

# Construction and Analysis of the Model of Energy Metabolism in *E. coli*

Zixiang Xu<sup>1,2</sup>, Xiao Sun<sup>2\*</sup>, Jibin Sun<sup>1\*</sup>

**1** Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China, **2** State Key Laboratory of Bioelectronics, Southeast University, Nanjing, China

## Abstract

Genome-scale models of metabolism have only been analyzed with the constraint-based modelling philosophy and there have been several genome-scale gene-protein-reaction models. But research on the modelling for energy metabolism of organisms just began in recent years and research on metabolic weighted complex network are rare in literature. We have made three research based on the complete model of *E. coli*'s energy metabolism. We first constructed a metabolic weighted network using the rates of free energy consumption within metabolic reactions as the weights. We then analyzed some structural characters of the metabolic weighted network that we constructed. We found that the distribution of the weight values was uneven, that most of the weight values were zero while reactions with abstract large weight values were rare and that the relationship between  $w$  (weight values) and  $v$  (flux values) was not of linear correlation. At last, we have done some research on the equilibrium of free energy for the energy metabolism system of *E. coli*. We found that  $E_{out}$  (free energy rate input from the environment) can meet the demand of  $E_{in}^{ch}$  (free energy rate dissipated by chemical process) and that chemical process plays a great role in the dissipation of free energy in cells. By these research and to a certain extend, we can understand more about the energy metabolism of *E. coli*.

**Citation:** Xu Z, Sun X, Sun J (2013) Construction and Analysis of the Model of Energy Metabolism in *E. coli*. PLoS ONE 8(1): e55137. doi:10.1371/journal.pone.0055137

**Editor:** Stephen S. Fong, Virginia Commonwealth University, United States of America

**Received:** November 2, 2012; **Accepted:** December 27, 2012; **Published:** January 30, 2013

**Copyright:** © 2013 Xu 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.

**Funding:** Support for this work was provided by "National Basic Research Program of China (2011CBA00804)" and the Open Research Fund of State Key Laboratory of Bioelectronics (2011E12). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: xzx21c@163.com

## Introduction

Since various 'Omics' datasets are becoming available, biology has transitioned from a data-poor to a data-rich environment. Systems biology has become a rapidly growing field as well [1]. Genome-scale models of metabolism have only been analyzed with the constraint-based modelling philosophy [2,3]. Genome-scale network models of diverse cellular processes have been generated and there have been several genome-scale GPR (gene-protein-reaction) models [4–10]. An extensive set of methods for analyzing these genome-scale models have also been developed and have also been applied to study a growing number of biological problems [12,13]. But research on the energy metabolism of organisms just begin in recent years, such as FVA (flux variability analysis) [14,15] and EBA (energy balance analysis) [16–18], and so on. All these methods are depended on the modelling of energy metabolism system of organisms. Data of Gibbs free energy of formation of every compound and Gibbs free energy change of every reaction is the core in this kind of modelling. Up to now, the most detailed genome-scale GPR model is the iAF1260 version of *E. coli* [5], but the modelling of it's energy metabolism is still incomplete [5]. There are 2381 reactions (not including the reaction defined for growth) and 1039 metabolites in *E. coli* iAF1260, and apart from 304 EX\_ & DM\_ reactions (The text 'EX\_' denotes an exchange reaction for a metabolite that can enter or leave the extra-cellular compartment. 'DM\_' reactions are similar and signify compounds that the degradation pathway is unknown), the reconstructed reaction number is 2077 [5]. By the newest group

contribution method (GCM),  $\Delta_r G$  of 1996 reactions (96%) and  $\Delta_f G$  of 872 compounds (84%) can be estimated [19–22]. There leaves Gibbs free energy change ( $\Delta_r G$ ) of 81 reactions (4%) and Gibbs free energy of formation ( $\Delta_f G$ ) of 167 compounds (16%) unknown for *E. coli* iAF1260 [5,19]. We have complemented, by computational method, the remaining unknown  $\Delta_r G$  (including the standard Gibbs free energy change  $\Delta_r G^\circ$  and the free energy change of reaction at 1 mM concentrations for all species  $\Delta_r G^m$ ) and  $\Delta_f G$  (just the standard Gibbs free energy formation  $\Delta_f G^\circ$ ). Energy metabolism models of other organisms, as we know, do not exist up to now. Research on metabolic weighted complex network are also rare in literature except that Almaas has used flux value as the weight of metabolic network [23].

In this paper, we have done three research on the complete model of *E. coli*'s energy metabolism. First, we construct a metabolic weighted network using the rates of free energy consumption in metabolic reactions as the network weights. Then we did some research on some structural characters of the metabolic weighted network we constructed. At last, we did some research on the equilibrium of free energy for the energy metabolism system of *E. coli*.

## Materials and Methods

Before constructing the metabolic weighted network, we complement the remained unknown  $\Delta_r G$  and  $\Delta_f G$  in *E. coli* iAF1260. Then we construct the model using the rates of free energy consumption in metabolic reactions as the network weights.

