Research Article　　　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Analyzing Intracellular flux Distributions of *Thermoanaerobacter* sp. X514 under Various Substrate Conditions by Gibbs Sampling

Jing Yang[1], Xiaofang Ling[1], Tian Xie[1], Lishan Yao[1] and Visakan Kadirkamanathan[2]*

[1]*Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong, P.R.China*
[2]*Department of Automatic Control & Systems Engineering, University of Sheffield, Sheffield, South Yorkshire, United Kingdom*

Abstract

Thermoanaerobacter sp. X514, capable of fermenting most hexose and pentose to produce ethanol, is of critical importance in industrial bioethanol production. This paper provides a detailed investigation of the intracellular metabolic flux flowing activities of X514 under different conditions, including sole glucose substrate condition, sole xylose substrate condition and mixed glucose and xylose substrates condition by means of Gibbs sampling algorithm and stoichiometric metabolic flux balances. Statistical analysis of the results show that all the flux distributions exhibit Gaussian or truncated Gaussian distributions under the assumption that noise of the system is Gaussian distribution. Pentose phosphate pathway becomes more active when xylose is the sole substrate whereas glycolysis pool is more active under sole glucose substrate condition. Major changes of the fluxes activities occur in the connection flux between glycolysis pool and pentose phosphate metabolite pool, indicating the balances between the need for NADPH, the requirement of ATP as well as the necessity of pyruvate under different substrate conditions. Generation of pyruvate reaches its maximum when sole glucose is fed into the system, indicating the preference of the system on glucose consumption. The whole metabolic flux distributions illustrate vivid information about the central metabolic map of X514.

Keywords: *Thermoanaerobacter* sp. X514; Metabolic flux distribution; Gibbs sampling

Introduction

It is of paramount importance to find alternative energy type so as to replenish the decreasing fossil energy. Biofuel is currently recognized as one of the promising options to replace traditional fossil fuel [1]. Cellulosic biofuel, which generates biofuel from a wide array of feedstock, has attracted a lot of attention mainly because the sources are inexpensive and plentiful while its production is environmentally benign [2]. However, the technology that converts cellulose to bioethanol is at a stage that is far from commercialization. At present, only 40% energy in breweries can be converted to ethanol [3]. Therefore, improvement in cellulosic bioethanol technology is critical and urgent. In terms of the configurations during enzyme hydrolysis and sugar fermentation, traditional bioethanol processes contain Simultaneous Saccharification and Fermentation (SSF) and Separate Hydrolysis and Fermentation (SHF) [4]. Although SSF was proved to give a 13% higher overall ethanol yield than SHF [4], the high cost of enzymes used in cellulosic hydrolysis makes bioethanol uncompetitive to the traditional fossil fuel. Consolidated BioProcessing (CBP), which features cellulase production, cellulose hydrolysis and fermentation in one step, is viewed as the most promising way to convert cellulose to bioethanol at a minimum cost [5].

Microorganisms that can be applied in CBP are not yet available. Currently, the combination of a cellulolytic strain of *Clostridium* species and a saccharolytic strain of *Thermoanaerobacter* species which can produce cellulosic ethanol at high temperature (around 60℃) without the consumption of oxygen, provides an efficient means to rapidly convert cellulose to ethanol by CBP technology [5,6]. A number of thermophilic ethanolgenic anaerobes have been isolated from all over the world in recent years [7-9]. Those organisms exhibit some attractive properties such as: viable substrate utilization, high ethanol productivity, anaerobiosis and tolerance to high temperature (around 60℃) [10-12]. Despite the importance of thermophilic bacteria of the genus *Thermoanaerobacter* spp. in bioenergy production, the ethanol production abilities of these bacteria and the intracellular mechanisms

that stimulate or inhibit ethanol yields were rarely investigated in full detail except for the microorganism *T. ethanolicus* JW200 [13,14]. However, it was illustrated that *T. ethanolicus* JW200 produced ethanol as the major product only at substrate concentrations below 1.0% [15], which prevents its universal industrialization. In contrast, *Thermoanaerobacter* sp. X514, which was discovered as a metal-reducing thermophilic anaerobe [16] can ferment pentose and hexose to produce ethanol simultaneously. Research work on X514 is quite sparse with only a few references available as in [17-19]. In [19], 13C labeled tracer experiments were applied to study the central metabolism of X514. These studies focused on obtaining only partial flux ratios for the central metabolic pathway of X514. A thorough metabolic flux analysis of X514 is, therefore, necessary so as to provide detailed information about the intracellular pathways of X514. The case where multiple substrates are utilized simultaneously by X514 needs addressing especially since the central metabolic flux directions and magnitudes delineate the intracellular reaction activities and convey fundamental information about the overall metabolic system.

Metabolic Flux Analysis (MFA) has long been viewed as a common approach in metabolic systems analysis of microorganisms [20-22]. Software for MFA, for example, MetaFluxNet [23] and FiatFlux [24], are also freely available to researchers. However, standard MFA approaches, which act by formulating a user-defined cost function based upon a few constraints, are often questioned due to their prior assumption about the optimization target of the metabolic system of

***Corresponding author:** Visakan Kadirkamanathan, Department of Automatic Control & Systems Engineering, University of Sheffield, Sheffield, South Yorkshire, United Kingdom, E-mail: visakan@shef.ac.uk

any microorganism. In this paper, without any specified cost function, one special form of Markov Chain Monte Carlo (MCMC) algorithm, Gibbs sampling [25], is applied to the metabolic flux quantification of X514. Though MCMC has been applied in metabolic flux estimation based upon ^{13}C fractional enrichment data [26], none of similar approaches has been applied in MFA without extra carbon labeling constraints. In this paper, without resorting to any optimization assumption about the metabolic system of X514 and by utilizing prior knowledge of the metabolic network and measured data, the paper reveals that Gibbs sampling is able to provide intracellular flux distributions with least constraints about the system. The results show that the flux distributions of X514 exhibit Gaussian or truncated Gaussian formats under different substrate conditions and constraints, which proves that utilizing a single optimized estimation to represent the corresponding flux without considering its sample distributions might be misleading. The connection flux between glycolysis pool and pentose phosphate metabolite pool and the pyruvate-generating fluxes experience major change under different substrate conditions, pointing out future metabolic engineering direction of X514.

