Assessing the impact of physicochemical parameters in the predictive capabilities of thermodynamics-based stoichiometric approaches under mesophilic and thermophilic conditions

Claudio Tomi-Andrino[1,2,3], Rupert Norman[2], Thomas Millat[2], Philippe Soucaille[2,4,5,6], Klaus Winzer[2], David A. Barrett[1], John King[3], Dong-Hyun Kim[1]

[1]Centre for Analytical Bioscience, Advanced Materials and Healthcare Technology Division, School of Pharmacy, University of Nottingham, Nottingham, United Kingdom.

[2]Nottingham BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, BioDiscovery Institute, University of Nottingham, Nottingham, United Kingdom.

[3]Nottingham BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Mathematical Sciences, University of Nottingham, Nottingham, United Kingdom.

[4]INSA, UPS, INP, Toulouse Biotechnology Institute, (TBI), Université de Toulouse, Toulouse, France.

[5]INRA, UMR792, Toulouse, France.

[6]CNRS, UMR5504, Toulouse, France.

*Corresponding author

Email: dong-hyun.kim@nottingham.ac.uk

Abstract

Metabolic engineering in the post-genomic era is characterised by the development of new methods for metabolomics and fluxomics, supported by the integration of genetic engineering tools and mathematical modelling. Particularly, constraint-based stoichiometric models have been widely studied: (i) flux balance analysis (FBA) (*in silico*), and (ii) metabolic flux analysis (MFA) (*in vivo*). Recent studies have enabled the incorporation of thermodynamics and metabolomics data to improve the predictive capabilities of these approaches. However, an in-depth comparison and evaluation of these methods is lacking. This study presents a thorough analysis of four different *in silico* methods tested against experimental data (metabolomics and ^{13}C-MFA) for the mesophile *Escherichia coli* and the thermophile *Thermus thermophilus*. In particular, a modified version of the recently published matTFA toolbox has been created, providing a broader range of physicochemical parameters. In addition, a max-min driving force approach (as implemented in eQuilibrator) was also performed in order to compare the predictive capabilities of both methods.

Validating against experimental data allowed the determination of the best physicochemical parameters to perform the TFA for *E. coli*, whereas the lack of metabolomics data for *T. thermophilus* prevented from a full analysis. Results showed that analytical conditions predicting reliable flux distributions (similar to the *in vivo* fluxes) do not necessarily provide a good depiction of the experimental metabolomics landscape, and that the original matTFA toolbox can be improved. An analysis of flux pattern changes in the central carbon metabolism between ^{13}C-MFA and TFA highlighted the limited capabilities of both approaches for elucidating the anaplerotic fluxes. Finally, this study highlights the need for standardisation in the fluxomics community: novel approaches are frequently released but a thorough comparison with currently accepted methods is not always performed.

Keywords

Constraint-based modelling, fluxomics, metabolomics, thermodynamics.

Author summary

Biotechnology has benefitted from the development of high throughput methods characterising living systems at different levels (e.g. concerning genes or proteins), allowing the industrial production of chemical commodities (such as ethylene). Recently, focus has been put on determining reaction rates (or metabolic fluxes) in the metabolic network of certain microorganisms, in order to identify bottlenecks hindering their exploitation. Two main approaches can be highlighted, termed metabolic flux analysis (MFA) and flux balance analysis (FBA), based on measuring and estimating fluxes, respectively. While the influence of thermodynamics in living systems was accepted several decades ago, its application to study biochemical networks has been only recently enabled. In this sense, a multitude of different approaches constraining well-established modelling methods with thermodynamics has been suggested. However, physicochemical parameters are not properly adjusted to the experimental conditions, which might affect their predictive capabilities. In this study, we improved the reliability of currently available tools by exploring the impact of varying said parameters in the simulation of metabolic fluxes and metabolite concentration values. Additionally, our in-depth analysis allowed us to highlight limitations and potential solutions that should be considered in future studies.

Introduction

Metabolic engineering aims to improve microbial strains by considering comprehensive metabolic pathways in their entirety rather than overexpressing a single gene (1). To improve the strains, hypothesis-driven studies have attempted to rationally identify gene targets and to evaluate the effects of those changes in the network (2, 3). However, the complex nature of cellular metabolism and its regulation demands a holistic understanding, i.e. a data-driven approach (1-3). Combining metabolic engineering with systems biology and mathematical modelling allows for an optimisation of entire cellular networks considering further downstream processes at early stages (4).

This systematic framework exploits information regarding the metabolic state, which comprises the metabolome (set of low-molecular-weight metabolites (<1.5 kDa)) and the fluxome (or metabolic activity, distribution of rates of conversion/transport in the metabolic network) (5, 6). Kinetic modelling can yield metabolic fluxes from metabolomics data, but lack of high-quality enzymatic parameters and computational limitations (e.g. time-consuming processes) hinder its application (7-9). As an alternative, stoichiometric modelling provides a flux distribution without any kinetic or metabolomics information (10). Briefly, a metabolic (quasi) steady-state for intracellular concentration values (C) is assumed, so the stoichiometric matrix (S) (including the stoichiometric coefficients of metabolites in each reaction of the metabolic network) constrains the set of metabolic fluxes (υ) (11):

$$\frac{dC}{dt} = S \times \upsilon \cong 0 \qquad (1)$$

Two main approaches to solve this equation can be found: (i) flux balance analysis (FBA), normally applied to large models (genome-scale model, GSM) (12) or (ii) metabolic flux analysis (MFA), used for smaller metabolic networks (mainly the central carbon metabolism) (Table 1).

FBA solves the underdetermined system represented in Eq. 1 by maximising or minimising the value of an assumed objective function (12). A plethora of different objectives has been described in the literature (13). Three can be highlighted: maximisation of biomass yield ($Y_{X/S}$), maximisation of ATP yield, and minimisation of sum of fluxes, which have been suggested to compete in the regulation of bacterial metabolism (14). Hence, selecting an adequate one/multi-dimensional objective function when analysing a GSM will depend on the growth conditions to be simulated in FBA. In general, measured extracellular metabolic rates

(e.g. substrate uptake) are insufficient to properly describe the intracellular metabolic fluxes (11). In contrast, MFA is based on a least-squares-regression problem, normally solved by exploiting experimental mass isotopomer distribution (MID) of proteinogenic amino acids (^{13}C-MFA) (11). Since this approach requires fewer assumptions and uses more experimental information than FBA, ^{13}C-MFA is considered to be the *gold standard* in fluxomics (15). However, current applicability (central carbon metabolism), and technical/computational complexity (particularly for autotrophic growth (16)) limit its usage.

The set of constraints characterising stoichiometric modelling approaches (Eq. 1) is insufficient to guarantee thermodynamically feasible results in the flux solution space (17, 18). Both FBA and ^{13}C-MFA assume most reactions to be reversible (11, 19): in the first case directionalities are dictated by the optimal flux distribution (which depends on the *a priori* chosen objective function (12)), whereas in ^{13}C-MFA they are determined by the MIDs (20). The flux-force relationship (thermodynamic displacement from the equilibrium (21)) links thermodynamic potentials and fluxes (Eq. 2):

$$\Delta_r G' = \Delta_r G^{o'} + RT\ln Q = RT\ln(Q/k_{eq}) = -RT\ln(J^+/J^-) \quad (2)$$

where $\Delta_r G'$ and $\Delta_r G^{o'}$ are the Gibbs free energies of reactions (the latter referring to adjusted standard conditions), Q and k_{eq} are the ratio of products to reactant concentrations or activities (the latter at equilibrium) and (J^+/J^-) is the relative forward-to-backward flux (20).

Four main approaches exploiting thermodynamics data can be highlighted: (i) energy balance analysis (EBA), where pre-selecting $\Delta_r G'$ bounds leads to biased results (22), (ii) network-embedded thermodynamic (NET) analysis, that needs pre-assigned directionalities (e.g. obtained by FBA) and evaluates the thermodynamic consistency (23), (iii) max-min driving force (MDF), which needs a flux distribution as input data to predict metabolite concentration values (24), and (iv) thermodynamically-constrained FBA. Two methods can be found within the latter: thermodynamics-based flux analysis (TFA), and an optimization problem allowing to obtain a thermodynamically realizable flux-minimised (TR-fluxmin) solution. TFA directly yields a thermodynamically feasible FBA solution (e.g. by maximising $Y_{X/S}$) and simulated metabolomics data (18, 25). In contrast, TR-fluxmin is based on the minimisation of sum of fluxes in the system whilst applying a penalty score for *in silico* metabolite concentration values (19). Other recent approaches are based on alternative constraints, such as setting an upper limit on the Gibbs energy dissipation rate (26).

MDF and TFA are generally performed using eQuilibrator (24) and matTFA (18), respectively. Given the code availability for both tools, they were selected for this study. Three features from both methods should be highlighted: (i) unique values for temperature (25 ℃) are considered, (ii) salinity is not taken into account, and (iii) Gibbs free energy values are adjusted for ionic strength (I) using the extended Debye-Hückel equation (Table 1). In this sense, it should be noted that the cytosol of *E. coli* is normally in the interval 0.15 – 0.20 M (25) (and so, salinity is not null), and the fact that the extended Debye-Hückel equation is valid for $I < 0.1$ M (27). Previous attempts to adjust physicochemical parameters to *in vivo* conditions can be found in the literature, but they require extra assumptions and only provide information regarding reaction directionalities (28).

Table 1. Comparison of frequently used approaches in fluxomics.