At first, we draw the metabolic unweighted network of *E. coli*; we then calculate the flux distribution of *E. coli*\_iAF1260; the third, we calculate the weights of the metabolic weighted network of *E. coli* that we will construct; at last, we calculated the input and output of free energy about *E. coli*.

### Complement the Remained Unknown Free Energy Change of *E. Coli*\_iAF1260

**1) Infer the unknown standard Gibbs free energy of formation for 167 compounds.** The stoichiometric matrix,  $\mathbf{S}$ , is the center-piece of a mathematical representation of genome-scale metabolic networks. It represents each reaction as a column and each metabolite as a row, where each numerical element is the corresponding stoichiometric coefficient. In the calculation of  $\Delta_f G$  of compounds or  $\Delta_r G$  of reactions, we should use “reaction with marvin charges (pH7)” as the stoichiometry [5].

There are 1039 metabolites in the iAF1260 model, and if distinguishing the different compartments in the cell, i.e. [c] (cytoplasm), [e] (extracellular space), and [p] (periplasm), the total number of metabolites in the model is 1668. So there are 1668 rows in the stoichiometric matrix  $\mathbf{S}$ . For a certain compound even in different compartments, the free energy change of formation is the same. In these 1039 metabolites, Gibbs free energy of formation of 872 compounds (84%) can be estimated by group contribution method [19–22], while 167 compounds (16%) remained unknown for *E. coli*\_iAF1260. There are 2381 reactions in the iAF1260 model, so there are 2381 columns in the stoichiometric matrix  $\mathbf{S}$ . For those 2077 reconstructed reactions, Gibbs free energy change of 1996 reactions (96%) can be estimated by group contribution method [19–22], while 81 reactions (4%) remained unknown for *E. coli*\_iAF1260.

From the equation (1) and (2) of the paper [19], we can infer that

$$\Delta_r G'_{est} = \sum_{i=1}^m v_i \left( \sum_{j=1}^{N_{gr}} n_j \Delta_{gr} G'_j \right) \text{ (for each compound)} \quad (1)$$

$$\Delta_f G'_{est} = \sum_{j=1}^{N_{gr}} n_j \Delta_{gr} G'_j \text{ (for each compound)} \quad (2)$$

$$\Delta_r G'_{est} = \sum_{i=1}^m v_i (\Delta_f^i G'_{est}) \quad (3)$$

Where  $\Delta_r G'_{est}$  is the estimated  $\Delta_r G'^{\circ}$ ;  $v_i$  is the stoichiometric coefficient of species  $i$  in the reaction, and  $m$  is the number of species involved in the reaction;  $\Delta_{gr} G'_j$  is the contribution of group  $j$ ;  $n_j$  is the number of instances of group  $j$  in the molecular structure;  $N_{gr}$  is the number of groups for which  $\Delta_{gr} G'_j$  is known;  $\Delta_f G'_{est}$  is the estimated  $\Delta_f G'^{\circ}$ ;  $\Delta_f^i G'_{est}$  is the estimated Gibbs free energy of formation of  $i$ -th species. Equation (3) reflects the relationship between  $\Delta_f G'^{\circ}$  and  $\Delta_r G'^{\circ}$  for a reaction.

From those 167 compounds with unknown  $\Delta_f G'^{\circ}$ , we seek out their involved reactions. Gibbs free energy change of one reaction can not be calculated, if a compound with unknown free energy change of formation appears in it. But the case is not for all. In the reactions involving the structural group with unknown energy, the structural group appears on both sides of the reaction, which

means it cancels out of “group energy change” for the reaction. That is to say, while the compounds contain a structural group with unknown energy (such as “R” group, a pseudoatom) and appear in a reaction and the reaction does involve a change in the group, we can still calculate and estimate the energy change of the reaction, but we cannot estimate the formation energies of the compounds. Because a reaction with unknown  $\Delta_r G$  may include several compounds with unknown  $\Delta_f G$ , from equation (3), the values of those 81 unknown  $\Delta_r G$  and the values of those 167 unknown  $\Delta_f G$  may be interdependent. So we cannot infer the unknown  $\Delta_f G$  of compounds in a reaction just from the value of known  $\Delta_r G$  of this reaction, but we may infer the 167 unknown  $\Delta_f G$  of compounds from all of the values of known  $\Delta_r G$  by solving their simultaneous equations.

