Supporting information

This document provides information in support of our article "Bayesian regression facilitates quantitative modelling of cell metabolism".

The results of all reported Maud runs can be found at https://github.com/biosustain/Methionine_model/blob/main/results.

1 Maud's input format

Maud inputs are structured directories, somewhat inspired by the PEtab format (Schmiester et al. 2021). A Maud input directory must contain a toml (Preston-Werner, Tom and Gedam, Pradyun 2020) file called `config.toml` which gives the input a name, configures how Maud will be run and tells Maud where to find the other files, allowing these to have custom names. It must also include a file containing a kinetic model definition, a file specifying information about parameters and a file with information experiments. The required structure of these files is documented at https://maud-metabolic-models.readthedocs.io/en/latest/inputting.html. The input is validated against a Pydantic (Pydantic developers 2022) data model.

We chose to implement a custom input format despite the existence of standard formats in similar areas, including SBML (Keating et al. 2020) and PEtab (Schmiester et al. 2021). This choice was partly motivated by the need to ensure flexibility as Maud was developed, but there are also features of SBML and PEtab that make them structurally unsuitable in this context. Our requirements for an input format included that it be mathematics-free, so that all mathematical details are encapsulated in source code, and that it has a detailed, verifiable structure. These requirements made toml more attractive than SBML: toml is easier for humans to read and edit and can straightforwardly be validated using tools like Pydantic. Further, an SBML representation of our desired input would not contain differential equations. It would therefore not be interoperable with most SBML targeting software, which typically assumes that differential equations are available and does not know about Maud's structure.

2 Maud's kinetic model

2.1 Parameters

Table 1 shows all of Maud's unknown parameters along with their dimensions

Note that Maud's metabolic model includes some quantities that are not treated as parameters in its statistical model, including temperatures, compartment volumes and the formation energy of water. Maud treats these quantities as if they were known precisely: they can be configured by the user or default values can be used. Although in practice there can be considerable uncertainty regarding these quantities, we chose to disregard this uncertainty in the interest of simplicity.

Table 1: Parameters of Maud's statistical model

Parameter	Modelled quantity	Dimensions
$\Delta_f G$	Formation energy	metabolites
k_M	Michaelis Menten constants	Substrates of all enzyme/reactions and products of reversible enzyme/reactions
k_I	Inhibition constants	Inhibiting metabolite/compartments of enzyme/reactions exhibiting competitive inhibition
k_{cat}	Rate constants	Enzyme/reactions
L_0	Transfer constants	Allosteric interactions
e_T	T dissociation constants	Modifying metabolites of allosteric inhibitions
e_R	R dissociation constants	Modifying metabolites of allosteric activations
$k_{cat\,pme}$	Rate constants of phosphorylation modifying enzymes	Phosphorylation modifying enzymes
v_{drain}	Drain fluxes	Drains, experiments
$Enzyme$	Enzyme concentrations	Enzymes, experiments
$C_{unbalanced}$	Unbalanced metabolite/compartment concentrations	Unbalanced metabolite/compartments, experiments
C_{pme}	Phosphorylation modifying enzyme concentrations	Phosphorylation modifying enzymes, experiments

Parameter	Modelled quantity	Dimensions
ψ	Membrane potentials	Experiments

Solving the steady state problem for a given set of parameters in an experiment yields a vector $C_{balanced}$ of balanced metabolite concentrations. These are combined with the balanced metabolite concentrations $C_{unbalanced}$ to produce a vector C_{mic} with a concentration for each metabolite/compartment combination.

$\Delta_f G$ parameters can optionally be fixed; this can be useful for computational purposes, as for example to avoid estimating the formation energy of a metabolite about which there is no available information due to it only participating in irreversible reactions.

