

# Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data

## Supplementary material

Wolfram Liebermeister<sup>1\*</sup>, Edda Klipp<sup>1</sup>

<sup>1</sup>Computational Systems Biology, Max Planck Institute for Molecular Genetics,  
Ihnestraße 63-73, 14195 Berlin, Germany

Email: Wolfram Liebermeister - lieberme@molgen.mpg.de; Edda Klipp - klipp@molgen.mpg.de;

\*Corresponding author

**Table S1 - List of mathematical symbols**

Symbol	Name
$N = (n_{il})$	stoichiometric matrix
$W = (w_{li})$	regulation matrix
$s = (s_i)$	vector of stationary metabolite concentrations
$j = (j_l)$	vector of stationary fluxes
$v_l(\cdot)$	kinetic law of reaction $l$
$k_i^G$	energy constant for metabolite $i$
$k_l^V$	velocity constant for reaction $l$
$k_{li}^M$	reactant constant for reaction $l$ and metabolite $i$
$k_{li}^A$	activation constant for reaction $l$ and metabolite $i$
$k_{li}^I$	inhibition constant for reaction $l$ and metabolite $i$
$k_{\pm l}^{\text{cat}}$	maximal turnover rates (forward and backward) for reaction $l$
$E_l$	enzyme concentration for reaction $l$
$v_{\pm l}^{\text{max}}$	maximal velocities (forward and backward) for reaction $l$
$\theta$	vector of all model parameters (logarithmic)
$x$	vector of all measured kinetic parameters (logarithmic)
$y$	vector of all measured metabolic data (non-logarithmic)
$R_\theta^x$	sensitivity matrix between system parameters and kinetic data
$R_\theta^y$	sensitivity matrix between system parameters and metabolic data
$\bar{\theta}_{(0)}, C_{(0)}$	mean vector and covariance matrix for prior of $\theta$
$\bar{\theta}_{(1)}, C_{(1)}$	mean vector and covariance matrix for first posterior of $\theta$
$\bar{\theta}_{(2)}, C_{(2)}$	mean vector and covariance matrix for second posterior of $\theta$
$\hat{\theta}$	expansion point for $\theta$
$\hat{s}^{(m)}$	vector of initial concentrations for computing the steady state $m$

## The threonine model

### Network structure

We downloaded the model in SBML format [1] from the model database JWS online [2]. The chemical reactions are catalysed by aspartate kinase (AK), aspartate semialdehyde dehydrogenase (ASD), homoserine dehydrogenase (HDH), homoserine kinase (HK), threonine synthase (TSY). In our model, the aspartate kinase reaction is represented by a single enzyme, AKI, while the isoenzyme AKIII was neglected. Protons and water also participate in the pathway but are not explicitly described by the model. The stoichiometric matrix  $N$  of the network reads

$$N = \begin{pmatrix} \text{AK} & \text{ASD} & \text{HDH} & \text{HK} & \text{TSY} \\ \hline 1 & -1 & \cdot & \cdot & \cdot \\ \cdot & 1 & -1 & \cdot & \cdot \\ \cdot & \cdot & 1 & -1 & \cdot \\ \cdot & \cdot & \cdot & 1 & -1 \\ \cdot & 1 & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & 1 \\ -1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot & \cdot \\ \cdot & -1 & -1 & \cdot & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot \\ -1 & \cdot & \cdot & -1 & \cdot \end{pmatrix} \begin{array}{l} \text{4-phospho-l-aspartate} \\ \text{l-aspartate 4-semialdehyde} \\ \text{l-homoserine} \\ \text{o-phospho-l-homoserine} \\ \text{orthophosphate} \\ \text{l-threonine} \\ \text{l-aspartate} \\ \text{NADP+} \\ \text{NADPH} \\ \text{ADP} \\ \text{ATP} \end{array} \quad (1)$$

In steady state, production and consumption of the internal compounds (aspartyl-phosphate, aspartate semialdehyde, homoserine, and P-homoserine) are balanced, which implies that all five reactions have the same velocity. The transposed regulation matrix  $W^T$  reads

$$W = \begin{pmatrix} \text{AK} & \text{ASD} & \text{HDH} & \text{HK} & \text{TSY} \\ \hline \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & \cdot & -1 & -1 & \cdot \\ \cdot & \cdot & -1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & -1 & \cdot \end{pmatrix} \begin{array}{l} \text{4-phospho-l-aspartate} \\ \text{l-aspartate 4-semialdehyde} \\ \text{l-homoserine} \\ \text{o-phospho-l-homoserine} \\ \text{orthophosphate} \\ \text{l-threonine} \\ \text{l-aspartate} \\ \text{NADP+} \\ \text{NADPH} \\ \text{ADP} \\ \text{ATP} \end{array} \quad (2)$$

Note that ATP acts both as a substrate and as an inhibitor of the homoserine kinase (HK) reaction.