Now we infer these unknown  $\Delta_f G'^{\circ}$  from the known  $\Delta_r G'^{\circ}$  data of those involved reactions. We use vector  $\mathbf{X}$  (with dimension  $167 \times 1$ ) to indicate the values of those unknown  $\Delta_f G'^{\circ}$  of 167 compounds; Use vector  $\mathbf{P}$  (with dimension  $1668 \times 1$ ) to indicate the values of  $\Delta_f G'^{\circ}$  of entire 1668 compounds in the model of *E. coli*\_iAF1260; Let the value of  $\mathbf{P}(i)$ ,  $i$ -th sub-variable of  $\mathbf{P}$ , be 0, if the  $\Delta_f G'^{\circ}$  of  $i$ -th compound is unknown; Use vector  $\mathbf{F}$  (with dimension  $2381 \times 1$ ) to indicate the values of  $\Delta_r G'^{\circ}$  of entire 2381 reactions in the model of *E. coli*\_iAF1260. From equation (3), we can obtain the following equation

$$\mathbf{S}^T \cdot (\mathbf{T} \cdot \mathbf{X} + \mathbf{P}) = \mathbf{F} \quad (4)$$

Where  $\mathbf{T}$  (with dimension  $1668 \times 167$ ) is the transfer matrix from vector  $\mathbf{X}$  to the vector indicating the values of  $\Delta_f G'^{\circ}$  of entire 1668 compounds in the model of *E. coli*\_iAF1260;  $\mathbf{S}$  (with dimension  $1668 \times 2381$ ) is the stoichiometric matrix of *E. coli*\_iAF1260,  $\mathbf{S}^T$  is its transpose. Further we obtain

$$\mathbf{S}^T \cdot \mathbf{T} \cdot \mathbf{X} = \mathbf{F} - \mathbf{S}^T \cdot \mathbf{P} \quad (5)$$

Write it as

$$\bar{\mathbf{S}} \cdot \mathbf{X} = \bar{\mathbf{F}} \quad (6)$$

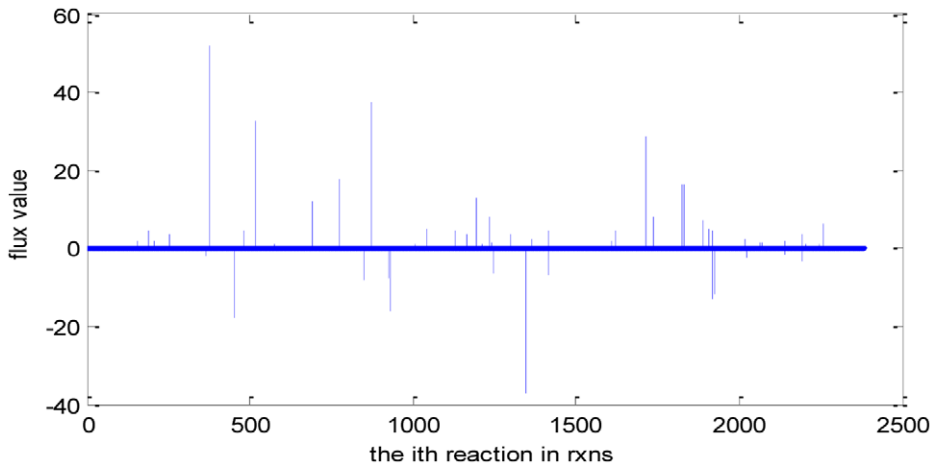
Where  $\bar{\mathbf{S}} = \mathbf{S}^T \cdot \mathbf{T}$ ,  $\bar{\mathbf{F}} = \mathbf{F} - \mathbf{S}^T \cdot \mathbf{P}$

The dimension of vector  $\bar{\mathbf{F}}$  is 2381, and there are 1996  $\bar{\mathbf{F}}(i)$  with known  $\Delta_r G'^{\circ}$ . But there are 244 rows of  $\bar{\mathbf{S}} \cdot \mathbf{X}$  with unknown  $\Delta_r G'^{\circ}$  in the corresponding 1996 rows of  $\bar{\mathbf{F}}(i)$ , for the product of the corresponding rows of  $\bar{\mathbf{S}}$  and  $\mathbf{X}$  includes the sub-variables of  $\mathbf{X}$  and while the remained 1752 rows of  $\bar{\mathbf{S}} \cdot \mathbf{X}$  do not include the sub-variables of  $\mathbf{X}$ . So we can use these 244 rows from  $\bar{\mathbf{S}} \cdot \mathbf{X} = \bar{\mathbf{F}}$  to get a new equation

$$\hat{\mathbf{S}} \cdot \hat{\mathbf{X}} = \hat{\mathbf{F}} \quad (7)$$

Where the dimensions of matrix  $\hat{\mathbf{S}}$  and  $\hat{\mathbf{F}}$  are respectively  $244 \times 167$  and  $244 \times 1$ . By solving equation (7), we can obtain the  $\Delta_f G'^{\circ}$  values of those 167 compounds which are unknown previously. Although equation (7) is not an exact equation (the row rank of matrix  $\hat{\mathbf{S}}$  is not equal to the column rank of matrix  $\hat{\mathbf{S}}$ ), its solution is of least-squares.

