Second Law of Thermodynamics applied to Metabolic Networks

R. Nigam[1] and S. Liang[2*]
[1]Geballe Laboratory for Advanced Materials,
Stanford University, Stanford, CA 94305-4085, USA
[2]NASA Ames Research Center, Moffett Field, CA, USA

1 Abstract

We present a simple algorithm based on linear programming, that combines Kirchoff's flux and potential laws and applies them to metabolic networks to predict thermodynamically feasible reaction fluxes. These law's represent mass conservation and energy feasibility that are widely used in electrical circuit analysis. Formulating the Kirchoff's potential law around a reaction loop in terms of the null space of the stiochiometric matrix leads to a simple representation of the law of entropy that can be readily incorporated into the traditional flux balance analysis without resorting to non-linear optimization. Our technique is new as it can easily check the fluxes got by applying flux balance analysis for thermodynamic feasibility and modify them if they are infeasible so that they satisfy the law of entropy. We illustrate our method by applying it to the network dealing with the central metabolism of *Escherichia coli*. Due to its simplicity this algorithm will be useful in studying large scale complex metabolic networks in the cell of different organisms.

*Corresponding author. E-mail: shoudan.liang@nasa.gov

2 Introduction

Completely sequenced genomes open the door for wholistic understanding of metabolic networks. Study of metabolic networks is important in predicting the behaviour of the cell as a result of interactions between the different reactions occuring within the cell (Bailey, 1991; Palsson, 2000; Stephanopoulos et al. 1998). For example the flux balance analysis (FBA) has proved to be useful for studying the steady state metabolic flux inside the cell (Varma and Palsson, 1994). This approach is especially a valuable tool in the absence of knowledge of detailed kinetic parameters of reactions inside the cell. Many authors have studied reaction networks from a very theoretical stand point (Oster et al. 1973; Peusner, 1986; Balabanian and Bickart, 1981), and also applied principles from non-equilibrium thermodynamics (Nicolis and Prigogine, 1977; Westerhoff and van Dam, 1987), but it has only been in the last few years, that these techniques have been applied to biological systems (Bonarius et al., 1996, 1997; Pramanik and Keasling, 1997). Recently there have been attempts to further constrain the fluxes that are obtained by FBA, to satisfy the second law of thermodynamics by a method called energy balance analysis (EBA) (Beard et al. 2002). The EBA eliminates thermodynamically infeasible fluxes that are got by applying FBA alone (Price et al. 2002).

In this paper we incorporate the energy feasibilty constraint that is akin to Kirchoff's potential law (Strang, 1986) into the linear FBA theory. This places additional constraints on the FBA solution space and eliminates thermodynamically infeasible fluxes that do not satisfy the loop law. We illustrate our method by applying it to a part of the metabolic network of *Escherichia coli* (Delgado and Liao, 1997). This method will be useful in predicting the behaviour of large scale networks, especially in making predicitions of gene regulation and thermodynamic chemical potentials of the different chemical reactions that go on inside the cell.

Previous attempts to formulate the EBA method resulted in a non-linear optimization problem (Beard et al. 2002), which can lead to errors if the method doesnot converge. The novelty of our approach is that we can solve it using linear programming, and that we can generate many solutions and test them for feasibilty. The approach taken in Beard et al. 2002 and Price et al. 2002 was to set fluxes to zero to satisfy the energy feasibilty constraints. This is a very restrictive approach as it misses out non-zero solutions. Our

method on the other hand first checks the fluxes without setting them to zero for feasibility and only sets fluxes to zero, if a sign transformation fails to make the flux distribution feasible. We also show mathematically that setting the flux of a reaction to zero, leads to no change in the chemical potential for that reaction. In Beard et al. 2002, nonzero change in the chemical potentials are computed for reactions with zero flux. They attribute reactions with zero fluxes to infinite resistance (or zero conductance). The objective of this paper is to present a simple technique that can be applied to large scale metabolic networks to predict the various internal fluxes and the corresponding change in the chemical potentials. To summarize we have developed a technique to test the thermodynamic feasibility of non-zero fluxes, and also to constructively generate feasible solutions. To accomplish this goal we develop some formal machinery using concepts from linear algebra.

