bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Glycolysis revisited: from steady state growth to glucose pulses

David Lao-Martil[1] | Joep P.J. Schmitz[2] | Bas Teusink[3] | Natal A.W. van Riel[1,4]

[1]Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Noord-Brabant, 5612AE, The Netherlands

[2]DSM Biotechnology Center, Delft, Zuid-Holland, 2613AX, The Netherlands

[3]Systems Biology Lab, Vrije Universiteit Amsterdam, Amsterdam, Noord-Holland, 1081HZ, The Netherlands

[4]Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Noord-Holland, 1105AZ, The Netherlands

Correspondence
Natal A.W. van Riel, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Noord-Brabant, 5612AE, the Netherlands
Email: n.a.w.v.riel@tue.nl

Runnign title
Glycoysis: steady state and glucose pulses

Funding information
Dutch Research Council (NWO). Project number: 737.016.00

Abstract

Kinetic metabolic models of central metabolism have been proposed to understand how *Saccharomyces cerevisiae* navigates through nutrient perturbations. Yet, these models lacked important variables that constrain metabolism under relevant physiological conditions and thus have limited operational use such as in optimization of industrial fermentations. In this work, we developed a physiologically informed kinetic model of yeast glycolysis connected to central carbon metabolism by including the effect of anabolic reactions precursors, mitochondria and the trehalose cycle. A parameter estimation pipeline was developed, consisting of a divide and conquer approach, supplemented with regularization and global optimization. We show how this first mechanistic description of a growing yeast cell captures experimental dynamics at different growth rates and under a strong glucose perturbation, is robust to parametric uncer-

Abbreviations: [E]], enzyme concentration; ADP, adenosine diphosphate; ALD, aldolase; AMP, adenosine monophosphate; ATP, adenoside triphosphate; CCM, central carbon metabolism; CFD, computational fluid dynamics; CO_2, carbon dioxide; *E.coli*, *Escherichia coli*; ENO, enolase; F6P, fructose 6-phosphate; FF, feast famine; G6P, glucose 6-phosphate; GP, glucose pulse; GSM, genome-scale model; K_{cat}, catalytic constant; K_m, m_{ATP}, ATP maintenance rate; michaelis constant; NAD, nicotinamide-adenine-dinucleotide; O_2, oxygen; ODE, ordinary differential equation; *P.chrysogenum*, *Penicillium chrysogenum*; P/O ratio, phosphate/oxygen ratio; PDC, pyruvate decarboxylase; PEP, phosphoenolpyruvate; PFK, phosphofructokinase; PGI, phosphoglucoisomerase; PPP, pentose phosphate pathway; PTM, post-translational modifications; q_{ATP}, ATP production rate; q_{CO2}, carbon dioxide transport rate; q_{O2}, oxygen transport rate; RQ, respiratory quotient; *S.cerevisiae*, *Saccharomyces cerevisiae*; SS, steady state; sum_{AXP}, sum of ATP, ADP and AMP; TCA cycle, Tricarboxylic acid cycle; TPS1, trehalose phosphate synthase 1; V_{max}, maximum reaction rate; VVUQ, verification, validation and uncertainty quantification.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tainty and explains the contribution of the different pathways in the network. Our work suggests that by combining multiple types of data and computational methods, complex but physiologically representative and robust models can be achieved.

Keywords — Glycolysis / Growing Cell / Kinetic Modeling / Parameter Estimation / *Sacchamoyces cerevisiae*

Introduction

Saccharomyces cerevisiae, commonly known as baker's yeast, is a prominent cell factory for the biotechnology industry (Nielsen *et al*, 2013). In large-scale fermentations, dynamic gradients expose yeast cells to rapid nutrient changes in their extracellular environment, which in turn will impact intracellular metabolic regulation (Enfors *et al*, 2001; Haringa *et al*, 2016). To understand this dynamic stress response process, mechanistic models representing intracellular carbon metabolism kinetics have been developed (van Eunen *et al*, 2012; Smallbone *et al*, 2013; van Heerden *et al*, 2014). However, upscaling these models from lab to industry has proven challenging and restricted to simplified mechanistic growth models with little information on intracellular dynamics (Wang *et al*, 2020; Sarkizi Shams Hajian *et al*, 2020).

This could be partly explained because models lacked important physiological information. In the fields of pharmacology and medicine, validation of physiological realism is a necessary step in development of credible mechanistic model (Pruett *et al*, 2020). To develop physiologically representative yeast models, several aspects are to be considered. Current state of the art models study glycolysis in isolation (van Eunen *et al*, 2012; Smallbone *et al*, 2013; van Heerden *et al*, 2014), but should consider the complex cellular context it interacts with, for example by representing the trehalose and tricarboxylic acid (TCA) cycles.

The situation regarding biomass synthesis is an important example. Whereas maximization of biomass synthesis is the physiological function to which genome scale models (GSM) are optimized (Yasemi and Jolicoeur, 2021), it has been neglected in yeast kinetic networks, resulting in models that cannot represent a growing cell and thus poorly represent an important physiological state that comes with importable demands (van Heerden *et al*, 2015). Considering sink reactions for glycolytic intermediates is of upmost importance, as they were shown to be central in achieving the steady state (Teusink *et al*, 2000).

Moreover, variables alien to carbon flux regulate glycolysis as well. For instance, oxygen concentration determines if respiration or fermentation is performed (Otterstedt *et al*, 2004) and is routinely quantified in both lab and industrial scale bioreactors (Canelas *et al*, 2011; Haringa *et al*, 2016), but has not been used to constrain dynamic models. Furthermore, cofactors NAD and ATP are required to carry out multiple glycolytic reactions, but their availability cannot be taken for granted. ATP homeostasis is challenged under strong dynamic perturbations, where the inosine salvage pathway is used as transient store (Walther *et al*, 2010). In addition, ATP required for maintenance changes under different growth conditions (Chen and Nielsen, 2019). Still, a yeast model considering the effect of these variables does not exist.

Recently published data and tools make implementation of yeast models in a physiological context more feasible. The first perturbations studies in yeast were restricted to a single glucose pulse perturbation of 1 g L^{-1} (Theobald *et al*, 1997), but experimental quantification has been extended to larger perturbations that result in profound intracellular dynamics, including possible nonlinear effects of stress response (Walther *et al*, 2010; van Heerden *et al*, 2014). Furthermore, data on different growth steady states (SS) (van Eunen *et al*, 2012; Canelas *et al*, 2011) and industrially

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

relevant feast/famine (FF) regimes (Suarez-Mendez *et al*, 2014) has also been generated. New experimental variables have been quantified, such as most metabolic species in central carbon metabolism (CCM) (Canelas *et al*, 2011), dynamic flux profiles (Suarez-Mendez *et al*, 2017), and proteome composition (Chen and Nielsen, 2019; Elsemman *et al*, 2022).

These data can now be used to quantify model parameters within a physiological range. Previous work showed that kinetic constants measured *in vitro* in isolated enzymes assays might not be representative of *in vivo* behavior (Teusink *et al*, 2000; Davidi and Milo, 2017). To overcome this issue, model parameters can be directly estimated to fit *in vivo* metabolomic and fluxomic data (Davidi and Milo, 2017). Few works estimated part of the yeast kinetic model parameter set in this way (Smallbone *et al*, 2013; Pritchard and Kell, 2002; Van Riel *et al*, 1998), but the data available has notably increased since then.

To address the abovementioned challenges, here we conceived a kinetic model of yeast glycolysis and trehalose cycle that considers the physiological effect of growth rate, gas exchange, ATP synthesis and maintenance, by developing a state-of-the-art approach for parameter identification. We show how this first mechanistic description of a growing yeast cell captures experimental dynamics at different growth rates and under a strong glucose perturbation, is robust to parametric uncertainty and explains the contribution of the different pathways in the network. Our work suggests that combining multiple types of data and computational methods, complex but physiologically representative and robust models can be achieved, and simultaneously points at specific locations in glycolytic models where regulation mechanisms are missing.

Results

| An integrated, physiologically informed kinetic model of yeast glycolysis

Existing yeast glycolysis models studied this pathway in isolation (Smallbone *et al*, 2013; van Heerden *et al*, 2014). To transit from pathways in isolation, to pathways embedded in a growing cell, inclusion of physiological variables informative of the bioreactor environment is crucial (Lao-Martil *et al*, 2022). The required data to implement these variables is now available for *S. cerevisiae*. Carbon flux distribution in central metabolism and secretion of O_2 and CO_2 were determined at different dilution rates in chemostat (Canelas *et al*, 2011). Additionally, growth and non-growth -associated maintenance (GAM and NGAM, respectively) were determined and extensively implemented in GSMs and can be used to estimate the ATPase activity (Beeftink *et al*, 1990; Lu *et al*, 2019).

These variables have been included in the model developed in this work. Carbon metabolism includes glycolysis, trehalose cycle, glycerol branch and sink reactions to biomass and other by products (Fig 1). Production of CO2 (q_{CO2}) consisted of the sum of what is released in the pyruvate decarboxylase (PDC) reaction and all the carbon moles that are processed in the mitochondria. The uptake of oxygen (q_{O2}) was estimated using the q_{CO2} and the experimental respiratory quotient (RQ) value. Finally, the synthesis of ATP (q_{ATP}) consisted of glycolytic production and mitochondrial contribution, estimated considering q_{O2} and P/O ratio (Verduyn *et al*, 1991). The maintenance consumption of ATP (m_{ATP}) was estimated using the GAM and NGAM determined in GSMs (Lu *et al*, 2019).