## Hypothetical kinetics

For testing our method, we created a model of the threonine pathway with a hypothetical convenience kinetics. The original parameters were chosen similar to the experimentally determined kinetics in Chassagnole et al. [3]. The values are listed in table S2. For the calculations, the remaining  $k_{li}^M$ ,  $k_{li}^A$ , and  $k_{li}^I$  values that do not correspond to a stoichiometric or regulatory connection were set to the arbitrary value of 1, that is, a logarithmic value of 0. Artificial experimental values for the kinetic parameters were generated by adding Gaussian noise (of width  $\ln 10$ ) to the original logarithmic enzyme parameters.

To create artificial metabolic data, we randomly chose concentrations for the fixed metabolites and the enzymes by multiplying the standard values with independent log-normal random numbers. We computed steady states for five such random assignments and repeated this ten times. The resulting steady state quantities (concentrations, fluxes, enzyme concentrations) were used as artificial data. As a prior for the logarithmic system parameters, we chose an uncorrelated Gaussian distribution with mean values and standard deviations as given in table 1 (see also methods section).

Time series for the original model and the sampled models were computed as follows: we chose the steady-state concentrations of the original model as initial conditions for all simulations. For the second half of the time series, the concentrations of the external metabolite aspartate was increased by a factor of 50.

**Table S2 - Kinetic parameters of the hypothetical threonine pathway model**

$G_{P-1-aspartate}^{(0)}$	3.94	kJ/mol/RT	$k_{1,1}^M$	0.017	mM	$k_{AKA}^V$	21.81	s <sup>-1</sup>
$G_{aspartatesemiald.}^{(0)}$	4.28	kJ/mol/RT	$k_{2,1}^M$	0.022	mM	$k_{ASD}^V$	0.02	s <sup>-1</sup>
$G_{homoserine}^{(0)}$	3.21	kJ/mol/RT	$k_{2,2}^M$	0.11	mM	$k_{HDH}^V$	0.34	s <sup>-1</sup>
$G_{P-homoserine}^{(0)}$	-1.47	kJ/mol/RT	$k_{3,2}^M$	0.24	mM	$k_{HK}^V$	0.0064	s <sup>-1</sup>
$G_{orthophosphate}^{(0)}$	-3.57	kJ/mol/RT	$k_{3,3}^M$	3.39	mM	$k_{TSY}^V$	0.017	s <sup>-1</sup>
$G_{threonine}^{(0)}$	-4.80	kJ/mol/RT	$k_{5,3}^M$	0.11	mM	$k_{1,6}^I$	0.167	mM
$G_{aspartate}^{(0)}$	-5.17	kJ/mol/RT	$k_{4,4}^M$	0.31	mM	$k_{3,6}^I$	0.097	mM
$G_{NADP+}^{(0)}$	-1.83	kJ/mol/RT	$k_{5,4}^M$	0.2	mM	$k_{5,6}^I$	1.09	mM
$G_{NADPH}^{(0)}$	1.83	kJ/mol/RT	$k_{2,5}^M$	10.2	mM	$k_{3,7}^I$	10	mM
$G_{ADP}^{(0)}$	-1.10	kJ/mol/RT	$k_{4,5}^M$	0.2	mM	$k_{5,11}^I$	4.35	mM
$G_{ATP}^{(0)}$	1.10	kJ/mol/RT	$k_{4,6}^M$	0.2	mM	$C_{orthophosphate}$	5	mM
			$k_{1,7}^M$	0.97	mM	$C_{threonine}$	3.49	mM
			$k_{2,8}^M$	0.144	mM	$C_{aspartate}$	1.34	mM
			$k_{3,8}^M$	0.067	mM	$C_{NADP+}$	0.63	mM
			$k_{2,9}^M$	0.029	mM	$C_{NADPH}$	0.56	mM
			$k_{3,9}^M$	0.037	mM	$C_{ADP}$	0.17	mM
			$k_{1,10}^M$	0.25	mM	$C_{ATP}$	1.31	mM
			$k_{5,10}^M$	0.2	mM			
			$k_{1,11}^M$	0.98	mM			
			$k_{5,11}^M$	0.072	mM			

The numerical subscripts refer to the numbering of substances and reactions in the transposed stoichiometric matrix (see above).