2.2 Rate equations

As discussed in the main text, Maud's kinetic model decomposes into factors contributing to the flux in a metabolic network in an experiment as shown in equation (1). For succinctness, and since Maud's model assumes that there are no interactions between experiments, we omit any notation referring to experiments below. We also omit any reference to the network's drain reactions: these are modelled as being exactly determined by the values of the parameter vector v_{drain}.

$$F(C;\theta) = Enzyme \cdot k_{cat} \cdot Reversibility \cdot Saturation \cdot Allostery \tag{1}$$

The term *Enzyme* in equation (1) is a vector of non-negative real numbers representing the concentration of the enzyme catalysing each reaction.

The term k_{cat} in equation (1) is a vector of non-negative real numbers representing the amount of flux carried per unit of saturated enzyme.

The term *Reversibility* in equation (1) is a vector of real numbers capturing the impact of thermodynamic effects on the reaction's flux, as shown in equation (2).

$$\begin{aligned} Reversibility &= 1 - \exp\left(\frac{\Delta_r G + RT \cdot S^T \ln(C_{mic})}{RT}\right) \\ \Delta_r G &= S^T \Delta_f G + nF\psi \end{aligned} \tag{2}$$

The terms in (2) have the following meanings:

- T is the temperature in Kelvin (a number),
- R is the gas constant (a number),

- $\Delta_r G$ is a vector representing the Gibbs free energy change of each reaction in standard conditions,
- $\Delta_f G$ is a vector representing the standard condition Gibbs free energy change of each metabolite's formation reaction, or in other words each metabolite's 'formation energy'.
- n is a vector representing the number of charges transported by each reaction.
- F is the Faraday constant (a number)
- ψ is a vector representing each reaction's membrane potential (these numbers only matter for reactions that transport non-zero charge)

Note that, for reactions with zero transported charge, the thermodynamic effect on each reaction is derived from metabolite formation energies. This formulation is helpful because, provided that all reactions' rates are calculated from the same formation energies, they are guaranteed to be thermodynamically consistent.

The term n accounts for both the charge and the directionality. For instance, a reaction that exports 2 protons to the extracellular space in the forward direction would have -2 charge. If a negatively charged molecule like acetate is exported in the forward direction, n would be 1.

Note that this way of modelling the effect of transported charge does not take into account that the concentration gradient used by the transport is that of the dissociated molecules. Thus, this expression is only correct for ions whose concentration can be expressed in the model only in the charged form; e.g., protons, K^+, Na^+, Cl^-, etc.

The term *Saturation* in equation (1) is a vector of non-negative real numbers representing, for each reaction, the fraction of enzyme that is saturated, i.e. bound to one of the reaction's substrates. To describe saturation we use equation (3), which is taken from Liebermeister, Uhlendorf, and Klipp (2010). Additionally, this term captures competitive inhibition: as competitive inhibitor concentration increases, the saturation denominator increases, effectively decreasing the saturation of the substrate on the total enzyme pool. Conversely, as the substrate concentration increases this term approaches 1.

$$Saturation_r = a \cdot \text{free enzyme ratio} \tag{3}$$

$$a = \prod_{\text{s substrate}} \frac{C^s_{mic}}{k^{rs}_M}$$

$$\text{free enzyme ratio} = \begin{cases} \prod_{\text{s sustrate}} (1 + \frac{C^s_{mic}}{k^{rs}_M})^{S_s r} + \sum_{\text{c inhibitor}} \frac{C^c_{mic}}{k^{rc}_I} & r \text{ irreversible} \\ -1 + \prod_{\text{s sustrate}} (1 + \frac{C^s_{mic}}{k^{rs}_M})^{S_s r} + \sum_{\text{c inhibitor}} \frac{C^c_{mic}}{k^{rc}_I} + \prod_{\text{p product}} (1 + \frac{C^p_{mic}}{k^{rp}_M})^{S_p r} & r \text{ reversible} \end{cases}$$

The term *Allostery* in equation (1) is a vector of non-negative numbers describing the effect of allosteric regulation on each reaction. Allosteric regulation happens when binding to a certain molecule changes an enzyme's shape in a way that changes its catalytic behaviour. We use equation (4) to describe this phenomenon, following the generalised MWC approach described

in Monod, Wyman, and Changeux (1965), Changeux (2013), Popova and Sel'kov (1975) and Popova and Sel'kov (1979).