The model consists of a series of nonlinear ordinary differential equations (ODE), where each mass balance describes a metabolite concentration, that changes over time according to reaction rates. Reactions that cannot be described by a single enzyme, such as the rates in which glycolytic intermediates are taken up to form biomass, are described by phenomenological expressions that closely resemble experimental data. The change in enzyme concentrations at different dilution rates found in (van Hoek *et al*, 2000) is adjusted for each enzyme. Model parameters were estimated to fit experimental data (see later sections). For further details on model implementation, parameters and

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

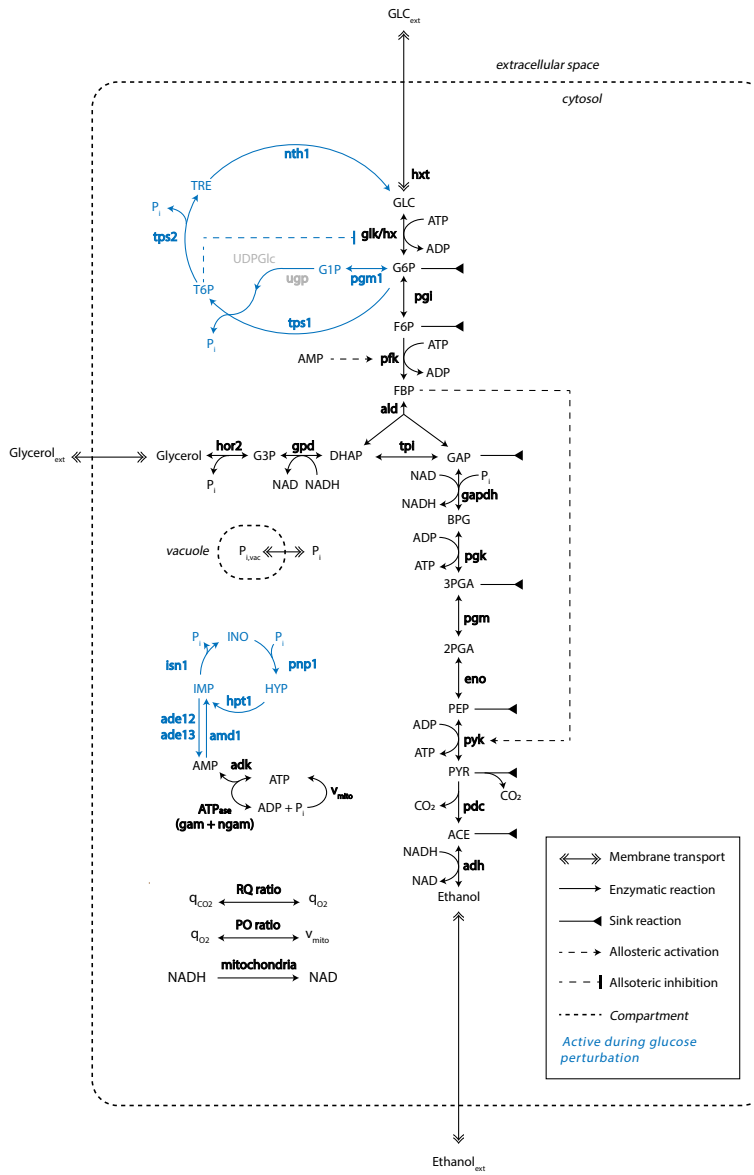


FIGURE 1 An updated central carbon metabolism *S. cerevisiae* model considering variables of physiological relevance. Black colored reactions are active for all simulations, blue only during the glucose perturbation and gray are lumped. Metabolites are connected by the reactions in the model (continuous lines), which are catalysed by enzymes (bold). Allosteric regulation is shown by semi-continuous arrows and compartments by semi-continuous lines.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

kinetic expressions, see the materials and methods and supporting information.

The model reproduced physiological properties of a growing cell in a bioreactor

The resulting model predicted experimental dynamics, differentiating respiratory and respirofermentative metabolism. At growth rates below 0.2 h^{-1}, carbon uptake was directed to the sink reactions, which account for biomass synthesis and TCA cycle (Fig 2B). Above this threshold, metabolism gradually shifted to ethanol production which eventually became the main carbon product. In line with this, q_{CO2} closely resembled q_{O2} at low growth rates but became predominant above 0.2 h^{-1}, discerning respiratory from respirofermentative metabolism (Fig 2C). Likewise, ATP was produced by respiratory metabolism at low growth rates and partially fermentative at high growth rates (Fig 2D). Maintenance reaction rate was close to a GAM of 40 mmol of ATP per g of biomass during the respiratory state but increased above this level as fermentation became predominant. A mismatch occurred at 0.2 h^{-1}, when carbon flux was notable underestimated. We will discuss this later.

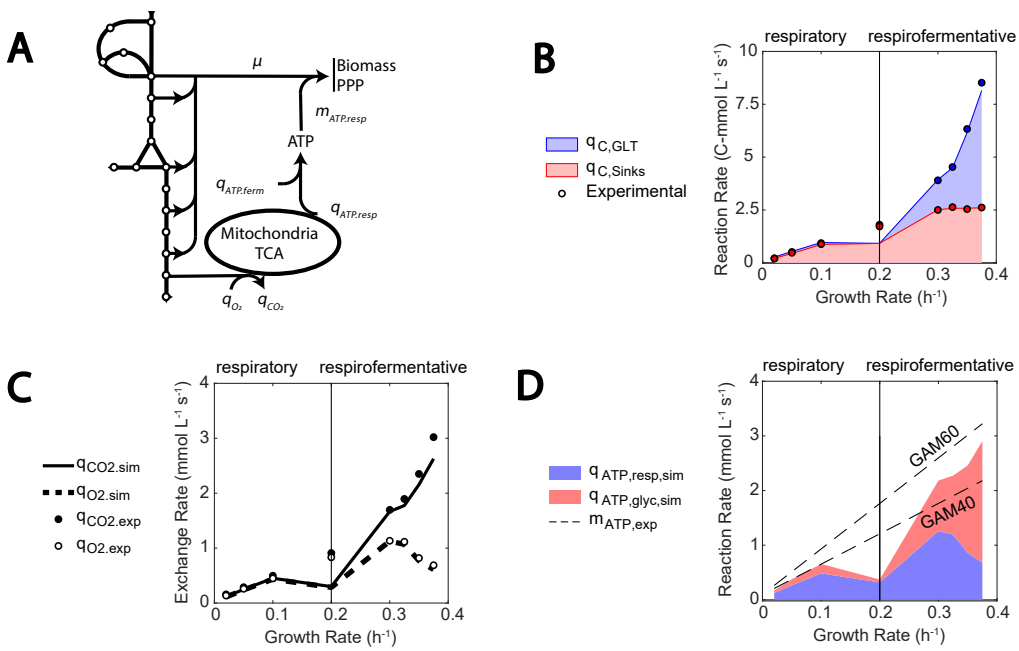


FIGURE 2 The model reproduces physiological properties of a growing cell in a bioreactor. (A) Simplified model scheme highlighting the role of sinks for anabolic reaction precursors, PPP and mitochondria (g6p, f6p, gap, pep, pyr, ace). (B) Carbon flux entering the sink reactions (red) compared to the total carbon uptake (blue). (C) Simulated exchange rate of CO_2 (continuous line) and O_2 (semi-continuous line). Experimental data shown as bullet points. (D) ATP produced by respiration (blue area) and fermentation (red area). Semi continuous lines point at the theoretical values calculated using the NGAM and GAM (40 and 60 mmol g_{DW}^{-1}) found in genome scale representations (Lu *et al*, 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In conclusion, the agreement between model simulations and experimental data suggests that the model can work within the physiologically relevant region studied here. The implementation of the growth rate dependency seems essential. For instance, if only the ATP produced by glycolysis is considered, the overall maintenance needs cannot be matched at the different growth rates. This opens the possibility for kinetic models to be used in new setups, such as different growth rates or respiration/fermentation regimes, by using variables such as gas exchange rate as constraints, rather than as validation.

| Parameter estimation: A problem decomposition approach supplemented by regularization, parameter space sampling and cost function balancing

Quantification of the parameter set was performed by optimizing the model fit to multiple experimental setups and data types (van Heerden *et al*, 2014; Canelas *et al*, 2011; van Hoek *et al*, 2000). Nonetheless, the high number of unknown kinetic parameters made this task challenging due to multiple local optimum and ill-conditioning. To deal with non-identifiability problems in large kinetic models, multiple optimization methods have been proposed and evaluated (Villaverde *et al*, 2019). For instance, the so-called divide and conquer method aimed to analyze undetermined biochemical networks despite parameter uncertainty (Kotte and Heinemann, 2009). This problem decomposition approach splits the model in smaller parts where parameters are estimated and later assembled in a complete network.

To develop the S. cerevisiae model in this study, this problem decomposition method was adapted and supplemented with global optimization and regularization to improve convergence, akin to the approach described in (Gábor and Banga, 2015). Global optimization algorithms meet the danger presented by local minima by starting parameter estimation from a set of initial parameter guesses (Moles *et al*, 2003). Regularization methods reduce ill-conditioning by introducing prior knowledge on the parameters (Engl *et al*, 2009). Additionally, to control that overfitting did not occur, weighting factors were used to leverage the components of the cost function.

The parameter estimation pipeline is illustrated for the well-studied enzyme enolase (ENO) in (Fig 3). In the first stage, the divide and conquer problem decomposition approach was applied to glycolytic enzymes. As suggested in (Kotte and Heinemann, 2009), steady state data was used (Canelas *et al*, 2011; van Hoek *et al*, 2000). A regularization factor was added in the cost function, so that parameters resembled prior knowledge, provided that the model error was not increased (Fig 3A). The benefits of global optimization and regularization can be appreciated for parameters $V_{max,ENO}$ and $K_{M,F6P,PFK}$ in Fig 3B and 3C, respectively. A parameter for phosphofructokinase (PFK) is shown since this is the most complex enzyme in the pipeline, and results differed according to enzyme kinetics complexity. While for ENO most initial parameter samples ended in the same parameter value as the literature one, the regularization factor brought estimated paraemters closer to literature values (Fig 3B). For PFK, global sampling revealed the existence of parameter dependencies (Fig 3C), and convergence towards literature values was achieved upon regularization. Since enzymatic kinetic constants vary within a large range, the parameters estimated were transformed from linear to logarithmic scale, as this has shown to be more reliable, also with global optimization methods (Kreutz, 2016; Villaverde *et al*, 2019).

