Bayesian inference of the sites of perturbations in metabolic pathways via Markov chain Monte Carlo

Jayawardhana, Bayu; Kell, Douglas B.; Rattray, Magnus

Published in:
Bioinformatics

DOI:
10.1093/bioinformatics/btn103

Publication date:
2008

Supplementary Materials of

Bayesian Inference of the Sites of Perturbations in Metabolic Pathways via Markov Chain Monte Carlo

Bayu Jayawardhana, Douglas B. Kell and Magnus Rattray

1 METHODS

1.1 Prior distributions

In general, the support of the prior distribution function p_θ and p_{x0} must contain all possible values of θ or x_0. If a component θ_i in the vector of system parameter θ or a component $x_{0,i}$ in the vector of initial condition x_0, is independent of other parameters or other initial conditions, and admits only positive real values, then a gamma or lognormal distribution function can be assigned as the prior distribution of θ_i or $x_{0,i}$. Multivariate distribution functions must be used if there are dependency among several components in θ or x_0.

The same principle applies to the prior distribution of the initial conditions of conserved variables. Since the conserved variables have to satisfy the constraint (2), we parameterize their initial conditions as follows: For the j-th constraint,

$$x_{0,i} = a_i C_j, \quad \forall i \in \mathfrak{A}_j \quad \text{where} \quad \sum_{i\in\mathfrak{A}_j} a_i = 1, \qquad \text{(SUPPL. 1)}$$

and $x_{0,i}$ is the i-th component of the initial condition x_0. Dirichlet distribution can be used to define the distribution of $a:=\{a_i \mid i\in\mathfrak{A}_j\}$ and any positive distribution function may define the distribution of C_j. The Dirichlet distribution for K variates has probability density function given by

$$\mathrm{p}(z_1,\ldots,z_K) = \frac{\Gamma(\sigma)\delta\left(1-\sum_{i=1}^K z_i\right)}{\prod_{i=1}^K \Gamma(\mu_i\sigma)} \prod_{i=1}^K z_i^{\mu_i\sigma-1}, \qquad \text{(SUPPL. 2)}$$

where Γ is the gamma function, $(\mu_1,\ldots,\mu_K)$ is the mean of the distribution and the constant σ controls the spread or denseness of the distribution. Note that by setting $\sigma = 1$, uniform distribution is obtained and is defined on the simplex $\mu_1+\ldots+\mu_K = 1$.

Hence, the distribution function of conserved variables corresponding to the j-th constraint using the parameterization (SUPPL. 1) and using Dirichlet distribution function is given by

$$\mathrm{p}_{C_j}(C_j)\frac{\Gamma(\sigma)}{\prod_{i\in\mathfrak{A}_j}\Gamma(a_i\sigma)}\prod_{i\in\mathfrak{A}_j} x_{0,i}^{a_i\sigma-1}$$

The biochemical reaction kinetics database, such as SABIO-RK (Rojas, et al., 2007), provides valuable information for the prior distribution of parameters. It can show both kinetics parameters and their uncertainties as reported in the literature.

1.2 Metropolis-Hastings algorithm and convergence measure

As described in Section 2.1 we can draw samples $w = (\theta, y_{ext}, x_0)$ from the target distribution $p_{target}(w)$ as defined in (6) and marginalizing the samples over x_0 and y_{ext}, to get samples from $p(\theta|\mathfrak{D}_{int}, \mathfrak{D}_{ext})$.

Metropolis-Hastings algorithm can be used to generate Markov chains that will eventually produce samples from the target distribution $w(n) \sim p_{target}(w)$ for large n (Hastings, 1970; Metropolis, et al., 1953). Given the current sample $w(n)$, the next sample $w(n+1)$ is proposed from a proposal distribution $p_{proposal}(.|\, w(n))$ with the acceptance probability:

$$\mathrm{p}_{acceptance}(w(n+1)\mid w(n)) := \min\left(1, \frac{\mathrm{p}_{target}(w(n))\mathrm{p}_{proposal}(w(n+1)\mid w(n))}{\mathrm{p}_{target}(w(n+1))\mathrm{p}_{proposal}(w(n)\mid w(n+1))}\right) \qquad \text{(SUPPL. 3).}$$

Therefore, the Metropolis-Hastings algorithm is summarized as follows:

1. Initialize the initial sample $w(0)$ and set sampling time $n = 0$;
2. Sample a point v from $p_{proposal}(.|w(n))$ and sample a point u from a uniform distribution defined on (0,1);
3. If $u \le p_{acceptance}(v|\, w(n))$ then $w(n+1) = v$, otherwise set $w(n+1) = w(n)$.
4. Increment n and repeat the procedure from step 2.