	^{13}C-MFA	FBA	MDF	TFA
Metabolic network size	small	GSM	GSM	GSM
Flux distribution	generated	generated	input	generated
Uptake rate	Yes	Yes	-	Yes
Specific growth rate, μ (h^{-1})	-	Yes	-	Yes
Gibbs free energy of formation (ΔG_f°)	-	-	Experimental (29), or CCM (30)	Experimental (29), or GCM (31)
Temperature, t (℃)	-	-	25	
Ionic strength, I (M)	-	-	input	0.25
Salinity, S (g/kg)	-	-	-	-
Adjustment method	-	-	Extended Debye-Hückel	
Parameter A	-	-	T-dependent	
Metabolite concentration values	-	-	Constraint or predicted	
Problem formulation	least square regression (11)	LP (12)	LP (24)	MILP (18)

^{13}C-MFA, ^{13}C metabolic flux analysis; CCM, component contribution method; FBA, flux balance analysis; GCM, group contribution method; GSM, genome-scale model; LP, linear programming; MDF, max-min driving force; MILP, mixed-integer linear programming; TFA, thermodynamics-based flux analysis.

This study was based on determining the impact of varying and adjusting the physicochemical parameters (t, I and S) on the predictive capabilities of thermodynamic-based fluxomics/metabolomics approaches under mesophilic and thermophilic growth conditions. In order to do so, a modified matTFA was developed by increasing the number of parameters and parameter values that were originally considered (18). To validate the results, a comparison with published ^{13}C-MFA and metabolomics data was performed.

Finally, flux pattern changes between *in vivo* and *in silico* fluxes in the central carbon metabolism were analysed, with a particular focus on the anaplerotic reactions. Intermediates

participating in the tricarboxylic acid (TCA) cycle are used for biosynthesis of amino acids (which is robust against changes in concentrations (32, 33)), so a continuous replenishment by anaplerosis is necessary (34). The 'anaplerotic node' consists of carboxylation/decarboxylation reactions including phosphoenolpyruvate, pyruvate, oxaloacetate, and malate (35). Given the fact similar MIDs (from proteinogenic amino acids) can be obtained from different precursors, ^{13}C-MFA has been noted to show a limited capability to elucidate fluxes around the anaplerotic node (32, 36, 37). An approach improving the resolution consisting in also measuring MIDs from intracellular intermediates has been suggested (32), but it is not commonly performed (11). Hence, ^{13}C-MFA data for *E. coli* and *T. thermophilus* (38, 39) was assumed as the *gold standard* in this study, as stated above (15).

Materials and Methods

Metabolic network, mapping of metabolic fluxes and experimental data

Two growth conditions (mesophilic and thermophilic) represented by two species were selected: *Escherichia coli*, widely used in biotechnology, and *Thermus thermophilus*, an extreme thermophile with the potential to become a non-model metabolic engineering platform (40). For *E. coli*, simulations were performed with the commonly used GSM *i*JO1366 (str. K-12 substr. MG1655), as available in BiGG Models (41). This model has proven to predict phenotypes in a wide range of growth conditions (42), and was explored with the original matTFA (18). In the case of *T. thermophilus*, the GSM *i*TT548 for the strain HB27 was used (43), downloaded from (http://darwin.di.uminho.pt/models/models). The metabolic networks were mapped on to previously published ^{13}C-MFA data (S1-S2 Tables). For the sake of consistency, metabolomics and fluxomics data were obtained from the same experiment when possible (Table 2).

Table 2. Bioprocessing, metabolomics and fluxomics (^{13}C-MFA) experimental data.

	***Escherichia coli* K-12**	***Thermus thermophilus* HB8**
Glucose uptake rate (mmol/gDCW-h)	2.93	3.7
Specific growth rate (h^{-1})	0.20	0.22 ± 0.02
$Y_{X/S}$ (gDCW/g)	0.38	0.33 ± 0.02
Temperature (°C)	37	72
Metabolomics	Yes (S2 Dataset)	-
Fluxomics (^{13}C-MFA)	Yes (S1 Table)	Yes (S2 Table)
Ref.	(18, 38)	(39)

It is important to note that for *E. coli* the same strain was used for both the GSM and the ^{13}C-MFA, whereas for *T. thermophilus* strain HB27 was used for constructing the GSM, and HB8 for the ^{13}C-MFA. The *E. coli* cells were grown in glucose-limited chemostats, whereas batch culture was used for *T. thermophilus* instead. GAM, growth-associated maintenance; NGAM, non-growth-associated maintenance; $Y_{X/S}$, biomass yield.

TFA required a higher glucose uptake rate than the experimental one (S1 Appendix), which provoked a difference between predicted and experimental growth rate (which is equal to the dilution rate in a continuous culture). Since the biomass elemental composition does not significantly vary due to changes in the dilution rate (44), biomass reactions remained unchanged in the model (45), and the energetic requirements were assumed to be constant for both bacteria (S1 Appendix). Using the default constraints from the metabolic networks also allowed comparing the results with previously published ones.

In order to achieve compatibility with the COBRA toolbox (46) and matTFA (18), some changes were applied to GSM *i*TT548: (i) the names of the metabolites were adapted to the

convention used in matTFA and associated to *metSEED_IDs* to enable access to the thermodynamics database in matTFA (S1 Dataset) (18), and (ii) the fields *CompartmentData*, *metCompSymbol* and *rev* were created in the model.

Modified matTFA (mod-matTFA) and parameters included in the analysis

The original matTFA toolbox uses unique values for t and I (18), and S is not taken into account (Table 1). To address this potential deficiency, a modified matTFA was created (mod-matTFA) as described below (Table 3). For reproducibility (47), the complete list of files used in this study was collected in S3 Table.

Table 3. Parameters considered in mod-matTFA.

Parameters	Mesophilic conditions (*E. coli*)	Thermophilic conditions (*T. thermophilus*)
Temperature, t (℃)	(0): 25 (1): 37	(0): 25 (1): 72
Ionic strength, I (M)	(0): 0 (1): 0.25	(0): 0.25 (1): 0.50
Salinity, S (g/kg)	(0): 0 (1): 13.74	(0): 13.74 (1): 27.10
Adjustment method	(0): Extended Debye-Hückel equation (1): Davies equation	(0): Extended Debye-Hückel equation (1): Davies equation
Parameter A	(0): T-dependent (1): T,S-dependent	(0): T-dependent (1): T,S-dependent
Metabolite concentration values	(0): Default matTFA (1): experimental data	(0): Default matTFA (1): -

Values 0/1 refer to the binary codification for the full factorial design (S4-S5 Tables). It is important to note that in the case of *E. coli*, 2^6 combinations were tested, whereas the lack of metabolomics data for *T. thermophilus* meant only 2^5 different tests were available. There is a 'default matTFA' constraint regarding set concentrations values for cofactors (AMP, ADP and ATP) as included in the original matTFA code. 'Experimental data' refers to the use of published metabolomics data (S2 Dataset), setting the lower and upper bound for the simulation as 90-110% of the concentration values.

Since I affects the Gibbs energy of formation, an adjustment from the reference state ($\Delta_f G_j^o$) was needed to obtain the standard transformed Gibbs energy of formation ($\Delta_f G_j'^0$) (29). In the original matTFA (18) and other studies (24, 26) the extended Debye-Hückel equation was used to adjust the Gibbs free energy values, with a proven validity for $I < 0.1$ M (27) (Eq. 3). The parameter B was assumed to be constant, with a value of 1.6 $mol^{-1/2}L^{1/2}$ (25, 29). Mod-matTFA also explored the impact of using the Davies equation ($\beta = 0.3$) (Eq. 4) as an alternative adjustment approach, with a tested validity for $I < 0.5$ M (27).

$$\Delta_f G'^{0}_j(I) = \Delta_f G^o_j + N_H(j)RT\ln(10)\mathrm{pH} - RT\left(\frac{A\sqrt{I}}{1+B\sqrt{I}}\right)\left(z_j^2 - N_H(j)\right) \quad (3)$$

$$\Delta_f G'^{0}_j(I) = \Delta_f G^o_j + N_H(j)RT\ln(10)\mathrm{pH} - RT\left(\frac{A\sqrt{I}}{1+\sqrt{I}} - \beta I\right)\left(z_j^2 - N_H(j)\right) \quad (4)$$

Both formulas include terms correcting the pH and I, where $N_H(j)$ is the number of hydrogen atoms in species j, R is the gas constant, T is the absolute temperature and z_j refers to the charge of the species (29). Applying the Gibbs-Helmholtz equation would be necessary to account for temperature different from standard conditions, i.e. 25 ℃, but the lack of measured changes in enthalpy (ΔH^o) for all the metabolites prevents from doing so (48). Hence, variations from 25 ℃ to 37 ℃ or to 72 ℃ were assumed to be small, as shown elsewhere (49). The parameter A is normally assumed to be constant (25) or calculated using a temperature-dependent function (Eq. 5) (18, 24), and the impact of using a temperature/salinity-dependent function (Eq. 6) (48) was also tested in this study (Fig. 1).

$$A\ (\mathrm{mol}^{-1/2}\mathrm{kg}^{1/2}) = 1.10708 - 1.54508\times10^{-3}T + 5.95584\times10^{-6}T^2 \quad (5)$$

$$A\ (\mathrm{mol}^{-1/2}\mathrm{kg}^{1/2}) = \frac{F^3\sqrt{2\varepsilon_0 R^3}}{4\pi\varepsilon_0 N_A}\times\left(\frac{\rho_{sw}(t,S)}{(\varepsilon_{sw}(t,S)T)^3}\right)^{1/2} \quad (6)$$

where the first term includes physical constants (Faraday's constant (F), vacuum permittivity (ε_0), gas constant (R) and Avogadro's number (N_A)), and the second the temperature (both in K, T, and in ℃, t), and salinity (S) dependent functions to calculate the density (ρ_{sw}) (50) and the relative permittivity (ε_{sw}) (51) for seawater (S3 Table). It should be noted that the function to calculate the density for seawater like solutions was used for the thermophile (t = 72 ℃) beyond the limit of applicability (t < 40 ℃).

