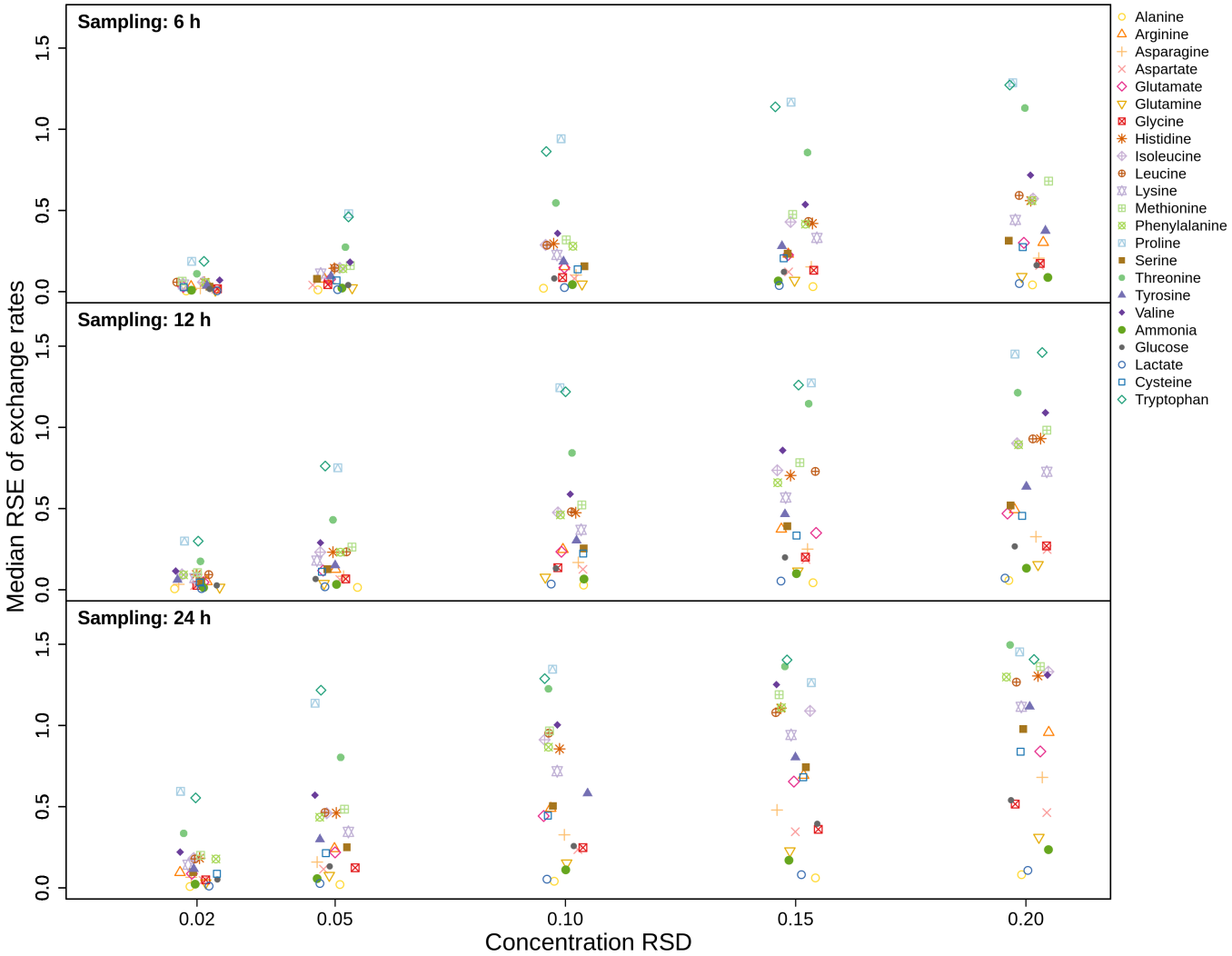


FIGURE A3 Relationship of the concentration RSDs and median RSEs of the exchange rates for cell line `K1par-8mMAP`.