A variant of the Metropolis-Hastings that we used in our simulation is the single-component Metropolis-Hastings (Spiegelhalter, et al., 1996). Let W be the size of the vector w. Let the the proposal distribution $p_{proposal}$ be decomposed into W proposal distributions, $p^i_{proposal}$, $i = 1,\ldots,W$, such that the domain of w can be spanned by the basis defining the range space of every $p^i_{proposal}$. For example, if $w\in\mathbb{R}^W$, $p^i_{proposal}$ is chosen such that the range space of each $p^i_{proposal}$ defines an orthogonal subspace in $\mathbb{R}^W$. Define the i-th acceptance probability by

$$\mathrm{p}^i_{acceptance}(w(n+1)\mid w(n)) := \min\left(1, \frac{\mathrm{p}_{target}(w(n))\mathrm{p}^i_{proposal}(w(n+1)\mid w(n))}{\mathrm{p}_{target}(w(n+1))\mathrm{p}^i_{proposal}(w(n)\mid w(n+1))}\right)$$

Define $\mathrm{mod}(a,W) := a - W\lfloor a/W\rfloor$ where $a > 0$ in an integer and $\lfloor.\rfloor$ is the floor function. The single-component Metropolis-Hastings algorithm can be given as follows:

1. Initialize the initial sample $w(0)$ and set sampling time $n = 0$;
2. Sample a point v from $p^i_{proposal}(.|w(n))$ where $i = \mathrm{mod}(n,W)$ and sample a point u from a uniform distribution defined on (0,1);
3. If $u \le p^i_{acceptance}(v|w(n))$ then $w(n+1) = v$, otherwise set $w(n+1) = w(n)$.
4. Increment n and repeat the procedure from step 2.

Several parallel Markov chains with different initial conditions can be simulated to assess the convergence of the chains to the target distribution using the variance-based convergence measure proposed by Gelman (1996). Although it evaluates the convergence of a single component w_i in the vector w, the convergence of the whole components is checked by monitoring the convergence of each component.

Suppose that M parallel chains each with N samples have been generated. The sequences of i-th component is denoted by $(w^j_i(n))$, n = 1,…,N, j=1,…,M. The measure uses the between-sequence variance B and the within-sequence variances W defined by:

$$B = \frac{N}{M-1}\sum_{j=1}^{M}(\overline{w}_i^j - \overline{w}_i)^2 \quad W = \frac{1}{M}\sum_{j=1}^{M}\sigma_{i,j}^2 \quad \text{where}$$

$$\overline{w}_i^j = \frac{1}{N}\sum_{n=1}^{N} w_i^j(n), \overline{w}_i = \frac{1}{M}\sum_{j=1}^{M}\overline{w}_i^j \text{ and } \sigma_{i,j}^2 = \frac{1}{N-1}\sum_{n=1}^{N}(w_i^j(n) - \overline{w}_i^j)^2.$$

The convergence measure is given by

$$\hat{R} = \frac{\text{var}(w_i)}{W}, \text{ where } \text{var}(w_i) = \frac{N-1}{N}W + \frac{1}{N}B.$$

Gelman proposed that the parallel chains is run until the value of $\sqrt{\hat{R}}$ is less than 1.2 (or approximately, $\hat{R}$ is less than 1.4). Once the parallel chains have converged, the samples from each chains can be combined.