**2) Calculate the unknown Gibbs free energy change of 81 reactions.** Conversely, we now use the obtained  $\Delta_f G'^{\circ}$  data to



**Figure 1. Flux distribution of *E. coli* iAF1260.** X-axis indicating every reaction in **rxns** (the order is as the same as in **rxns**, total 2382) and y-axis indicating the value of its corresponding flux (unit is mmol gDW<sup>-1</sup>h<sup>-1</sup>). **Rxns** is the reaction set in the model. doi:10.1371/journal.pone.0055137.g001

calculate the 81 unknown  $\Delta_r G'^{\circ}$ . The method is to substitute the solution value (defined as  $\mathbf{X}_0$ ) of  $\hat{\mathbf{X}}$  which we got from equation (7) to the equation (4), and by a simple calculation, we got the vector  $\mathbf{F}_0$  of  $\Delta_r G'^{\circ}$

$$\mathbf{S}^T \cdot (\mathbf{T} \cdot \mathbf{X}_0 + \mathbf{P}) = \mathbf{F}_0 \quad (4')$$

Now all the sub-variables of  $\mathbf{F}_0$  are known, so we can now obtain the  $\Delta_r G'^{\circ}$  values of those 81 reactions which are unknown previously.

**3) Adjust  $\Delta_r G'^{\circ}$  (the standard Gibbs free energy change) to  $\Delta_r G^m$  (the free energy change of reaction at 1mM concentrations).** The 1M reference state for the metabolite concentrations on which  $\Delta_r G'^{\circ}$  is based does not accurately reflect the metabolite concentrations found in the cell (approximately 1 mM). Thus, we should computationally adjust all estimated  $\Delta_r G'^{\circ}$  to the free energy change of reaction at 1 mM concentrations for all species,  $\Delta_r G^m$ . The relationship between  $\Delta_r G^m$  and  $\Delta_r G'^{\circ}$  is as follows [5,19].

$$\Delta_r G^m = \Delta_r G'^{\circ} + \Delta G_{\text{Transport}} + RT \sum_{i=1}^{PR} n_i \ln(0.001) \quad (8)$$

**Table 1.** Our computation result comparing with Ref. [5].

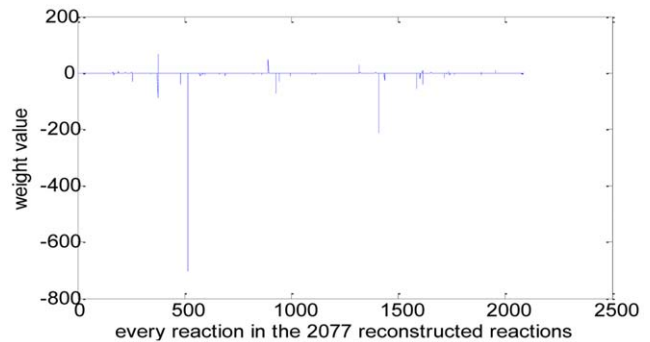
	A	B	C	D
A	81	244	1752	1996
B	can't compare	consistent	consistent	consistent

Note: Row A – Number in *E. coli* iAF1260; Row B – Our computation result comparing with Ref. [5].  
 Column A – reactions with unknown  $\Delta_r G$ ;  
 Column B – reactions with known  $\Delta_r G$  but involving compounds with unknown  $\Delta_f G$ ;  
 Column C – reactions with known  $\Delta_r G$  but not involving compounds with unknown  $\Delta_f G$ ;  
 Column D – total reactions with known  $\Delta_r G$ .  
 doi:10.1371/journal.pone.0055137.t001

Where  $R$  is the universal gas constant;  $T$  is the temperature assumed to be 298 K;  $n_i$  is the stoichiometric coefficient of compound  $i$  in the reaction ( $n_i$  is negative for reactants and positive for products);  $PR$  is the set of products and reactants in this reaction. Note also that for H<sub>2</sub>, we should substitute 0.000034 for 0.001; For O<sub>2</sub>, we should substitute 0.000055 for 0.001; For H<sub>2</sub>O and H<sup>+</sup>, we should not include these compounds in the concentration portion of the calculation at all [5,19]. Here, all of the  $\Delta_r G^m$  values reported in our work have included the energy contribution of the transmembrane electrochemical potential and proton gradient for all reactions involving transport across the cytoplasmic membrane.