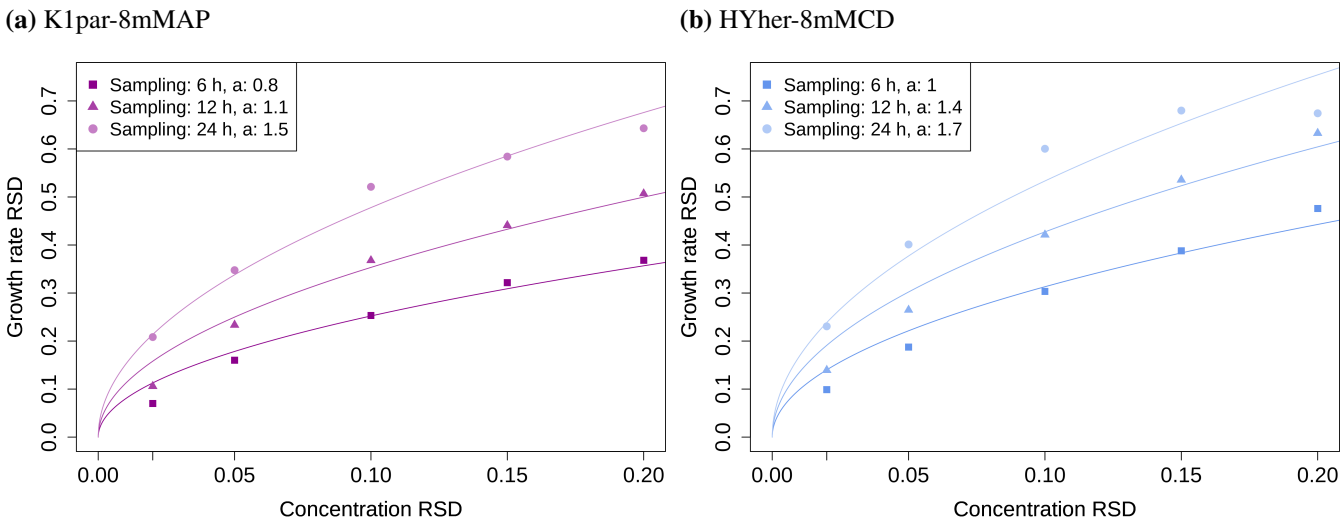


FIGURE A4 Related to Figure 3 . RSDs of growth rate predictions as a function of concentration RSDs at different sampling frequencies. The growth rate RSDs were fitted with the function $a\sqrt{x}$, where x are the concentration RSDs and a is a fitting parameter (the fitted values are shown in the legend). Panel a shows results for the cell line `K1par-8mMAP`, panel b for `HYher-8mMCD`.

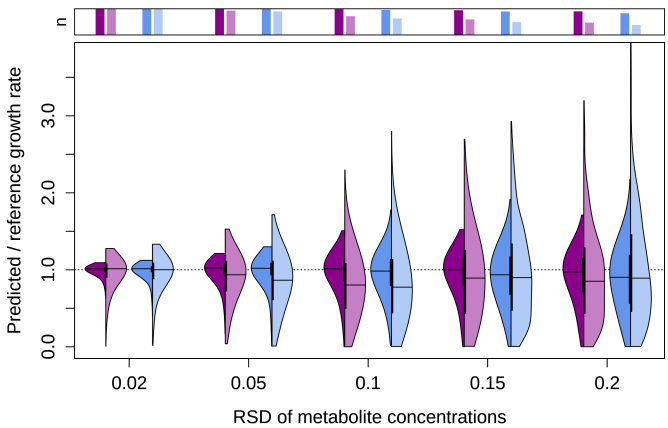


FIGURE A5 FBA predictions at different concentration RSDs for two cell lines (purple: `K1par-8mMAP`, blue: `HYher-8mMCD`). Compared to Figure 3 , not only metabolite concentrations, but also the cell concentrations are perturbed (with RSD of 6%). The left halves of the violin plots correspond to sampling every 6 h, the right halves every 24 h. The barplots above the plots indicate the fraction of feasible FBA solutions. The apparent cutoffs on the top at lower RSDs are artifacts due to the visualization.