1.3 Proposal distributions

The proposal distribution is crucial in the acceptance or the rejection of a proposed move in the MCMC sampler. Given the state of current MCMC sample $w(n)$, the proposal distribution function $p_{proposal}(.|w(n))$ can be any distribution function as long as the proposal can be reversed. The choice determines the convergence rate of the sampler to the stationary posterior distribution. Ideally, it belongs to the same class of functions as the posterior distribution function. However, since the posterior distribution function contains an indeterminate function g, it is impossible to find an analytic proposal distribution that approximates the posterior distribution.

It is more feasible to assign the proposal according to possible values of parameter θ or initial condition x_0 and control the acceptance/rejection ratio by adaptively tuning the parameter of proposal distribution, such as σ in (SUPPL. 2) whenever a Dirichlet distribution is used. Thus the assignment of proposal distribution follows a similar line to the prior distribution. For the conserved variables, the parameterization (SUPPL. 2) can be used with Dirichlet distribution as a proposal distribution for a and with lognormal distribution as a proposal distribution for C_j.

Most proposal distributions have parameters μ that control the distribution function shape. These parameters can be adaptively tuned during the Markov process such that the acceptance rate of the proposed move is within a desirable interval in order to prevent the chain moves being localized. Typically, a rejection rate within 60% to 70% suffices to guarantee the chains will not be confined to the neighborhood of a local minimum. Change of parameters in the proposal distribution can also improve the convergence rate of the chains (Spiegelhalter, et al., 1996). One way to do this is by changing the proposal distribution according to the distribution of the current chains $w(n)$, n=1,…,N. For example, if Gaussian distributions are used as the proposal distribution, the covariance matrix can be updated by computing the covariance matrix of the current chains to improve the convergence speed and by scaling the covariance matrix to control the acceptance rate. Similarly, the parameter σ in the Dirichlet distribution (SUPPL. 2) can be manipulated to maintain the convergence rate.

2 SIMULATIONS SETUP

2.1 Glycolytic Pathways

We also refer to the Section 3.1 of the main paper for the model of glycolytic pathways and some simulation setup for the MCMC. Table 1 gives the sixteen parameter datasets used in Section 3.1 with a set of initial conditions that can give steady-state solution.

The Markov chain moves according to the single-component Metropolis-Hastings algorithm as described in Section 1.2 of this Supplementary Material.

Table 1. Sixteen parameters and initial conditions used to generate sixteen observation datasets