In the second stage, the developed models were assembled in a complete glycolysis model with trehalose cycle and cofactor metabolism. The weighting factors in the cost function were balanced ad hoc, until model fit for multiple data could not be further improved. This resulted in pareto fronts, such as the one for ENO (Fig 3D), where the fit for phosphoenolpyruvate (PEP) concentration in steady state and the 110 Mm glucose pulse was balanced (Fig 3E and 3F, respectively). See the supplementary information for an overview of the divide and conquer parameter estimation results, enzyme by enzyme.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

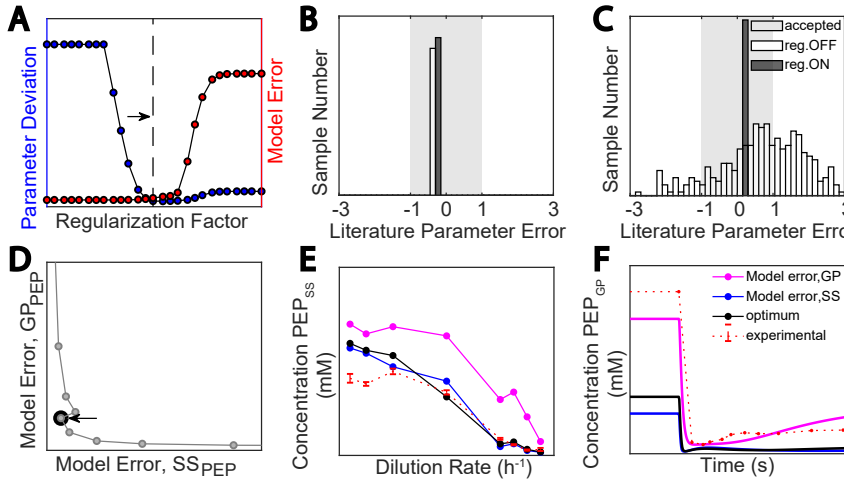


FIGURE 3 A problem decomposition approach supplemented by regularization, parameter sampling space and weighting data types. (A) Regularization approach: the regularization factor (x-axis), literature parameter deviation (blue, left y-axis) and model error (red, right y-axis) (B and C) Parameter estimation case for an identifiable (Enolase, K_{cat}) and weakly identifiable parameter (Phosphofructokinase, K_{F6P}), respectively. Histogram plot showing the estimated parameter deviation from the literature value (x-axis) and the number of initial parameter samples in the y-axis. Non-regularized and regularized cases in white and dark bars, respectively. The shaded area covers the range of one order of magnitude of deviation from the literature parameter value. (D) Pareto front commonly obtained when fitting two types of data simultaneously: case of PEP concentration. Model error for steady state and glucose perturbation concentrations of PEP (x-axis and y-axis, respectively). (E and F) Model error, PEP steady state and glucose perturbation concentrations, respectively.

| Combining different data types was central for physiological system identification

Determining a model parameter set, or system identification, is a data-intensive process (Almquist *et al*, 2014). To identify the model properties, excitation experiments such as impulse or step response and time variant and invariant data are used. Considering different experimental data types when building a kinetic metabolic model contributes to an accurate physiological description of the system (Peskov *et al*, 2012). In the case of yeast fermentations, early studies focused on the transient metabolic response to a glucose pulse (Theobald *et al*, 1997), but implementations have broadened to study the role of the trehalose cycle during higher residual glucose concentrations (van Heerden *et al*, 2014) and steady state dynamics were analyzed at different growth rates (Canelas *et al*, 2011). In the process, data types have also increased from describing metabolomics to fluxomics and even enzymatic constraints (van Hoek *et al*, 2000; Elsemman *et al*, 2022).

Therefore, the benefits of combining data types were examined for the developed model, in the areas highlighted in (Fig 4A). Simulations were compared when changing enzyme concentrations were considered as shown in (van Hoek *et al*, 2000) (data displayed in the supporting information), to when they were fixed at the dilution rate value of 0.1 h^{-1} (Fig 4B). This analysis showed that inclusion of proteome changes improved the predicted fluxes substantially, especially at higher dilution rates, as the glycolytic flux is increased. This is supported by the relevance attributed to the proteome at different growth rates using GSMs (Sánchez *et al*, 2017; Elsemman *et al*, 2022).

Furthermore, parameters were estimated to fit a single or combination of experimental variables, to show the

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

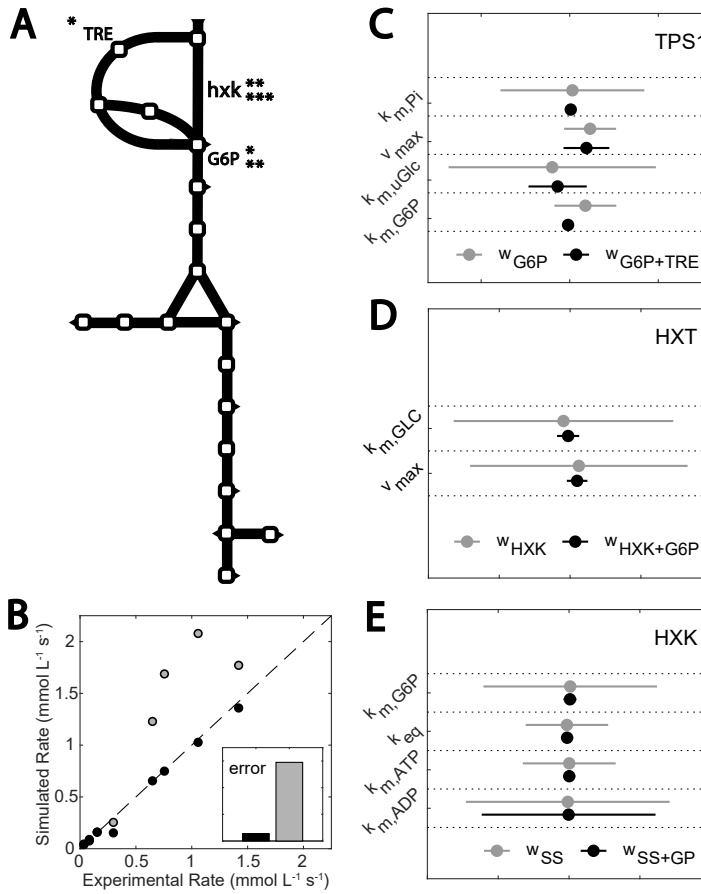


FIGURE 4 Contribution of different data sets to model identification. (A) Simplified model diagram: variables included in the TPS1 (*), HXT (**) and HXK (***) parameter estimation examples. (B) Simulated HXT reaction rates plotted against experimental rates. Including dilution rate dependency in enzyme concentrations (black) or keeping them constant at the experimental value found for 0.1 h-1. Confidence intervals obtained for the parameter estimates when: (C) G6P or the combination G6P+trehalose was used in TPS1 parameter estimation, (D) reaction rates (HXK) or the combination reaction rate and metabolite concentration (G6P) was used in HXT parameter estimation and (E) SS or the combination SS+GP data was included in HXK parameter estimation.

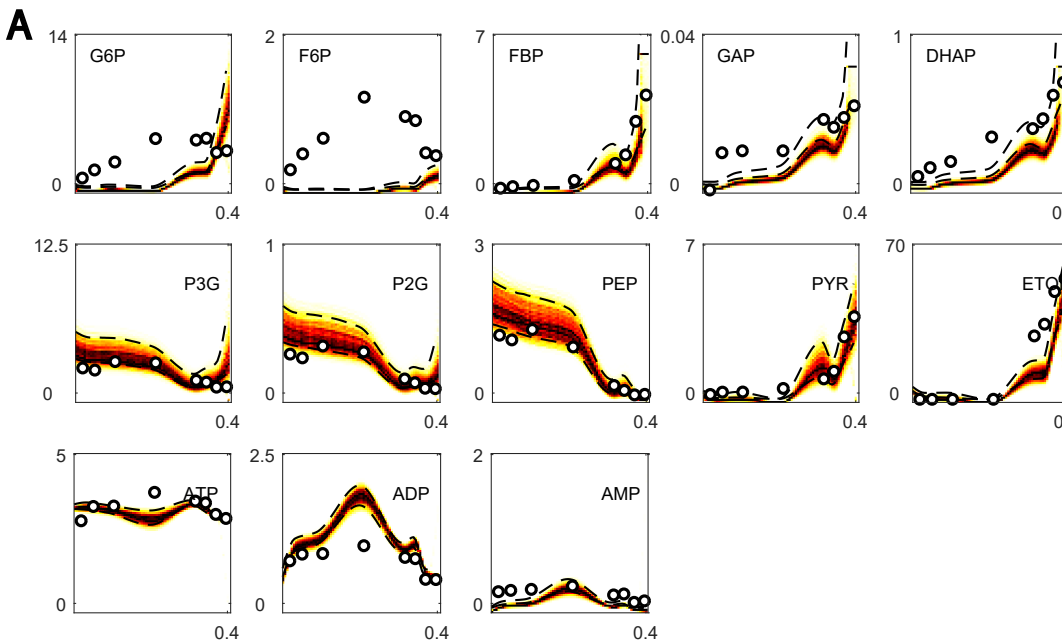
benefits of using multiple data types when fitting the model. For instance, reduced confidence intervals were obtained for TPS1 kinetic constants if trehalose concentrations were used in combination with glucose 6-phosphate (G6P) (Fig 4C), and similar results were obtained combining metabolomic and fluxomic, or steady state and glucose perturbation data (Fig 4D and 4E, respectively). This tendency highlights how the identified system becomes more accurate, on top of more representative of cell dynamics. This is relevant, given that kinetic yeast CCM models have often been only developed with a single type of data, most usually a glucose perturbation for which only some metabolites were measured.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

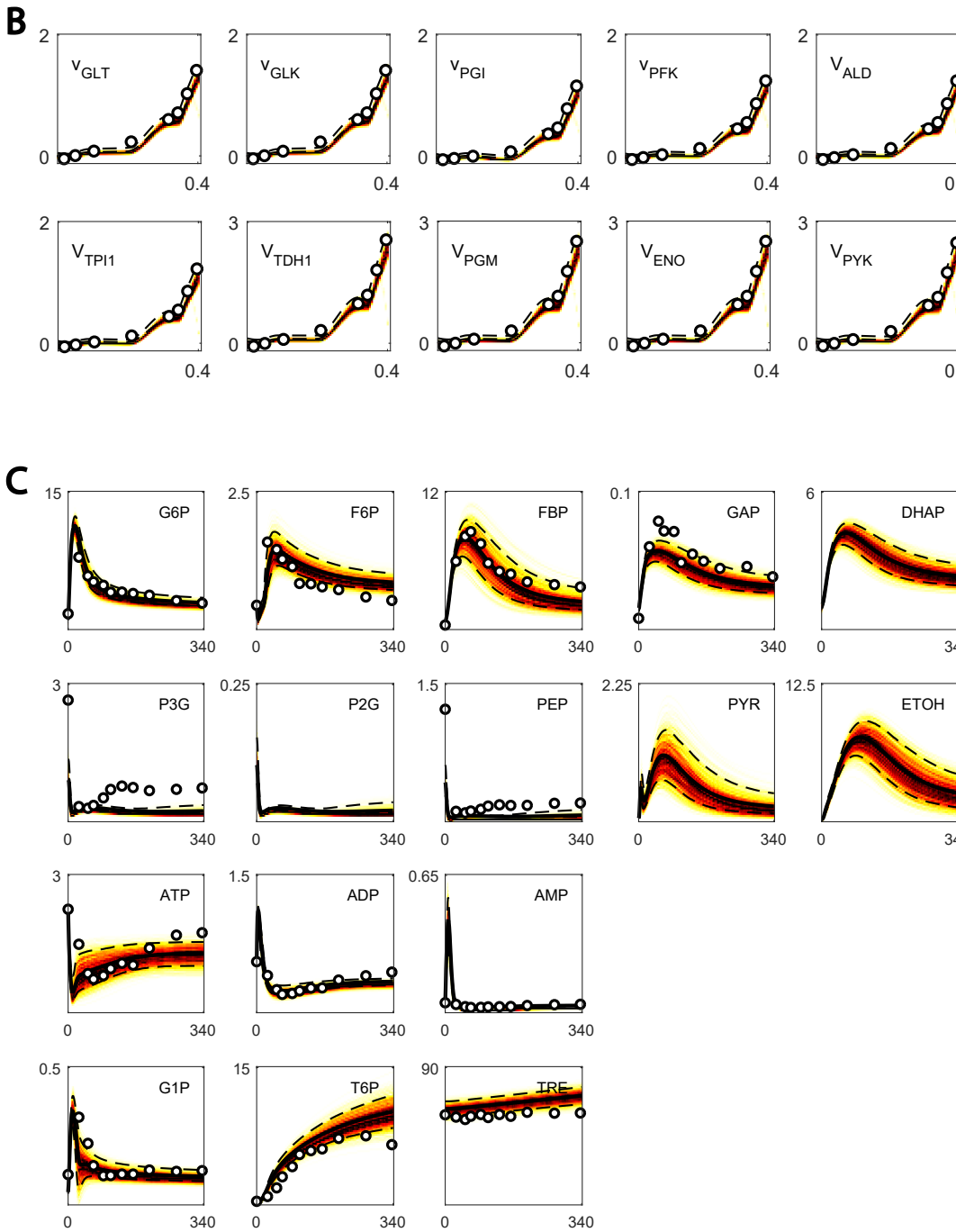
Ensemble modelling suggested that the model was robust to parameter heterogeneity