The paper is organized as follows: the experiment setup and measurement are described in Section 2. Section 3 presents the Gibbs sampling algorithm. Section 4 illustrates the outcomes of the presented algorithm. Section 5 provides discussions and future prospects of the work.

Materials and Methods

Strains and media

X514 (ATCC BAA-938) was received from Institute for Environmental Genomics, University of Oklahoma, United States. Central metabolic map of X514 is shown in Figure 1. X514 was incubated under a N2 (99.99%) headspace in a flask in the mineral medium which is a modified version of the one used for *Thermoanaerobacter pseudoethanolicus* ATCC 33223 by Wiegel [27] at 60°C. The mineral medium contains (per litre of distilled water) 4.2 g $Na_2HPO_4 \cdot 12H_2O$, 1.5 g KH_2PO_4, 1.0 g NH_4Cl, 0.2 g $MgCl_2 \cdot 6H_2O$, 10.0 ml Wolfe's mineral solution [28] and 1.0 ml 0.1% (m/v) resazurin. The mineral medium was autoclaved for 20 minutes at 115°C. Reducing agents which were composed of 5 g/l $Na_2S_2O_4$ and 3 g/l $NaHCO_3$ were injected into the medium by a syringe after the basal medium was sterilized and cooled to room temperature. In sole substrate condition, 2.0 g/l of glucose or xylose was injected into the mineral medium, whereas in the multiple substrate condition, 1.0 g/l glucose and 1.0 g/l xylose were injected into the mineral medium. All chemicals were purchased from Amresoco (USA), USB (USA), Sigma (USA) or Sinopharm (China) with high purity [29].

Analytical methods

X514 was sampled firstly at 12 hours, then the samples were taken every 8 hours until 68 hours and the last sample was taken at 84 hours. The Optical Density (OD) was measured by Cary 50 spectrophotometer (Varian, USA) at the wavelength of 600 nm and calibrated to Cell Dry Weight (CDW). The relationship between OD and CDW is illustrated in Figure 2 (plot F). Concentrations of the substrates, glucose and xylose, and the two main products, acetate and lactate, were measured by ionic chromatography ics-3000 (Dionex, USA). Concentrations of ethanol were measured by Megazyme Enzyme Kit (Megazyme, Ireland). All experiments were performed in triplicate cultures.
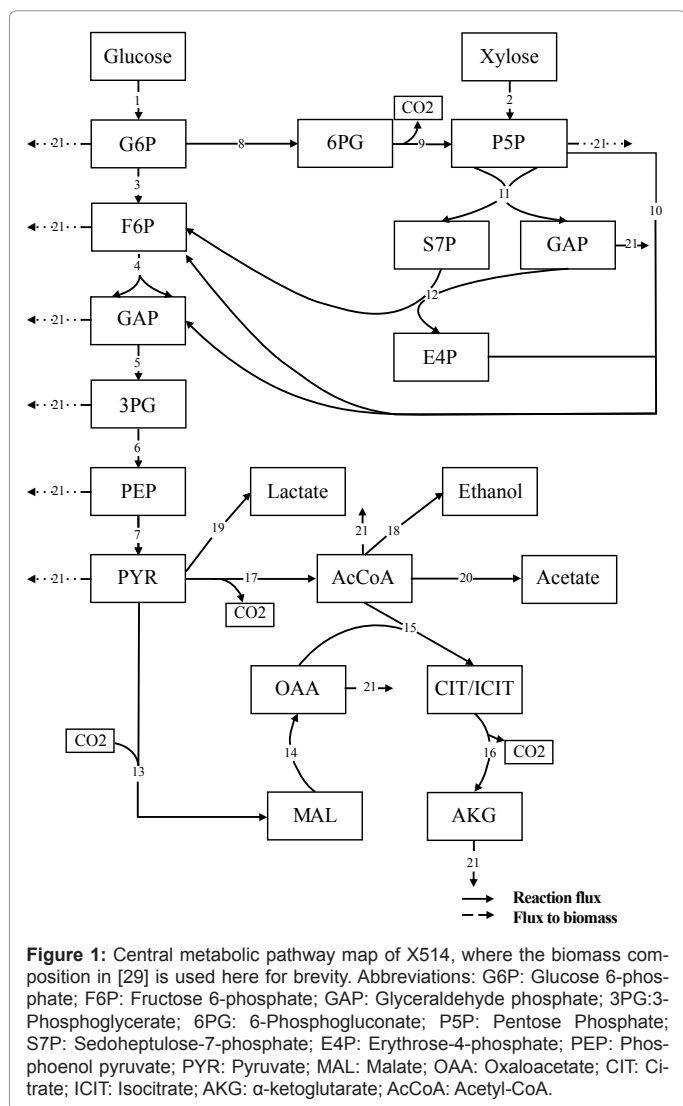
Extracellular flux quantification

Assuming that the cell has exponential growth with a constant growth rate μ, according to the mass balance condition, the cell mass X, the substrate concentration C_s and the product concentration C_p at time t satisfy the following equations [30]:

$$\frac{dX(t)}{dt} = \mu X(t)$$
$$\frac{dC_s(t)}{dt} = -v_s X(t) \quad\quad (1)$$
$$\frac{dC_p(t)}{dt} = v_p X(t)$$

where v_s and v_p are substrate consumption rate and product formation rate, respectively. From equation (1), it can then be derived that:

$$X(t) = X(0)\exp(\mu t) \quad\quad (2)$$

where $X(0)$ is initial cell mass. Replacing $X(t)$ in equation (1) by equation (2) and integrating,