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10	Case 11	Case 12	Case 13	Case 14	Case 15	Case 16
V1	101.3	81.3	121.3	101.3	303.9	101.3	117.2683	104.9292	106.8582	121.4975	128.1376	120.8088	101.7018	114.7746	104.8378	104.7551
V2	670.5	770.5	670.5	670.5	670.5	670.5	722.2568	294.2046	731.5657	734.8304	690.7653	550.8387	179.1309	861.5725	303.2597	670.5
V3	1933	1633	2433	1933	1933	1933	4404.236	586.9121	368.8674	625.2543	9157.505	1270.638	608.3028	10722.49	2448.627	1636.628
V4	121.5	101.5	181.5	121.5	121.5	121.5	476.2175	149.9992	154.3231	203.8541	139.8317	284.3603	587.364	401.885	139.3763	188.3947
V5	101	81	161	111	101	101	106.2359	95.76396	112.3662	136.4772	531.96	384.8218	123.7587	133.7628	111.4681	122.6127
V6	2336	1336	3336	2336	2336	2336	1530.545	1291.498	1072.684	1910.2	2005.513	1712.204	1838.168	1471.81	1133.351	1222.284
V7	3298	2298	4298	3298	3298	3298	1805.233	4120.604	4512.33	2302.527	5099.889	4936.259	3260.414	1645.973	4214.43	3975.674
V8	2291	1291	2991	2291	2291	2291	1195.713	1205.501	2491.115	1264.377	537.5697	819.5761	909.0141	1524.612	1237.29	1072.316
V9	2423	1423	3223	2423	2423	2423	2185.293	2326.269	1296.852	9873.841	1072.615	1494.455	2058.951	3322.948	7311.557	1583.496
V10	240.4	180.4	290.4	240.4	240.4	240.4	273.6475	217.2014	262.2954	517.4553	501.0479	297.2893	271.5892	246.5318	237.3816	251.9329
V11	700.5	650.5	700.5	700.5	700.5	700.5	1716.369	746.525	905.6832	947.8167	1034.941	816.0751	747.1872	1121.721	682.1825	739.4743
V12	869.9	809.9	969.9	869.9	869.9	869.9	777.4968	1274.402	482.5397	876.7966	525.5381	5939.067	576.7195	836.2826	2227.776	2504.769
V13	50.2	40.2	70.2	50.2	50.2	50.2	1963.495	242.0179	144.4956	3914.104	83.61441	781.5596	84.12415	54.78029	908.821	900.7167
V14	47.2	37.2	57.2	47.2	47.2	470.2	84.14668	84.68681	90.11679	157.7769	60.799	94.62859	47.22591	50.04119	176.3519	76.77777
GII(0)	0.087	0.087	0.087	0.087	0.045	10	0.028689	0.028689	0.557415	0.564208	0.093	0.020577	0.017221	0.247204	0.050874	0.197929
ATP(0)	2.521	2.521	2.521	2.521	2.521	0.02	2.358634	2.358634	0.149494	2.197799	2.130685	1.906482	1.771163	0.9209	3.088413	2.669857
G6P(0)	2.353	2.353	2.353	2.353	2.353	0.02	1.82771	1.82771	52.00647	34.68389	0.936371	2.615373	0.615499	3.249371	0.313617	0.413272
ADP(0)	1.284	1.284	1.284	1.284	1.284	0.02	0.721978	0.721978	2.629772	1.79213	1.316746	0.953176	1.245218	0.440995	0.578617	1.092415
F6P(0)	0.598	0.598	0.598	0.598	0.598	0.598	0.550244	0.550244	0.365891	4.030142	0.088311	1.40405	1.04304	0.844291	1.006192	0.078823
F16BP1(0)	5.592	5.592	5.592	5.592	5.592	5.592	2.753098	2.753098	2.846785	5.106613	1.895737	24.79862	23.92652	86.81573	6.379008	13.0038
AMP1(0)	0.294	0.294	0.294	0.294	0.294	0.294	0.099449	0.099449	0.51041	0.032934	0.490734	1.591179	0.658209	1.942555	0.191166	0.330205
DHAP1(0)	0.816	0.816	0.816	0.816	0.816	0.816	0.845203	0.845203	2.012857	0.907349	1.254938	5.225015	0.199077	1.000312	2.360834	0.780832
GAP1(0)	0.037	0.037	0.037	0.037	0.7	0.37	0.038028	0.038028	0.020483	0.039073	0.055029	0.509213	0.026436	0.023888	0.010745	0.004574
NAD1(0)	1.172	1.172	1.172	1.172	1.172	1.172	0.753487	0.753487	0.141701	0.107777	1.572251	0.930924	0.531203	0.803878	0.752978	0.75
BPG1(0)	0.1	0.1	0.1	1	0.1	0.1	0.001611	0.001611	0.600464	0.052335	0.029515	0.077441	0.057695	0.217367	3.954795	0.1
NADH1(0)	0.418	0.418	0.418	0.418	0.418	0.418	0.035726	0.035726	0.992455	0.927402	0.00404	0.520476	0.40119	1.000468	0.349737	0.267843
P3G1(0)	0.8741	0.8741	0.8741	0.8741	0.8741	0.8741	0.86441	0.86441	2.933414	1.748035	3.459581	0.760726	2.07922	2.147956	0.584236	0.482995
P2G1(0)	0.122	0.122	0.122	0.122	0.122	0.122	0.118288	0.118288	0.473084	0.181156	0.007501	0.043145	0.019263	0.372297	0.186714	0.785471
PEP1(0)	0.1	0.1	0.1	0.1	0.1	0.1	0.112586	0.112586	0.03375	0.007228	0.244833	0.019015	0.029812	0.1034	0.006869	0.058282
PYR1(0)	1.862	1.862	1.862	1.862	1.862	1.862	1.248962	1.248962	1.692495	5.761615	1.693989	1.278919	1.218052	0.726678	2.17099	4.009405
AcAld1(0)	0.251	0.251	0.251	0.251	0.251	0.251	0.263344	0.250695	16.77693	0.057305	1.689921	0.041459	0.780978	0.225717	0.004577	0.251
EtOH1	50	50	50	50	50	50	144.327	77.98366	30.70488	77.16282	31.54835	95.36344	80.45193	21.70103	36.93222	118.6237
F26BP1	0.02	0.02	0.02	0.02	0.02	0.02	0.010211	0.00489	0.029298	0.017619	0.038774	0.003531	0.00217	0.027614	0.004377	0.02
Glc1	50	50	50	50	0.5	50	8.619485	26.20229	37.45999	67.60121	15.2616	36.39508	69.36499	81.7448	25.12868	57.78456
Glycerol	0.15	0.15	0.15	0.15	0.15	15	0.119468	0.034983	0.153478	0.022433	0.538828	0.01757	0.039994	0.014668	0.194563	0.03077