Kinetic metabolic models are generally designed as deterministic descriptions of an average cell in the bioreactor, represented by a unique parameter set (Almquist *et al*, 2014). Nonetheless, heterogeneity between cells of the same population or between strains implies a certain degree of metabolic diversity (van Heerden *et al*, 2014; Nidelet *et al*, 2016; Rodrigues *et al*, 2021). These Bayesian dynamics can be represented by means of ensemble modelling, in which parameters are sampled from a variability distribution and are often used to account for uncertainty (Tan *et al*, 2011; Jia *et al*, 2012; Tiemann *et al*, 2013). In this work, a Monte Carlo approach was taken to assess how robust were the modelled glucose perturbation response and steady state dynamics. The models were generated by adding a 10% random noise to the parameter set .

Model simulations were found to be notably robust within the conditions tested. Predictions showed certain quantitative but little qualitative deviation (Fig 5). Steady state concentrations and fluxes were especially robust, as it has been found for other kinetic networks (Tran *et al*, 2008; Tan *et al*, 2011; Christodoulou *et al*, 2018). Variability was more pronounced during the glucose perturbation, which could be explained by the notably higher extracellular glucose concentrations (110 mM) experienced during the strong perturbation, when compared to the steady state range (0.13-3.30 mM).



bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

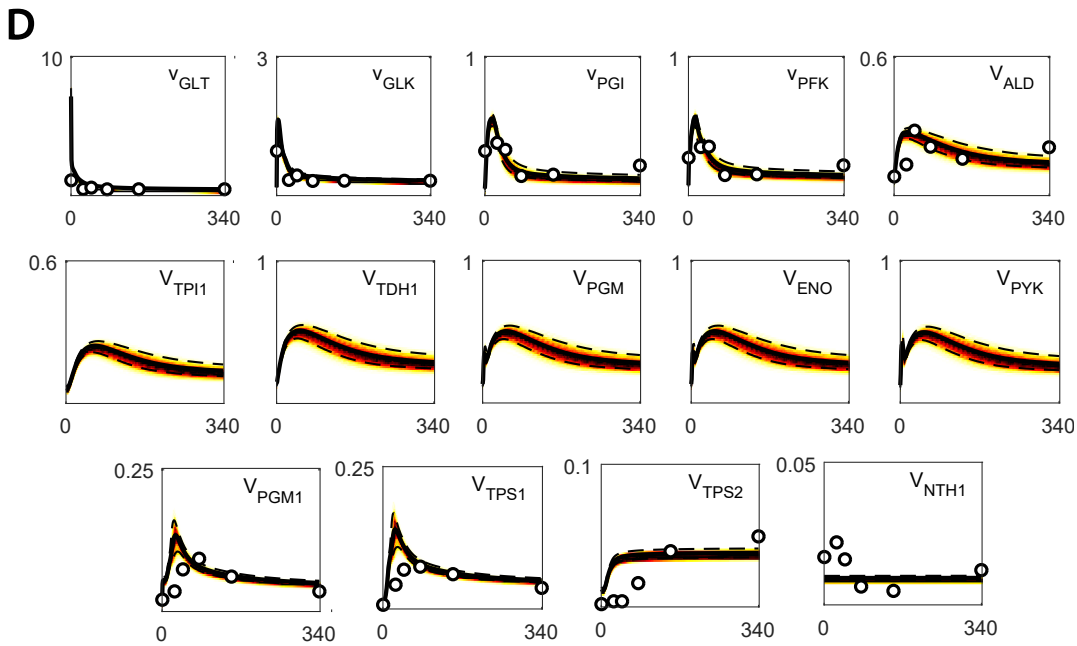


FIGURE 5 The model was robust to parameter heterogeneity. Simulations with the ensemble of models and the experimental data. Random noise was added to generate up to 1000 models in the ensemble. This noise was sampled from a random distribution within 10% of the value of each parameter. The model ensemble is shown as heat maps, where darker regions indicate higher simulation agreement. Experimental data is shown is bullet points. The semi-continuous lines indicate the region where 90% of simulations were found. (A) Steady state concentrations (mM) and dilution rate (h^{-1}) in the Y. and X-axes, respectively. (B) Steady state reaction rates (mM s^{-1}) and dilution rate (h^{-1}) in the Y. and X-axes, respectively. (C) Glucose perturbation concentrations (mM) over time (s) in the Y. and X-axes, respectively. (D) Glucose perturbation reaction rates (mM s^{-1}) over time (s) in the Y. and X-axes, respectively. To see the simulations of each metabolite and reaction rate in the model, see supplementary materials.

| Problem decomposition revealed a mismatch between literature and *in vivo* parameters

A concern in the literature is the accuracy of the estimated enzymatic kinetic constants. Previous work showed that kinetic constants measured *in vitro* in isolated enzymes assays might not be representative of *in vivo* behavior (Teusink *et al*, 2000; Davidi and Milo, 2017). This was improved by standardizing *in vivo*-like assay conditions, but only some glycolytic parameters were updated (Van Eunen *et al*, 2010; Smallbone *et al*, 2013). In this work, the abovementioned problem decomposition approach was used to assess this experimental bias and to adjust the parameters to fit the data.

Simulating the experimental steady state reaction rates with the models generated in the problem decomposition step revealed a notable bias if the parameter values found in the literature were used, even if *in vivo*-like experimental parameters were considered (Fig 6). These fits were improved using the parameter estimation pipeline described in (Fig 3). In line with (Davidi and Milo, 2017), this highlights the need of cross validating experimental data generated in individual enzyme assays with *in vivo* data and suggest that even with improved media conditions, an impactful bias exists. See the supplementary information for a complete comparison between the estimated and the literature

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

parameter set.

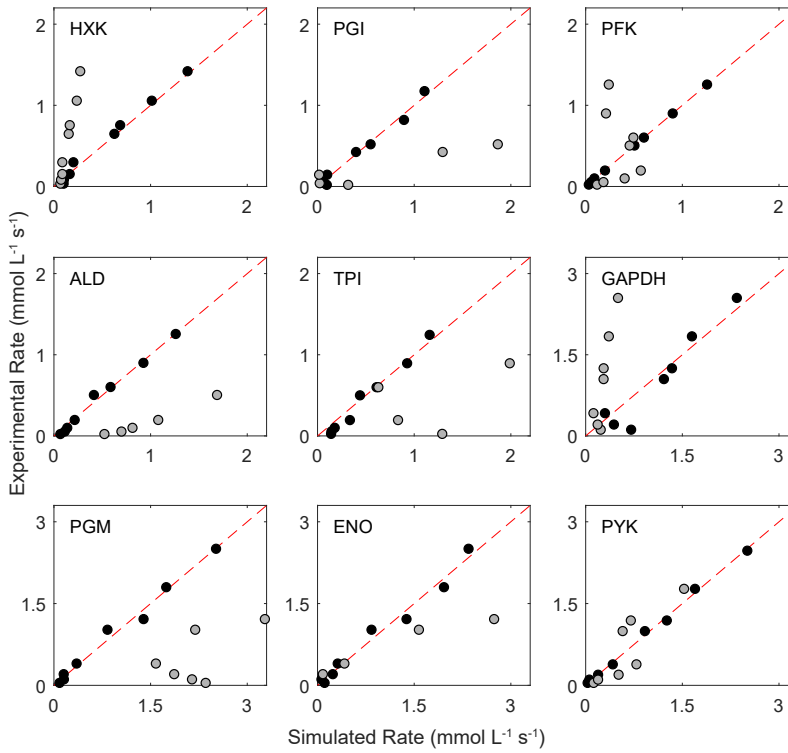


FIGURE 6 Problem decomposition reveals a mismatch between literature and *in vivo* parameters. Experimental reaction rates plotted against simulated reaction rates. Gray dots show the model fit using literature parameters. Black dots show the fit with the developed models. The red dashed line shows the location of a perfect fit. Each plot shows a different enzyme in glycolysis. Upper and lower glycolysis in top and bottom row, respectively.

| Large deviations from literature parameters suggested missing knowledge around PFK kinetics

When the estimated parameters were compared to the literature, most deviated by less than 0.5-fold from the literature value. Nonetheless, the kinetic constants which deviated the most belonged to PFK kinetics or the two enzymes found before and after this reaction in glycolysis, phosphoglucose isomerase (PGI) or aldolase (ALD) (Table 1). Furthermore, the model fit agreed for most experimental variables, except for the SS concentrations of metabolites G6P and fructose 6-phosphate (F6P), for which simulations disagreed both qualitatively and quantitatively. This could be solved if PFK kinetics parameters were changed in a growth rate dependent manner, for instance, by adding a 'fudge factor' to the maximum reaction rate (see supplementary information), which could suggest that there is a missing regulation on PFK kinetics . Even though well studied, the kinetics of this enzyme are known to be notable com-

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

plex (Teusink *et al*, 2000; Gustavsson *et al*, 2014). For instance, the concentrations of F26bP, a modulator of PFK was not present in the data and a constant value was assumed in our simulations.

TABLE 1 Largest deviations between estimated and literature values are found in PGI, PFK and ALD enzymes. Only deviations above 0.75-fold change from the literature value are displayed. Fold change defined as: Fold change = estimated parameter value / literature parameter value.