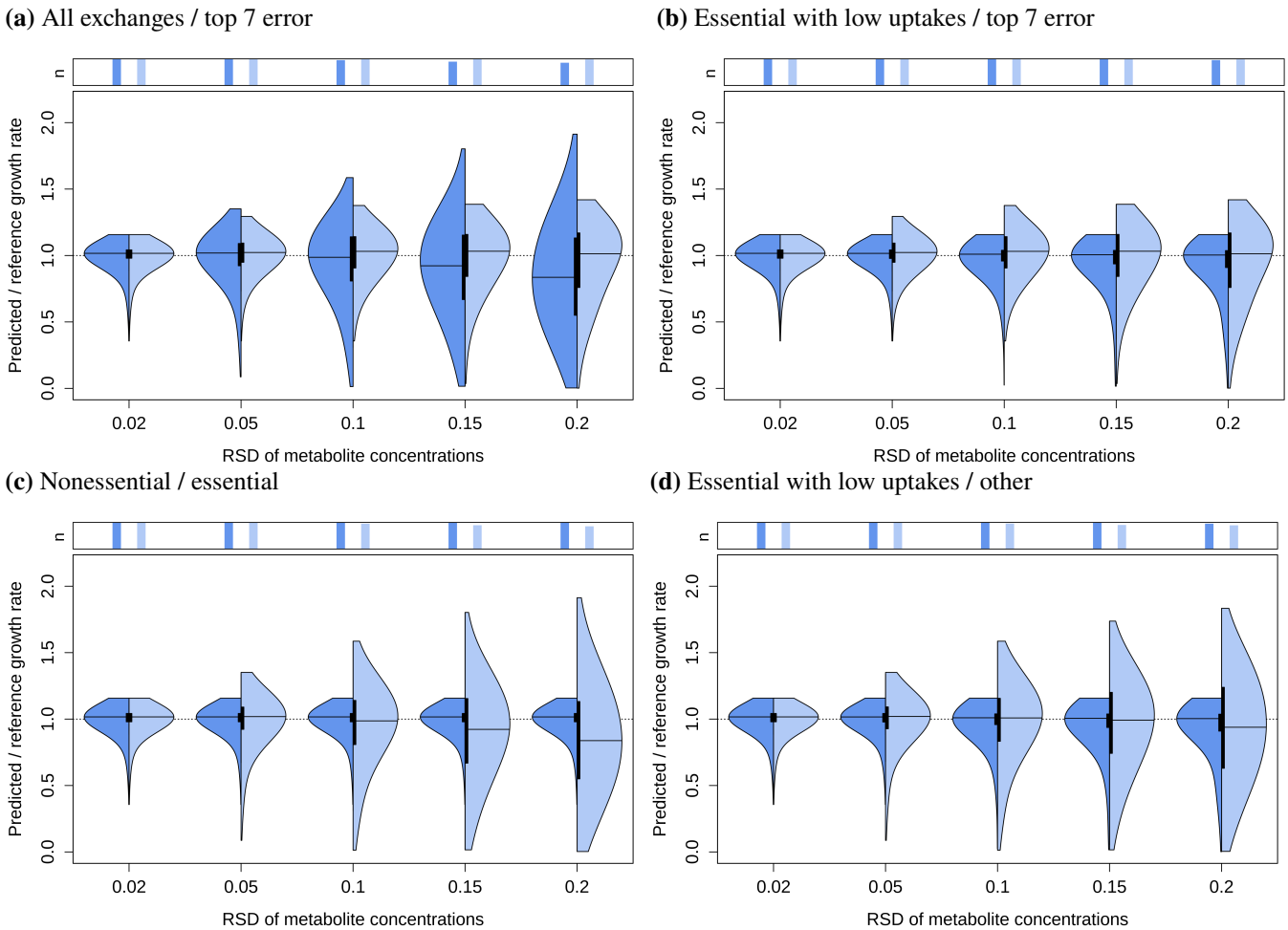


FIGURE A6 Related to Figure 4 . FBA predictions when different groups of metabolites are measured accurately. In both, panel a and b, the right halves of the violin plots show results when the RSDs of the metabolites with top 7 highest RSEs of the rates are set to 2% (tryptophan, isoleucine, threonine, methionine, proline, leucine, lysine). In panel a, the left halves show results when RSDs of all metabolites are varied (the same as the left halves in Figure 3) and in panel b when RSDs of the amino acids with low normalized uptakes are set to 2%. In panel c, left halves, the RSDs of the essential AA concentrations are set to 2%, while all others are varied; in the right halves, the other way around. In panel d, left halves, the RSDs of essential AAs with uptakes $< 1.5\times$ biomass requirements (arginine, leucine, lysine, methionine, phenylalanine, tryptophan, cysteine) are