## An algorithm for computing the posterior distribution

### First step: integration of kinetic parameters

In the first estimation step, only the system parameters  $\theta^{\text{kin}}$  have to be fitted. We indicate the respective quantities by a superscript (like in  $\bar{\theta}_{(0)}^{\text{kin}}$ ). The algorithm reads

1. Parse the model (for instance, in SBML format [1] ) and construct the network matrices  $N$  and  $W$ .
2. Parse the available experimental data for kinetic parameters and collect them in a vector  $x^*$ . Parse or assume experimental error levels  $\sigma_x$  and construct the corresponding covariance matrix  $C_x$ .
3. Construct the sensitivity matrix  $R_{\theta^{\text{kin}}}^x$  from  $N$  and  $W$  (see [4]).
4. Choose a Gaussian prior distribution for the system parameters. Construct the vector  $\bar{\theta}_{(0)}^{\text{kin}}$  and the matrix  $C_{(0)}^{\text{kin}}$  from mean values and standard deviations of parameter types (for numerical values, see table 1).
5. Compute  $\bar{\theta}_{(1)}^{\text{kin}}$  and  $C_{(1)}^{\text{kin}}$  for the kinetics-based posterior, using eqn. (10) and the matrix  $R_{\theta^{\text{kin}}}^x$  instead of  $R_{\theta}^x$ .

To avoid costly matrix inversions, we compute mathematical expressions of the form  $A^{-1}x$  (with a sparse matrix  $A$  and a vector  $x$ ) by Gaussian elimination (“left matrix divide” in MATLAB).

### Second step: integration of metabolic data

The algorithm for the second step, the metabolics-based parameter estimation, is more complicated. Here we assume that metabolic quantities have been measured for a couple of steady states (indicated by  $m$ ). Each steady state results from certain concentrations of enzymes and fixed metabolites, which are included in the model as model parameters  $\theta^{\text{met}}$  (logarithmic values). Moreover, for each steady state, we have to guess initial concentrations for all metabolites. In the end, the conserved moieties of the initial concentrations must be compatible with the conserved moieties of the experimental concentrations. In the algorithm, we start with parameters and initial concentrations that correspond to a chemical equilibrium. Then we linearise the model, compute the posterior mode, and move the current parameter set towards this mode, using a relaxation factor  $\lambda$ .

After running the algorithm for the first estimation step described above, we continue as follows:

1. Choose a relaxation parameter  $\lambda$  (e.g.,  $\lambda = 0.2$ ) and a convergence parameter  $\varepsilon$  (e.g.,  $\varepsilon = 10^{-5}$ ).
2. Compute a reduced stoichiometric matrix  $N_R$  and the corresponding link matrix  $L$  (see [5]) such that  $N = L N_R$  and  $N_R$  has full row rank.
3. Append additional empty columns to  $R_{\theta^{\text{kin}}}^x$  (for each enzyme and each external metabolite, in each steady state) to obtain the complete matrix  $R_{\theta}^x$ . Append the prior means and the covariance matrix of the metabolic parameters  $\theta^{\text{met}}$  to  $\bar{\theta}_{(1)}^{\text{kin}}$  and  $C_{(1)}^{\text{kin}}$ , respectively. This yields  $\bar{\theta}_{(1)}$  and  $C_{(1)}$ .
4. For each steady state  $m$ ,
  - Parse the metabolic data, construct the data vectors  $s_{(m)}^*$  (concentrations),  $j_{(m)}^*$  (fluxes), and  $E_{(m)}^*$  (enzyme concentrations). If data values are missing, the corresponding elements are not listed in the vectors. Construct the corresponding vectors of measurement noise levels  $\sigma_{s(m)}^*$ ,  $\sigma_{j(m)}^*$ ,  $\sigma_{u(m)}^*$ .