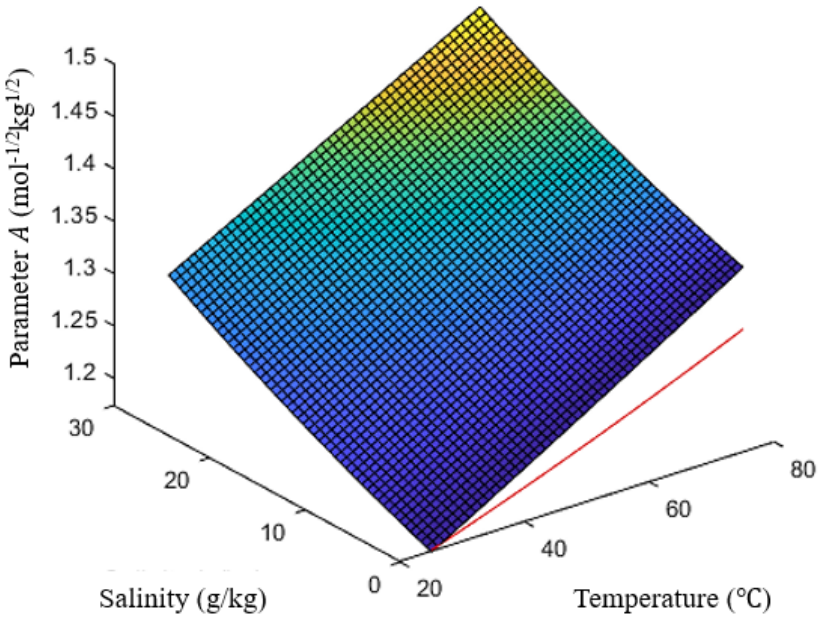

Fig. 1. Calculation of the parameter A. The red line refers to the temperature-dependent function (Eq. 5), whereas the surface is the temperature/salinity-dependent function (Eq. 6).

In general, consistency in units between parameters A ($mol^{-1/2}kg^{1/2}$) and B ($mol^{-1/2}L^{1/2}$) is achieved by assuming 1 kg = 1 L. In this study, an expression for seawater (Eq. 7) (52) was used to estimate a salinity value by considering a buoyant density (ρ) for bacterial cells of 1.11 kg/L (53). For I, values of 0.25 M (upper level for *E. coli*) (18) and 0.50 M (upper level for *T. thermophilus*) were used (Table 3).

$$I(\mathrm{M}) \times \rho(\mathrm{kg/L}) = \frac{19.92 \times S}{1000 - 1.005 \times S} \quad (7)$$

Assessment of fluxomics and metabolomics predictive capabilities

Two different growth conditions (mesophilic and thermophilic) were analysed using two bacteria (*E. coli* and *T. thermophilus*), respectively. Mod-matTFA was allowed to consider a broader range of parameters: 6 for *E. coli* and 5 for *T. thermophilus*, which yielded 64 and 32 different combinations of parameter levels (Table 3). Constraints regarding substrate uptake rate, specific growth rate and energetic requirements were applied as explained in S1 Appendix, and maximisation of $Y_{X/S}$ was selected as objective function. It is important to note that lower and upper boundaries for uptake rates for other macronutrients (such as O_2) were applied as originally constrained in the metabolic networks. To compare the *in silico* fluxes from FBA and TFA with *in vivo* ^{13}C-MFA values (or estimated and experimental metabolite concentration values), a goodness-of-fit analysis based on the Pearson correlation coefficient (r) was performed, as shown in (54). In particular, MATLAB's in-built *corrcoef* function was used.

In the mesophilic case (*E. coli*), the 64 tests were ranked according to two criteria: (i) correlation coefficient at the fluxomics level, and (ii) correlation coefficient at the metabolomics level. In order to assess the concordance of the results, the non-parametric Kendall's W statistics was performed (S3 Table), where a value of 0 means no agreement of ranking position with respect to each criterion, and a value of 1 indicates total agreement. In contrast to the parametric equivalent (Spearman's rank correlation coefficient), Kendall's W accounts for tied ranks (55). Finally, a joint ranking after weighting the ranking position according to each criterion was considered (the higher the score, the better the correlation in both the fluxomics and metabolomics levels).

Prediction of metabolite concentration values with an MDF-based approach (*E. coli*)

Two main distinctions between matTFA and eQuilibrator can be highlighted: (i) the necessity of a flux distribution as input in the latter (24), and (ii) the definition of the problem, which focuses on the MDF framework (24) (Table 1). In this study, the predicted flux distributions from FBA and TFA were analysed using an in-house MDF script based on the eQuilibrator API (Fig. 2), as explained in S2 Appendix. Since metabolites were needed to be named and identified after the Kyoto Encyclopedia of Genes and Genomes (KEGG) (56), a conversion from the GSM *i*JO1366 (42) was performed by using The Chemical Translation Service (57), followed by a manual curation (S3 Dataset).

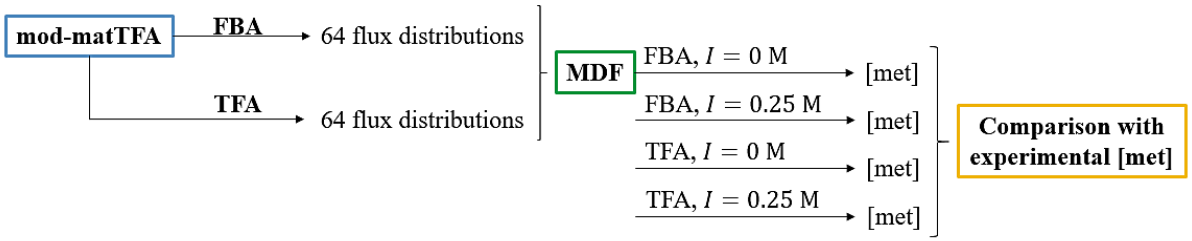

Fig. 2. Workflow to analyse the predictive capabilities of the MDF-based approach. pH = 7 in all cases. [met], metabolite concentration values.

Results

In this study three questions were addressed: (i) how good available thermodynamic-based approaches in predicting metabolic fluxes and metabolite concentrations values are, (ii) whether there is room for improvement by widening the range of physicochemical parameters that are taken into account, and (iii) how reliable the predicted fluxes in the anaplerotic node are.

To tackle these problems, the published matTFA (18) toolbox was modified as shown in S3 Table to include more parameters and a broader range of parameters (Table 3). Two growth conditions represented by two species were selected: *E. coli*, as a widely used organism in biotechnology (mesophile) and *T. thermophilus*, a potential non-model metabolic engineering platform (thermophile). The metabolic network for *E. coli* provided with the original toolbox was used, whereas modifications were necessary to adapt the published GSM for *T. thermophilus* (43). FBA and TFA analysis were performed (64 tests for *E. coli* and 32 for *T. thermophilus*), by assuming maximisation of biomass yield as the objective function. Results were tested against available experimental data (^{13}C-MFA (38, 39) and metabolomics (38)) by calculating the Pearson correlation coefficient, and Kendall's W to determine the agreement between criteria (only for *E. coli*). In addition, a MDF approach was tested against experimental metabolomics data to assess its predictive capabilities in comparison with mod-matTFA. Finally, flux pattern changes between *in vivo* and *in silico* fluxes in the anaplerotic node were compared to identify potential limitations in the predictive capabilities.

Simulation of metabolic fluxes and metabolite concentration values under mesophilic growth conditions (*E. coli*)

The widely used GSM *i*JO1366 (42) was selected for the mod-matTFA analysis, and results were compared with experimental data (metabolomics, fluxomics and bioprocessing data) (Table 2) to evaluate the predictive capabilities of mod-matTFA (S4 Dataset). Particularly, 6 parameters with 2 levels each were tested (Table 3), yielding 64 runs (Fig. 3).