set to 2%, while all others are varied (the same as in panel b left halves); in the right halves, the other way around. For simplicity, the data is shown only for cell line `HYher-8mMCD` at the sampling frequency of every 6 h. The barplots above the plots indicate the fraction of feasible FBA solutions. The apparent cutoffs on the top are artifacts due to the visualization.

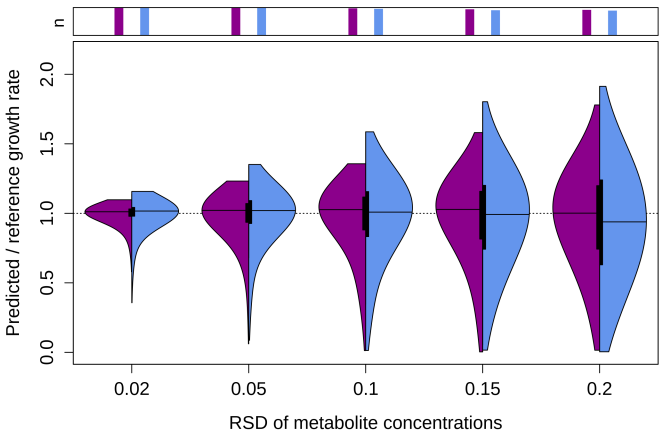


FIGURE A7 Related to Figure 4 d and Figure A6 d. FBA predictions when the RSDs of essential AAs with low uptakes (< 1.5× biomass requirements) are varied (same as in the right halves of Figure 4 d and Figure A6 d) and all other uptakes are left unconstrained. Purple: `K1par-8mMAP`, blue: `HYher-8mMCD`. For simplicity, the data is shown only for the sampling frequency of every 6 h. The barplots above the plot indicate the fraction of feasible FBA solutions.