Figure 1: Central metabolic pathway map of X514, where the biomass composition in [29] is used here for brevity. Abbreviations: G6P: Glucose 6-phosphate; F6P: Fructose 6-phosphate; GAP: Glyceraldehyde phosphate; 3PG:3-Phosphoglycerate; 6PG: 6-Phosphogluconate; P5P: Pentose Phosphate; S7P: Sedoheptulose-7-phosphate; E4P: Erythrose-4-phosphate; PEP: Phosphoenol pyruvate; PYR: Pyruvate; MAL: Malate; OAA: Oxaloacetate; CIT: Citrate; ICIT: Isocitrate; AKG: α-ketoglutarate; AcCoA: Acetyl-CoA.
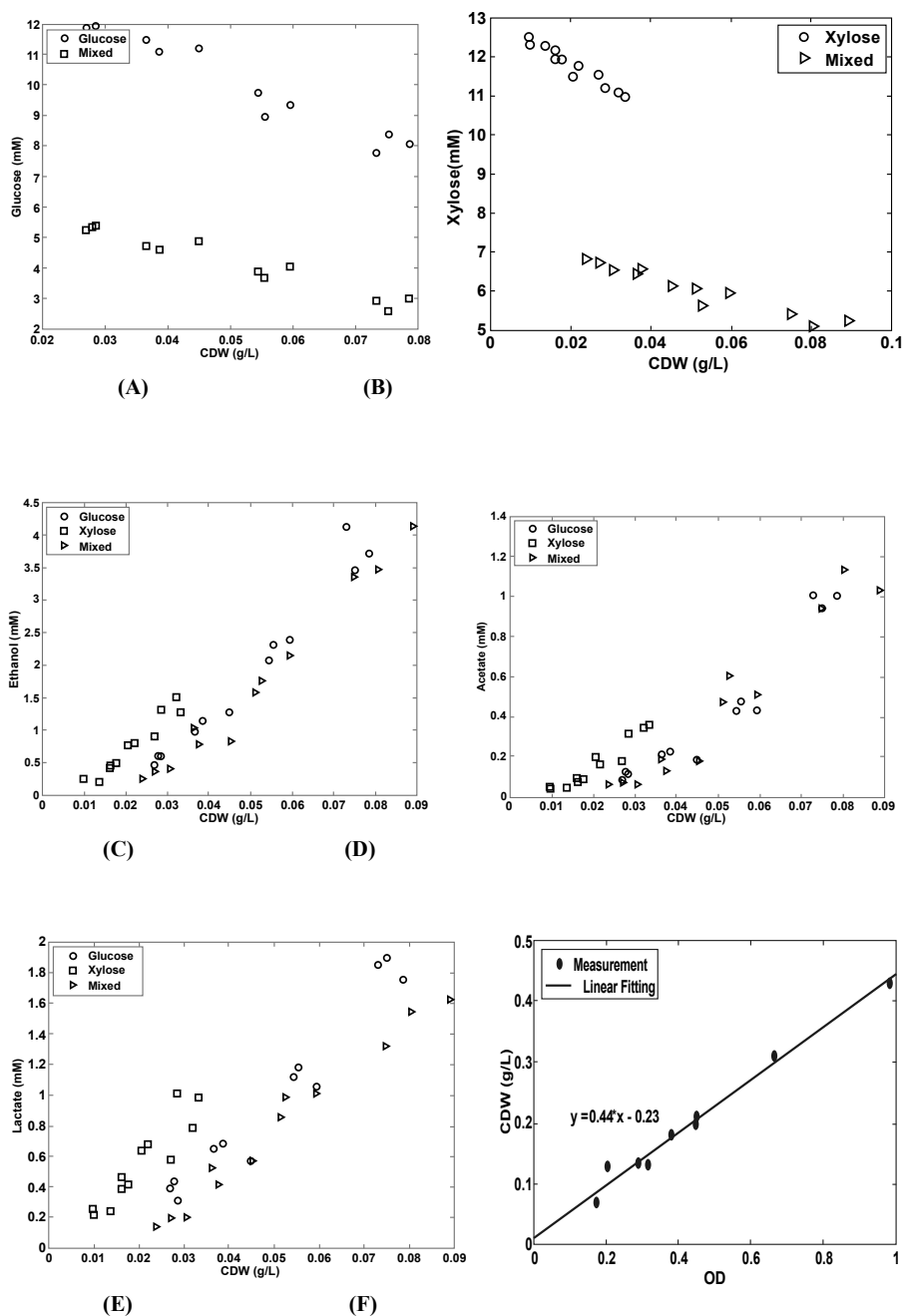
Figure 2: The measured concentrations of glucose (A), xylose (B), ethanol (C), acetate (D) and lactate (E) under various substrate conditions and the linear regression between OD_{600} and CDW (F).

$$C_s(t) = -\frac{v_s}{\mu}X(t) + c_s$$

$$C_p(t) = -\frac{v_p}{\mu}X(t) + c_p$$

where c_s and c_p are constants. The extracellular fluxes v_s and v_p can then be estimated by linear regression plots of $C_s(t)$, $X(t)$ and $C_p(t)$, $X(t)$, respectively.

Gibbs sampling algorithm for metabolic flux distribution analysis

Based on the assumption that the metabolic system is on a quasi-steady state, the following relationship holds [31]

$$Sv = d \qquad (3)$$

where $S \in R^{m \times n}$ is the stoichiometric matrix, $v \in R^n$ are fluxes, including intracellular and extracellular fluxes, $d \in R^m$ is a column

vector whose elements are the flux quantities when the corresponding rows in S involve extracellular fluxes, or zero otherwise. m is the number of metabolites in the metabolic system studied and n is the number of fluxes involved in these reactions between different metabolites.

Considering the noise associated with the measured quantities and the possibility of unknown pathways existing in the studied metabolic system, equation (3) can be modified to include uncertain terms as,

$$Sv + w = d \qquad (4)$$

where the noise w is assumed to be Gaussian distribution with mean 0 and variance $\Sigma \in R^{m \times m}$, i.e. $w \sim N(0, \Sigma)$. According to Bayes theorem, the posterior distribution of v, $\pi(v)$ is then,

$$\pi(v) = p(v \mid d) \propto p(d \mid v) prior(v) \qquad (5)$$

and the likelihood distribution $p(d \mid v)$ can be derived from equation (4):

$$p(d|v) \propto \exp\left(-\frac{1}{2}(d - Sv)^T \Sigma^{-1}(d - Sv)\right) \qquad (6)$$

The prior distribution of v, $prior(v)$, involves the upper and lower constraints of these quantities and expert knowledge about specified fluxes, for example, the bidirectionality or unidirectionality of a reaction, which can be formulated by the expression $Uv \geq f$. Here $U \in R^{l \times n}$ is a matrix specified values of 1 or -1 when the corresponding flux has lower or upper constraints, and 0, otherwise. l is the number of constraints and f is the vector containing all the corresponding constraints. $prior(v)$ can then be expressed as,

$$prior(v) = \Phi(Uv - f) \qquad (7)$$

where $\Phi(\cdot)$ is the Heaviside function taking the value of 1 when $Uv \geq f$, and vanishing otherwise. Incorporating prior distribution in equation (7) and the likelihood distribution in equation (6) into equation (5), the posterior distribution of can then be formulated as,

$$\pi(v) \propto \exp\left(-\frac{1}{2}(d - Sv)^T \Sigma^{-1}(d - Sv)\right)\Phi(Uv - f) \qquad (8)$$