3 Flux Balance Analysis: Law of Flux Conservation

In a metabolic network inside the cell, several reactions catalyzed by different enzymes occur in concert. To study the rates of these reactions we resort to flux balance analysis. Applying the law of mass balance, the concentration of metabolites and the reaction rates (or fluxes) are related through the stoichiometry of the system. Using steady state approximation for the mass balance, the flux balance analysis has been formulated as a linear program by many authors (e.g. Varma and Palsson, 2000), in which there is a linear objective function, which could for example be to maximize growth, maximize ATP production, minimize glucose intake etc. This is written as a linear combination of the fluxes.

$$Z = \boldsymbol{d}^T \boldsymbol{f} \tag{1}$$

$$S\boldsymbol{f} = 0 \tag{2}$$

$$\boldsymbol{l} \leq \boldsymbol{f} \leq \boldsymbol{u} \tag{3}$$

where, $\boldsymbol{f} \in \mathcal{R}^n$ is the vector of n fluxes, $S \in \mathcal{R}^{m \times n}$ is a stoichiometric matrix, m is the number of reactants in the network. All vectors by default will be

column vectors. Also, $\boldsymbol{d}$, $\boldsymbol{l}$ and $\boldsymbol{u}$ are vectors $\in \mathcal{R}^n$ of objective function coefficients, lower and upper bound constraints on the fluxes respectively, and $\mathbf{0}$ is a zero vector of size m. In equation (3) the inequalities are componentwise for the vectors. The vector of objective function coefficients have to be determined experimentally, but in most cases has only one non-zero component corresponding to the flux one is trying to optimize.

In the above formulation the number of fluxes n exceeds the number of metabolites m in the cell. So a convenient way to solve the above system of underdetermined equations is to resort to linear programming. Due, to the degeneracy in the problem, there are an infinite number of solutions possible that satisfy all of the constraints and optimize the chosen objective function.

4 Energy Balance Analysis: Second law of Thermodynamics

According to the second law, fluxes must flow from reactants of higher chemical potential to ones of lower chemical potential. This is the direction in which the entropy of the reaction increases (Qian et al. 2002). Since the FBA analysis gives rise to an infinite number of fluxes, many of these flux distributions violate the second law and hence are infeasible. From a network topology point of view, it is the presence of cycles in the flux direction, that violate the law of production of entropy. Applying Kirchoff's loop law, one gets rid of these entropy violating cycles.

From S we remove the columns corresponding to boundary fluxes and keep only the columns of non-redundant internal fluxes. The resulting matrix $G \in \mathcal{R}^{m \times n_i}$, where, n_i is the number of internal fluxes in the network. Using the singular value decomposition (Strang, 1986) one can find the null space matrix N of G. The matrix $N \in \mathcal{R}^{n_i \times n_l}$ consists of n_l orthonormal basis vectors of the null space of G. The dimension of the null space of G gives the number of independent and irreducible loops n_l in the network (Strang, 1986). By this we mean that a single basis loop cannot be decomposed into smaller loops. By taking linear combinations of these basis loops we can generate bigger and compund cycles. The singular value decomposition (SVD) is not unique for the G matrix, so the orthonormal basis for the null space of G is not unique.

Instead of obtaining the orthonormal basis for the null space of G from the SVD, one can obtain the rational basis for the null space from the reduced row echelon form of G. This basis is unique as the reduced echelon form of G is unique.

Associated with each internal flux f_i is a chemical potential difference $\Delta\mu_i$. These potential differences satisfy a law similar to the Kirchoff's loop law in electrical circuits, namely (Beard et al. 2003),

$$K\Delta\boldsymbol{\mu} = 0 \tag{4}$$

where, $K = N^T \in \mathcal{R}^{n_l \times n_i}$ is a matrix whose rows are the basis vectors of the null space of G, and $\Delta\boldsymbol{\mu} \in \mathcal{R}^{n_i}$ is a vector of chemical potential differences for the internal fluxes in the cell.