As a complementary section to the Section 3.1 of the main paper, details of the log-mean values and the log-variance values for the prior distributions for V_{max} are listed in Table 2. The interval of uniform distributions for the prior distributions of the initial conditions and the concentrations of external glucose, F26BP, glycerol and ethanol are listed in Table 3. Note that we use the same prior distributions for all sixteen cases.

For the conserved variables, the proposal distribution is assigned according to the parametrization in (SUPPL. 1) with the parameters C_j and a. Dirichlet distribution is used for the proposal distribution of a.

Based on the model and parametrization of conserved variables, $w = (\theta, y_{ext}, x_0^{\text{non-conserved}}, C_1, C_2, a_{ATP}, a_{ADP}, a_{AMP}, a_{NAD}, a_{NADH}) \in \mathbb{R}_+^{34} \times \{z \in \mathbb{R}^3 \mid z_1+z_2+z_3 = 1\} \times \{z \in \mathbb{R}^2 \mid z_1+z_2 = 1\}$ where $\mathbb{R}_+$ is the space of positive real value.

The proposal distribution for parameters, external observable variables, C_j and the initial concentrations of non-conserved metabolites are multivariate lognormal distribution where log value of the current sample is used as the log-mean and the log-covariance matrix is updated every 500,000 samples according to the covariance of the log of MCMC samples obtained so far from the three parallel chains. For the first 500,000 samples, the log-covariance matrix is set to diagonal matrix where each diagonal entry is listed in Table 4.

Suppose that Σ is the log-covariance matrix of the samples (θ, y_{ext}, $x_0^{\text{non-conserved}}$, C_1, C_2) and Σ^i is the i-th eigenvector of Σ. In this case, for $i = 1,\ldots,34$, we use $p^i_{proposal}$ given by a multivariate lognormal distribution where the log space is defined in the direction of $(\Sigma^{iT},0,0,0,0,0)^T$. In other words, let $z \sim \mathcal{N}(0,\sigma_i)$ where $\mathcal{N}$ is the Gaussian distribution with zero mean and variance σ_i, the proposed sample v is given by $v = \exp([\Sigma^{iT},0,0,0,0,0]^T z)$. Here, σ_i is the parameter of the proposal distribution $p^i_{proposal}$ that is adaptively set to control the rejection rate. Initially, σ_i is set to one and then, in every update of Σ, σ_i is reset to one tenth of the i-th eigenvalue of Σ. For $i = 35,\ldots,37$, the i-th proposal distribution is given by $[0^{1x34}, z_1, z_2, z_3, 0, 0]^T$ where 0^{1xN} denotes the zero vector with dimension of N and (z_1, z_2, z_3) are sampled from a Dirichlet distribution with the denseness parameters $\sigma_{adenine}$. For $i = 35,\ldots,37$, the i-th proposal distribution is given by $[0^{1x37}, 0, 0, 0, z_1, z_2]^T$ where (z_1, z_2) are sampled from a Dirichlet distribution with the denseness parameters $\sigma_{pyridine}$. The parameters $\sigma_{pyridine}$ and $\sigma_{adenine}$ is set initially to 50 and adaptively changed based on the rejection rate.

Table 2. The log-mean and log-variance for the prior distribution of $V_{\max}$.