Parameter	Enzyme	Fold Change	Reference
$K_{m,DHAP}$	ALD	-1.90	(Teusink *et al*, 2000)
$K_{m,F6P}$	PGI	1.37	(Smallbone *et al*, 2013)
K_{AMP}	PFK	-0.98	(Teusink *et al*, 2000)
$K_{m,G6P}$	PGI	0.89	(Smallbone *et al*, 2013)
$K_{m,ATP}$	PFK	0.88	(Teusink *et al*, 2000)
$K_{i,FBP}$	PFK	0.77	(Teusink *et al*, 2000)
K_{cat}	PGI	-0.76	(van Eunen *et al*, 2012)

| Dilution rate dependent glucose transport kinetics

Transmembrane glucose transport kinetics are known to change with dilution rate. For instance, changing affinities had to be considered to model different dilution rates (Postma *et al*, 1989). Thereafter, multiple transporters (HXT1-7) were found to be notably expressed at different growth rates (Diderich *et al*, 1999). The kinetics of these transporters were found to be divergent and isoenzyme dependent, both in terms of affinity constants (K_m) and maximum reaction rates (V_{max}) (Bosdriesz *et al*, 2018; Maier *et al*, 2002; Reifenberger *et al*, 1997). Therefore, we investigated whether a unique parameter set could explain the entire data, or dilution rate dependency had to be considered on the kinetic constants.

The simulations with the unique parameter set fit the data to a notable degree, but with certain deviation (Fig 7A). Uptake was overestimated at dilution rates below 0.2 h^{-1}, resulting in fermentative excess where residual glucose concentrations remained under 0.20 mM (Fig 7B) (Postma *et al*, 1989). Then, this fit could be improved by allowing HXT V_{max} to change in a dilution rate dependent manner (Fig 7A). As expected, V_{max} was found to decrease and increase at low and high dilution rates, respectively (Fig 7C). Nonetheless, the change at low growth rates could also be explained by a decreased enzyme affinity (see supplementary information) or a combination of both V_{max} and K_{cat} changes, in line with (Diderich *et al*, 1999). Interestingly, a data point that could not be fit was the dilution rate 0.2 h^{-1}, given the low residual glucose concentration. As suggested in (Postma *et al*, 1989), this could be caused by intracellular changes on the driving force.

| Glycolysis was regulated by other pathways in central carbon metabolism

Glycolysis is a pathway that does not act in isolation but that is influenced by other metabolic routes in CCM. Therefore, here we investigated how the different pathways in the model regulate glycolysis. Flux towards the different glycolytic intermediate sinks was dilution rate dependent. At low growth rates, most carbon entered the sink reactions, while fermentation was predominant at higher growth rates (Fig 8A). Nonetheless, the model did overestimate fermentative

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

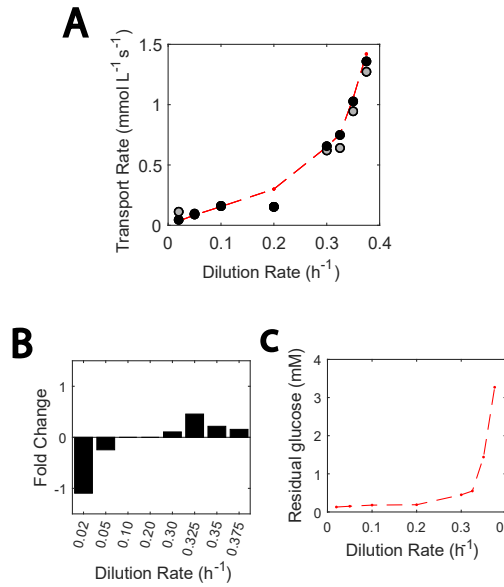


FIGURE 7 Adjustments in the HXT V_{max} can describe most growth rate-dependent transporter kinetics. (A) Reaction rates plotted against dilution rates. Gray markers show the model fit if a single set of kinetic constants for the hexose transporter is used. Black markers highlight how this fit is improved if the maximum reaction rate is fit for every growth rate. Experimental data point are shown in red markers, united by the semi-continuous red line. (B) Residual glucose concentration (mM) plotted against dilution rate, and (C) Fold change required for V_{max} at each dilution rate.

flux at lower growth rate, due to small deviations of the already low glucose transport flux.

Furthermore, the effect of knock outs resulted in expected dynamics. Simulating the lack of oxygen by knocking out mitochondrial reactions resulted in an increase in fermentation and less NADH mitochondrial respiration lowered glycolytic intake, since the network compensated by over-activation of the glycerol branch (Fig 8B and 8C). Then, knocking out the trehalose cycle during glucose perturbation resulted in an imbalance between upper and lower glycolysis reaction rates, in line with (van Heerden *et al*, 2014) (Fig 8D) and a knock out of the inosine salvage pathway meant that no ATP paradox would be observed (Fig 8E).

| Missing regulation could explain glycolytic response under a 20 g L^{-1} glucose pulse

At this point, we investigated how our model could simultaneously explain the response to a strong glucose perturbation of a 20 g L^{-1} glucose pulse experiment (van Heerden *et al*, 2014), the moderate perturbations of 0.4 g L^{-1} from FF experiments (Suarez-Mendez *et al*, 2017), and the steady state data from (Canelas *et al*, 2011).

We identified two type of dynamics depending on if PFK and ALD reaction rates balanced each other (Fig 9A). By analyzing the experimental reaction rate data, we could see that during the moderate glucose perturbations in (Suarez-Mendez *et al*, 2017), the glycolytic response of PFK and ALD occurred hand in hand, resulting in a transient increase of ATP (Fig 9B). Meanwhile, during the stronger perturbation in (van Heerden *et al*, 2014) the response of ALD was slower (Fig 9C), leading to more ATP being consumed in upper glycolysis than recycled in lower glycolysis, and a

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

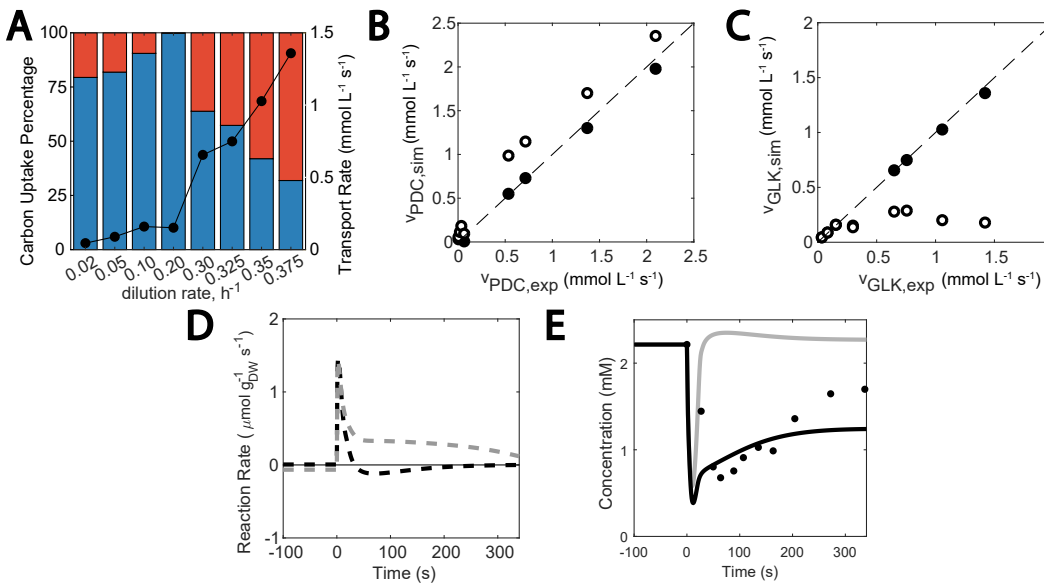


FIGURE 8 Contribution of different modules to glycolytic regulation. (A) Carbon flux distribution at different growth rates: flux towards sink reactions (blue), and towards ethanol and glycerol synthesis (red). The black dots show the simulated glucose uptake rate. (B) Effect of knocking out mitochondrial flux: Simulated PDC reaction rate plotted against experimental PDC reaction rate. Reference model (black) and effect of knockout carbon flux to respiration (white). (C) Effect of no oxygen: Simulated HXK reaction rate plotted against experimental HXK reaction rate. Reference model (black) and effect of knocking out NADH mitochondrial respiration (white). (D) Imbalance between upper and lower glycolysis (calculated as vGLK – 2*vGAPDH) occurring if the trehalose cycle is knocked out. The black and gray dashed lines show the *WT* and Trehalose cycle-knocked out strain, respectively. (E) Sensitivity of the ATP paradox to inosine salvage pathway. When the inosine salvage pathway is active (black), a decay in ATP concentrations can be observed, while knocking the cycle (gray) leads to a conservation of ATP steady state concentration.

consequent temporary decrease in ATP.

To understand what differed between the two conditions, we modelled both and found that they could not be simultaneously understood with a single parameter set. A pareto analysis was performed to find out which parameters needed to change to fit one and the other dataset. This resulted in the pareto front described in (Fig 9D). If the steady state fructose 1,6-bisPhosphate (FBP) concentrations were reproduced, ALD flux rapidly increased in response to the glucose perturbation. Another solution along the Pareto front was obtained for which the increase in ALD flux was slower , resembling experimental data, but then, the fit for FBP was not achieved (Fig 9E and 9F, respectively). Interestingly, this slower response resulted from apparent substrate affinity constants in upper glycolysis strongly deviating from the literature value and de facto caused enzymes to require a higher buildup of substrate before carrying out their reaction.