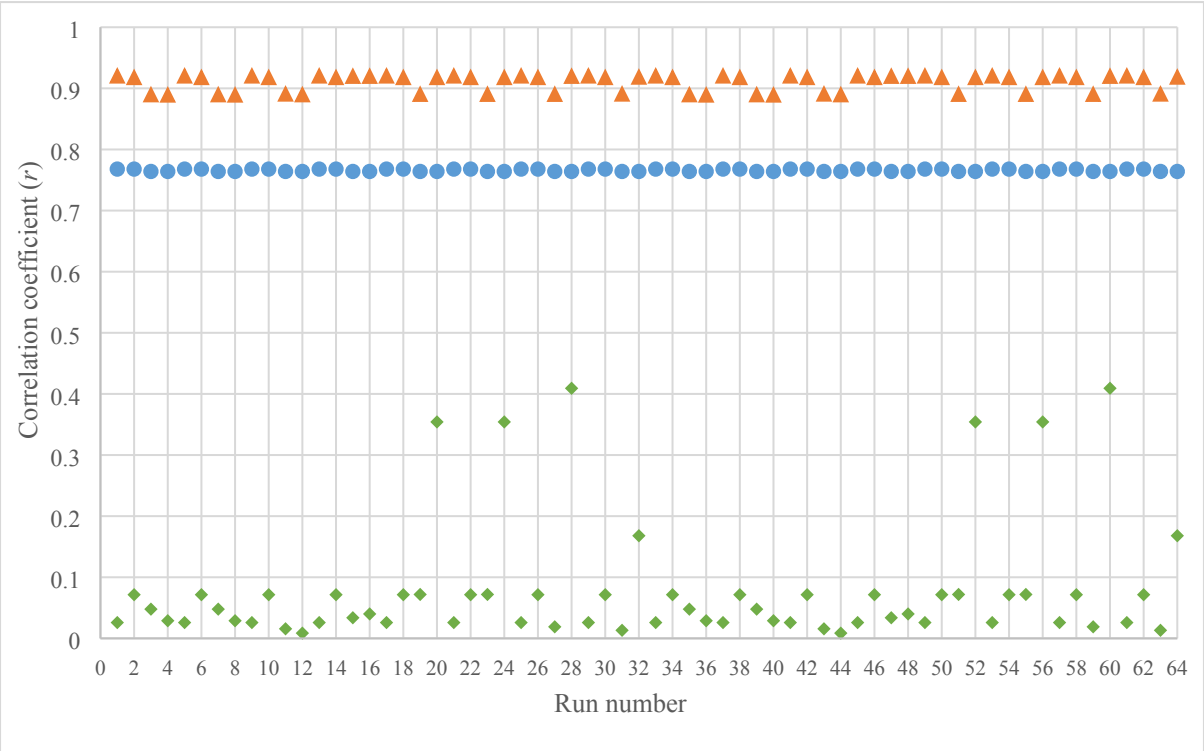

Fig. 3. Goodness-of-fit analysis for predicted flux distributions under mesophilic growth conditions. ● = FBA (after fixing directionalities) against ^{13}C-MFA, ▲ = TFA against ^{13}C-MFA, ♦ = simulated metabolite concentrations values against experimental. Run #3 recreates the analytical conditions as performed in the original matTFA toolbox.

Correlation coefficients for FBA in all runs were $r \approx 0.76$. For TFA, values were either $r \approx 0.91$ or $r \approx 0.89$, where the latter never happened when $I = 0$ M. Metabolite concentration values range in the interval $0 < r < 0.42$. Due to the mismatch between experimental and modelling conditions, $v_{glucose\ uptake}$ had to be set at a value higher than uptake rate (8.16 instead of 2.93 mmol/gDCW-h), as explained in S1 Appendix. Regarding μ, it was higher than the experimental value for FBA and TFA (0.69 and 0.80 h^{-1} versus 0.20 h^{-1}). Hence, the predicted $Y_{X/S}$ values were 0.47 and 0.55 g DCW/g glucose respectively, which differ from the experimental yield (0.38 g DCW/g glucose). For *E. coli*, a $Y_{X/S}^{max} = 0.54$ C-mol glucose/C-mol biomass (0.48 g DCW/g glucose, assuming 70% of water content (58)) has been suggested (59), which is not far from the predicted values. It should be noted that the FBA was performed after fixing directionalities and considering some thermodynamic constraints (18), rather than a traditional FBA (12).

The concordance analysis retrieved a Kendall's $W \approx 0.43$, showing that a high correlation between experimental and simulated metabolic fluxes did not necessarily mean a high correlation between experimental and simulated metabolite concentration values. In order

to identify the run with the best predictive capability at both levels, a joint ranking was performed (Table 5).

Table 5. Runs with the highest score in the joint ranking.

Rank sum	63.5		60.5				57.5		55.5		53.5		14.5
Correlation coefficient TFA vs. ^{13}C-MFA	0.92		0.92				0.92		0.92		0.92		0.89
Correlation coefficient metabolomics	0.41		0.35				0.17		0.04		0.03		0.05
Run number	28	60	20	24	52	56	32	64	16	48	15	47	3*
t (°C) (0 = 25, 1 = 37)	1	1	1	1	1	1	1	1	1	1	0	0	0
I (M) (0 = 0, 1 = 0.25)	1	1	1	1	1	1	1	1	1	1	1	1	1
S (g/kg) (0 = 0, 1 = 13.74)	0	0	0	1	0	1	1	1	1	1	1	1	0
Parameter A (0 = t-dependent, 1 = t/S-dependent)	1	1	0	0	0	0	1	1	1	1	1	1	0
Adjustment method (0 = DH, 1 = Davies)	1	1	1	1	1	1	1	1	0	0	0	0	0
[met] (0 = default, 1 = experimental values)	0	1	0	0	1	1	0	1	0	1	0	1	0

Davies, Davies equation; DH, extended Debye-Hückel equation; [met], metabolite concentration values. Values of 0 and 1 in the headers refer to the binary codification from the full factorial design (S4 Table). *Run #3 represents the analytical conditions from the original matTFA, added here as a reference. The complete ranking can be found in S4 Dataset. There is a 'default matTFA' constraint regarding set concentrations values for cofactors (AMP, ADP and ATP) as included in the original matTFA code. 'Experimental values' refers to the use of published metabolomics data (S2 Dataset). Correlation coefficient values were approximated to the third floating number for ranking purposes.

The quality of predicted flux distributions was overall high: run #3 showed approximately the same correlation coefficient as run #28 (0.89 and 0.92), whilst differing at the metabolomics level (0.05 and 0.41). Hence, varying the physicochemical parameters affected mainly the simulation of the metabolome (Fig. 3). The nature of ^{13}C-MFA only allows determination of flux distributions in the central carbon metabolism by considering amino acid synthesis (11), which has been noted to be very robust against changes in the intermediate metabolite concentrations (32). In addition, the recent discovery of non-enzymatic metabolism-like reactions suggests that current metabolic networks evolved from prebiotic reaction sequences so that a well-established flux distribution in the central pathways can be expected (60).

The best results were achieved by using I = 0.25 M, as done in the original matTFA toolbox. Adjusting t at 37 °C along with using the Davies equation produced an improvement from 4% to 17% at the metabolomics level, without affecting the fluxomics predictive capabilities (4th and 3rd top values, respectively). Interestingly, the runs with the highest joint

score did not consider S, but did use the *t*/S-dependent function for the parameter *A*. Using experimental metabolomics data (38) did not improve the correlation coefficient at the fluxomics or metabolomics level in any run. However, it enabled performing a concordance analysis which showed that a predicted flux distribution with a high correlation coefficient against experimental fluxomics data did not guarantee a strong link between predicted and experimental metabolite concentrations. Consequently, this allowed identifying the set of physicochemical parameters with the highest predictive capability, an assessment that has not been performed in the literature. It has been shown exploiting metabolomics data becomes particularly useful for determining flux patterns when the uncertainty in predicted ΔG_f° is low (49). It should be noted that in matTFA, Gibbs free energy values are relaxed when no feasible solution is found (18) so that the constraining power of experimental metabolite concentration values is reduced.

Simulation of metabolic fluxes under thermophilic growth conditions (*T. thermophilus*)

A GSM for *T. thermophilus* HB27 (43) along with experimental measurements (fluxomics and bioprocessing data) for *T. thermophilus* HB8 (Table 2) were used to assess the fluxomics predictive capabilities of the mod-matTFA (S5 Dataset). Particularly, 5 parameters with 2 levels each were tested (Table 3), yielding 32 runs (Fig. 4).

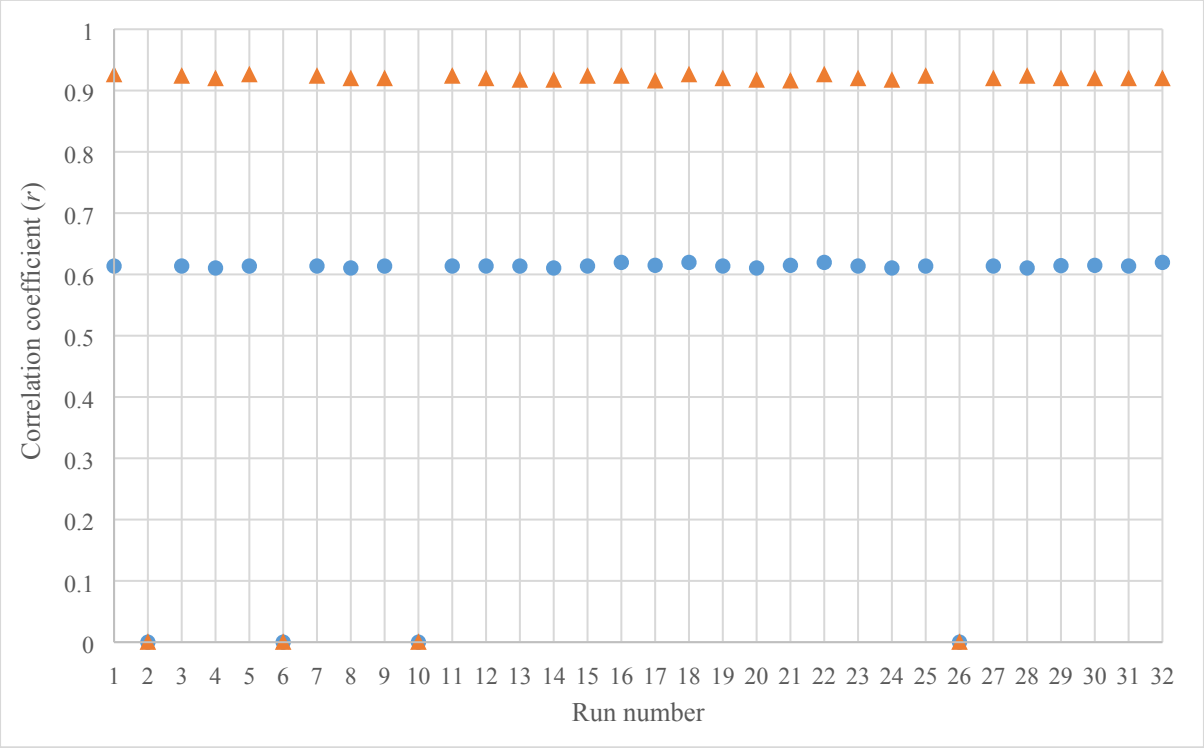

Fig. 4. Goodness-of-fit analysis for predicted flux distributions under thermophilic growth conditions. ● = FBA (after fixing directionalities) against ^{13}C-MFA, ▲ = TFA against ^{13}C-MFA. Run #1 recreates the analytical conditions as performed in the original matTFA toolbox.

