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

**Jing Yang**, Xiaofang Ling, Tian Xie, Lishan Yao and Visankan Kadirkamanathan*

1Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong, P.R.China
2Department 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°C) without the consumption of oxygen, provides an efficient means to rapidly convert cellulose to ethanol by CBP technology [5,6]. A number of thermophilic ethanologenic 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°C) [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: Visankan Kadirkamanathan, Department of Automatic Control & Systems Engineering, University of Sheffield, Sheffield, South Yorkshire, United Kingdom, E-mail: visankan@shef.ac.uk

Received March 27, 2012; Accepted May 03, 2012; Published May 07, 2012


Copyright: © 2012 Yang J, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
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 13C 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 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 Na2HPO4·12H2O, 1.5 g KH2PO4, 1.0 g NH4Cl, 0.2 g MgCl2·6H2O, 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 Na2S2O4 and 3 g/l NaHCO3 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 Amresco (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 $\mu$, 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]:

\[
\frac{dX(t)}{dt} = \mu X(t)
\]

\[
\frac{dC_s(t)}{dt} = -v_s X(t)
\]

\[
\frac{dC_p(t)}{dt} = v_p X(t)
\]

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

\[
X(t) = X(0) \exp(\mu t)
\]

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; PSP: Pentose Phosphate; S7P: Sedoheptulose-7-phosphate; E4P: Erythrose-4-phosphate; PYR: Pyruvate; MAL: Malate; OAA: Oxaloacetate; CIT: Citrate; ICIT: Isocitrate; AKG: α-ketoglutarate; AcCoA: Acetyl-CoA.

---
where \( c_i \) 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]

\[
Sv = d
\]

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

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

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

$$ \pi(v) = p(v | d) \propto p(d | v) \pi_{\text{prior}}(v) \quad (5) $$

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

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

The prior distribution of $v$, $\pi_{\text{prior}}(v)$, 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 $\delta v \geq f$. Here $U \in \mathbb{R}^{m \times m}$ 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. $\pi_{\text{prior}}(v)$ can then be expressed as,

$$ \pi_{\text{prior}}(v) = \Phi(Uv - f) \quad (7) $$

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 $v$ can then be formulated as,

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

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 $S = [S_{ij} \quad S_{ij} \quad S_{ij}]$, where $S_{ij}$ is the matrix or vector composed of the $i^{th}$ to $(i-1)^{th}$ column of the matrix $S$, $S_{ij}$ is the $j^{th}$ column of the matrix $S$, respectively. And the vector $v = [v_i \quad v_i \quad v_i]^T$, where $i$, $v_i$, and $v_i$ represent the $i^{th}$ to $(i-1)^{th}$, $i^{th}$ and $(i-1)^{th}$ to the last item of the vector $v_i$, respectively, so that,

$$ Sv = [S_{ij} \quad S_{ij} \quad S_{ij}][v_i \quad v_i \quad v_i]^T $$

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

$$ p(v_i | \cdot) \propto \exp\left(-\frac{1}{2}(v_i - \mu)^T \Sigma(v_i - \mu)\right)\pi_{\text{prior}}(v_i) \quad (9) $$

where

$$ \mu_i = S_i^T(S_{ij}v_i - S_{ij}v_i - d) $$

$$ \Sigma_i = S_i^T \Sigma S_i $$

and $\pi_{\text{prior}}(v_i)$ is the constraint information of the $i^{th}$ flux, $v_i$. Conditional distribution of $v_i$ is then a (truncated) Gaussian distribution with mean $\mu_i$ and variance $\Sigma_i$. 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.

### 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); Calculate $\mu_i$ and $\Sigma_i$, according to equation (10); Generate the sample $\theta$ from the Gaussian distribution $N(\mu_i, \Sigma_i)$ within the constraint $\Phi(\cdot)$; Update $v_i(k)$ with the sample $\theta$, $v_i(k)$; end end |

Table 1: Basic procedure for Gibbs sampling algorithm.
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 X 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+CO2 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 reaction flux reaches its maximum 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 fluxes, but they also illustrate statistical formats of the estimations.
Figure 3: Intracellular flux distributions of X514 under different substrates, where the horizontal labels represent different substrate conditions. 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
enzyme constraints to analyze metabolic fluxes and then use the results sampling algorithm together with available expert experiences about metabolic engineering feedback information to provide more accurate enzyme constraints and apply these constraints together with Gibbs sampling to further understand the studied system.

Acknowledgement

This work is financially supported by National Natural Science Foundation of China (NSFC) and Royal Society, United Kingdom [grant numbers 30800202, 31011130157] and Scientific Research Foundation for the Returned Overseas Chinese Scholars, respectively. Professor Jizhong Zhou from Institute for Environmental Genomics, University of Oklahoma, United States, provided the strain Thermoanaerobacter sp. X514 used in this work. We also thank Prof Jian Xu, Prof. Qiu Cui, Dr. Chenggang Xu, Ms Yueong Ji from Bioenergy Resource Centre, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Science, for their helps and useful discussions during the work.

References


This article was originally published in a special issue, Medical statistics: Clinical and experimental research handled by Editor(s). Dr. Herbert Pang, Duke University, USA.