The second law ensures that the entropy increases in each reaction i and hence the direction of flux f_i is from metabolites of higher chemical potential to one of lower chemical potential and can be expressed as (Beard et al. 2002).

$$f_i\Delta\mu_i < 0 \tag{5}$$

Equations (4) and (5) are thermodynamic feasibility constraints that are applied in addition to the flux balance constraints. Equation (5) is a nonlinear constraint which when incorporated into the FBA makes the problem a non-linear programming problem. In this paper we propose a simpler algorithm to solve the problem as a linear programming problem.

In addition to the above constraints one imposes upper and lower bound constraints on $\Delta\boldsymbol{\mu}$.

$$\boldsymbol{\beta} \leq \Delta\boldsymbol{\mu} \leq \boldsymbol{\alpha} \tag{6}$$

Where, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} \in \mathcal{R}^{n_i}$ represent the lower and upper bounds on the change in chemical potential $\Delta\boldsymbol{\mu}$, and the inequality is componentwise.

In the next section we describe the condition for checking the presence of cyclic fluxes, and we introduce some notation here. We will use upper-case indices to denote sets for example, let F be the set of all fluxes in the network,R be the set of unrestricted fluxes, $F^{\geq 0}$ be the set of non-negative fluxes, $F^{<0}$ and $F^{>0}$ be the set of negative and positive fluxes respectively. Fluxes like $f_i \in F$ means that flux f_i is a flux, r_i is an unrestricted flux, $f_i^{\geq 0}$ is

a non-negative flux etc. The matrix N can be written in terms of its column vectors as $N = [N_{*1}....N_{*k}....N_{*i}....N_{*n_l}]$, where N_{*k} is the kth column vector of N. Also $N_{*k} = [n_{1k}, n_{2k}, ..., n_{ik}, ..., n_{n_ik}]^T$, where n_{ik} is the (i,k) th entry of the matrix N.

5 No cycle feasibility constraint

In this section we introduce a simple test to detect the presence of loops in a metabolic network that violate the second law of thermodynamics. To do so we take advantage of the directionality of the flow of fluxes in the cycle. The number of rows n_l of the K matrix gives the number of loops or cycles in the network (Strang, 1986). These loops are the basis cycles, as combining them gives bigger cycles. If in any row j of the K matrix K_{j*} all the entries (there should be more than one non-zero entry) of the jth row are of the same sign, corresponding to the set of positive fluxes $F_j^{>0}$ for the jth cycle, then the flux distribution is thermodynamically infeasible. If some unrestricted flux r_i belonging to the cycle j is negative, it can be transformed to be positive by reversing the sign of the corresponding entries in the ith column of the G and K matrices, namely G_{*i} and K_{*i} respectively. If after this transformation, any of the rows of the K matrix still have the same sign, then the flux distribution is thermodynamically infeasible, and we detect the presence of a cycle. By satisfying this condition we can get rid off the non-linear constraint in equation (5), and hence we transform the non-linear problem to a linear one.

It should be noted that this no cycle feasibility condition is equivalent to solving the FBA problem with constraints (4) and (5). It can be seen that to satisfy equation (4), for a single row of the K matrix corresponding to a single cycle, atleast one of the $\Delta\mu_i$ should be of a different sign from the rest of the other components in the $\boldsymbol{\Delta\mu}$ vector. This when combined with equation (5), prevents the formation of energy violating loops in the network, and hence satisfies the second law of thermodynamics. This no cycle feasibility condition is only applicable to basis loops, which can be easily observed in the echelon basis.

Lemma 1: (sign transformation lemma)
Transforming the unrestricted internal flux $-r_i$ to r_i changes the sign of the ith column K_{*i} of matrix K.