Altogether , this could imply that a mechanism is missing that regulates glycolytic response under strong perturbations by, for instance, transiently decreasing ALD reaction rates. Since parameter values are closer to the literature when the SS data is fit, which also results in a fast ALD response during FF, this seems to suggest that our model is

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

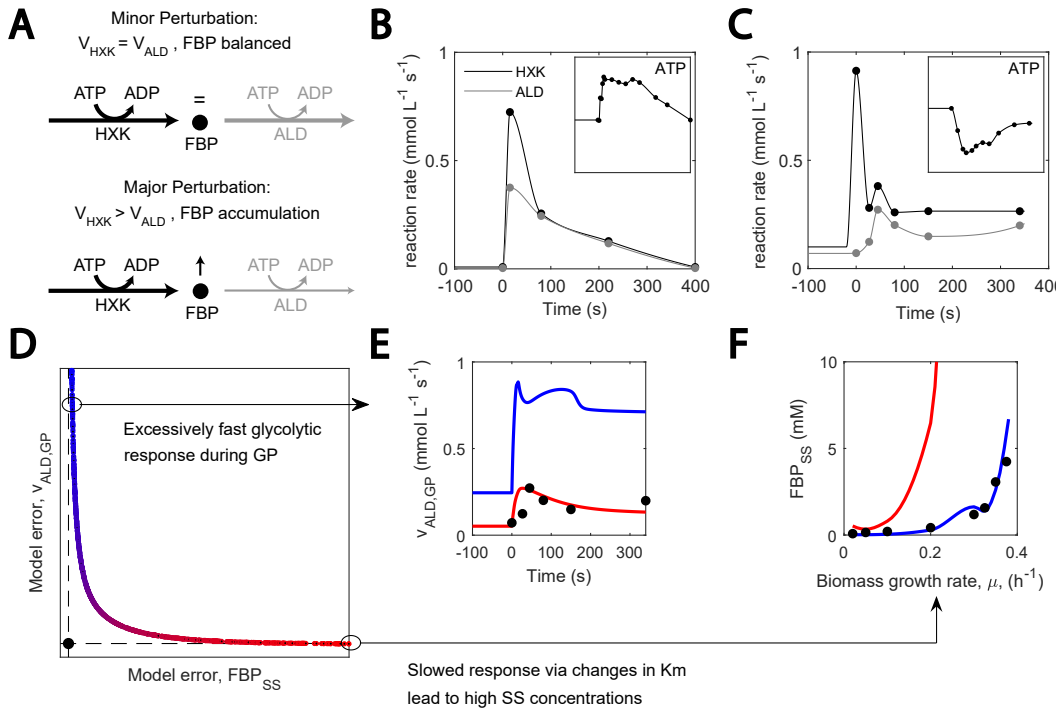


FIGURE 9 A transient imbalance between ALB and FBP can be explained by missing regulation. (A) A large glucose perturbation triggers FBP accumulation, (B-C) HXK and ALD reaction rate and ATP concentration upon glucose perturbation under feast famine regime, 0.4 mM (Suárez-Mendez et al., 2014), and single, 110 mM glucose perturbation (van Heerden et al., 2014), (D) Pareto front observed when fitting ALD reaction rate during GP and FBP SS concentrations. Red and blue colors show the zone with lower model error for ALD and FBP, respectively. (E-F) Model fits for ALD reaction rate during GP and FBP SS concentrations, respectively. Colors correspond with (D).

best suited to fit physiology from SS and FF regime.

Discussion

How cells respond and adapt to environmental changes while balancing internal needs prevails as a fundamental question in biology. Mechanistic models based on enzyme kinetics are important to cope with the complexity of the underlying molecular circuitry. For these models to be useful, however, they require a proper intracellular embedding of central carbon metabolism in the context of the larger metabolic system and should not be studies in isolation.

Hence, a physiologically informed kinetic model of this pathway was developed. It contained detailed descriptions of the enzymatic mechanisms that compose glycolysis, trehalose cycle and glycerol branch and extended this knowledge with coarse-grained and growth rate-dependent phenomenological descriptions of glycolytic intermediate demand, gas exchange, mitochondrial activity and ATPase activity (Fig 1). This more was then used to characterize metabolic dynamics under both a 110 mM glucose pulse and at different growth rates at steady state (Fig 2). Smaller models exist for *Escherichia coli* and *Penicillium chrysogenum* metabolic models (Chassagnole *et al*, 2002; Tang *et al*,

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2017), but to our knowledge this is the first time a kinetic CCM model can realistically describe a growing *S. cerevisiae* cell. Therefore, new circumstances are now open for exploration with a unique model. For instance, by constraining the model to anaerobic conditions or by studying the shift between respiratory and respirofermentative metabolism and glucose perturbations at different growth rates (Wiebe *et al*, 2008; Ishtar Snoek and Yde Steensma, 2007; van Eunen *et al*, 2012).

One of the core developments of this work is a parameter estimation pipeline suited to identify complex systems and using large data sets (Fig 3). This approach made use of different resources: (i) divide and conquer model decomposition resulted in smaller scale systems easier to identify that the full system, (ii) global sampling determined whether multiple local minima existed, (iii) regularization improved convergence when parameters were still underdetermined and (iv) leveraging the weights of the components of the cost function to avoid overfitting. As a result, this method increases parameter identifiability, and we believe that it is well-suited for implementation in other cell factory models. Compared to other state of the art parameter estimation methods (Gábor and Banga, 2015; Villaverde *et al*, 2019; Smallbone *et al*, 2013; Raue *et al*, 2015; Steiert *et al*, 2016), this pipeline makes use of the divide and conquer approach, combined with regularization for parameters where physiological values are known. Nonetheless, limitations of this approach are the lack of reliable parameter estimates confidence intervals, given that the regularization bias added is affected by experimental noise, and the need for leveraging the weights of these data in the cost function.

In order to develop this complex and physiologically representative model, multiple data types needed to be combined. A collection of datasets containing metabolomic, fluxomic and proteomic data at steady state and during glucose perturbations was considered to make the complete model identifiable (van Hoek *et al*, 2000; Canelas *et al*, 2011; van Heerden *et al*, 2014). Considering this data was required and increased the accuracy of the parameter estimates (Fig 4). We therefore suggest making use of growingly available proteomics data to constrain the parameter solution space, as it is the case for GSM (Chan *et al*, 2017; Sánchez *et al*, 2017; Elsemman *et al*, 2022).

Ensemble modelling suggested that the network dynamics are robust to parameter heterogeneity. A modelling approach based on Monte Carlo sampling of parameter values was used to represent the heterogeneity in a cell population. This showed how dynamics and steady states were resilient to parameter value deviations (Fig 5). Assays such as this start with the premise that phenotypes arise from model structure and are performed to represent population heterogeneity (Gunawardena, 2010; Oguz *et al*, 2017; Nijhout *et al*, 2019). Therefore, given the heterogeneity in the industrial fermenter (Haringa *et al*, 2016), we propose that this ensemble modelling perspective is a more physiologically realistic representation than a single parameter set. Nonetheless, one limitation of this approach is the range to which parameters are disturbed, given that cell-specific measurements in a population are very limited (Newman *et al*, 2006).

Furthermore, we advocate to estimate parameters and supplement with literature knowledge for determination of *in vivo* enzymatic kinetic constants. Enzymatic kinetic constants were traditionally estimated *in vitro* for isolated enzymes (Teusink *et al*, 2000; van Eunen *et al*, 2012; Smallbone *et al*, 2013). However, using these to simulate experimental data showed how a notable bias still exists (Fig 6). Hence, we propose to re-estimate parameters to fit the *in vivo* metabolomic and fluxomic data, in line with (Davidi and Milo, 2017), and use the *in vitro* determined constants as reference to which the cost function is regularized. The resulting improved data fit with generally little deviation from the literature parameters implies that, even though *in vitro* studies are good estimators, they need adjustment to be implemented in a full-scale model.

When estimated parameters widely deviated from the literature values or when steady state data was not properly fit, we identified uncertainty in our model that reflected possible unknown biology. The estimated parameters that deviated the most from the literature value were from PFK or surrounding enzymes PGI and ALD (Table 1). Together

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with the fact that steady state concentrations of G6P and F6P are only fit in the model decomposition simulations (Fig 6) and not in the complete model simulations (Fig 5), this hints at uncertainty surrounding PFK reaction kinetics, not covered in the complex regulation already considered for this reaction (Teusink *et al*, 2000; Gustavsson *et al*, 2014). Furthermore, another area where steady state data is not perfectly fit is the HXT reaction. HXT kinetics could be explained by changes in enzyme concentration and isoenzyme-specific affinity constants, in line with (Reifenberger *et al*, 1997; Diderich *et al*, 1999). Nonetheless, the consistent lack of fit at 0.2 h^{-1} points at a missing mechanism acting intracellularly (Postma *et al*, 1989).

Finally, missing regulation could explain the transient imbalance occurring between PFK and ALD upon strong glucose perturbation (Fig 9). When explaining this imbalance with the same model structure, we found that a subset of parameters had to be adjusted, leading to a lower ALD reaction rate. Nonetheless, in the industrial bioreactor, prolonged exposure to glucose concentrations higher than 100 mM as in (van Heerden *et al*, 2014) is rare (Haringa *et al*, 2017). A more realistic representation is the FF setup, in which this imbalance did not take place under moderate perturbations in FF experiments. This could hint at influence by glucose level (Suarez-Mendez *et al*, 2014) and suggest that there is a missing phenomenon in the model, which becomes active upon large glucose perturbations. We believe that this could be due to either a regulation cascade, such as post translation modifications or cAMP/PKA pathway (Tripodi *et al*, 2015; Tamaki, 2007), given that cAMP buildup was detected during perturbation with high residual concentration but not for low (Botman *et al*, 2019; Suarez-Mendez *et al*, 2014). Consequently, the synthesis of fructose-2,6-bisphosphate by the activated PKA pathway could be regulating could be regulating PFK activity. In addition, the pH decay observed during the glucose perturbation (van Heerden *et al*, 2014) could also be relevant, given the sensitivity of glycolytic enzyme constants to this variable (Van Leemputte *et al*, 2020; Luzia *et al*, 2022). Nonetheless, to further proof these hypothesis, more experimental testing would be needed.

Here, we presented a model that describes yeast glycolysis as part of the more comprehensive central carbon metabolism, and not in isolation. This model was built using data from the metabolome, fluxome and proteome; to which extent gene regulation influenced the metabolic dynamics here studied remains an open question. At this point, gradients in the industrial bioreactor can be considered thanks to the transport rates and yields included in the model (Haringa *et al*, 2016; Nadal-Rey *et al*, 2021), bringing it closer to industrial applications. We believe that complex intracellular models like the hereby presented could soon be linked to bioreactor dynamics simulations, an area that has been so far restricted to simplified cell models (Tang *et al*, 2017; Sarkizi Shams Hajian *et al*, 2020).

Materials and methods

| Experimental data sets used to develop the model

Three experimental datasets were used in model development: (i) steady state metabolic concentrations and fluxes at different chemostat dilution rates (0.025 - 0.375 h^{-1}) (Canelas *et al*, 2011), (ii) concentrations and fluxes during a single glucose perturbation of 20 g L^{-1} (van Heerden et al., 2014) and (iii) glycolytic enzymes activity also at several chemostat dilution rates (van Hoek *et al*, 2000). Data in (Canelas *et al*, 2011; van Heerden *et al*, 2014) was obtained with the haploid yeast Saccharomyces cerevisiae CEN PK 113-7D strain, while (van Hoek *et al*, 2000) used the strain DS28911.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Model description

The kinetic metabolic model in this work included individual enzyme reaction descriptions for glycolysis, glycerol branch and trehalose cycle, whose kinetics were taken from (van Heerden *et al*, 2014), (Smallbone *et al*, 2011) and (Cronwright *et al*, 2002), respectively. The reaction of UDP-glucose phosphorylase in the trehalose cycle was lumped due to lack of experimental data. Cofactor metabolism reactions and sinks for anabolic precursors were lumped, and the sink reactions and inosine salvage pathway were adapted from (Chassagnole *et al*, 2002) and (Walther *et al*, 2010), respectively.