$$
\begin{aligned}
Allostery_r &= \frac{1}{1 + L_0^r \cdot (\text{free enzyme ratio}_r \cdot \frac{Qtense}{Qrelaxed})^{subunits}} \\
Qtense &= 1 + \sum_{\text{i inhibitor}} \frac{C^i_{mic}}{e^{ri}_T} \\
Qrelaxed &= 1 + \sum_{\text{a activator}} \frac{C^a_{mic}}{e^{ra}_R}
\end{aligned}
\tag{4}
$$

The parameter L_0 in equation (1) is called the transfer constant, and the parameter vectors e_T and e_R are called tense and relaxed dissociation constants respectively.

Finally, the term *Phosphorylation* in equation (1) captures the important effect whereby enzyme activity is altered due to a coupled process of phosphorylation and dephosphorylation. This description achieves a similar behaviour to the MWC formalism for describing allosteric regulation, but using the rates of phosphorylation and dephosphorylation rather than concentrations of metabolites.

$$
\begin{aligned}
Phosphorylation_r &= (\frac{\alpha}{\alpha + \beta})^{subunits} \\
\alpha &= \sum_{\text{p phosphoylator}} k^p_{cat\,pme} \cdot C^p_{pme} \\
\beta &= \sum_{\text{d dephosphoylator}} k^d_{cat\,pme} \cdot C^d_{pme}
\end{aligned}
\tag{5}
$$

3 Methionine case study

3.1 Dataset generation

Starting with the model in Saa and Nielsen (2016), we extracted values for enzyme concentrations, boundary conditions and fluxes. We used these values to generate MCMC samples using Maud using the priors specified in section Section 3.2. When this was finished, we selected one sample with relatively high log probability to use as a ground truth in our case study. These parameter values are shown below in table Table 2. We manually inspected the parameter values to screen for any obviously implausible values; we did not find any of these.

3.2 Prior distributions compared with true parameter values

Table 2 shows the prior distributions we used for independent parameters. The first two columns show the 1% and 99% quantiles of each marginal prior distribution. True parameter value are shown in column three, and the last column shows the z-score on log scale of the true parameter value according the marginal prior distribution. As can be seen from the table, there are 7 parameters for which the true value is outside the 1%-99% range.

Table 2: Parameter specification, marginal prior distributions and true parameter values used in our case study.

parameter name	1% prior quantile	99% prior quantile	true value	prior Z-score of true value
$e_R^{CBS1,ametc}$	3.430e-06	0.002480	9.3e-05	0.004
$e_R^{GNMT1,ametc}$	3.000e-05	0.002000	2.000e-05	-2.787
$e_R^{MAT3,ametc}$	1.000e-04	0.001000	3.170e-04	0.003
$e_R^{MAT3,met-Lc}$	4.500e-04	0.000800	6.000e-04	0.000
$e_R^{MTHFR1,ahcysc}$	1.120e-07	0.000081	2.000e-06	-0.101
$e_T^{GNMT1,mlthfc}$	1.120e-05	0.008050	2.290e-04	-0.136
$e_T^{MTHFR1,ametc}$	1.120e-07	0.000081	1.500e-05	0.549306
k_{cat}^{AHC1}	1.200e+02	400.000000	2.340e+02	0.179861
k_{cat}^{BHMT1}	6.000e+00	35.000000	1.380e+01	-0.135
k_{cat}^{CBS1}	1.000e+01	188.000000	7.020e+00	-2.887
k_{cat}^{GNMT1}	7.000e-01	60.000000	1.050e+01	0.352083
k_{cat}^{MAT1}	8.200e-02	59.100000	7.900e+00	0.44375
k_{cat}^{MAT3}	5.890e-01	424.000000	1.990e+01	0.080556
$k_{cat}^{METH-Gen}$	4.840e-01	349.000000	1.160e+00	-1.209
$k_{cat}^{MS1kcatMS1}$	1.000e+00	3.300000	1.770e+00	-0.091
k_{cat}^{MTHFR1}	1.300e+00	4.200000	3.170e+00	0.183333
k_{cat}^{PROT1}	1.590e-01	0.222000	2.650e-01	0.41875
$k_I^{GNMT1,ahcysc}$	2.000e-06	0.001400	5.300e-05	0.010
$k_I^{MAT1,ametc}$	3.000e-04	0.000400	3.470e-04	0.014