Proof: From the flux conservation equation $S\boldsymbol{f} = \mathbf{0}$, one can partition the flux vector $\boldsymbol{f}$ into internal flux vector $\boldsymbol{x}$ and boundary flux vector $\boldsymbol{y}$. The columns of the stiochimetric matrix S can likewise be partitioned into columns G corresponding to the internal fluxes and columns H corresponding to the boundary fluxes. That is, $S = [G \;\; H]$. The flux conservation equation can be rewritten as $G\boldsymbol{x} = -H\boldsymbol{y}$.
Since N is the null space matrix of G, we have $GN = 0$, where 0 is a $(m \times n_l)$ matrix of zeros. Consider the ith component of the unrestricted internal flux r_i, which is a component of $\boldsymbol{x}$, that is negative. In the matrix vector product $G\boldsymbol{x}$, r_i multiplies the ith column G_{*i} of matrix G. A negative r_i value can be made positive by transferring the negative sign to all the elements in the column of G_{*i}. By this process the value of the matrix vector product $G\boldsymbol{x}$ is unaffected. Hence G_{*i} becomes $-G_{*i}$. From the equation $GN = 0$ we have $GN = [G_{*1}....G_{*i}....G_{*n_i}][N_{*1}....N_{*i}....N_{*n_l}]$, where G_{*1}, N_{*1} are the first columns of the G and N matrix respectively. This matrix-matrix product can be written compactly as $[G_{*i}n_{i1}....G_{*i}n_{in_l}]$, where, there is implict summation on the repeated index i. From this it is clear that transforming G_{*i} to $-G_{*i}$ changes the sign of the ith row of matrix N, that is entries n_{i1}, ..., n_{in_l} change sign. Since by construction $K = N^T$, the ith row of matrix N corresponds to the ith column K_{*i} of matrix K, which then changes sign.

By use of lemma 1 we transform all the unrestricted negative internal fluxes to reference positive fluxes, from which we can test the cycle condition for fluxes.

Lemma 2: (zero transformation lemma)
If the entries of the ith column of the matrix G, G_{*i}, corresponding to the ith internal flux x_i, which *only* belongs to the jth cycle are set equal to zero then the jth row of the K matrix K_{j*} is zero everywhere except at the ith column, that is entry K_{ji} is nonzero. Hence $\Delta\mu_i = 0$.
Proof: Since N is the null space matrix of G, we have $GN = 0$, where 0 is a $(m \times n_l)$ matrix of zeros. From $GN = 0$, we have $[G_{*i}n_{i1}....G_{*i}n_{ij}....G_{*i}n_{in_l}] = 0$. Here there is summation on the repeated index i. The jth cycle corresponds to the jth row of K and hence the jth column of N matrix. Hence consider the equation corresponding to the jth column of N matrix

$$G_{*i}n_{ij} = \mathbf{0} \tag{7}$$

where, there is summation over the repeated index i and $\mathbf{0} \in \mathcal{R}^m$ is a column

vector of m zeros, and n_{ij} is (i,j)th element of the N matrix.

$$G_{*1}n_{1j} + ... + G_{*n_i}n_{n_ij} = -G_{*i}n_{ij} \tag{8}$$

where on the left side of the above equation we exclude the ith index as it is brought to the right. Without loss of generality let $n_{ij} = 1$, hence

$$G_{*1}n_{1j} + ... + G_{*n_i}n_{n_ij} = -G_{*i} \tag{9}$$

Since G_{*i} is a zero vector, we have

$$G_{*1}n_{1j} + ... + G_{*n_i}n_{n_ij} = 0 \tag{10}$$

From the above equation (10) we see on the left hand side that, $G_{*1}n_{1j}, ..., G_{*n_i}n_{n_ij}$ are atmost $n_i - 1$ nonzero vectors, which were a part of the jth cycle along with $G_{*i}n_{ij}$ (which corresponds to the ith internal flux), setting $G_{*i} = 0$ (is like setting $x_i = 0$) breaks the cycle, and the rest of the nonzero column vectors, $G_{*1}, ..., G_{*n_i}$ in the broken cycle are linearly independent by construction, since we only included non-redundant internal fluxes, and since atmost n_i internal fluxes are present in the jth cycle corresponding to the jth row of the K matrix K_{j*}, that are linearly dependent, breaking the cycle makes the rest of the fluxes in the jth cycle linearly independent by construction. The only way equation (9) holds is when $n_{1j} = ... = n_{n_ij} = 0$. Hence the jth colomn of the matrix N, N_{*j} is zero except for the n_{ij} element which is nonzero. Since $K = N^T$, the jth row of K, K_{j*} is zero except $K_{ji} = n_{ij}$. Therefore, from equation (4) we see that $\Delta\mu_i = 0$.