Due to the fact that it is difficult to sample from $\pi(v)$ directly, Gibbs sampling is adopted here to generate statistical samples from $\pi(v)$. Gibbs sampling works by generating samples from the full conditional distribution of each component v_i in v, and accepts these samples unconditionally [32]. Taking the stoichiometric matrix $S = [S_{i-} \ S_i \ S_{i+}]$, where S_{i-} is the matrix or vector composed of the 1st to $(i-1)^{th}$ column of the matrix S, S_i is the ith column of S, S_{i+} is the matrix or vector with $(i-1)^{th}$ column to the last column of the matrix S, respectively. And the vector $v = [v_{i-} \ v_i \ v_{i+}]^T$, where v_{i-}, v_i and v_{i+} represent the 1st to $(i-1)^{th}$, i^{th} and $(i-1)^{th}$ to the last item of the vector v, respectively, so that,

$$Sv = [S_{i-} \ S_i \ S_{i+}][v_{i-} \ v_i \ v_{i+}]^T$$

full conditional distribution of v_i, $p(v_i \mid \cdot)$ can then be derived:

$$p(v_i \mid \cdot) \propto \exp\left(-\frac{1}{2}(v_i - \mu_i)^T \Sigma_i (v_i - \mu_i)\right) prior(v_i) \qquad (9)$$

where

$$\mu_i = S_i^T(S_{i+}v_{i+} - S_{i-}v_{i-} - d)$$
$$\Sigma_i = S_i^T \Sigma S_i \qquad (10)$$

and $prior(v_i)$ is the constraint information of the ith flux, v_i. Conditional

distribution of v_i is then a (truncated) Gaussian distribution with mean μ_i and variance Σ_i. The Gibbs sampling algorithm used to generate a Markov Chain for the flux v is described in Table 1, where the maximum iteration is set as 10,000 by trials.

In order to test the convergence of the generated Markov Chains, three chains are generated by Gibbs sampling algorithm with random initial fluxes, simultaneously. Convergence of the chains with the burn-in period removed is examined by comparing the variances between chains, ϕ_b, and the variances within each chain, ϕ_w [33]. Assuming that the variance for the ith chain with j samples is ϕ_{ij}, then the average of the variances within each chain $\overline{\phi_i}$ and the average of the variances between all the chains $\overline{\phi}$ are,

$$\overline{\phi_i} = \frac{1}{s}\sum_{j=1}^{s} \phi_{ij}$$
$$\overline{\phi} = \frac{1}{t}\sum_{i=1}^{t} \overline{\phi_i} \qquad (11)$$

where s and t are the length of each chain and the number of chains, respectively. Given

$$\phi_b = \frac{s}{t-1}\sum_{i=1}^{t}(\overline{\phi_i} - \overline{\phi})^2$$
$$\phi_w = \frac{1}{t}\sum_{i=1}^{t}\left(\frac{1}{s-1}\sum_{j=1}^{s}(\phi_{ij} - \overline{\phi_i})^2\right) \qquad (12)$$

then

$$Q = \sqrt{\left(\frac{s-1}{s} + \frac{\phi_b}{s\phi_w}\right)} \qquad (13)$$

As the simulation converges, Q should be decreasing and be close to 1. In the paper, an empirical threshold of 1.5 is taken, as a measure of the convergence. So that when the estimated Q is lower than 1.5, it is deemed that the chains have converged.

Results

Measured concentrations of the substrates and the products for sole glucose condition, sole xylose condition and mixed glucose and xylose condition are shown in Figure 2. By exponential fitting of the growth curve of X514, and linear regression between the concentrations of substrates and CDW, as well as the concentrations of products and CDW, extracellular fluxes are estimated and shown in Table 2. It is

MaxIter = 10,000;
Initialize the flux v(0) with random value generator;
for k = 1:MaxIter
v(k) = v(k -1);
for l = 1:n
Calculate μ_i and Σ_l according to equation (10);
Generate the sample θ from the Gaussian distribution
$N(\mu_l, \Sigma_l)$ within the constraint $\Phi(\cdot)$;
Update $v_l(k)$ with the sample θ, $v_l(k)$;
end
end

Table 1: Basic procedure for Gibbs sampling algorithm.

Substrate (mmol/g)	Glucose	Xylose	Mixed
$v_{glucose}$	6.43	0	3.96
v_{xylose}	0	5.01	2.44
$v_{ethanol}$	5.42	4.41	5.6
$v_{lactate}$	2.42	2.56	2.2
$v_{acetate}$	1.45	1.15	1.64

Table 2: Estimated fluxes for the substrates and products under glucose, xylose and mixed substrates conditions.

then clear that the consumption rates for glucose and xylose under sole substrate condition are both much faster than their rates in the mixed substrate condition, whereas the generation rates of the three major products are pretty much the same in all the three cases, and the productions rates of lactate are always faster than that of acetate.

Gibbs sampling is applied to estimate the flux distributions of X514. The source code is implemented in Matlab (MathWorks). Considering the measurement noise nature and existence of unknown pathways in the metabolic system, the system variance matrix Σ is assumed to be a diagonal matrix with unique value 1 on the diagonal. The selection of a diagonal matrix as the variance matrix indicates that interactive disturbances between metabolites are ignored and each metabolite is freely active in the reaction pool. And the specified value for system variance represents the magnitude of the disturbance of unknown metabolic pathways to metabolic reactions. All fluxes are assumed unidirectional except the fluxes in the TCA cycle, pentose phosphate metabolite pool, and the flux G6P→F6P. G6P→6PG and 6PG→P5P+CO$_2$ are assumed to be unidirectional under all cases, thanks to the fact that NADPH, most of which has to be generated from the reaction, is required during the whole cell life cycle. The flux G6P→F6P is assumed in reversed direction under sole xylose substrate condition, due to the requirement of NADPH generated in its neighbor reaction G6P→6PG. Under mixed substrates condition, the same reaction is assumed flowing from G6P→F6P, based upon the fact that glucose consumption is normally in priority position in mixed glucose and xylose substrates condition. Maximum constraints for the fluxes are set as 50 or 0 and the lower constraints for the fluxes are set as 0 or -50 depending on their unidirectionality or bidirectionality characteristics. From paper [19], it is proved that under glucose substrate condition, the flux split ratio between G6P→6PG is less than 3%, the flux for this reaction is then set as 3% of the original glucose uptake rate under sole glucose substrate condition.