Parameters	Log-mean	Log-variance
$V^1_{\max}$	4.5643	0.4805
$V^2_{\max}$	5.4116	1.2069
$V^3_{\max}$	5.8171	3.2104
$V^4_{\max}$	5.2003	0.4805
$V^5_{\max}$	5.76	1.2069
$V^{7,f}_{\max}$	7.0676	0.4805
$V^{7,r}_{\max}$	8.7765	0.4805
$V^8_{\max}$	7.1546	0.4805
$V^9_{\max}$	7.8267	0.4805
$V^{10}_{\max}$	5.8861	0.4805
$V^{11}_{\max}$	6.9847	0.4805
$V^{12}_{\max}$	5.1552	2.5903
$V^{13}_{\max}$	6.6846	7.6872
$V^{14}_{\max}$	3.912	0.4805

Note that $\exp(\sqrt{0.4805}) \approx 2$, $\exp(\sqrt{1.2069}) \approx 3$, $\exp(\sqrt{2.5903}) \approx 5$, $\exp(\sqrt{3.2104}) \approx 6$ and $\exp(\sqrt{7.6872}) \approx 16$.

Table 3. The interval of uniform distribution for the prior distribution of initial conditions and the concentrations of external glucose, F26BP, glycerol and ethanol (excluding the conserved variables).

x_0	Interval	x_0	Interval	x_0	Interval
[Gli](0)	(0.01,10)	[P3G](0)	(0.01,10)	[Glo]	(0.01,1000)
[G6P](0)	(0.001,10)	[P2G](0)	(0.01,10)	[Glycerol]	(0.01,100)
[F6P](0)	(0.01,10)	[PEP](0)	(0.001,10)		
[F16BP](0)	(0.001,10)	[PYR](0)	(0.01,10)		
[DHAP](0)	(0.01,10)	[AcAld](0)	(0.01,10)		
[GAP](0)	(0.001,10)	[EtOH]	(0.01,10)		
[BPG](0)	(0.01,10)	[F26BP]	(0.001,10)		

[Gli] is the internal glucose, [Glo] is the external glucose.

Table 4. The diagonal entries $\Sigma^{i,i}$ of the initial log-covariance matrix Σ used in the proposal distribution for (θ, y_{ext}, $x_0^{\text{non-conserved}}$, C_1, C_2).

Corresponding sample	$\Sigma^{i,i}$	Corresponding sample	$\Sigma^{i,i}$	Corresponding sample	$\Sigma^{i,i}$
$V^1_{\max}$	0.0048	$V^{13}_{\max}$	0.0769	[P2G](0)	0.0048
$V^2_{\max}$	0.0121	$V^{14}_{\max}$	0.0048	[PEP](0)	0.0048
$V^3_{\max}$	0.0321	Glycogen flux	0.0048	[PYR](0)	0.0048
$V^4_{\max}$	0.0048	Trehalose flux	0.0048	[AcAld](0)	0.0048
$V^5_{\max}$	0.0121	[Gli](0)	0.0048	[EtOH]	0.0048
$V^{7,f}_{\max}$	0.0048	[G6P](0)	0.0048	[F26BP]	0.0048
$V^{7,r}_{\max}$	0.0048	[F6P](0)	0.0048	[Glo]	0.0048
$V^8_{\max}$	0.0048	[F16BP](0)	0.0048	[Glycerol]	0.0048
$V^9_{\max}$	0.0048	[DHAP](0)	0.0048	C_1	0.0048
$V^{10}_{\max}$	0.0048	[GAP](0)	0.0048	C_2	0.0048
$V^{11}_{\max}$	0.0048	[BPG](0)	0.0048		
$V^{12}_{\max}$	0.0259	[P3G](0)	0.0048		

We use the same strategies for all sixteen cases. For each case, we run three parallel chains where each chain uses the single-component Metropolis-Hastings with the setting described above. The simulations took about 500,000 to 1,000,000 samples to get to the convergence of the three chains using the convergence measure presented in Section 1.2 of this Supplementary Material. A thinning with the size of 39 is applied to the MCMC samples to get less correlated samples. (A *thinning* with the size of M is a procedure of subsampling the samples with the period of M samples.)