**Unweighted Network of *E. Coli* iAF1260**

The general features of *E. coli* iAF1260 are given in Ref. [5]. Two SBML (systems biology markup language) format files to the model *E. coli* iAF1260 can be downloaded from the supplementary information of Ref. [5]. The *in silico* model that we used is *E. coli* iAF1260\_flux1.xml. SBML file properties are also given in Ref. [5]. The dimensions of **rxns**, **mets**, and **genes** are respectively 2382, 1668, 1261. The minimal media of *in silico* model is also an important aspect. The computational minimal



**Figure 2. Weight value distribution of the metabolic network of *E. coli* iAF1260.** X-axis indicates every reaction in the reconstructed reactions (the order is as the same as in **rxns**, total 2077) and y-axis indicates the value of its corresponding weight. **rxns** is the reaction set in the model. doi:10.1371/journal.pone.0055137.g002

**Table 2.** *w* scopes, number of reactions (NR) and their percentages.

<i>w</i> scopes	<-200	-200~-100	-100~-50	-50~0	0	0~40	40~100	>100
NR	2	0	3	248	1753	69	2	0
%	0.1	0	0.14	11.94	84.40	3.32	0.1	0

doi:10.1371/journal.pone.0055137.t002

media of *E. coli*\_iAF1260 is also included in the supplementary information of Ref. [5]. In the method of constraint-based analysis, the biomass objective function (BOF) should be defined. The BOF was generated by defining all of the major and essential constituents that make up the cellular biomass content of *E. coli* [5]. Gene-protein-reaction associations embodied in **rxnGene-Mat** matrix, which is a matrix with as many rows as there are reactions in the model and as many columns as there are genes in the model. The *i*th row and *j*th column contains a one if the *j*th gene in **genes** is associated with the *i*th reaction in **rxns** and zero otherwise. The simulation condition (the nutrients and the uptake rates of the nutrients) of this paper is the same as in the file.

**Flux Distribution of *E. Coli*\_iAF1260**

We now calculate the flux distribution of *E. coli*\_iAF1260. The computational method we use is flux balance analysis (FBA) [11], one of the fundamental genome-scale phenotypic calculations, which can simulate cellular growth. FBA is based on linear optimization of an objective function, which typically is biomass formation. Given an uptake rate for key nutrients and the biomass composition of the cell (usually in mmol component gDW<sup>-1</sup> and defined in the biomass objective function), the maximum possible growth rate of the cells can be predicted *in silico*.

$$\max Z = v_{\text{growth}} \tag{9}$$

Subject to

$$S \cdot v = 0 \tag{10}$$

$$\alpha_i < v_i < \beta_i \tag{11}$$

Where **S** is the stoichiometric matrix, and  $\alpha_i$  and  $\beta_i$  define the bounds through each reaction  $v_i$ . The flux range was set arbitrarily high for all internal reactions so that no internal reaction restricted the network, with the exception of irreversible reactions, which have a minimum flux of zero. The inputs to the system were restricted to a minimal media. We use the COBRA toolbox [11] to

**Table 3.** *w* scopes, number of reactions (NR) and reaction names (RM).

<i>w</i> scopes	<-200	-100~-50	>40
NR	2	3	2
RM	CYTBO3_4ppNADH16pp	ATPM GLCptspp PDH	ATPS4rpp GAPD

doi:10.1371/journal.pone.0055137.t003

carry out this computation of FBA. The flux distribution of *E. coli*\_iAF1260 is illustrated in **Fig. 1**.

**Metabolic Weighted Network Construction for *E. Coli*\_iAF1260**

By the newest group contribution method,  $\Delta_r G$  of 1996 reactions (96%) and  $\Delta_f G$  of 872 compounds (84%) can be estimated [19–22]. We have complemented, by computation method, the remained unknown  $\Delta_r G$  (including the standard Gibbs free energy change  $\Delta_r G^\circ$  and the free energy change of reaction at 1 mM concentrations for all species  $\Delta_r G^m$ ) and  $\Delta_f G$  (just the standard Gibbs free energy formation  $\Delta_f G^\circ$ ). So a complete set of the data of free energy changes for reactions in *E. coli* can be obtained. .

Now we can construct a metabolic weighted network. There is not a standard manner of determining reaction edge weights and Almaas has used flux value as the weight of metabolic network [23]. Here we use the rates of free energy dissipation in metabolic reactions as the network weights. For  $\Delta_r G^m$  of a reaction is the free energy dissipation in unit mol while flux is the passed mol number in unit time (as second). So the rate of free energy dissipation in every reaction is the multiplying product of the flux in this reaction and the free energy change of this reaction.

$$w_i = \Delta_r G_i^m \cdot v_i \tag{12}$$

Where  $w_i$  is the weight of *i*-th edge (i.e. reaction) of metabolic network,  $\Delta_r G_i^m$  is the free energy change of *i*-th reaction and  $v_i$  is the flux value of *i*-th reaction.