With available extracellular flux data shown in Table 2, distributions of the intracellular fluxes and extracellular fluxes are estimated and illustrated in Figure 3 and 4, respectively. Convergence of Gibbs sampling algorithm is tested by running three chains simultaneously with random starting values. The parameter Q (Materials and Methods) is calculated and compared to prespecified threshold, 1.5, during the simulation to ensure convergences of these chains. Mean of the samples of flux v_2 of the three chains in 10,000 runs under glucose substrate condition is taken as an example and illustrated in Figure 5. It is clear that the three chains have eventually converged to the same value.

In order to justify the intracellular flux estimation results obtained by using Gibbs sampling approach, common Flux Balance Analysis (FBA) method is applied to quantify the intracellular fluxes of X514 with the same constraints and the objective of maximizing biomass synthesis. The Matlab program linprog is used to get the optimization results. Comparisons of the estimation results from FBA and the mean estimation results from Gibbs sampling of intracellular fluxes under

different substrate conditions are shown in Figure 6. It is obvious that the two results match closely. However, FBA method alone has to rely on predefined objective function and can only provide single estimations of the intracellular fluxes, whereas Gibbs sampling approach is able to provide the distributions of all intracellular fluxes without accounting on any assumptions of system preferences.

From the Gibbs sampling results, it is clear that all the fluxes exhibit Gaussian or truncated Gaussian distribution formats, where the truncations obviously originate from the prior constraints information imposed on these fluxes. Thorough comparison of the flux distributions and their statistical information under various substrate conditions indicate that the flux G6P→F6P experiences critical changes under these conditions. Mean of the reaction flux reaches its maximum under glucose substrate condition, whereas under mixed substrate condition, its mean value is in-between that of the other two conditions, indicating the variation of the active degree of the glycolysis pool under different substrate conditions. The mean values of the reactions G6P→6PG and 6PG→P5P+CO$_2$, which are the major link reactions between the glycolysis pool and pentose phosphate pathway, increase gradually from the glucose substrate condition to xylose substrate condition and mixed substrates condition, implying the requirement for NADPH becomes much more intense along with the import of xylose into the metabolic system. Generation of pyruvate from glycolysis pathways is at its maximum when glucose is adopted as the sole substrate, whereas this value is gradually decreasing under the mixed substrate condition and sole xylose substrate condition, implying the system's priority on glucose consumption. The pentose phosphate metabolite pool is much more active when xylose is used as the sole substrate. The flux leading to TCA cycle, i.e. PYR→MAL, is quite small under all the three conditions. The reactions in the TCA cycle are much stable and the changes are negligible compared to these reactions in the other two metabolic pools. With the aim of maximizing ethanol yield, sole glucose substrate seems still the best candidate for bioethanol generation compared to sole xylose substrate condition and mixed glucose and xylose substrates condition (Table 3).

The importance of prior constraints on metabolic flux distributions is illustrated in Figure 7, where the posterior flux distribution of the flux G6P→6PG under sole glucose substrate condition with or without the unidirectional prior constraint on it is compared. It is clear that the prior constraint information significantly affect the mean and the variance of the flux. Compared to standard MFA which can only provide an optimized estimation of intracellular metabolic fluxes based upon a user-defined objective function, MFA by means of Gibbs sampling can incorporate the prior constraints and generate the posterior statistical formats of these fluxes, which are more comprehensive and straightforward.