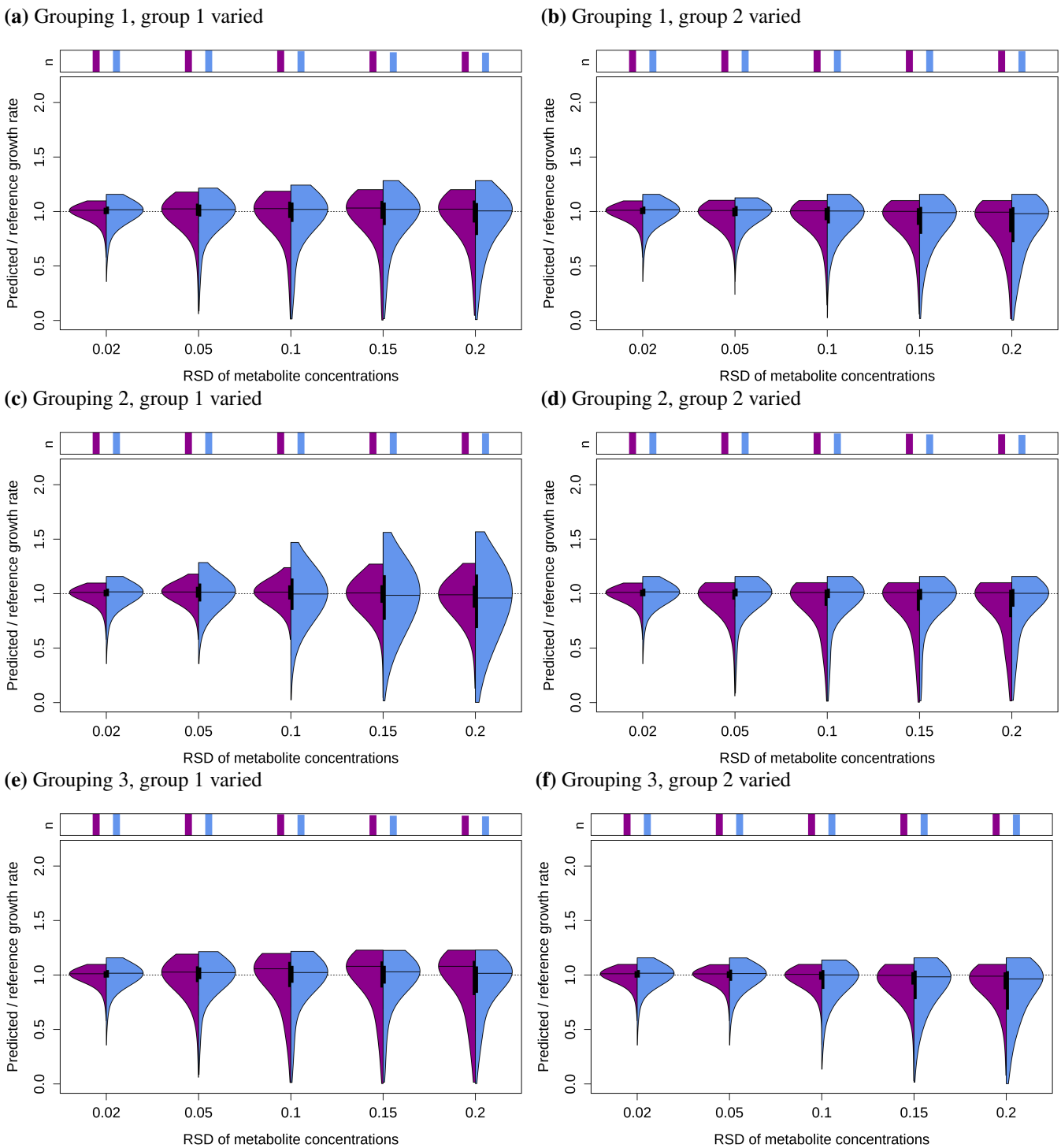


FIGURE A8 Related to Figure 4 . The metabolites are randomly split into two groups and the RSDs of group 1 are varied, while keeping the RSDs of group 2 at 2% (and the other way around). The random grouping was repeated 3 times for each cell line (labelled as Grouping 1-3). Purple: `K1par-8mMAP`, blue: `HYher-8mMCD`. For simplicity, the data is shown only for the sampling frequency of every 6 h. The barplots above the plot indicate the fraction of feasible FBA solutions.