The results for both FBA and TFA showed consistency between runs, with $r \approx 0.6$ and $r \approx 0.9$ respectively, using a $v_{glucose\ uptake}$ equivalent to 110% of an experimental value (S1 Appendix). Even though the specific growth rate was constrained in the interval 0.11 to 0.60, predicted values (0.25 and 0.29 h^{-1} for FBA and TFA) were similar to the published value of 0.22 h^{-1} (Table 2). The average predicted $Y_{X/S}$ for FBA (≈ 0.38 g DCW/g glucose) and TFA (≈ 0.44 g DCW/g glucose) proved to be close to the experimental value (≈ 0.33 g DCW/g glucose). As explained for *E. coli*, matTFA/mod-matTFA performs the FBA after fixing directionalities, which depends on thermodynamic parameters. Hence, runs #2, #6, #10 and #26 (both with $T = 72$ ℃ and $I = 0.50$ M in common) resulted from some fixed directionalities so that no feasible solution could be found in FBA and TFA. Since the lack of metabolomics data prevented from further studying the predictive capabilities at both levels, the impact of adjusting the physicochemical parameters to an environment with high salt content and temperature could not be assessed. However, it should be noted that in general, predicted metabolic fluxes in the central carbon metabolism by TFA showed a good correlation coefficient with *in vivo* data, as in the previous case.

Comparison of metabolomics predictions of TFA with an MDF approach (*E. coli*)

MDF-based methods are limited by the fact that they cannot generate flux distributions, so they depends on other approaches to provide that information. eQuilibrator (an user-friendly online MDF-based tool (24)) can predict metabolite concentrations values from a given flux distribution, instead of calculating both at the same time as matTFA does. The 64 flux distributions previously obtained were used as input data for an in-house MDF script (pH = 7.0 with $I = 0$ M or $I = 0.25$ M) (Fig. 2), and the correlation coefficient between predicted metabolite concentration values and experimental metabolomics data was calculated. It should be noted that Gibbs free energy of formation values (ΔG_f°) in the thermodynamic databases for matTFA and eQuilibrator were not exactly the same (Table 1), so that this test focused on comparing their predictive capabilities using eQuilibrator as it is available online.

Overall, MDF showed a better predictive capability than TFA, based on a lower variation between runs calculated with different physicochemical parameters (standard deviations lower than 0.05). For flux distributions obtained by FBA after fixing directionalities,

$r \approx 0.38$ were obtained for all runs when considering both I values. Similarly, $r \approx 0.45$ was achieved for TFA (S6 Dataset), which indicates a lower sensitivity to variations than TFA (Fig. 3). Thus, we believe that eQuilibrator has proven to be ideal for small metabolic networks or parts of pathways, whereas TFA-based approaches should be used when analysing GSM. In this sense, differences in the problem definition (Table 1) should be further studied to identify potential strategies allowing to improve TFA-based approaches.

Flux pattern changes between *in vivo* and *in silico* fluxes in the central carbon metabolism

In order to evaluate changes in reaction directionalities, the available *in vivo* fluxes were tested against their equivalents in the simulated TFA flux distributions (S1-S2 Tables). Overall, the 'anaplerotic node' (Fig. 5) is particularly affected. For *E. coli*, changes in the flux pattern were found for 12/40 of the central carbon metabolism reactions from ^{13}C-MFA (Table 6), out of which three changed between the TFA runs (FBA, PYK and TALA). In the case of *T. thermophilus*, 14/38 mapped reactions showed a different sign from the predicted using matTFA (Table 7).

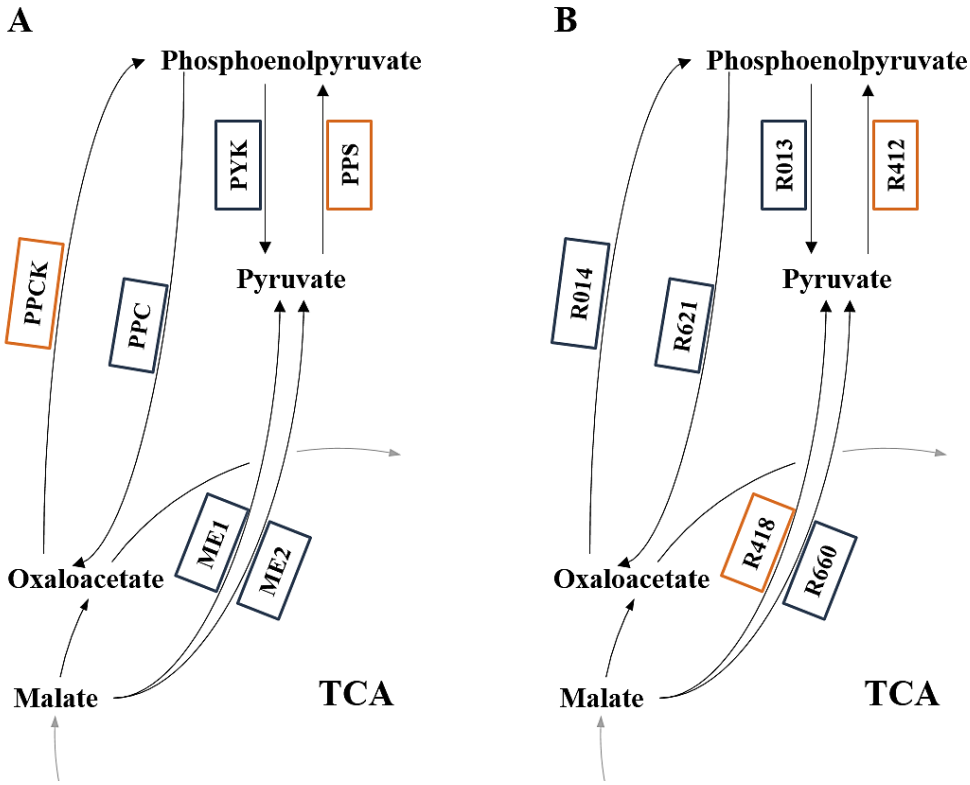

Fig. 5. Anaplerotic node for *E. coli* (A) and *T. thermophilus* (B). Set of carboxylation/decarboxylation reactions including phosphoenolpyruvate, pyruvate, oxaloacetate, and malate. Arrows indicate the expected direction of carbon fluxes. Boxes refer to reactions: blue when they are defined in both the GSM and the metabolic network used for ^{13}C-MFA, and orange when they are exclusively considered in the GSM. In the latter case no mapping was possible (S1-S2 Tables).

Discrepancies in flux pattern between methods are caused by both differences in the structure of the metabolic networks and the way the problem is defined (Table 1). On the one hand, *i*JO1366 includes 8 reactions concerning the anaplerotic node and the glyoxylate shunt (S4 Dataset): PPC and PPCK (between phosphoenolpyruvate and oxaloacetate), PYK and PPS (between phosphoenolpyruvate and pyruvate), ME1 and ME2 (between pyruvate and malate) (Fig. 5), and finally ICL and MALS (from isocitrate to malate, via glyoxylate). In contrast, the metabolic network used for the ^{13}C-MFA did not consider PPCK and PPS (S1 Table), which could have affected the determination of fluxes to/from phosphoenolpyruvate. Since ^{13}C-MFA

is based on lumped reaction, branched pathways are not taken into account (11). Thus, having a smaller range of alternative pathways than FBA/TFA may affect the estimation of flux values.

Table 6. Flux pattern changes between ^{13}C-MFA data and matTFA predictions in *E. coli.*

Reaction (GSM)	Definition (GSM)	Definition (^{13}C-MFA)	Direction (^{13}C-MFA)	Corrected direction (^{13}C-MFA)	Direction (TFA)
ACALD	acald_c + coa_c + nad_c ↔ accoa_c + h_c + nadh_c	AcCoA → Ethanol	+	-	0
ACKr	ac_c + atp_c + h_c ↔ actp_c + adp_c	AcCoA → Acetate	0	0	+
ALCD2x	etoh_c + nad_c ↔ acald_c + h_c + nadh_c	AcCoA → Ethanol	+	-	+
FBA	fdp_c ↔ dhap_c + g3p_c	F1,6P → DHAP + G3P	+	+	0/+
ICL	icit_c → glx_c + succ_c	ICT → Glyoxylate + SUC	+	+	0
ME1	mal-L_c + nad_c → co2_c + nadh_c + pyr_c	MAL → PYR + CO2	+	+	0
ME2	mal-L_c + nadp_c → co2_c + nadph_c + pyr_c	MAL → PYR + CO2	+	+	0
PFK	atp_c + f6p_c <=> adp_c + fdp_c	F6P -> F1,6P	+	+	0/+
PTAr	accoa_c + h_c + pi_c ↔ actp_c + coa_c	AcCoA → Acetate	0	0	-
PYK	adp_c + pep_c ↔ atp_c + pyr_c	PEP → PYR	+	+	0/+
SUCOAS	atp_c + coa_c + succ_c ↔ adp_c + pi_c + succoa_c	2-KG → SUC + CO2	+	+	-
TALA	g3p_c + s7p_c ↔ e4p_c + f6p_c	S7P + G3P ↔ E4P + F6P	+	+	-/0/+

Where +, flux in the forward direction; -, flux in the reverse direction; 0, no flux. *Corrected direction*, refers to the adjustments due to differences in the definition of the reaction between ^{13}C-MFA and GSM (S1 Table). For example the case of ALCD2x: *in vivo* flux (^{13}C-MFA) suggests production of ethanol, whereas the *in silico* one (GSM/TFA) predicts consumption of ethanol. Since reactions are defined in opposite directions, a correction becomes necessary. Discrepancy between corrected directions and predicted ones allowed an automated identification of flux pattern changes.