Discussion

Simple estimation of intracellular fluxes by traditional metabolic flux analysis approaches is often insufficient to provide critical information about a metabolic system. Contrary to standard MFA, which has to rely on system constraints and user-defined cost function to obtain an optimum solution of intracellular fluxes, in this paper, Gibbs sampling is applied to metabolic flux quantification. Without resorting to any objective function, the distributions of the intracellular fluxes are generated by means of deliberately constructed Markov Chains. Not only do the distributions provide means and variances of intracellular fluxes, but they also illustrate statistical formats of the estimations. The Gaussian or truncated Gaussian formats of these intracellular flux
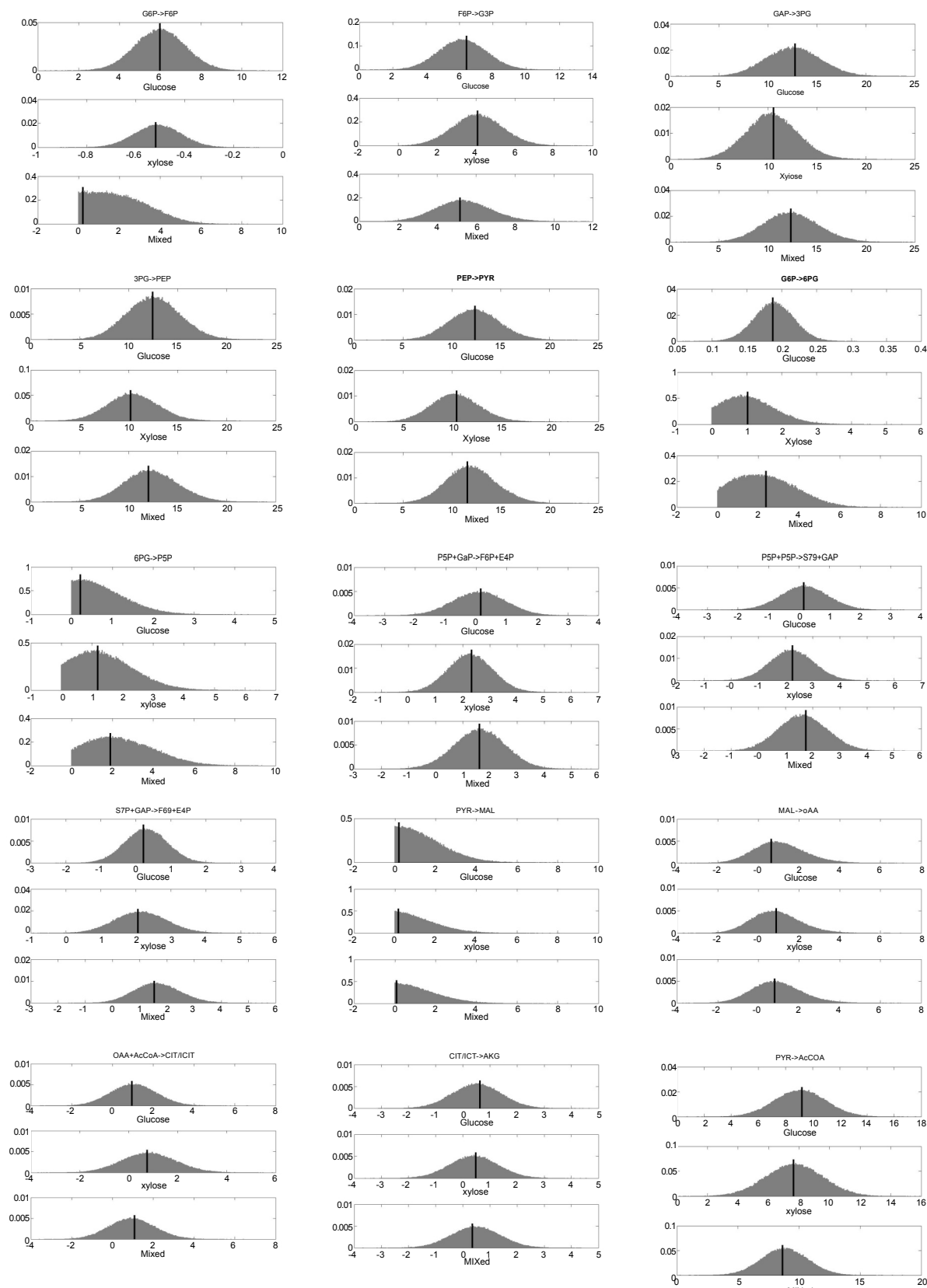
Figure 3: Intracellular flux distributions of X514 under different substrates, where the horizontal labels represent different substrate condition. The black vertical lines are estimated mean of the intracellular flux distributions and the title of each plot represents the corresponding reaction.
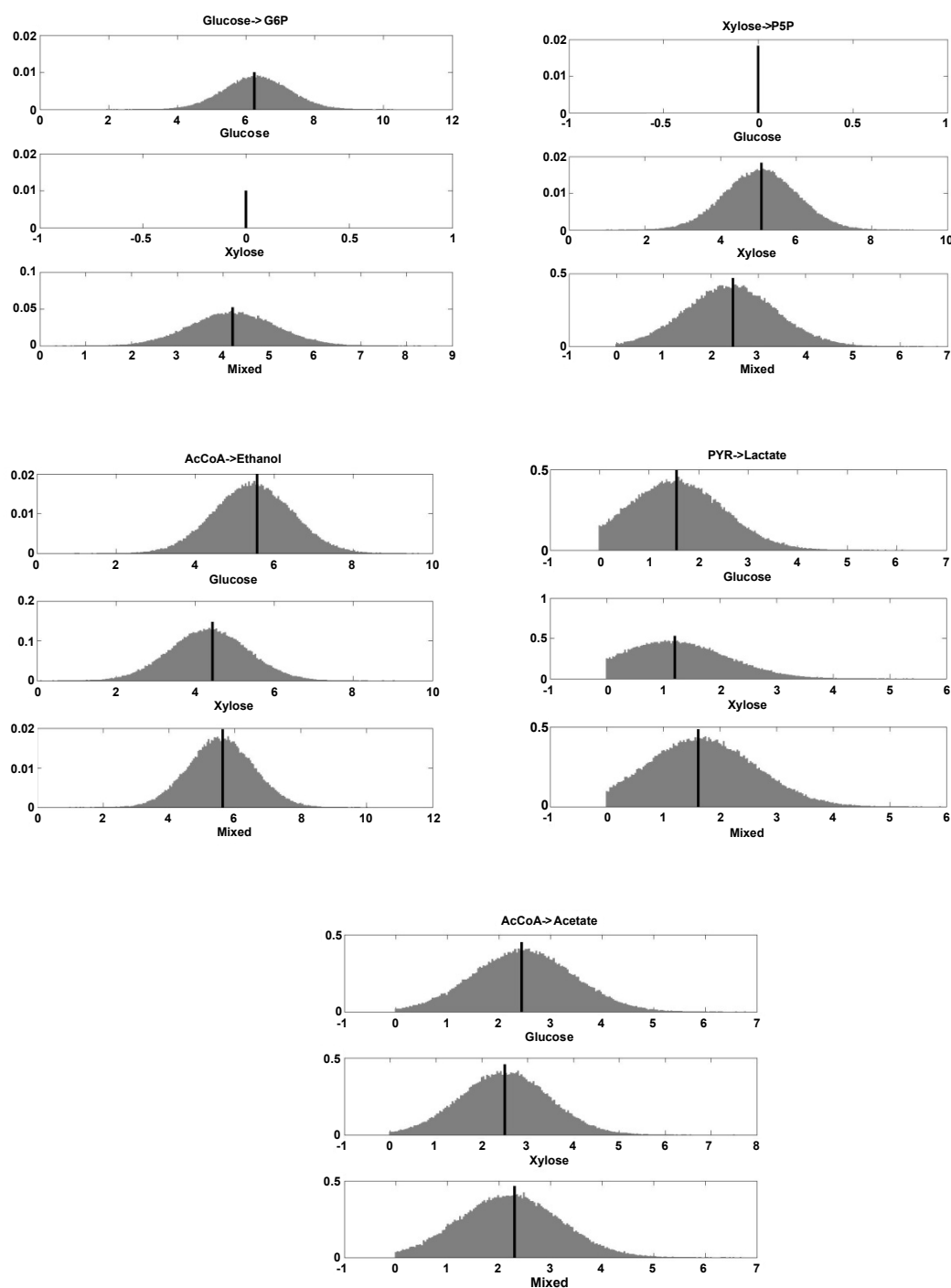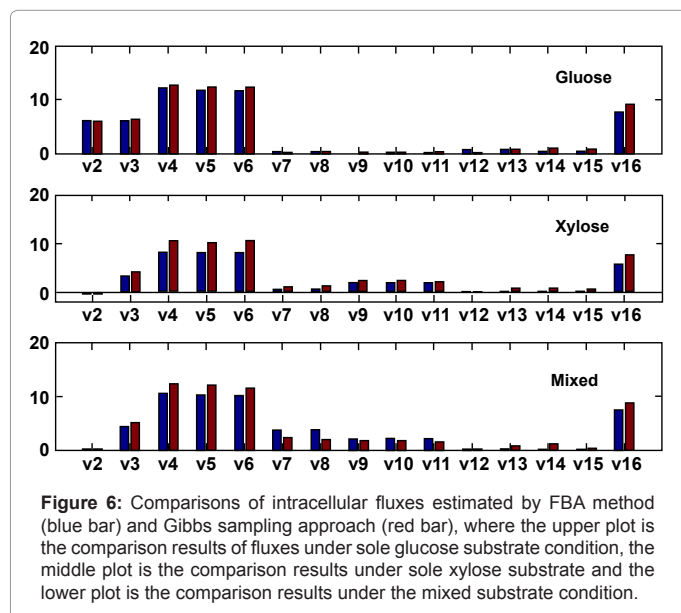
Figure 4: Extracellular flux distributions of X514 under different substrates, where the horizontal labels represent different substrate condition. The black vertical lines are estimated mean of the intracellular flux distributions and the title of each plot represents the corresponding reaction.
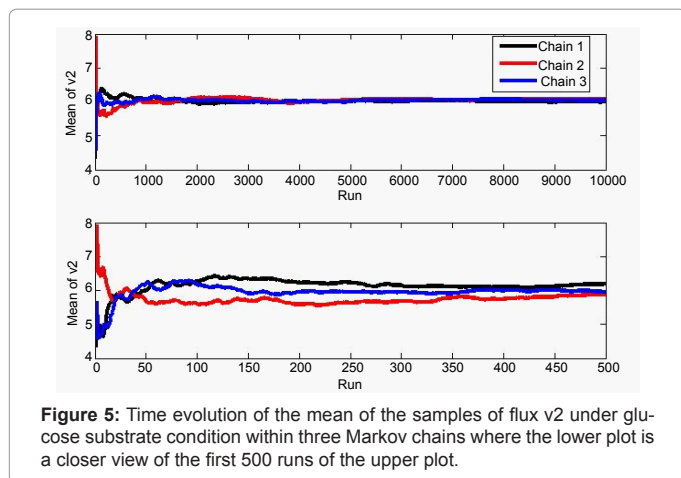
distributions address the importance of taking sample distributions into account during metabolic system analysis, as is the missing part in standard MFA. Analysis of the intracellular flux distributions of X514 illustrate that the connection flux between glycolysis pool and pentose phosphate metabolite pool as well as the pyruvate-generating fluxes are critical fluxes experiencing major changes during ethanol fermentation under different substrate condition. However, similar information will be difficult to obtain from standard MFA. Future metabolic engineering work in X514 will need to focus on these pathways.
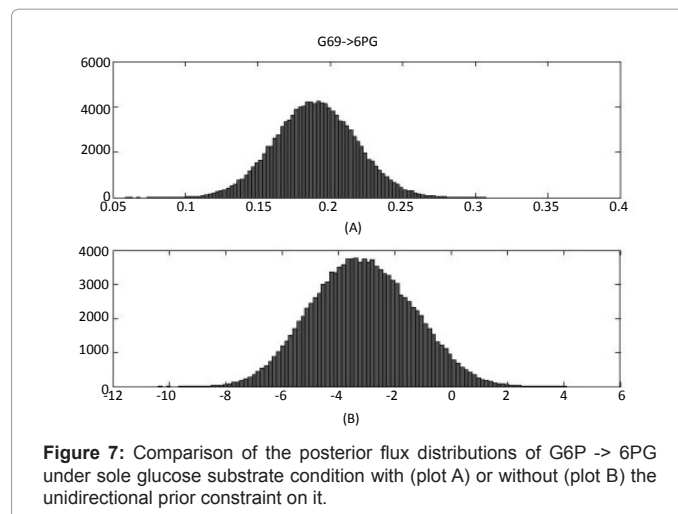
However, one has to note that the distribution formats of Gibbs sampling are sensitive to the constraints of enzyme activities. indicating the importance of prior biological knowledge in metabolic flux analysis. Therefore, the future work will have to be twofold, one is to apply Gibbs