parameter name	1% prior quantile	99% prior quantile	true value	prior Z-score of true value
$k_I^{METH-Gen,ahcysc}$	1.000e-06	0.000030	6.000e-06	0.021
$k_M^{AHC1,ahcysc}$	5.220e-05	0.037600	2.320e-05	-2.050
$k_M^{AHC1,adnc}$	1.670e-07	0.000120	5.660e-06	0.081944
$k_M^{AHC1,hcys-Lc}$	1.580e-07	0.000114	1.060e-05	0.318056
$k_M^{BHMT1,hcys-Lc}$	1.200e-05	0.000032	1.980e-05	0.049
$k_M^{BHMT1,glybc}$	4.720e-05	0.034000	8.460e-03	0.659028
$k_M^{CBS1,hcys-Lc}$	1.000e-06	0.000025	4.240e-05	3.090
$k_M^{CBS1,ser-Lc}$	2.000e-06	0.000004	2.830e-06	0.004
$k_M^{GNMT1,ametc}$	1.300e-05	0.009400	5.200e-04	0.1375
$k_M^{GNMT1,ahcysc}$	4.100e-07	0.000295	1.100e-05	0.000
$k_M^{GNMT1,glyc}$	5.480e-05	0.039500	2.540e-03	0.189583
$k_M^{GNMT1,sarcsc}$	3.730e-09	0.000003	1.000e-07	0.000
$k_M^{MAT1,met-Lc}$	1.400e-05	0.000720	1.070e-04	0.074
$k_M^{MAT1,atpc}$	5.270e-05	0.038000	2.030e-03	0.125694
$k_M^{MAT3,met-Lc}$	4.470e-05	0.032200	1.130e-03	-0.029
$k_M^{MAT3,atpc}$	5.270e-05	0.038000	2.370e-03	0.179167
$k_M^{METH-Gen,ametc}$	7.000e-06	0.000013	9.370e-06	-0.135
$k_M^{MS1,5mthfc}$	3.320e-06	0.002390	6.940e-05	-0.124
$k_M^{MS1,hcys-Lc}$	1.000e-06	0.000003	1.710e-06	-0.054

parameter name	1% prior quantile	99% prior quantile	true value	prior Z-score of true value
$k_M^{MTHFR1,mlthfc}$	7.500e-05	0.000088	8.080e-05	-0.158
$k_M^{MTHFR1,nadphc}$	1.600e-05	0.000028	2.090e-05	-0.105
$k_M^{PROT1,met-Lc}$	4.500e-05	0.000085	4.390e-05	-2.507
L_0^{CBS1}	3.730e-02	26.800000	1.030e+00	0.017
L_0^{GNMT1}	3.730e-02	26.800000	1.310e+02	0.3875
L_0^{MAT3}	3.730e-03	2.680000	1.080e-01	0.037
L_0^{MTHFR1}	1.120e-01	80.500000	3.920e-01	-1.018

$\Delta_f G$ parameters for most metabolites were fixed; those that were modelled as unknown had a multivariate normal prior distribution derived from eQuilibrator (Beber et al. 2021).

The values for $\Delta_f G$ parameters, as well as all other model parameters, can be found by inspecting the file `priors.toml` which is online at https://github.com/biosustain/Methionine_model/blob/main/data/methionine/priors.toml.