System of ordinary differential equations and reaction rate equations

The model consisted of a series of ordinary differential equations representing the mass balances for each metabolite modelled in the cytosol, except glucose and inorganic phosphate, also modelled in the extracellular space and vacuole. Moiety conservations were used for the total sum of adenosine and nicotinamide adenine nucleotides (ATP + ADP + AMP and NAD + NADH, respectively) as in (Smallbone *et al*, 2013). The ATP + ADP + AMP moiety conservation was not considered during the single glucose perturbation response, where the inosine salvage pathway was included as a pool. When available, the dilution rate-dependent protein activity change (van Hoek *et al*, 2000) was considered by adjusting the V_{max}.

Enzymatic reaction rate kinetics were obtained from previous works, where reversible Michaelis Menten kinetics dominated (van Heerden *et al*, 2014; Smallbone *et al*, 2011; Cronwright *et al*, 2002). Exceptions to this were hill type kinetics for pyruvate kinase and decarboxylase (van Eunen *et al*, 2012), phosphofructokinase and facilitated diffusion for membrane transport (Teusink *et al*, 2000). Allosteric regulation acted both by activation and (competitive) inhibition. Hydrolysis reactions were modelled as irreversible. Sink reactions were modelled by phenomenological expressions that closely resembled experimental data in (Canelas *et al*, 2011). Reaction rates were expressed in (mM s^{-1}). For a detailed description, see supplementary information.

Simulation setup

Simulations were performed in agreement with the experimental setup. To confirm simulation stability, the model was first simulated for 3000 seconds using the experimental concentrations at 0.1 h^{-1} dilution rate. Then, the steady states at different dilution rates were modelled in parallel simulations where the residual glucose concentration was changed to the value in (Canelas *et al*, 2011), for 3000 seconds. The trehalose cycle was not modelled in steady state. Anabolic sink reaction rates, ATP maintenance and the mitochondrial activity were adjusted in a dilution rate-dependent manner as well. For the single glucose perturbation, residual glucose was increased to 110 mM (van Heerden *et al*, 2014) and the inosine salvage pathway was made active in a step manner. This simulation lasted 340 seconds. Matlab version 9.3.0.713579 (R2017b) was used. A summary of the differences between simulating steady states at different growth rates of the 110 mM single glucose perturbation can be seen in Table 2. Detailed expplanation on these can be found in the supplementary information.

Literature parameter values

Initial estimates of kinetic constants were obtained from the literature. For glycolytic enzymes, (van Eunen *et al*, 2012) and (Smallbone *et al*, 2013) were the most recent estimates of V_{max} and K_m. If a parameter was not available, it was

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TABLE 2 Summary of main differences in the simulation setup of the steady state and the single glucose perturbation.

Variable	Steady state	Single Glucose Perturbation
Trehalose cycle	Inactive. Its steady state flux in included in the G6P-sink reaction	Active.
Inosine salvage pathway	Inactive.	Active.
sumAXP (= ATP + ADP + AMP)	Changes in the total sum of ATP + ADP + AMP are considered by setting the initial concentrations of the adenosine nucleotides to the experimental concentration at the beginning of each stead state simulation.	The inosine salvage pathway acts as a sink of adenosine nucleotides.
Enzyme concentra-tion	A ratio between the experimental activity at a given dilution rate and the value at $0.1h^{-1}$ is calculated and then multiplied to the V_{max}.	This ratio equals 1.
Sink reactions	A phenomenological expression is derived to make these reaction rates resemble experimental data.	The kinetics at $0.1\ h^{-1}$ is used.
ATPase activity	ATPase is adjusted in a growth rate dependent manner, using GAM and NGAM.	The reaction constant is estimated to fit the data.
mitoATP	The growth rate-dependent mitochondrial activity is considered using the PYR sink reaction. This is implemented by changes in the reaction constant.	The reaction constant is estimated to fit the data.
mitoN-ADH	The growth rate-dependent mitochondrial activity is considered using the PYR sink reaction. This is implemented by changes in the reaction constant.	The reaction costant is estimated to fit the data.

taken from (Teusink *et al*, 2000). Glycerol branch and trehalose cycle parameters were retrieved from (Cronwright *et al*, 2002) and (Smallbone *et al*, 2011), respectively. For the specific parameter values used, see supplementary materials.

| In vivo parameter estimation and cost function development

Parameters in the model were estimated to fit the in vivo experimental data. The lsqnonlin solver in the Matlab optimization toolbox (version 9.3.0.713579, R2017b), which uses an interior reflective Newton method (Coleman and Li, 1996), was used to minimize the error between experimental and simulated data. To overcome the effects of ill conditioning and parameter dependencies in this large parameter set (Gábor and Banga, 2015), a model decomposition approach, also known as divide-and-conquer, was used (Kotte and Heinemann, 2009). This was first implemented to individual reactions in glycolysis. Once parameters were estimated for them, parameters were estimated for other pathways. The global solution space was explored by means of multi-start deterministic local searches (Villaverde

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

et al, 2019). This model fit was supplemented with L1-type regularization (Steiert *et al*, 2016; Dolejsch *et al*, 2019), where kinetic constants were biased to resemble the experimental measurement, as long as data fit was adequate. Steady state data (Canelas *et al*, 2011) was fit first and, afterwards, single glucose perturbation (van Heerden *et al*, 2014).

Acknowledgements

We thank our colleagues from the Yeast 3M team, S.Aljoscha Wahl, Koen J.A. Verhagen, Laura Guilherme Luzia, Evelina Tutucci and Johann van Heerden, for the insightful discussions.

Author contributions

David Lao-Martil: Conceptualization; Validation; Investigation; Methodology; Writing - originak draft; Writing - review and editing. **Joep Schmitz:** Conceptualization; Investigation; Methodology; Writing - review and editing. **Bas Teusink:** Conceptualization; Writing - review and editing. **Natal van Riel:** Conceptualization; Writing; Methodology - review and editing.

Conflict of interest

The authors declare that they have no conflict of interest.

Data, software and model availability

This work was developed in MATLAB and model implementations are also available in Python language and SBML format. The data used and model developed in this work are available in the github repository from the author (github.com/DavidLaoM/y3m1_ss_gp) and the CBio group (github.com/Computational-Biology-TUe).

References

Almquist J, Cvijovic M, Hatzimanikatis V, Nielsen J, Jirstrand M (2014) Kinetic models in industrial biotechnology–improving cell factory performance. *Metabolic engineering* **24**: 38–60

Beeftink H, Van der Heijden R, Heijnen J (1990) Maintenance requirements: energy supply from simultaneous endogenous respiration and substrate consumption. *FEMS Microbiology Ecology* **6**: 203–209

Bosdriesz E, Wortel MT, Haanstra JR, Wagner MJ, De La Torre Cortés P, Teusink B (2018) Low affinity uniporter carrier proteins can increase net substrate uptake rate by reducing efflux. *Scientific reports* **8**: 1–9

Botman D, O'Toole TG, Goedhart J, Bruggeman FJ, van Heerden JH, Teusink B (2019) A yeast FRET biosensor enlightens cAMP signalling. *bioRxiv* : 831354

Canelas AB, Ras C, ten Pierick A, van Gulik WM, Heijnen JJ (2011) An in vivo data-driven framework for classification and quantification of enzyme kinetics and determination of apparent thermodynamic data. *Metabolic engineering* **13**: 294–306

Chan SH, Cai J, Wang L, Simons-Senftle MN, Maranas CD (2017) Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics* **33**: 3603–3609

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M (2002) Dynamic modeling of the central carbon metabolism of Escherichia coli. *Biotechnology and bioengineering* **79**: 53–73

Chen Y, Nielsen J (2019) Energy metabolism controls phenotypes by protein efficiency and allocation. *Proceedings of the National Academy of Sciences* **116**: 17592–17597

Christodoulou D, Link H, Fuhrer T, Kochanowski K, Gerosa L, Sauer U (2018) Reserve flux capacity in the pentose phosphate pathway enables Escherichia coli's rapid response to oxidative stress. *Cell systems* **6**: 569–578

Coleman TF, Li Y (1996) A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM Journal on Optimization* **6**: 1040–1058

Cronwright GR, Rohwer JM, Prior BA (2002) Metabolic control analysis of glycerol synthesis in Saccharomyces cerevisiae. *Applied and Environmental Microbiology* **68**: 4448–4456

Davidi D, Milo R (2017) Lessons on enzyme kinetics from quantitative proteomics. *Current opinion in biotechnology* **46**: 81–89

Diderich JA, Schepper M, van Hoek P, Luttik MA, van Dijken JP, Pronk JT, Klaassen P, Boelens HF, de Mattos MJT, van Dam K, *et al* (1999) Glucose uptake kinetics and transcription of HXTGenes in chemostat cultures of Saccharomyces cerevisiae. *Journal of Biological Chemistry* **274**: 15350–15359

Dolejsch P, Hass H, Timmer J (2019) Extensions of l1 regularization increase detection specificity for cell-type specific parameters in dynamic models. *BMC bioinformatics* **20**: 1–13

Elsemman IE, Rodriguez Prado A, Grigaitis P, Garcia Albornoz M, Harman V, Holman SW, van Heerden J, Bruggeman FJ, Bisschops MM, Sonnenschein N, *et al* (2022) Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *Nature communications* **13**: 1–12

Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B, Krüger E, Schweder T, Hamer G, O'beirne D, *et al* (2001) Physiological responses to mixing in large scale bioreactors. *Journal of biotechnology* **85**: 175–185

Engl HW, Flamm C, Kügler P, Lu J, Müller S, Schuster P (2009) Inverse problems in systems biology. *Inverse Problems* **25**: 123014

Gábor A, Banga JR (2015) Robust and efficient parameter estimation in dynamic models of biological systems. *BMC systems biology* **9**: 1–25

Gunawardena J (2010) Models in systems biology: the parameter problem and the meanings of robustness. *Elements of computational systems biology* **1**: 21–47

Gustavsson AK, van Niekerk DD, Adiels CB, Kooi B, Goksör M, Snoep JL (2014) Allosteric regulation of phosphofructokinase controls the emergence of glycolytic oscillations in isolated yeast cells. *The FEBS journal* **281**: 2784–2793

Haringa C, Deshmukh AT, Mudde RF, Noorman HJ (2017) Euler-Lagrange analysis towards representative down-scaling of a 22 m3 aerobic S. cerevisiae fermentation. *Chemical Engineering Science* **170**: 653–669

Haringa C, Tang W, Deshmukh AT, Xia J, Reuss M, Heijnen JJ, Mudde RF, Noorman HJ (2016) Euler-Lagrange computational fluid dynamics for (bio) reactor scale down: an analysis of organism lifelines. *Engineering in life sciences* **16**: 652–663