**Calculation of Input and Output of Free Energy in *E. Coli*\_iAF1260**

For an open system at nonequilibrium steady state, from the theory of system science, its free energy rate dissipated by the system,  $E_{in}$ , is in absolute value equal to the free energy rate input by the environment,  $E_{out}$ .

$$E_{out} + E_{in} = 0 \tag{13}$$

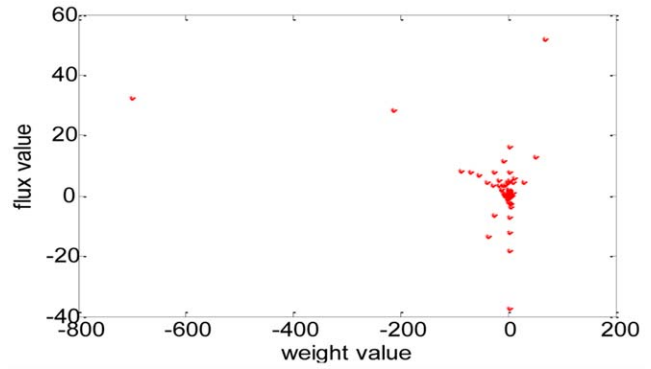
We also distinguish the free energy rate  $E_{in}$  to  $E_{in}^{ch}$  and  $E_{in}^{ph}$  respectively dissipated by chemical process and by physical process that take place in the cell (eq. 14a), while the free energy rate input from environment and through physical process can be neglected (eq. 14b).

$$E_{in} = E_{in}^{ch} + E_{in}^{ph} \tag{14a}$$

**Table 4.** Reaction names (RM) and their corresponding related genes.

RM	ATPM	ATPS4rpp	CYTBO3_4pp	GLCptspp	GAPD	NADH16pp	PDH
Equations	[c] : atp+h2o → adp+h+pi	adp[c]+(4) h[pi]+p[c]<=>atp[c]+(3) h[c]+h2o[c]	(4) h[c]+(0.5) o2[c]+q8h2[c] -> (4) h[pi]+h2o[c]+q8[c]	[c] : bglycogen+pi -> g1p	[c] : g3p+nad+pi <=> 13dpg+h+nadh	(4) h[c]+nadh[c]+q8[c] -> h[pi]+nad[c]+q8h2[c]	(3)[c] : coa+nad+pyr -> accoa+co2+ nadh
genes	b3736, b3737, b3738, b3739, b3731, b3732, b3733, b3734, b3735	b0429, b0430, b0431, b0432	b2415, b2416, b2417, b2418, b2419, b1621, b1101, b1817, b1818, b1818	b1779	b2276, b2277, b2278, b2279, b2280, b2281, b2282, b2283, b2284, b2285, b2286, b2287, b2288	b0114, b0115, b0116	

doi:10.1371/journal.pone.0055137.t004



**Figure 3.** The scatter diagram ( $w$ ,  $v$ ). X-axis indicates  $w$  and y-axis indicates  $v$ .

doi:10.1371/journal.pone.0055137.g003

$$E_{out} = E_{out}^{ch} + E_{out}^{ph} \approx E_{out}^{ch} \quad (14b)$$

For *E. coli*, we have known all of its reactions, the values of corresponding  $\Delta_r G^m$  and the values of corresponding flux, so we can calculate its  $E_{in}^{ch}$  and  $E_{out}$  using the following equation (15) and (16).

$$E_{in}^{ch} = \sum_i^{R_{in}} \Delta_r G_i^m \cdot v_i \quad (15)$$

Where  $R_{in}$  is the set of reactions of the metabolic network excluding EX\_ & DM\_ reactions (The text ‘EX\_’ denotes an exchange reaction for a metabolite that can enter or leave the extra-cellular compartment. ‘DM\_’ reactions are similar and signify compounds that the degradation pathway is unknown),  $\Delta_r G_i^m$  is the free energy change of  $i$ -th reaction in  $R_{in}$  and  $v_i$  is the flux value of  $i$ -th reaction in  $R_{in}$ .

$$E_{out} \approx E_{out}^{ch} = \sum_j^{R_{out}} \Delta_r G_j^m \cdot v_j \quad (16)$$

Where  $R_{out}$  is the set of EX\_ & DM\_ reactions of the metabolic network,  $\Delta_r G_j^m$  is the free energy change of  $j$ -th reaction in  $R_{out}$  and  $v_j$  is the flux value of  $j$ -th reaction in  $R_{out}$ .

**Table 5.** Equilibrium of free energy.

	$E_{out}$	$E_{in}^{ch}$	$E_{in}^{ph}$
value	1890.1	-1424.7	-465.4
%	100%	75.37%	24.63%

doi:10.1371/journal.pone.0055137.t005

## Results and Discussion

### Complement the Remained Unknown Free Energy Change of *E. Coli*\_iAF1260

With our method, we obtain Gibbs free energy change ( $\Delta_r G$ ) of 81 reactions (see **Table S1**) and Gibbs free energy of formation ( $\Delta_f G$ ) of 167 compounds (see **Table S2**) which are previously unknown for *E. coli*\_iAF1260. We add our computed  $\Delta_f G^\circ$  of those 167 compounds to the former known  $\Delta_f G^\circ$  of 872 compounds, and obtain a complete set of  $\Delta_f G^\circ$  of *E. coli*\_iAF1260. The entire  $\Delta_f G^\circ$  values of *E. coli*\_iAF1260 are consistent with the known  $\Delta_r G^\circ$  of 1996 reactions (see **Table S2** and **Table 1**). So we conclude that our computed  $\Delta_f G^\circ$  of those 167 compounds can also agree with the unknown  $\Delta_r G^\circ$  of 81 reactions. Up to now, there is no experimental data in literatures to test the  $\Delta_f G^\circ$  values of those 167 compounds and the  $\Delta_r G^\circ$  values of those 81 reactions.