From lemma 2 we see that when a particular internal flux is zero (we can zero out its respective column in the G matrix), then the corresponding change in chemical potential is zero, which is physically quite intuitive. This can be incorporated as an EBA constraint. Lemma 2 could be applied to overlapping cycles which share fluxes (two cycles are overlapping if they share atleast one flux, that is a column of the K matrix corresponding to the rational basis of the null space has two non-zero entries), in this case some cycles may be broken while some others may be created. It is therefore best to remove one cycle at a time by zeroing out fluxes that only belong to a single cycle, and are not shared among cycles. This is again got by observing the columns of the K matrix (corresponding to the rational basis of the null

space) that contain only a single nonzero entry corresponding to the flux that belongs to a single cycle.

These lemmas provide an automated way of reducing the K matrix in large scale networks with many overlapping loops that share the same fluxes. The K matrix will also be called the feasibilty matrix. In the next section these concepts will be used in the EBA linear programming algorithm.

6 EBA Algorithm

1) Solve the FBA for the flux distribution.
2) If all the fluxes are positive and the K matrix satisfies the no cycle feasibility constraint, then the fluxes obtained from the FBA are thermodynamically feasible. We also find the corresponding $\Delta\mu$ vector that satisfies constraints (4), (5) and (6).
3) If some of the internal fluxes are negative then apply the sign transformation lemma 1 to change the sign of the entries of the columns of the K matrix corresponding corresponding to these negative fluxes. Then if the K matrix satisfies the no cycle feasibility condition we get a feasible flux vector. The corresponding components of the $\boldsymbol{\Delta\mu}$ vector that satisfies constraints (4), (5) and (6) is computed.
4) If after the above three steps, the K matrix is infeasible, we identify fluxes that are not shared by cycles. Corresponding to these fluxes, there is only one nonzero entry in their respective columns, which indicates the cycle in which the flux is present. For example we zero out any one flux x_i in some cycle j and apply the zero transformation lemma 2 to both the G and K matrix.
In the K matrix then remove the jth row and ith column corresponding to the jth cycle and ith flux, and constrain $\Delta\mu_i = 0$ and $x_i = 0$. Also remove the ith column from the G matrix. After this do steps (1)-(3) on the reduced system. If feasible compute $\boldsymbol{\Delta\mu}$ vector that satisfies constraints (4), (5) and (6).
5) If not feasible step (4) can be repeated for each cycle one at a time and checking feasiblity until they are exchausted.

In the above algorithm, computation of $\boldsymbol{\Delta\mu}$ is decoupled from the computaion of the flux distribution. In the MATLAB code however the two can

be combined into an equivalent linear programming formulation, the details are not discussed in this paper. Also, when trying to identify non-shared fluxes, it is best to use the rational basis of the null space for the K matrix.

7 Application of EBA Algorithm

In this section we consider the example discussed in Beard et al (2002). The set of reactions are shown below.
rxn 1: A + 2B $<--->$ C
rxn 2: C + D $<--->$ 2A + 2B
rxn 3: A + B $<--->$ 2D
rxn 4: A + C $<--->$ B + 3D
rxn 5: B $<--->$ D

This reaction network has 5 internal fluxes corresponding to the 5 reactions, and associated with each internal flux is a change in chemical potential. In addition to these there are 4 boundary or exchange fluxes corresponding to the 4 metabolites A, B, C and D.

For this reaction network we wish to determine the maximum steady-state production of reactant D, for a given maximal input flux of reactant C. This problem assumes that reactant C is the only available input substrate, and that its value is set equal to 1, and that for A and B their values are set equal to 0.

The stoichiometric matrix G for the four metabolites A, B, C and D corresponding to the four rows respectively and the five reactions corresponding to the five columns of G respectively, is

$$G = \begin{bmatrix} -1 & 2 & -1 & -1 & 0 \\ -2 & 2 & -1 & 1 & -1 \\ 1 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & 3 & 1 \end{bmatrix}$$

Also, the stoichiometric matrix H for the four metabolites as rows and the four exchange fluxes corresponding to the four metabolites respectively as the four columns is

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

The full stoichiometric matrix is $S = [G \quad H]$ and the flux distribution $\boldsymbol{f} \in \mathcal{R}^9$ that maximizes the production of D, and that satisfies the flux conservation constraint $S\boldsymbol{f} = \mathbf{0}$. The five internal fluxes corresponding to the five internal reactions and the boundary flux corresponding to D are unrestricted, while the other three boundary fluxes corresponding to A, B are less than or equal to 0, and that for C it is less than or equal to 1.