3.3 Computation

We conducted adaptive Hamiltonian Monte Carlo sampling for the full and missing-data datasets. For the full dataset we obtained 1000 post-warmup samples each from 4 independent Markov chains after 1000 warm-up samples and "hot-starting" with a mass metric output by a previous model run.

For the missing-data dataset XXX.

4 Laplace approximation case study

To compare MCMC sampling with Laplace approximation we used a different model with fewer parameters and state variables. This model was chosen because we were not able to generate results for our methionine model using Laplace approximation. The simpler case still serves to illustrate the general issues with approximating the posterior distributions of Bayesian kinetic models using the Laplace method, and that the associated numerical instability is another reason to prefer other methods where possible.

The full Maud input folders used for our Laplace approximation case study can be found at https://github.com/biosustain/Methionine_model/tree/main/data/example_ode and https://github.com/biosustain/Methionine_model/tree/main/data/example_ode_laplace.

To generate Laplace samples we used Maud's Laplace mode.

5 References

Beber, Moritz E., Mattia G. Gollub, Dana Mozaffari, Kevin M. Shebek, and Elad Noor. 2021. "eQuilibrator 3.0 – a Platform for the Estimation of Thermodynamic Constants." *arXiv:2103.00621 [q-Bio]*, February. http://arxiv.org/abs/2103.00621.

Changeux, Jean-Pierre. 2013. "50 Years of Allosteric Interactions: The Twists and Turns of the Models." *Nature Reviews. Molecular Cell Biology* 14 (12): 819–29. https://doi.org/10.1038/nrm3695.

Keating, Sarah M, Dagmar Waltemath, Matthias König, Fengkai Zhang, Andreas Dräger, Claudine Chaouiya, Frank T Bergmann, et al. 2020. "SBML Level 3: An Extensible Format for the Exchange and Reuse of Biological Models." *Molecular Systems Biology* 16 (8): e9110. https://doi.org/10.15252/msb.20199110.

Liebermeister, Wolfram, Jannis Uhlendorf, and Edda Klipp. 2010. "Modular Rate Laws for Enzymatic Reactions: Thermodynamics, Elasticities and Implementation." *Bioinformatics* 26 (12): 1528–34. https://doi.org/10.1093/bioinformatics/btq141.

Monod, J, J Wyman, and J P Changeux. 1965. "On the Nature of Allosteric Transitions: A Plausible Model." *Journal of Molecular Biology* 12 (May): 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.

Popova, S V, and E E Sel'kov. 1975. "Generalization of the Model by Monod, Wyman and Changeux for the Case of a Reversible Monosubtrate Reaction SR,TP." *FEBS Letters* 53 (3): 269–73. https://doi.org/10.1016/0014-5793(75)80034-2.

———. 1979. "[Description of the Kinetics of the Two Substrate Reactions S1+S2 Goes to and Comes from S3+S4 by a Generalized Monod, Wyman, Changeux Model]." *Molekuliarnaia Biologiia* 13 (1): 129–39. https://www.ncbi.nlm.nih.gov/pubmed/156878.

Preston-Werner, Tom and Gedam, Pradyun. 2020. "TOML Specification 1.0.0-Rc.1." https://toml.io/en/v1.0.0-rc.1/.

Pydantic developers. 2022. "Pydantic." https://pypi.org/project/pydantic/.

Saa, Pedro A, and Lars K Nielsen. 2016. "Construction of Feasible and Accurate Kinetic Models of Metabolism: A Bayesian Approach." *Scientific Reports* 6 (July): 29635. https://doi.org/10.1038/srep29635.

Schmiester, Leonard, Yannik Schälte, Frank T. Bergmann, Tacio Camba, Erika Dudkin, Janine Egert, Fabian Fröhlich, et al. 2021. "PEtab—Interoperable Specification of Parameter Estimation Problems in Systems Biology." *PLOS Computational Biology* 17 (1): 1–10. https://doi.org/10.1371/journal.pcbi.1008646.