2.2 Pyruvatic Pathways

The setting of simulation in the pyruvate pathways of *L. lactis* is similar to that in the glycolytic pathways of *S. cerevisiae*. In particular, the single-component Metropolis-Hastings algorithm is used in the simulation.

The prior distributions for $V_{\max}$ are uniform distributions defined on (0,20000). The interval of uniform distributions for the prior distributions of the initial conditions and the concentrations of external glucose, lactate, acetoin, O_2, phosphate, ethanol and butanediol are listed in Table 5. Note that we use the same prior distributions for the three cases.

For the conserved variables, the proposal distribution is assigned according to the parametrization in (SUPPL. 1) with the parameters C_j and a. Dirichlet distribution is used for the proposal distribution of a.

Based on the model and parametrization of conserved variables, $w = (\theta, x_0^{\text{non-conserved}}, C_1, C_2, C_3, a_{ATP}, a_{ADP}, a_{NAD}, a_{NADH}, a_{CoA}, a_{AcCoA}) \in \mathbb{R}_+^{30}\times\Omega\times\Omega\times\Omega$ where $\mathbb{R}_+$ is the space of positive real value and $\Omega := \{z\in\mathbb{R}^2 \mid z_1+z_2 = 1\}$.

For $i = 1,\ldots,30$, the proposal distribution $p^i_{proposal}$ is similar to that in the glycolytic pathways. Suppose that Σ is the log-covariance matrix of the samples (θ, $x_0^{\text{non-conserved}}$, C_1, C_2, C_3) and Σ^i is the i-th eigenvector of Σ. We use $p^i_{proposal}$ given by a multivariate lognormal distribution where the log space is defined in the direction of $(\Sigma^{iT},0,0,0,0,0,0)^T$. The log-covariance matrix is set to di-

agonal matrix for the first 500,000 samples where each diagonal entry is set to 0.0048.

For $i = 31,32$, the i-th proposal distribution is given by $[0^{1x30}, z_1, z_2, 0^{1x4}]^T$ where (z_1, z_2) are sampled from a Dirichlet distribution with the denseness parameters $\sigma_{adenine}$. For $i = 33,34$, the i-th proposal distribution is given by $[0^{1x32}, z_1, z_2, 0^{1x2}]^T$ where (z_1, z_2) are sampled from a Dirichlet distribution with the denseness parameters $\sigma_{pyridine}$. For $i = 35,36$, the i-th proposal distribution is given by $[0^{1x34}, z_1, z_2]^T$ where (z_1, z_2) are sampled from a Dirichlet distribution with the denseness parameters σ_{CoA}. The parameters $\sigma_{pyridine}$, $\sigma_{adenine}$ and σ_{CoA} is set initially to 50 and adaptively changed based on the rejection rate.

Table 5. The interval of uniform distribution for the prior distribution of initial conditions and the concentrations of external glucose, lactate, acetoin, O_2, phosphate, ethanol and butanediol (excluding the conserved variables).

x_0	Interval	x_0	Interval	x_0	Interval
[pyruvate](0)	(0.0001,10)	C_2	(1,20)	O_2	(0.002,2)
[AcP](0)	(0.0001,10)	C_3	(0.1,2)	PO_4	(0.1,100)
[AcO](0)	(0.0001,10)	[lactate]	(0.01,10)		
[AcLac](0)	(10^{-6},1)	[halfglucose]	(0.1,100)		
[AcetIn](0)	(10^{-6},1)	[Ac]	(0.01,10)		
[AcetOut](0)	(10^{-6},1)	[EtOH]	(0.01,10)		
C_1	(0.5,10)	[Butanediol]	(0.0001,0.1)		

[AcP] is the Acetyl phosphate, [AcO] is the acetaldehyde, [AcLac] is the Acetolactate, [AcetIn] is the internal acetoin, [AcetOut] is the external acetoin, [Ac] is the acetate and [EtOH] is the ethanol.

3 SIMULATIONS DATA

Table 6 describes the comparison of credible interval computed based on full measurement and based on minimal measurement.

Table 6. Credible interval for full and minimal informations in Case A.