It is important to know free energies for all metabolites and reactions in *E. coli* by using our method. First of all, the reason why GCM can't calculate the free energies for all metabolites and reactions in *E. coli* or other organisms is that the free energies of some molecular substructures are present in organic-inorganic complexes involving iron, nickel, or cobalt for which the new group contribution method has not been designed [19]. So if we use large scale free energy datasets, not confined to *E. coli*, such as free energies for reaction of KEGG [19], we will get free energies for more metabolites which can't be calculated by the GCM in ref. [19]. Even more, we can estimate some of the free energies of some molecular substructures in organic-inorganic complexes. So the method in our paper will directly contribute to GCM. At the same time, free energies for reactions are useful reference in determine the directions of reactions in cell [5,17] and can also be used as constraints in FBA [18], so if we know all the free energies of reactions for an organism, we can better carry out these tasks.

### Some Structural Characters of the Metabolic Weighted Network of *E. Coli*\_iAF1260

**1) Uneven distribution of the weight values of the metabolic weighted network of *E. coli*\_iAF1260.** We can calculate the weight values of the metabolic network of *E. coli*\_iAF1260 using the above equation (12) (see **Table S2**). We can easily find that the distribution of the weight values is uneven and that most of the weight values are near zero while reactions with abstract large weight values are rare, illustrating in **Fig. 2** and **Table 2**. The reason for the uneven distribution of weight values maybe lies in the uneven distribution of fluxes and the uneven distribution of Gibbs free energy change of reactions. From the uneven distribution of weight values, we can learn that there just are some main channels of free energy dissipation in the physiological process of *E. coli*. **Table 3** has illustrated some reaction channels which have large weight values and Table 4 gives the functions of these reactions.

**2) Reactions of large weight values and their related genes.** **Table 3** shows high  $w$  scopes, corresponding reaction number within these scopes and these reaction names, and we call these reactions the highly-dissipative reactions in the energy metabolism of *E. coli*. We examined into these large weights and found that they were the result of joint action from flux and free energy dissipation, while their  $\Delta_r G^m$  values were not the highest level. **Table 4** gives all of the genes related to these reactions and the rules among genes in these reactions (rules are defined as the relationship among genes catalyzing a reaction such as "AND, OR, NOT" and these rules can be found in Supplementary

Information 1 of [5]), and we find that all of them are not essential genes from the literature [5]. This is important in the energy metabolism of *E. coli*, the deletion or loss of one gene will not result in death, and this may be due to the result of evolution.

**3) Correlation between the weight values and the flux values.** By comparing the distribution of the weight values (**Fig. 2**) with the distribution of fluxes (**Fig. 1**), we can also find that they are not consistent. **Fig. 3** is the scatter diagram ( $w, v$ ), 2077 data pairs in total. Many data pairs are superposition and locate at the same place. From the diagram, we can easily find that the relationship between  $w$  and  $v$  is not of linear correlation. So we can't say that a reaction with high flux has a corresponding high weight and vice versa, and in other words, we can't say that a reaction with high flux will dissipate more free energy. In fact, many different flux values correspond to the same weight value.

Although there is no consistency between flux values and energetic weights, energetic weights are very useful and important in determine the distribution of reaction fluxes. We have defined energetic weights as the rates of free energy dissipations, i.e. the multiplying product of the reaction fluxes and the free energy changes of reactions. Free energy dissipation can be regarded as the counterpart of entropy increase. Based on maximum entropy production principle (MEPP), the authors of the paper have developed a method to improve the prediction accuracy of flux balance analysis [24].

### Equilibrium of Free Energy in the Energy Metabolism of *E. Coli*\_iAF1260

There are 2077 reactions in  $R_{in}$  and 304 reactions in  $R_{out}$  of *E. coli*\_iAF1260. The values of  $E_{in}^{ch}$  and  $E_{out}$  which we calculated are respectively  $-1424.7$  and  $1890.1$ , see **Table 5**. So  $E_{out}$  can meet the demand of  $E_{in}^{ch}$  and  $E_{out}$  is a little more than  $E_{in}^{ch}$ . The absolute difference between  $E_{out}$  and  $E_{in}^{ch}$  is an estimation of  $E_{in}^{ph}$ , and we can find that it takes about a quarter of  $E_{in}^{ch}$ , so we can conclude that chemical process plays a great role in the dissipation of free energy in cells while physical process can not be ignored.