- Determine which of the metabolites are described by fixed values in the model and have been measured in steady state  $m$ . Collect their measured concentrations in a vector  $\hat{s}_{(m)}^*$ . Let  $\tilde{N}_{(m)}$  denotes the corresponding stoichiometric matrix. Determine a left-kernel matrix  $G_{(m)}$  of  $\tilde{N}_{(m)}$ . Compute the vector of conserved moieties  $f_{(m)} = G_{(m)}\hat{s}_{(m)}^*$
  - Initialise the vector  $\hat{s}_{(m)}$  of initial concentrations with a value of 1 for each element.
5. Collect the metabolic data  $s_{(m)}^*, j_{(m)}^*, E_{(m)}^*$  for all steady states  $m$  in a vector  $y^*$  and construct the diagonal covariance matrix  $C_y$  for the measurement noise.
  6. Initialise all model parameters (elements of  $\hat{\theta}$ ) with 1.
  7. Compute all steady state quantities  $y(\hat{\theta})$  and the corresponding response coefficients matrix  $R(\hat{\theta})$ , based on the initial values  $\hat{s}_{(m)}$  (computation see below).
  8. Iterate until convergence:
    - Compute the posterior mode  $\hat{\theta}^{\text{mode}}$  of the current linearised model from eqn. (13). Compute the change  $\Delta\hat{\theta} = \hat{\theta}^{\text{mode}} - \hat{\theta}$  and update  $\hat{\theta} \rightarrow \hat{\theta} + \lambda \Delta\hat{\theta}$ .
    - For each steady state  $m$ , compute a change for the initial concentrations  $\Delta\hat{s}_{(m)} = G_{(m)}^T \left( G_{(m)} G_{(m)}^T \right)^{-1} (f_{(m)} - G_{(m)}\hat{s}_{(m)})$  and update  $\hat{s}_{(m)} \rightarrow \hat{s}_{(m)} + \lambda \Delta\hat{s}_{(m)}$ . This improves the match to the experimental conserved moieties.
    - Compute again all steady state quantities  $y(\hat{\theta})$  and the corresponding response coefficients matrix  $R(\hat{\theta})$ , based on the new parameters  $\hat{\theta}$  and the new initial values  $\hat{s}_{(m)}$ . If no steady state is found, choose a smaller  $\lambda$  and repeat the last iteration step.
  9. As convergence criterion, we require that

$$\max_k \left| \frac{\hat{\theta}_k^{\text{new}} - \hat{\theta}_k^{\text{old}}}{\hat{\theta}_k^{\text{new}} + \hat{\theta}_k^{\text{old}}} \right| < \varepsilon$$

10. Accept the final  $\hat{\theta}$  as  $\bar{\theta}_{(2)}$  and compute the corresponding covariance matrix  $C_{(2)}$  from eqn. (13).

### Calculation of steady state quantities

The computation of the steady state quantities works as follows:

1. For each steady state  $m$ :
  - Assume the system parameters  $\hat{\theta}^{\text{kin}}$ , metabolic parameters  $\theta_{(m)}^{\text{met}}$ , and initial concentrations  $\hat{s}_{(m)}$ .
  - Initialise the system with concentrations  $\hat{s}_{(m)}$ , find the nearest steady state  $s_{(m)}$  fulfilling  $0 = Nv(s_{(m)}, \theta^{\text{kin}}, \theta_{(m)}^{\text{met}})$ , and compute the fluxes  $j_{(m)} = v(s_{(m)}, \theta^{\text{kin}}, \theta_{(m)}^{\text{met}})$ . Compute the enzyme concentrations  $E_{(m)}$  directly from  $\theta_{(m)}^{\text{met}}$ .
  - Compute the response matrices  $R_{(m)}^S$  and  $R_{(m)}^J$ . The response matrix with respect to the logarithmic system parameters  $\theta$  is computed as

$$R_{(m)} = \hat{R}_{(m)} \text{diag}(x) R_{\theta}^x \quad (3)$$

from the matrix  $\hat{R}_{(m)}$  of non-normalised response coefficients  $\hat{R}_{im} = \partial y_i / \partial x_m$  (to be computed using using  $L$ ,  $N_R$ , and the elasticity matrices, see [5]).

2. Collect the concentration vectors  $s_{(m)}$ , the flux vectors  $j_{(m)}$ , and the enzyme concentrations  $E_{(m)}$  in a vector  $y$ , and the corresponding response coefficients matrices in a response coefficient matrix  $R$ . Keep only the elements in  $y$  and rows in  $R$  for which data are available.

## References

1. Hucka M, et al.: **The Systems Biology Markup Language (SBML): A Medium for Representation and Exchange of Biochemical Network Models.** *Bioinformatics* 2003, **19**(4):524–531.
2. Olivier B, Snoep J: **Web-based kinetic modelling using JWS Online.** *Bioinformatics* 2004, **20**(13):2143–2144.
3. Chassagnole C, Rais B, Quentin E, Fell DA, Mazat J: **An integrated study of threonine-pathway enzyme kinetics in Escherichia coli.** *Biochem J* 2001, **356**:415–423.
4. Liebermeister W, Klipp E: **Bringing metabolic networks to life: convenience rate law and thermodynamic constraints.** *Theor Biol Med Mod.*
5. Reder C: **Metabolic control theory: a structural approach.** *J Theor Biol* 1988, **135**:175–201.