The orthonormal null space vectors corresponding to matrix G are given in Beard et al. (2002) and can be computed easily using MATLAB, and are

$$\boldsymbol{k_1}^T = [-0.7163, -0.3205, 0.4710, -0.3958, -0.0752] \tag{11}$$

$$\boldsymbol{k_2}^T = [-0.3345, -0.4347, -0.6349, 0.1001, 0.5348] \tag{12}$$

The dimension of the null space of matrix G is 2. The K matrix consists of $\boldsymbol{k_1}^T$ and $\boldsymbol{k_2}^T$ as its rows.

Alternatively, one can use the rational null space vectors corresponding to matrix G, that are

$$\boldsymbol{k_{1_r}}^T = [2, 1, -1, 1, 0] \tag{13}$$

$$\boldsymbol{k_{2_r}}^T = [-1, -1, -1, 0, 1] \tag{14}$$

The K matrix then consists of $\boldsymbol{k_{1_r}}^T$ and $\boldsymbol{k_{2_r}}^T$ as its rows. This representation tends to give a clear picture about shared and non-shared fluxes among cycles. Here fluxes x_1, x_2 and x_3 are shared among the two cycles, whereas x_4 belongs to cycle 1 and x_5 belongs to cycle 2.

Optimal flux vector $\boldsymbol{f} = [\boldsymbol{x}^T, \boldsymbol{y}^T]^T \in R^9$ and change in chemical potential vector $\Delta\boldsymbol{\mu} \in R^5$ are obtained using the standard linear programming routine in MATLAB. They are

$$\boldsymbol{y} = [0, 0, 1, 3]^T \tag{15}$$

$$\boldsymbol{x} = [-0.154, 0.445, 0.643, 0.401, 0.956]^T \tag{16}$$

$$\Delta\boldsymbol{\mu} = [0.6178, -0.8158, -0.5801, -1.000, -0.7780]^T \tag{17}$$

Where $\boldsymbol{x} \in \mathcal{R}^5$ is a flux vector of the 5 internal fluxes and $\boldsymbol{y} \in \mathcal{R}^4$ is a flux vector of the 4 boundary fluxes. From equation (15) we observed that 3 units of D are produced per unit of C consumed. All the fluxes have been normalized with respect to one unit of C.

The optimal flux distribution obtained using our algorithm that satisfies the thermodynamic constraint is the same as the one obtained by Beard et al (2002). But according to Beard et al. it is infeasible thermodynamically. But according to our sign test we see that it is feasible and we also provide a corresponding $\boldsymbol{\Delta\mu}$ vector for the $\boldsymbol{x}$ internal flux distribution, that satisfies all the feasibility constraints given in equations (4) and (5). Since in this example $\boldsymbol{\Delta\mu}$ has unrestricted components, in constraint (6) the lower and upper bounds are -∞ and +∞ respectively. The formulation given in Beard et al. might have failed to converge to a $\boldsymbol{\Delta\mu}$ vector and hence the flux distribution was reported infeasible. Our criterian for the EBA test is quite robust and we can immediately tell if the flux distribution is thermodynamically infeasible, without carrying out any nonlinear optimization.

8 Analysis of *E. Coli* Central Metabolism

We use the stoichiometric matrix S of the model *E. coli* system from Table 1 (Delgado and Liao (1997)) for our FBA/EBA analysis. The reaction network contains 19 metabolites linked by 23 reactions (Figure 1 Delgado and Liao (1997). Out of these 23 fluxes there are 3 external or boundary fluxes and the rest 20 are internal fluxes. The network considered takes glucose as input and produces acetate and carbon dioxide. The energy and the metabolites involved in this process are used for the synthesis of proteins, DNA, RNA etc. We applied our algorithm to maximize the production of biomass flux, which is a linear combination of the different fluxes with experimentally determined stoichiometric coefficients, and are given in Neidhardt et al. (1990). These coefficients are for the conversion of key metabolites to biomass of *E. coli* propotoplasm. In the FBA optmization the internal fluxes are unrestricted, and only satisfy the flux balance constraint. The C02 and acetate fluxes come from the literature (Kleman and Strohl (1994); Shiloach et al. (1996)). Since only the relative rates matter, the glucose flux is set to 1, and all other fluxes are normalized with respect to it.