On the other hand, *in silico* flux distributions are the result of optimising the system according to the chosen objective function. Thus, FBA and TFA promote pathways with a lower energetic cost (when possible), as illustrated by the fact that PPCK (ATP-consuming reaction) carries no flux (S4 Dataset). In contrast, experimental data from *E. coli* grown on glucose has proven that both PPC and PPCK (which constitute a *futile cycle*) are active and

play a role in metabolic regulations (61). However, given the fact that ICL and ME1/ME2 do not generate any ATP, fluxes are shut down in the simulated flux distributions (as shown in (32)). In this sense, it should be noted that stochastic events or regulatory processes have been suggested to provoke a variation of the fluxes through PPCK and ME1/ME2 (62). FBA/TFA also faced problems regarding the overflow metabolism, represented by acetate production (PTAr and ACKr): acetate was consumed rather than produced (32).

Results were similar for *T. thermophilus*. The GSM (*i*TT548) comprises 9 anaplerotic reactions and the glyoxylate shunt (S5 Dataset): R014 and R621 (between phosphoenolpyruvate and oxaloacetate), R013 and R412/413 (between phosphoenolpyruvate and pyruvate), R660 (between pyruvate and malate), and finally R425 and R420 (between isocitrate to malate, via glyoxylate). In this case the PEP-carboxykinase activity (R014) was included in the metabolic network for ^{13}C-MFA (S2 Table). As for *E. coli*, this reaction carried no flux in the TFA (Table 7), and the pool of malate was also affected. Regarding the glyoxylate shunt, it should be noted that R425 (conversion of isocitrate into glyoxylate) carried no flux for both ^{13}C-MFA and TFA. However, the consumption of glyoxylate was activated in TFA (R420), which suggests that alternative pathways must have participated in the production of glyoxylate.

Table 7. Flux pattern changes between ^{13}C-MFA data and matTFA predictions in *T. thermophilus*.

Reaction (GSM)	Definition (GSM)	Definition (^{13}C-MFA)	Direction (^{13}C-MFA)	Corrected direction (^{13}C-MFA)	Direction (TFA)
R014	atp_c + oaa_c → adp_c + pep_c + co2_c	OAC + ATP → PEP + CO2	+	+	0
R016	atp_c + coa_c + ac_c → ppi_c + amp_c + accoa_c	AcCoA ↔ Ac + ATP (net)	-	+	0/+
R024	nad_c + coa_c + akg_c → nadh_c + co2_c + succoa_c	AKG → SucCoA + CO2 + NADH	+	+	0
R026	succ_c + fad_c ↔ fadh2_c + fum_c	Suc ↔ Fum + FADH2 (net)	+	+	-
R027	mal-L_c ↔ h2o_c + fum_c	Fum ↔ Mal (net)	+	-	-/0
R029	glc-D_c + q_c → g15lac_c + qh2_c	*G6P → 6PG + NADPH	+	+	0
R041	2ddg6p_c → g3p_c + pyr_c	KDPG → Pyr + GAP	+	+	0
R420	h2o_c + accoa_c + glx_c → h_c + coa_c + mal-L_c	Glyox + AcCoA → Mal	0	0	+
R621	pep_c + hco3_c ↔ pi_c + oaa_c	PEP + CO2 → OAC	+	+	-

R660	nadp_c + mal-L_c → pyr_c + co2_c + nadph_c	Mal → Pyr + CO2 + NADPH	+	+	0
R710	atp_c + glcn_c → adp_c + h_c + 6pgc_c	G6P → 6PG + NADPH	+	+	0
R713	h2o_c + g15lac_c → h_c + glcn_c	G6P → 6PG + NADPH	+	+	0
R714	6pgc_c → h2o_c + 2ddg6p_c	6PG → KDPG	+	+	0
R722	ac_e ↔ ac_c	Ac → Ac.ext	+	-	-/+

Where +, flux in the forward direction; -, flux in the reverse direction; 0, no flux. *Corrected direction*, refers to the adjustments due to differences in the definition of the reaction between ^{13}C-MFA and GSM (S2 Table). The directionality for R722 is the same: both the definition and the sign are opposed. *Glucose-6-P (G6P) is used instead of glucose (glc-D) due to an incongruence between the metabolic networks (S2 Table).

Even though flux pattern changes between predicted and experimentally determined intracellular fluxes were present, TFA offered a reliable prediction of intracellular fluxes (Figs. 3 and 4). This overall consistency has been noted in the literature by comparing an array of different objective functions and constraints (based on split ratios rather than on mapping on a reaction-by-reaction case) (13). A combination of both approaches to overcome their limitations and discerning flux space solutions has also been suggested (63, 64). However, fluxes concerning the TCA cycle, the glyoxylate shunt and acetate secretion have proven to be difficult to predict (13), as also shown in this study. Similarly, other reactions are also affected by the substrate uptake rate: ALCD2x becomes unidirectional at high glucose levels (26).

In addition, the nonlinear dependency of the anaplerotic fluxes on the growth rate has been reported in the literature, limiting the reliability of conclusions from experiments using single dilution rates (61, 62). Given the fact that substrate uptake rates had to be relaxed (S1 Appendix), predicted growth rates (as well as other fluxes) differed from the corresponding experimental ones (Table 2). Particularly, metabolic fluxes through the aforementioned futile cycle are expected under glucose-limited growth conditions (65), rather than being totally shut down (Fig. 5). In this sense, a higher degree of consistency between predicted and experimental flux distributions could have been achieved by (i) focusing on data from cultures with high dilution rates, so that futile cycle activity is lowered and the flux distribution becomes closer to the optimal solution, or (ii) applying further constraints to properly model the anaplerotic reactions (66). The first option is limited by the lack of published data at both the metabolomics and fluxomics levels for the same experiment, and the second one by the unavailability of the code (consequently it has not been widely used). In this sense, it was assumed that the high correlation coefficient achieved for TFA against *in vivo* fluxomics data ($r \approx 0.9$) was high

481 enough to enable the analyses on the impact of varying the physicochemical parameters in the
482 predictive capabilities.

Discussion

In the last two decades, biotechnology and systems biology have benefitted from the development of ^{13}C-MFA and FBA to measure and estimate intracellular metabolic fluxes in industrially relevant bacteria. Although the influence of thermodynamics in living systems has been considered since several decades ago, its application to study biochemical networks has been only recently enabled (22, 29). In this sense, a multitude of different approaches constraining well-established modelling approaches with thermodynamics have been suggested. Given their relevance and the code availability, this study focused on analysing TFA and MDF (performed by matTFA toolbox and eQuilibrator, respectively). Two main points were explored: (i) their reliability in predicting metabolic fluxes and metabolite concentration values, and (ii) the possibility of improvement by widening the range (and values) of certain physicochemical parameters. Towards this end, GSMs and *in vivo* fluxomics data from the mesophile *E. coli* and the thermophile *T. thermophilus* were selected.

Due to the interest in comparing results with the original matTFA, maximisation of biomass yield ($Y_{X/S}$) was selected as the objective function and energetic requirements maintained (S1 Appendix). Given the nature of ^{13}C-MFA, the validation of predicted fluxomes between different sets of physicochemical parameters could only consider fluxes in the central carbon metabolism. Overall, TFA provided more accurate flux distributions than FBA for both bacteria, even though substrate uptake rates for TFA had to be set higher than the experimental ones to obtain a solution (as set in the original matTFA toolbox). Surprisingly, different sets of physicochemical parameters did not produce changes in the reliability of the predicted flux distributions. We hypothesise that this was due to the proven robustness of metabolic fluxes in these pathways against changes in the metabolic state, as previously noted (32, 33).

Regarding the metabolomics level, our modified matTFA showed that widening the range of parameters and adjusting them to the experimental growth conditions improves the predictive capabilities of TFA. Hence, we suggest the adjustment of the physicochemical parameters when simulating mesophiles and thermophiles (away from biochemical standard conditions) should be considered. The best *in silico* metabolite concentrations profile had a correlation coefficient with experimental data of 41%, against the 5% from the conditions recreating the original matTFA (having in both cases ≈ 90% at the fluxomics level). We believe that a combination of several limitations and factors account for this upper achievable correlation coefficient with experimental concentration values. They can be listed at different

levels: (i) thermodynamic and physicochemical parameters, (ii) problem formulation and constrains, and (iii) suitability of available experimental data.

Apart from intrinsic uncertainties in the experimental or theoretical determination of Gibbs free energy values, it should be noted that it was not possible to account for deviations from standard conditions in temperature by using the Gibbs-Helmholtz equation. In addition, matTFA/mod-matTFA do not consider other relevant factors affecting the thermodynamic feasibility of metabolic pathways such as Mg complexation with metabolites, or compound dissociation into more than two protonated species (17, 18) (as shown in the file *calcDGspecies.m*).