Figure 5: Time evolution of the mean of the samples of flux v2 under glucose substrate condition within three Markov chains where the lower plot is a closer view of the first 500 runs of the upper plot.

Figure 6: Comparisons of intracellular fluxes estimated by FBA method (blue bar) and Gibbs sampling approach (red bar), where the upper plot is the comparison results of fluxes under sole glucose substrate condition, the middle plot is the comparison results under sole xylose substrate and the lower plot is the comparison results under the mixed substrate condition.

Flux	Glucose	Xylose	Mixed
G6P→F6P	5.99±1.23	-0.52±0.10	0.20±2.15
F6P→GAP + GAP	6.42±1.43	4.08±1.29	5.18±1.42
GAP→3PG	12.77±2.79	10.57±2.64	12.34±2.82
3PG→PEP	12.44±2.72	10.18±2.57	12.01±2.75
PEP→PYR	12.35±2.57	10.48±2.44	11.60±2.56
G6P→6PG	0.19±0.03	1.02±0.84	2.38±1.61
6PG→P5P +CO_2	0.22±0.94	1.19±1.17	1.89±1.99
P5P + E4P→F6P + GAP	0.15±0.87	2.33±0.83	1.62±0.94
P5P + P5P→S7P + GAP	0.16±0.78	2.26±0.81	1.75±0.91
S7P + GAP→F6P + E4P	0.22±0.61	2.06±0.74	1.54±0.84
PYR + CO_2→MAL	0.20±1.90	0.18±1.65	0.08±1.75
MAL→OAA	0.63±1.24	0.88±1.19	0.80±1.22
OAA + AcCoA→CIT/ICIT	0.96±1.12	0.76±1.11	1.08±1.13
CIT/ICIT→AKG	0.63±0.90	0.48±0.90	0.36±0.91
PYR→AcCoA + CO_2	9.21±1.90	7.64±1.87	8.64±1.91

Table 3: Mean and standard deviation of the estimated intracellular fluxes.

sampling algorithm together with available expert experiences about enzyme constraints to analyze metabolic fluxes and then use the results to perform metabolic engineering work; another one is to use the

Figure 7: Comparison of the posterior flux distributions of G6P -> 6PG under sole glucose substrate condition with (plot A) or without (plot B) the unidirectional prior constraint on it.

metabolic engineering feedback information to provide more accurate enzyme constraints and apply these constraints together with Gibbs sampling to further understand the studied system.