The G matrix is formed by considering the columns of the following in-

ternal fluxes from Table 1 of Delgado and Liao (1997):
$\boldsymbol{x} = [J_{pgi}, J_3, J_{pep}, J_{pyk}, J_{pdh}, J_{ace}, J_8, J_{ict}, J_{11}, J_{12}, J_{ppc}, J_{14}, J_{15}, J_{16}, J_{tkt}, J_{tal}, J_{resp}, J_{atp}, J_{biomass}, J_{glyox}]$, and the H matrix is formed from the columns of the external fluxes $\boldsymbol{y} = [J_{gluc}, q_{CO_2}, q_{ace}]$.
The null space of the G matrix is of dimension 1, hence the K matrix consists of one row. This also indicates the presence of a single loop in the network.

$K = [0, 0, 0, 0.24, 0.24, 0, 0, -0.24, -0.25, -0.24, -0.25, 0, 0, 0, 0, 0, -0.24, -0.73, 0, 0.24]$. We can also use the rational null space vector $\boldsymbol{k}_{1r}^T$ of G to construct the matrix K as before
$\boldsymbol{k}_{1r}^T = [0, 0, 0, 1, 1, 0, 0, -1, -1, -1, -1, 0, 0, 0, 0, 0, -1, -3, 0, 1]$.
From the above equation we see the following 9 fluxes which form a cycle: $[J_{pyk}, J_{pdh}, J_{ict}, J_{11}, J_{12}, J_{ppc}, J_{resp}, J_{atp}, J_{glyox}]$. These correspond to the non-zero entries in the row of the K matrix and also in the $\boldsymbol{k}_{1r}$ vector.

The optimal flux vector $\boldsymbol{f} = [\boldsymbol{x}^T, \boldsymbol{y}^T]^T \in R^{23}$ and change in chemical potential vector $\boldsymbol{\Delta\mu} \in R^{20}$ satisfying the flux balance and thermodynamic constraints computed by our algorithm are:

$\boldsymbol{y} = [1, 2.2, 0.3]^T$

$\boldsymbol{x} = [0.87, 0.85, 1.58, 1.92, 2.71, 0.27, 0.56, -1.03, -1.11, -1.11, -1.38, 0.12, 0.09, 0.03, 0.03, 0.03, 1.80, 2.95, 0.0001, 1.59]^T$

$\boldsymbol{\Delta\mu} - [-8.25, \ 8.25, \ 8.25, \ 6.27, -6.27, -8.25, -8.25, 10.95, 10.95, 10.95, 10.95, -8.25, -8.25, -8.25, -8.25, -8.25, -10.74, -17.30, -8.25, -6.27]^T$

Where $\boldsymbol{x} \in \mathcal{R}^{20}$ is a flux vector of the 20 fluxes including the 19 internal fluxes and the biomass flux. and $\boldsymbol{y} \in \mathcal{R}^3$ is a flux vector of the 3 boundary fluxes. In this example we have rounded the numerical values to two decimal places. The optimized biomass flux is $J_{biomass} = 7.27 \times 10^{-5} \approx 0.0001$ per unit of glucose consumed.

To see if the $\boldsymbol{x}$ vector of internal fluxes is thermodynamically feasible, we apply the sign transformation lemma to the negative flux components x_8, x_9, x_{10} and x_{11}, and see that the K matrix satisfies the feasibility criteria discussed in the previous sections.