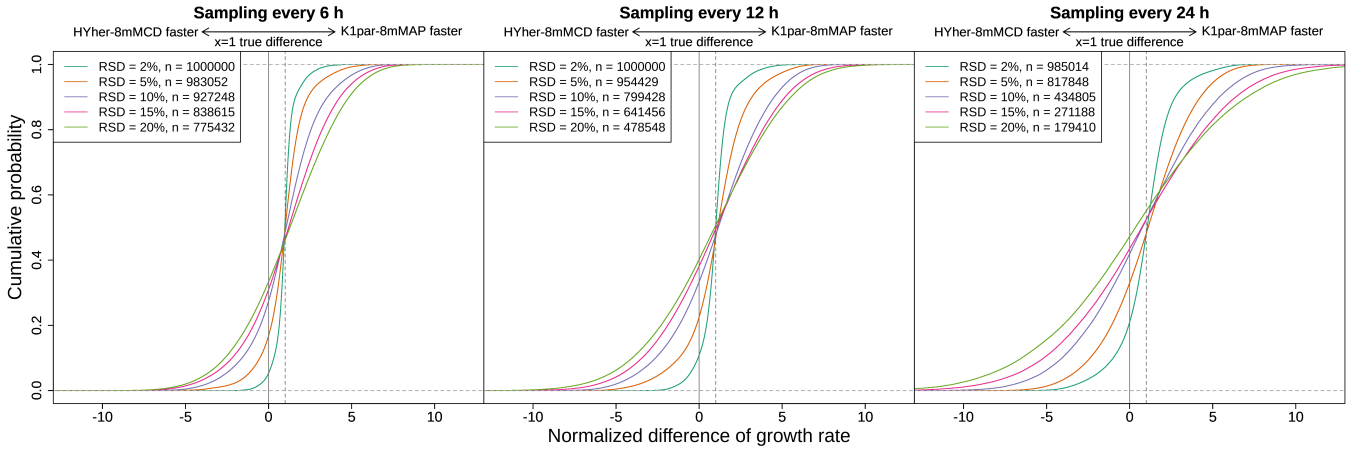


FIGURE A9 Related to Figure 6 a. The differences in growth rates predicted by FBA for two cell lines were normalized by the expected difference and plotted as a cumulative distribution. "RSD" refers to the error of the metabolite concentration data and "n" indicates the number of comparisons where FBA solutions were feasible for both cell lines.

(a) 6 replicates **(b)** 100 replicates

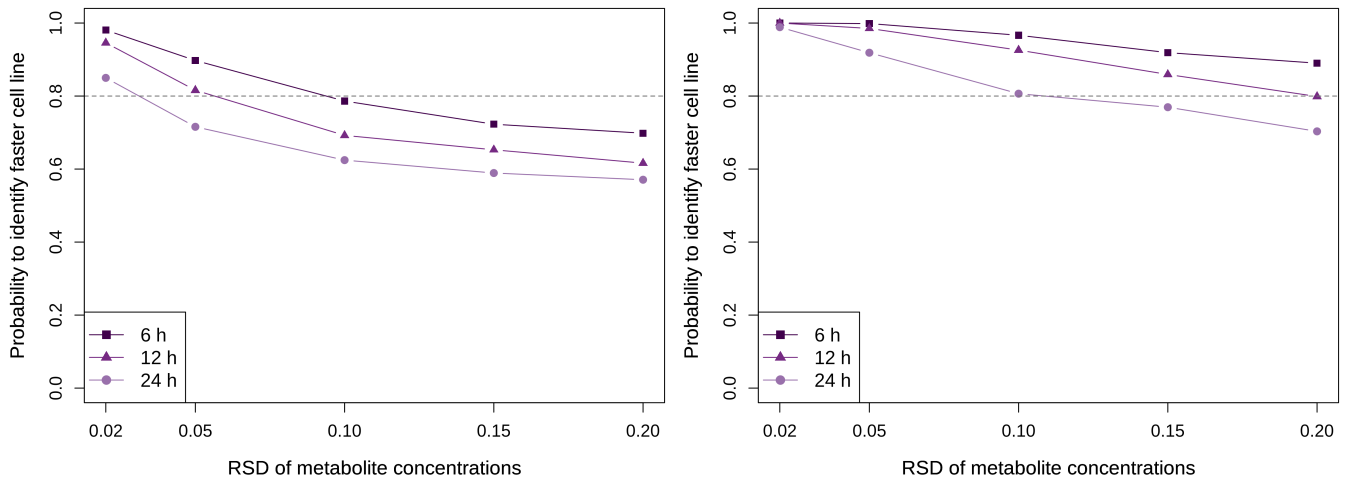


FIGURE A10 Percentage of FBAs where faster growing cell line is correctly identified when 6 (panel a) or 100 (panel b) replicates are analysed at each time point (instead of 3 replicates as in Figure 6 a).