References

1. Lynd LR, de Brito Cruz CH (2010) Make way for ethanol. Science 330: 1176.

2. Waltz E (2008) Cellulosic ethanol booms despite unproven business models. Nat Biotechnol 26: 8-9.

3. Service RF (2010) Is there a road ahead for cellulosic ethanol? Science 329: 784-785.

4. Ohgren K, Bura R, Lesnicki G, Saddler J, Zacchi G (2007) A comparison between simultaneous saccharification and fermentation and separate hydrolysis and fermentation using steam-pretreated corn stover. Process Biochem 42: 834-839.

5. Lynd LR, van Zyl WH, McBride JE, Laser M (2005) Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol 16: 577-583.

6. Ng TK, Ben-Bassat A, Zeikus JG (1981) Ethanol production by thermophilic bacteria: fermentation of cellulosic substrates by cocultures of Clostridium thermocellum and Clostridium thermohydrosulfuricum. Appl Environ Microbiol 41: 1337-1343.

7. Xue Y, Xu Y, Liu Y, Ma Y, Zhou P (2001) Thermoanaerobacter tengcongensis sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China. Int J Syst Evol Microbiol 51: 1335-1341.

8. Onyenwoke RU, Kevbrin VV, Lysenko AM, Wiegel J (2007) Thermoanaerobacter pseudethanolicus sp. nov., a thermophilic heterotrophic anaerobe from yellowstone national park. Int J Syst Evol Microbiol 57: 2191-2193.

9. Larsen L, Nielsen P, Ahring BK (1997) Thermoanaerobacter mathranii sp. nov., an ethanol-producing, extremely thermophilic anaerobic bacterium from a hot spring in Iceland. Arch Microbiol 168: 114-119.

10. Carere CR, Sparling R, Cicek N, Levin DB (2008) Third generation biofuels via direct cellulose fermentation. Int J Mol Sci 9: 1342-1360.

11. Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, et al. (2008) How biotech can transform biofuels. Nat Biotechnol 26: 169-172.

12. Taylor MP, Eley KL, Martin S, Tuffin MI, Burton SG, et al. (2009) Thermophilic

ethanologenesis: future prospects for second-generation bioethanol production. Trends Biotechnol 27: 398-405.

13. Lacis LS, Lawford HG (1988) Ethanol production from xylose by Thermoanaerobacter ethanolicus in batch and continuous culture. Arch Microbiol 150: 48-55.

14. Hild HM, Stuckey DC, Leak DJ (2003) Effect of nutrient limitation on product formation during continuous fermentation of xylose with Thermoanaerobacter ethanolicus JW200 Fe (7). Appl Microbiol Biotecnol 60: 679-686.

15. Wiegel J, Ljungdahl LG, Demain AL (1985) The importance of thermophilic bacteria in biotechnology. Critical Reviews in Biotechnology 3: 39-108.

16. Roh Y, Liu SV, Li G, Huang H, Phelps TJ, et al. (2002) Isolation and characterization of metal-reducing thermoanaerobacter strains from deep subsurface environments of the piceance basin, Colorado. Appl Environ Microbiol 68: 6013-6020.

17. Lin L, Song H, Ji Y, He Z, Pu Y, et al. (2010) Ultrasound-mediated DNA transformation in thermophilic gram-positive anaerobes. PLoS One 5: e12582.

18. He Q, Lokken PM, Chen S, Zhou J (2009) Characterization of the impact of acetate and lactate on ethanolic fermentation by Thermoanaerobacter ethanolicus. Bioresource Technol 100: 5955-5965.

19. Feng X, Mouttaki H, Lin L, Huang R, Wu B, et al. (2009) Characterization of the central metabolic pathways in Thermoanaerobacter sp. strain X514 via isotopomer-assisted metabolite analysis. Appl Environ Microbiol 75: 5001-5008.

20. Bailey JE (1998) Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. Biotechnol Prog 14: 8-20.

21. Varma A, Palsson BO (1994) Metabolic flux balancing: Basic concepts, scientific and practical use. Nat Biotechnol 12: 994-998.

22. Wiechert W (2001) 13C Metabolic Flux Analysis. Metabolic Engineering 3: 195-206.

23. Lee DY, Yun H, Park S, Lee SY (2003) Metafluxnet: the management of metabolic reaction information and quantitative metabolic flux analysis. Bioinformatics 19: 2144-2146.

24. Zamboni N, Fischer E, Sauer U (2005) Fiatflux - a software for metabolic flux analysis from 13C-glucose experiments. BMC Bioinformatics 6: 209.

25. Gilks WR , Wild P (1992) Adaptive rejection sampling for gibbs sampling, JR Stat Soc Series C (Appl Statist) 41: 337-348.

26. Kadirkamanathan V, Yang J, Billings SA, Wright PC (2006) Markov Chain Monte Carlo algorithm based metabolic flux distribution analysis on Corynebacterium glutamicum, Bioinformatics 22: 2681-2687.

27. Wiegel J, Ljungdahl LG, Rawson JR (1979) Isolation from soil and properties of the extreme thermophile Clostridium thermohydrosulfuricum. J Bacteriol 139: 800-810.

28. Wolin EA, Wolin MJ, Wolfe RS (1963) Formation of methane by bacterial extracts. J Biol Chem 238: 2882-2886.

29. Wu JR, Son JH, Seo HJ, Kim KH, Nam YK, et al. (2005) Metabolic flux analysis of Beijerinckia indica for PS-7 production. Biotechnology and Bioprocess Engineering 10: 91-98.

30. Provost A, Bastin G (2004) Dynamic metabolic modelling under the balanced growth condition. J Process Control 14: 717-728.

31. Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. Curr Opin Biotechnol 14: 491-496.

32. Casella G, George EI (1992) Explaining the gibbs sampler. The American Statistician 46: 167-174.

33. Gilks WR, Richardson S, Spiegelhalter D (1996) Markov chain Monte Carlo in practice. Chapman & Hall/CRC.