Regarding the problem formulation, although maximisation of $Y_{X/S}$ is the default objective function, recent studies have suggested that maximisation of the ATP yield and minimisation of the sum of fluxes are competing with the former (14). In this sense, TR-fluxmin also defines the problem as a MILP, but focuses on minimising the overall sum of fluxes (whilst optimising a chosen reaction) and considers soft/hard bounds for metabolite concentrations values to allow for relaxation (19). To the best of our knowledge, matTFA does not offer those options (18). Since the objective function determines the flux space solution (12), by extension it also affects the associated metabolite concentrations profile. Hence, experimental values might be outside the allowable metabolite space solution. Studying flux pattern changes on a reaction-by-reaction basis also allowed to confirm previously reported limitations from both ^{13}C-MFA and FBA/TFA with regards to the anaplerotic node (36, 37, 66). Consequently, metabolites in the node are expected to be directly affected. Potential solutions adding extra constraints have been suggested in the literature (66), but they have not been widely implemented.

Our results showed that using predefined ATP/ADP/AMP concentration values (as in the original matTFA) or constraining with experimental metabolomics data lead to the same predictive capabilities (Table 5), when maximising $Y_{X/S}$. In this sense, the possibility of achieving different metabolic space solutions when assuming another objective function cannot be ruled out, which stresses the necessity for accurate quantitative metabolomics data (6). For the matter of our analysis, it should be noted that pre-existing metabolite concentration values focusing on the central carbon metabolism were used. Alternatively, there are theoretical approaches based on sensitivity analysis to identify metabolites of interest to be considered during the experimental design (67). As a matter of fact, relative metabolite abundance data has been successfully combined with thermodynamics to improve flux prediction between differential physiological states (54). The impact of the inherent dynamics (cell cycle and cell

ageing) has been pointed out as a source of metabolic heterogeneity in clonal microbial populations (68). In a chemostat, cells are maintained at the exponential growth phase, but the cell cycle is not synchronised across single cells unless forced (69, 70). In *E. coli*, concentration values for NAD(P)H oscillate along the cell cycle (71), and ATP concentration values show an asymmetric distribution across single cells in a continuous culture (72). Hence, it can be assumed that a distribution of cells at different stages is achieved in steady state, so that experimental fluxomics and metabolomics data reflect a weighted average of the different flux distributions and metabolite concentration profiles from each stage. In this sense, it should not surprise that the predicted concentration values from one flux distribution (obtained by optimising just one objective function) differ from the average experimental profile. However, given the robustness of fluxes in the central carbon metabolism (33, 60), we do not expect this phenomenon to explain by itself deviations in metabolite concentration values in the central carbon metabolism. Instead, we believe that the predictive capabilities of this approach depend on all the previously stated limitations, as well as the fact that phenomena such as substrate tunnelling (according to which intermediates are not released into solution) (73) or cell size variations over the cell cycle (which directly affects the concentration values) (74) were not considered.

Regarding MDF, using the predicted fluxomes (FBA and TFA) as input data for an eQuilibrator-like approach (MDF-based) did not result in remarkably improved simulated metabolite concentration values. Thus, we believe a TFA-based approach should be used for analysing GSMs, and eQuilibrator to be used as a user-friendly biochemical calculator for smaller metabolic networks. Nevertheless, similarities and differences regarding the problem definition could be an interesting source to further develop the TFA framework.

This study proved that the predictive capabilities of thermodynamics-based stoichiometric approaches can be improved by adjusting the considered physicochemical parameters to the experimental conditions. Additionally, our study stressed out the necessity of performing an in-depth assessment of available methods in the fluxomics field. In particular, we believe interesting published potential solutions to known problems (e.g. elucidation of the anaplerotic fluxes) should be integrated with the widely used approaches. This should increase the degree of standardisation in the community, allowing to cross-validate novel strategies and improving the reliability of the simulated data.

Acknowledgements

We would like to thank Nicole Pearcy for her helpful comments.

References

1. Stephanopoulos G, Alper H, Moxley J. Exploiting biological complexity for strain improvement through systems biology. Nat Biotechnol. 2004;22(10):1261-7.
2. Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. BioEssays. 2003;26:99-105.
3. Dai Z, Nielsen J. Advancing metabolic engineering through systems biology of industrial microorganisms. Curr Opin Biotechnol. 2015;36:8-15.
4. Park JH, Lee SY, Kim TY, Kim HU. Application of systems biology for bioprocess development. Trends Biotechnol. 2008;26(8):404-12.
5. Toya Y, Shimizu H. Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. Biotechnol Adv. 2013;31(6):818-26.
6. Kim DH, Achcar F, Breitling R, Burgess KE, Barrett MP. LC-MS-based absolute metabolite quantification: application to metabolic flux measurement in trypanosomes. Metabolomics. 2015;11(6):1721-32.
7. Cortassa S, Caceres V, Bell LN, O'Rourke B, Paolocci N, Aon MA. From metabolomics to fluxomics: a computational procedure to translate metabolite profiles into metabolic fluxes. Biophysical Journal. 2015;108(1):163-72.
8. van Eunen K, Kiewiet JAL, Westerhoff HV, Bakker BM. Testing Biochemistry Revisited: How In Vivo Metabolism Can Be Understood from In Vitro Enzyme Kinetics. PLoS Comput Biol. 2012;8(4).
9. Tomar N, De RK. Comparing methods for metabolic network analysis and an application to metabolic engineering. Gene. 2013;521(1):1-14.
10. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO. Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol. 2009;7(2):129-43.
11. Antoniewicz MR. Methods and advances in metabolic flux analysis: a mini-review. J Ind Microbiol Biotechnol. 2015;42(3):317-25.
12. Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nat Biotechnol. 2010;28(3):245-8.
13. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Molecular Systems Biology. 2007;3(119).
14. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer M. Multidimensional optimality of microbial metabolism. Science. 2012;336:601-4.
15. Long CP, Antoniewicz MR. High-resolution 13C metabolic flux analysis. Nat Protoc. 2019.
16. Wiechert W, Noh K. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. Curr Opin Biotechnol. 2013;24(6):979-86.
17. Vojinović V, Von Stockar U. Influence of uncertainties in pH, pMg, activity coefficients, metabolite concentrations, and other factors on the analysis of the thermodynamic feasibility of metabolic pathways. Biotechnol Bioeng. 2009;103(4):780–95.
18. Salvy P, Fengos G, Ataman M, Pathier T, Soh KC, Hatzimanikatis V. pyTFA and matTFA a Python package and a Matlab toolbox for Thermodynamics-based Flux Analysis. Bioinformatics. 2018;1(3).
19. Hoppe A, Hoffmann S, Holzhütter H-G. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. BMC Syst Biol. 2007;1(23).
20. Park JO, Rubin SA, Xu YF, Amador-Noguez D, Fan J, Shlomi T, et al. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nat Chem Biol. 2016;12.
21. Ataman M, Hatzimanikatis V. Heading in the right direction: thermodynamics-based network analysis and pathway engineering. Curr Opin Biotechnol. 2015;36:176–82.
22. Beard DA, Babson E, Curtis E, Qian H. Thermodynamic constraints for biochemical networks. J Theor Biol. 2004;228:327-33.
23. Kümmel A, Panke S, Heinemann M. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. Mol Syst Biol. 2006.

643 24. Flamholz A, Noor E, Bar-Even A, Milo R. eQuilibrator - the biochemical
644 thermodynamics calculator. Nucleic Acids Res. 2012;40:D770–D5.
645 25. Henry CS, Broadbelt LJ, Hatzimanikatis V. Thermodynamics-Based Metabolic Flux
646 Analysis. Biophys J. 2007;92:1792–805.
647 26. Niebel B, Leupold S, Heinemann M. An upper limit on Gibbs energy dissipation
648 governs cellular metabolism. Nat Metab. 2019;1:125–32.
649 27. Simonin J-P. Thermodynamic consistency in the modeling of speciation in
650 selfcomplexing electrolytes. Ind Eng Chem Res. 2017;56:9721-33.
651 28. Fleming RM, Thiele I. von Bertalanffy 1.0: a COBRA toolbox extension to
652 thermodynamically constrain metabolic models. Bioinformatics. 2011;27(1):142-3.
653 29. Alberty RA. Thermodynamics of Biochemical Reactions: John Wiley & Sons; 2005.
654 30. Noor E, Haraldsdóttir HS, Milo R, Fleming RM. Consistent estimation of Gibbs
655 energy using component contributions. PLoS Comput Biol. 2013;9(7).
656 31. Jankowski K, Henry CS, Broadbelt LJ, Hatzimanikatis V. Group contribution method
657 for thermodynamic analysis of complex metabolic networks. Biophys J. 2008;95:1487-99.
658 32. Toya Y, Ishii N, Nakahigashi K, Hirasawa T, Soga T, Tomita M, et al. 13C-metabolic
659 flux analysis for batch culture of Escherichia coli and its pyk and pgi gene knockout mutants
660 based on mass isotopomer distribution of intracellular metabolites. Biotechnol Prog.
661 2010;26(4):975-92.
662 33. Costenoble R, Muller D, Barl T, van Gulik WM, van Winden WA, Reuss M, et al. 13C-
663 Labeled metabolic flux analysis of a fed-batch culture of elutriated Saccharomyces
664 cerevisiae. FEMS Yeast Res. 2007;7(4):511-26.
665 34. Owen OE, Kalhan SC, Hanson RW. The key role of anaplerosis and cataplerosis for
666 citric acid cycle function. J Biol Chem. 2002;277(34):30409-12.
667 35. Sauer U, Eikmanns BJ. The PEP-pyruvate-oxaloacetate node as the switch point for
668 carbon flux distribution in bacteria. FEMS Microbiol Rev. 2005;29(765-794).
669 36. Kappelmann J, Wiechert W, Noack S. Cutting the Gordian Knot: Identifiability of
670 anaplerotic reactions in Corynebacterium glutamicum by means of (13) C-metabolic flux
671 analysis. Biotechnol Bioeng. 2015;113(3).
672 37. Fischer E, Zamboni N, Sauer U. High-throughput metabolic flux analysis based on
673 gas chromatography–mass spectrometry derived 13C constraints. Anal Biochem.
674 2004;325(2):308-16.
675 38. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, et al. Multiple high-
676 throughput analyses monitor the response of E.coli to perturbations. Science.
677 2007;316(5824):593-7.
678 39. Cordova LT, Cipolla RM, Swarup A, Long CP, Antoniewicz MR. 13C metabolic flux
679 analysis of three divergent extremely thermophilic bacteria: Geobacillus sp. LC300,
680 Thermus thermophilus HB8, and Rhodothermus marinus DSM 4252. Metab Eng.
681 2017;44:182-90.
682 40. Crosby JR, Laemthong T, Lewis AM, Straub CT, Adams MW, Kelly RM. Extreme
683 thermophiles as emerging metabolic engineering platforms. Curr Opin Biotechnol.
684 2019;59:55-64.
685 41. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A
686 platform for integrating, standardizing, and sharing genome-scale models. Nucleic Acids
687 Res. 2016;44:D515–D22.
688 42. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive
689 genome-scale reconstruction of Escherichia coli metabolism—2011. Mol Syst Biol.
690 2011(535).
691 43. Lee N-R, Lakshmanan M, Aggarwal S, Song J-W, Karimi IA, Lee D-Y, et al. Genome-
692 scale metabolic network reconstruction and in silico flux analysis of the thermophilic
693 bacterium Thermus thermophilus HB27. Microb Cell Fact. 2014;13(61).
694 44. Folsom JP, Carlson RP. Physiological, biomass elemental composition and proteomic
695 analyses of Escherichia coli ammonium-limited chemostat growth, and comparison with
696 iron- and glucose-limited chemostat growth. Microbiology. 2015;161:1659-70.
697 45. J. CSH, Cai J, Wang L, Simons-Senftle MN, Maranas CD. Standardizing biomass
698 reactions and ensuring complete mass balance in genome-scale metabolic models.
699 Bioinformatics. 2017;33(22):3603-9.