9 Conclusions

In this paper we give a simple linear programming algorithm for the flux and energy balance analysis. Instead of applying a nonlinear thermodynamic feasibility constraint, as has been done previously, it uses the sign of the null

space to decide if the flux distribution of the metabolic network computed by flux balance anaysis satisfies the second law of thermodynamics. This technique is different from the previous approaches, as it is constructive, and can generate several feasible solutions for the metabolic network. A self consistent connection between setting a reaction flux to zero with respect to a corresponding change in the chemical potential for that reaction is developed. This can be used as a energy feasibility constraint for the reaction. We applied the method to a part of the metabolic network of *E. coli* and computed the fluxes and change in chemical potentials for the internal reactions. It should however be noted that FBA together with EBA are still not able to constrain the metabolic network completely and this leads to an infinity of flux and change in chemical potential distributions. More realistic bounds on the values of fluxes and change in chemical potentials are required to further constrain the system. This could be got from studying the biochemistry of several pathways. The connection between linear algebra and thermodynamics is exploited to formulate the law of entropy in metabolic networks. Furthermore, as the analysis and technique presented here are simple and straightforward, they can easily be applied to study large scale metabolic networks.

10 Acknowledgments

We like to thank Bob Laughlin for stimulating discussions. This research was supported by a NASA cooperative agreement NCC2-5475.

11 References

Bailey, J. E. 1991. Toward a science of metabolic engineering. *Science* 252:1668-75.

Balabanian, N. and T. A. Bickart. 1981. Linear Network Theory: Analysis, Properties, Design and Synthesis. Matrix, Beaverton, OR.

Beard, D. A., S. Liang, and H. Qian. 2002. Energy balance for analysis of complex metabolic networks. *Biophys. J.* 83:79-86.

Beard, D. A., H. Qian, and J. Bassingthwaighte. 2003. Stoichiometric foundation of large-scale biochemical system analysis.

Bonarius, H. P. J., V. Hatzimanikatis, K. P. J. Meesters, C. D. De Gooijer, G. Schmid and J. Tramper. 1996. Metabolic flux analysis of hybridoma cells in different culture media using mass balances. *Biotechnol. Bioeng.* 50:299-318.

Bonarius, H. P. J., G. Schmid and J. Tramper. 1997. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends Biotechnol.* 15:308-314.

Delgado, J., and J. C. Liao. 1997. Inverse Flux Analysis for Reduction of Acetate Excretion in *Escherichia coli. Biotechnol. Prog.* 13:361-367.

Kleman, G. L., and W. R. Strohl. 1994. Acetate metabolism by *Escherichia coli* in high-cell-density fermentation. *Appl. Environ. Microbiol.* 60:3952-3958.

Neidhardt, F. C., J. L. Ingraham and M. Schaechter. 1990. *Physiology of the Bacterial Cell. A Molecular Approach.* Sinauer Associates, Inc. Sunderland, MA.

Nicolis, G. and I. Prigogine. 1977. Self-organization in Nonequilibrium Systems. Wiley Intersci., New York.

Oster, G. F., A. S. Perelson and A. Katchalsky. 1973. Network thermodynamics: Dynamic modelling of biophysical systems. *Quart. Rev. Biophys.* 6:1-134.

Palsson, B. O. 2000. The challenges of *in silico* biology. *Nat. Biotechnol.* 18:1147-1150.

Peusner, L. 1986. Studies in Network Thermodynamics, Elsevier.

Pramanik, J., and J. D. Keasling. 1997. Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* 56:398-421.

Price, N. D., I. Famili, D. A. Beard and B. O. Palsson. 2002. Extreme Pathways and Kirchhoff's second law. *Biophys. J.* 83:2879-2882.

Qian, H., D. Beard, and S. Liang. 2002. Stoichiometric network theory for nonequilibrium biochemical systems. *Eur. J. Biochem.* 270:415-421.

Shiloach, J., J. Kaufman, A. S. Guillard and R. Fass. 1996. Effect of glucose supply strategy on acetate accumulation , growth and recombinant protein production by *Escherichia coli* BL21 (1DE3) and *Escherichia coli* JM109. *Biotechnol. Bioeng.* 49:421-428.

Stephanopoulos, G., A. Aristidou and J. Nielsen, 1998. Metabolic Engineering: Principles and Applications.

Strang, G. 1986. Introduction to Applied Mathematics Wellesley-Cambridge

Press, Wellesley, MA.
Westerhoff, H. V. and K. van Dam. 1987. Thermodynamics and Control of Biological Free-Energy Transduction. Elsevier. Amsterdam.
Varma, A., and B. O. Palsson. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* 60: 3724-3731.