Parameters	Credible Interval	
	Full information	Minimal information
V^1_{max}	(95.02 , 142.6)	(95.54 , 268.24)
V^2_{max}	(152.44 , 2598.82)	(147.3 , 1938.66)
V^3_{max}	(351.3 , 10575.69)	(86.31 , 8683.26)
V^4_{max}	(92.11 , 126.16)	(89.53 , 719.58)
V^5_{max}	(95.46 , 136.96)	(111.54 , 2702.47)
$V^{7,f}_{max}$	(1273.35 , 4554.52)	(325.65 , 4013.86)
$V^{7,r}_{max}$	(1546.9 , 15912.29)	(1780.56 , 17816.62)
V^8_{max}	(376.32 , 5290.68)	(313.05 , 4841.79)
V^9_{max}	(1046.81 , 7230.22)	(654.06 , 10018.24)
V^{10}_{max}	(216.88 , 307.98)	(152.71 , 1368.56)
V^{11}_{max}	(596.59 , 875.71)	(315.05 , 4028.74)
V^{12}_{max}	(381.76 , 4029.47)	(153.49 , 5587.67)
V^{13}_{max}	(28.35 , 85.48)	(23.83 , 108.79)
V^{14}_{max}	(47.96 , 287.65)	(39.58 , 264.87)

The unit for V_{max} is mmol(l-internal vol)$^{-1}$min^{-1}.

Most of the measurement data in Case B, C, D or E are close to those in Case A and have overlapping error bars (Figure 1). For Case B, only the measurement of AMP, glucose flux and pyruvate flux have error bars that do not overlap with those of Case A (the 'wild-type'). For Case C, only the measurement of F16BP, P3G, PEP, succinate flux, glucose flux and pyruvate flux that do not have overlapping error bars with Case A. Only the error bar of F6P, F16BP and pyruvate flux in Case D that do not intersect with those in Case A. For the last scenario, Case E, the error bars of NADH, P3G, F16BP, succinate flux, glucose flux and pyruvate flux do not overlap with those in Case A. Therefore, it is not easy to guess the changes in the parameters due to these overlapping of uncertainties in measurement data.

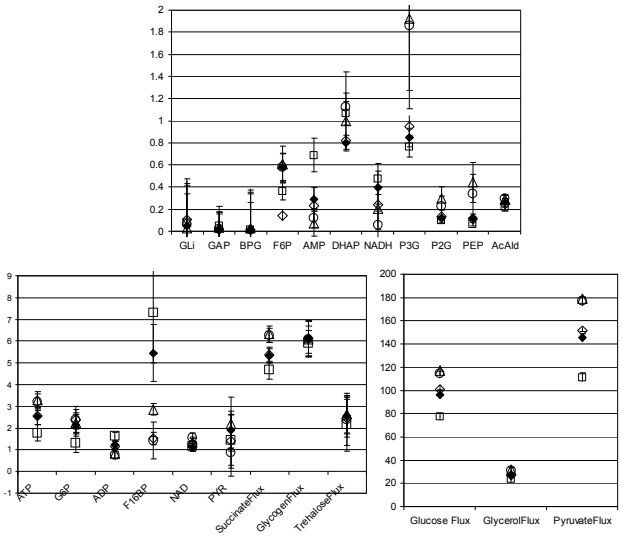

Figure 1. The mean value and standard deviation of simulated measurement data in glycolysis pathways for five different scenarios. The symbol (♦) – Case A, (□) – Case B, (△) – Case C, (◊) – Case D, (○) – Case E.

REFERENCES

Gelman, A. (ed) (1996) Inference and monitoring convergence. Chapman & Hall, London.

Hastings, W.K. (1970) Monte-Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika, 57, 97-&.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953) Equation of State Calculations by Fast Computing Machines, Journal of Chemical Physics, 21, 1087-1092.

Rojas, I., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A. and Wittig, U. (2007) SABIO-RK: a database for biochemical reactions and their kinetics, BMC Systems Biology, 1, S6.

Spiegelhalter, W.R., Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996) Markov chain Monte Carlo in practice. Chapman & Hall, London.