### Conclusions

In this paper, we constructed a metabolic weighted network by using the rates of free energy consumption within metabolic reactions as the network weights. We found several important and interesting results: 1) the distribution of the weight values was uneven; 2) the relationship between  $w$  (weight values) and  $v$  (flux values) was not of linear correlation; 3) by analyzing of the free-energy equilibrium for the energy metabolism system of *E. coli*, we found that it is chemical process other than physical process that plays a great role in the dissipation of free energy in cells. By these research and to a certain extend, we can understand more about the energy metabolism of *E. coli*.

In our next step, we will conduct a similar type of analysis for different organisms using some of the other readily available constraint-based models and run the baseline simulations for growth in different carbon substrate environments. We will also do FBA analysis including energetic weighting as an additional constraint to bias flux distributions.

### Supporting Information

**Table S1** Standard Gibbs free energy formation  $\Delta_f G^\circ$  of 167 compounds.  
(XLS)

**Table S2** Gibbs free energy change ( $\Delta_r G$ ) of 81 reactions (including the standard Gibbs free energy change  $\Delta_r G^\circ$  and the

free energy change of reaction at 1 mM concentrations for all species  $\Delta_r G^m$ , flux distribution and weight values. (XLS)

## References

- Bork P (2005) Is there biological research beyond Systems Biology? A comparative analysis of terms. *Molecular Systems Biology*. 1: Art. No.2005.0012.
- Price ND, Papin JA, Schilling CH, Palsson BO (2003) Genome-scale microbial *in silico* models: the constraints-based approach. *Trends in Biotechnology*. 21(4):162–169.
- Reed JL, Vo TD, Schilling CH, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*. 2(11): 886–897.
- Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genomescale model of *Escherichia coli* K-12 (JLR904 GSM/GPR). *Genome Biology*. 4: R54.
- Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*. 3: Art. No.121.
- Scott AB, Bernhard OP (2005) Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiology*. 5: Art. No.8.
- Thiele I, Vo TD, Price ND, Palsson BO (2005) Expanded metabolic reconstruction of *Helicobacter pylori* (iT341 GSM/GPR): an *in silico* genome-scale characterization of single- and double-deletion mutants. *Journal of Bacteriology*. 187 (16) : 5818–5830.
- Feist AM, Scholten JC, Palsson BO, Brockman FJ, Ideker T (2006) Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Molecular Systems Biology*. 2: Art. No.2006.0004.
- Duarte NC, Herrgård MJ, Palsson BO (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*. 14(7): 1298–1309.
- Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R (2007) Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *The Journal of Biological Chemistry*. 282 (39): 28791–28799.
- Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature Protocols*. 2(3): 727–738.
- Feist AM, Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nature Biotechnology*. 26(6):659–667.
- Di Ventura B, Lemerle C, Michalodimitrakis K, Serrano L (2006) From *in vivo* to *in silico* biology and back. *Nature*. 443 (7111): 527–533.
- Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*. 5: 264–276.
- Segura D, Mahadevan R, Juárez K, Lovley DR (2008) Computational and Experimental Analysis of Redundancy in the Central Metabolism of *Geobacter sulfurreducens*. *PLoS Computational Biology*. 4(2): e36.
- Beard DA, Liang SD, Qian H (2002) Energy Balance for Analysis of Complex Metabolic Networks. *Biophysical Journal*. 83: 79–86.
- Yang F, Qian H, Beard DA (2005) Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry. *Metabolic Engineering*. 7: 251–259.
- Beard DA, Babson E, Curtis E, Qian H (2004) Thermodynamic constraints for biochemical networks. *Journal of Theoretical Biology*. 228: 327–333.
- Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V (2008) Group Contribution Method for Thermo-dynamic Analysis of Complex Metabolic Networks. *Biophysical Journal*. 95: 1487–1499.
- Henry CS, Jankowski MD, Broadbelt LJ, Hatzimanikatis V (2006) Genome-Scale Thermodynamic Analysis of *Escherichia coli* Metabolism. *Biophysical Journal*. 90: 1453–1461.
- Henry CS, Broadbelt LJ, Hatzimanikatis V (2007) Thermodynamics-Based Metabolic Flux Analysis. *Biophysical Journal*. 92: 1792–1805.
- Hatzimanikatis V, Li CH, Ionita JA, Henry CS, Jankowski MD, et al. (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics*. 21(3): 1603–1609.
- Almaas E, Kovács B, Vicsek T, Oltvai ZN, Barabási AL (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*. 427(6997):839–843.
- Zhu Y, Song J, Xu Z, Sun J, Zhang Y, Li Y, Ma Y (2012) Development of thermodynamic optimum searching (TOS) to improve the prediction accuracy of flux balance analysis. *Biotechnology and Bioengineering*. published online: DOI: 10.1002/bit.24739.

## Author Contributions

Conceived and designed the experiments: ZX. Performed the experiments: ZX. Analyzed the data: ZX. Contributed reagents/materials/analysis tools: ZX JS XS. Wrote the paper: ZX.