Ishtar Snoek I, Yde Steensma H (2007) Factors involved in anaerobic growth of Saccharomyces cerevisiae. *Yeast* **24**: 1–10

Jia G, Stephanopoulos G, Gunawan R (2012) Ensemble kinetic modeling of metabolic networks from dynamic metabolic profiles. *Metabolites* **2**: 891–912

Kotte O, Heinemann M (2009) A divide-and-conquer approach to analyze underdetermined biochemical models. *Bioinformatics* **25**: 519–525

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Kreutz C (2016) New concepts for evaluating the performance of computational methods. *IFAC PapersOnLine* **49**: 63–70

Lao-Martil D, Verhagen KJ, Schmitz JP, Teusink B, Wahl SA, van Riel NA (2022) Kinetic Modeling of Saccharomyces cerevisiae Central Carbon Metabolism: Achievements, Limitations, and Opportunities. *Metabolites* **12**: 74

Lu H, Li F, Sánchez BJ, Zhu Z, Li G, Domenzain I, Marcišauskas S, Anton PM, Lappa D, Lieven C, *et al* (2019) A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nature communications* **10**: 1–13

Luzia L, Lao-Martil D, Savakis P, van Heerden J, van Riel N, Teusink B (2022) pH dependencies of glycolytic enzymes of yeast under in vivo-like assay conditions. *The FEBS Journal*

Maier A, Völker B, Boles E, Fuhrmann GF (2002) Characterisation of glucose transport in Saccharomyces cerevisiae with plasma membrane vesicles (countertransport) and intact cells (initial uptake) with single Hxt1, Hxt2, Hxt3, Hxt4, Hxt6, Hxt7 or Gal2 transporters. *FEMS yeast research* **2**: 539–550

Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome research* **13**: 2467–2474

Nadal-Rey G, McClure DD, Kavanagh JM, Cornelissen S, Fletcher DF, Gernaey KV (2021) Understanding gradients in industrial bioreactors. *Biotechnology Advances* **46**: 107660

Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**: 840–846

Nidelet T, Brial P, Camarasa C, Dequin S (2016) Diversity of flux distribution in central carbon metabolism of S. cerevisiae strains from diverse environments. *Microbial cell factories* **15**: 1–13

Nielsen J, Larsson C, van Maris A, Pronk J (2013) Metabolic engineering of yeast for production of fuels and chemicals. *Current opinion in biotechnology* **24**: 398–404

Nijhout HF, Best JA, Reed MC (2019) Systems biology of robustness and homeostatic mechanisms. *Wiley Interdisciplinary Reviews Systems Biology and Medicine* **11**: e1440

Oguz C, Watson LT, Baumann WT, Tyson JJ (2017) Predicting network modules of cell cycle regulators using relative protein abundance statistics. *BMC systems biology* **11**: 1–24

Otterstedt K, Larsson C, Bill RM, Ståhlberg A, Boles E, Hohmann S, Gustafsson L (2004) Switching the mode of metabolism in the yeast Saccharomyces cerevisiae. *EMBO reports* **5**: 532–537

Peskov K, Mogilevskaya E, Demin O (2012) Kinetic modelling of central carbon metabolism in Escherichia coli. *The FEBS journal* **279**: 3374–3385

Postma E, Alexander Scheffers W, Van Dijken JP (1989) Kinetics of growth and glucose transport in glucose-limited chemostat cultures of Saccharomyces cerevisiae CBS 8066. *Yeast* **5**: 159–165

Pritchard L, Kell DB (2002) Schemes of flux control in a model of Saccharomyces cerevisiae glycolysis. *European journal of biochemistry* **269**: 3894–3904

Pruett WA, Clemmer JS, Hester RL (2020) Physiological Modeling and Simulation—Validation, Credibility, and Application. *Annual review of biomedical engineering* **22**: 185–206

Raue A, Steiert B, Schelker M, Kreutz C, Maiwald T, Hass H, Vanlier J, Tönsing C, Adlung L, Engesser R, *et al* (2015) Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics* **31**: 3558–3560

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Reifenberger E, Boles E, Ciriacy M (1997) Kinetic characterization of individual hexose transporters of Saccharomyces cerevisiae and their relation to the triggering mechanisms of glucose repression. *European journal of biochemistry* **245**: 324–333

Rodrigues CIS, Wahl A, Gombert AK (2021) Aerobic growth physiology of Saccharomyces cerevisiae on sucrose is strain-dependent. *FEMS yeast research* **21**: foab021

Sánchez BJ, Zhang C, Nilsson A, Lahtvee PJ, Kerkhoven EJ, Nielsen J (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular systems biology* **13**: 935

Sarkizi Shams Hajian C, Haringa C, Noorman H, Takors R (2020) Predicting by-product gradients of baker's yeast production at industrial scale: A practical simulation approach. *Processes* **8**: 1554

Smallbone K, Malys N, Messiha HL, Wishart JA, Simeonidis E (2011) Building a kinetic model of trehalose biosynthesis in Saccharomyces cerevisiae. *Methods in Enzymology* **500**: 355–370

Smallbone K, Messiha HL, Carroll KM, Winder CL, Malys N, Dunn WB, Murabito E, Swainston N, Dada JO, Khan F, *et al* (2013) A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes. *FEBS letters* **587**: 2832–2841

Steiert B, Timmer J, Kreutz C (2016) L 1 regularization facilitates detection of cell type-specific parameters in dynamical systems. *Bioinformatics* **32**: i718–i726

Suarez-Mendez C, Ras C, Wahl S (2017) Metabolic adjustment upon repetitive substrate perturbations using dynamic 13C-tracing in yeast. *Microbial Cell Factories* **16**: 1–14

Suarez-Mendez CA, Sousa A, Heijnen JJ, Wahl A (2014) Fast "feast/famine" cycles for studying microbial physiology under dynamic conditions: a case study with Saccharomyces cerevisiae. *Metabolites* **4**: 347–372

Tamaki H (2007) Glucose-stimulated cAMP-protein kinase A pathway in yeast Saccharomyces cerevisiae. *Journal of bioscience and bioengineering* **104**: 245–250

Tan Y, Rivera JGL, Contador CA, Asenjo JA, Liao JC (2011) Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metabolic engineering* **13**: 60–75

Tang W, Deshmukh AT, Haringa C, Wang G, van Gulik W, van Winden W, Reuss M, Heijnen JJ, Xia J, Chu J, *et al* (2017) A 9-pool metabolic structured kinetic model describing days to seconds dynamics of growth and product formation by Penicillium chrysogenum. *Biotechnology and Bioengineering* **114**: 1733–1743

Teusink B, Passarge J, Reijenga CA, Esgalhado E, Van der Weijden CC, Schepper M, Walsh MC, Bakker BM, Van Dam K, Westerhoff HV, *et al* (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry* **267**: 5313–5329

Theobald U, Mailinger W, Baltes M, Rizzi M, Reuss M (1997) In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae: I. Experimental observations. *Biotechnology and bioengineering* **55**: 305–316

Tiemann CA, Vanlier J, Oosterveer MH, Groen AK, Hilbers PA, van Riel NA (2013) Parameter trajectory analysis to identify treatment effects of pharmacological interventions. *PLoS computational biology* **9**: e1003166

Tran LM, Rizk ML, Liao JC (2008) Ensemble modeling of metabolic networks. *Biophysical journal* **95**: 5606–5617

Tripodi F, Nicastro R, Reghellin V, Coccetti P (2015) Post-translational modifications on yeast carbon metabolism: regulatory mechanisms beyond transcriptional control. *Biochimica et Biophysica Acta BBA General Subjects* **1850**: 620–627

Van Eunen K, Bouwman J, Daran-Lapujade P, Postmus J, Canelas AB, Mensonides FI, Orij R, Tuzun I, Van Den Brink J, Smits GJ, *et al* (2010) Measuring enzyme activities under standardized in vivo-like conditions for systems biology. *The FEBS journal* **277**: 749–760

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.22.497165; this version posted June 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

van Eunen K, Kiewiet JA, Westerhoff HV, Bakker BM (2012) Testing biochemistry revisited: how in vivo metabolism can be understood from in vitro enzyme kinetics. *PLoS computational biology* **8**: e1002483

van Heerden JH, Bruggeman FJ, Teusink B (2015) Multi-tasking of biosynthetic and energetic functions of glycolysis explained by supply and demand logic. *BioEssays* **37**: 34–45

van Heerden JH, Wortel MT, Bruggeman FJ, Heijnen JJ, Bollen YJ, Planqué R, Hulshof J, O'Toole TG, Wahl SA, Teusink B (2014) Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science* **343**: 1245114

van Hoek P, van Dijken JP, Pronk JT (2000) Regulation of fermentative capacity and levels of glycolytic enzymes in chemostat cultures of Saccharomyces cerevisiae. *Enzyme and microbial technology* **26**: 724–736

Van Leemputte F, Vanthienen W, Wijnants S, Van Zeebroeck G, Thevelein JM (2020) Aberrant Intracellular pH Regulation Limiting Glyceraldehyde-3-Phosphate Dehydrogenase Activity in the Glucose-Sensitive Yeast tps1 Δ Mutant. *Mbio* **11**: e02199–20

Van Riel N, Giuseppin M, TerSchure E, Verrips C (1998) A Structured, Minimal parameter Model of the Central Nitrogen Metabolism in Saccharomyces cerevisiae: The Prediction of the Behaviour of Mutants. *Journal of theoretical biology* **191**: 397–414

Verduyn C, Stouthamer AH, Scheffers WA, van Dijken JP (1991) A theoretical evaluation of growth yields of yeasts. *Antonie Van Leeuwenhoek* **59**: 49–63

Villaverde AF, Fröhlich F, Weindl D, Hasenauer J, Banga JR (2019) Benchmarking optimization methods for parameter estimation in large kinetic models. *Bioinformatics* **35**: 830–838

Walther T, Novo M, Rössger K, Létisse F, Loret MO, Portais JC, François JM (2010) Control of ATP homeostasis during the respiro-fermentative transition in yeast. *Molecular systems biology* **6**: 344

Wang G, Haringa C, Noorman H, Chu J, Zhuang Y (2020) Developing a computational framework to advance bioprocess scale-up. *Trends in Biotechnology* **38**: 846–856

Wiebe MG, Rintala E, Tamminen A, Simolin H, Salusjärvi L, Toivari M, Kokkonen JT, Kiuru J, Ketola RA, Jouhten P, *et al* (2008) Central carbon metabolism of Saccharomyces cerevisiae in anaerobic, oxygen-limited and fully aerobic steady-state conditions and following a shift to anaerobic conditions. *FEMS Yeast research* **8**: 140–154

Yasemi M, Jolicoeur M (2021) Modelling cell metabolism: a review on constraint-based steady-state and kinetic approaches. *Processes* **9**: 322