700 46. Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, et al. Quantitative
701 prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.
702 Nat Protoc. 2011;6(9):1290-307.
703 47. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten Simple Rules for Reproducible
704 Computational Research. PLoS Comput Biol. 2013;9(10).
705 48. Atkins P, Paula Jd. Equilibrium electrochemistry. Atkins' physical chemistry. United
706 States: Oxford University Press; 2002.
707 49. Fleming RM, Thiele I, Nasheuer HP. Quantitative assignment of reaction
708 directionality in constraint-based models of metabolism: Application to Escherichia coli.
709 Biophys Chem. 2009;145:47-56.
710 50. Siedler G, Peters H. Physical properties (general) of sea water. Oceanography.
711 Landolt-Börnstein: Numerical data and functional relationships in science and technology.
712 V/3a. Berlin: Springer; 1986. p. 233-64.
713 51. Meissner T, Wentz FJ. The complex dielectric constant of pure and sea water from
714 microwave satellite observations. IEEE T Geosci Remote. 2004;42(9):1836-49.
715 52. Millero FJ, Leung WH. Thermodynamics of seawater at one atmosphere. Am J Sci.
716 1976;276:1035-77.
717 53. Baldwin WW, Myer R, Powell N, Anderson E, Koch AL. Buoyant density of
718 Escherichia coli is determined solely by the osmolarity of the culture medium. Arch
719 Microbiol. 1995;164:155-7.
720 54. Pandey V, Hadadi N, Hatzimanikatis V. Enhanced flux prediction by integrating
721 relative expression and relative metabolite abundance into thermodynamically consistent
722 metabolic models. PLoS Comput Biol. 2019;15(5).
723 55. Field A. Discovering statistic using SPSS. 3rd ed: SAGE Publications; 2009.
724 56. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic
725 Acids Res. 2000;28(1):27-30.
726 57. Wohlgemuth G, Haldiya PK, Willighagen E, Kind T, Fiehn O. The Chemical
727 Translation Service—a web-based tool to improve standardization of metabolomic reports.
728 Bioinformatics. 2010;26(20):2647-8.
729 58. Neidhardt FC, Curtiss R. Escherichia coli and Salmonella: cellular and molecular
730 biology. Second ed. Washington, D.C.: ASM Press; 1996.
731 59. von Stockar U, Maskow T, Liu J, Marison IW, Patiño R. Thermodynamics of microbial
732 growth and metabolism: An analysis of the current situation. J Biotechnol. 2006;121:517-
733 33.
734 60. Ralser M. An appeal to magic? The discovery of a non-enzymatic metabolism and
735 its role in the origins of life. Biochem J. 2018;475:2577-92.
736 61. Yang C, Hua Q, Baba T, Mori H, Shimizu K. Analysis of Escherichia coli anaplerotic
737 metabolism and its regulation mechanisms from the metabolic responses to altered
738 dilution rates and phosphoenolpyruvate carboxykinase knockout. Biotechnol Bioeng.
739 2003;84(2):129-44.
740 62. Nanchen A, Schicker A, Sauer U. Nonlinear dependency of intracellular fluxes on
741 growth rate in miniaturized continuous cultures of Escherichia coli. Appl Environ Microb.
742 2006;72(2):1164-72.
743 63. Williams TCR, Poolman MG, Howden AJM, Schwarzlander M, Fell DA, Ratcliffe RG,
744 et al. A genome-scale metabolic model accurately predicts fluxes in central carbon
745 metabolism under stress conditions. Plant Physiol. 2010;154:311-23.
746 64. Chen X, Alonso AP, Allen DK, Reed JL, Shachar-Hill Y. Synergy between (13)C-
747 metabolic flux analysis and flux balance analysis for understanding metabolic adaptation
748 to anaerobiosis in E. coli. Metab Eng. 2011;13(1):38-48.
749 65. Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, Szyperski T, et al. Metabolic flux
750 ratio analysis of genetic and environmental modulations of Escherichia coli central carbon
751 metabolism. J Bacteriol. 1999;181(21):6679-88.
752 66. Myoung Park J, Yong Kim T, Yup Lee S. Prediction of metabolic fluxes by
753 incorporating genomic context and flux-converging pattern analyses. P Natl Acad Sci USA.
754 2010;107(33):14931–6.
755 67. Kiparissides A, Hatzimanikatis V. Thermodynamics-based Metabolite Sensitivity
756 Analysis in metabolic networks. Metab Eng. 2017;39:117-27.

757 68. Takhaveev V, Heinemann M. Metabolic heterogeneity in clonal microbial
758 populations. Curr Opin Microbiol. 2018;45:30-8.
759 69. Goodwin BC. Synchronization of Escherichia coli in a chemostat by periodic
760 phosphate feeding. Eur J Biochem. 1969;10:511-4.
761 70. Massie TM, Blasius B, Weithoff G, Gaedke U, Fussmann GF. Cycles, phase
762 synchronization, and entrainment in single-species phytoplankton populations. P Natl Acad
763 Sci USA. 2010;107(9):4236-41.
764 71. Zhang Z, Milias-Argeitis A, Heinemann M. Dynamic single-cell NAD(P)H
765 measurement reveals oscillatory metabolism throughout the E. coli cell division cycle. Sci
766 Rep-UK. 2018;8(2162).
767 72. Yaginuma H, Kawai S, Tabata KV, Tomiyama K, Kakizuka A, Komatsuzaki T, et al.
768 Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative
769 single-cell imaging. Sci Rep-UK. 2014;4(6522).
770 73. Huang X, Holden HM, Raushel FM. Channeling of substrates and intermediates in
771 enzyme-catalyzed reactions. Annu Rev Biochem. 2001;70:149-80.
772 74. Barber F, Ho P-Y, Murray AW, Amir A. Details matter: noise and model structure
773 set the relationship between cell size and cell cycle timing. Front Cell Dev Biol. 2017;5(92).

774
775

Author contributions

Conceptualisation: Claudio Tomi-Andrino, Thomas Millat

Data curation: Claudio Tomi-Andrino

Formal analysis: Claudio Tomi-Andrino, Rupert Norman

Methodology: Claudio Tomi-Andrino, Rupert Norman, Thomas Millat, Philippe Soucaille

Project administration: Claudio Tomi-Andrino

Resources: Klaus Winzer, David A. Barrett, John King, Dong-Hyun Kim

Software: Claudio Tomi-Andrino, Rupert Norman, Thomas Millat

Supervision: Klaus Winzer, David A. Barrett, John King, Dong-Hyun Kim

Writing – original draft: Claudio Tomi-Andrino

Writing – review & editing: Claudio Tomi-Andrino, Rupert Norman, Thomas Millat, Philippe Soucaille, Klaus Winzer, David A. Barrett, John King, Dong-Hyun Kim

Supporting information

S1 Appendix. Energetic requirements and determination of analytical conditions.

S2 Appendix. Max-min driving force linear program in Python.

S1 Dataset. Dictionary of metabolites (*T. thermophilus*).

S2 Dataset. Metabolomics Keio database (*E. coli*).

S3 Dataset. Dictionary of metabolites (*E. coli* - from matTFA to eQuilibrator).

S4 Dataset. Results of the mod-matTFA analysis for *E. coli.*

S5 Dataset. Results of the mod-matTFA analysis for *T. thermophilus.*

S6 Dataset. Correlation coefficients of MDF-derived metabolite concentration values.

S1 Table. Mapping of metabolic fluxes (*E. coli*).

S2 Table. Mapping of metabolic fluxes (*T. thermophilus*).

S3 Table. List of files used in this study.

S4 Table. Full factorial design (*E. coli*).

S5 Table. Full factorial design (*T. thermophilus*).