Identification of Localized and Distributed Bottlenecks in Metabolic Pathways

Michael L. Mavrovouniotis

Institute for Systems Research and Department of Chemical Engineering
University of Maryland
College Park, Maryland 20742, U.S.A.

mlmavro@src.umd.edu Tel. (301) 405-6620 FAX (301) 405-6707

Abstract

The usual thermodynamic evaluation, based solely on the Standard Gibbs Energy of reaction, does not take into account the permissible ranges of concentrations of metabolites, and it faces further difficulties when, instead of isolated reactions, we are examining whole pathways. For pathways, we seek not only to decide *whether* they are feasible but also to pinpoint the pathway segment that causes any thermodynamic difficulties. We define a set of scaled quantities which reformulate the thermodynamic-feasibility problem for the whole pathway. We present an algorithm which analyzes individual reactions and selective construction of larger subpathways and uncovers localized and distributed thermodynamic bottlenecks of the biotransformation. This type of thermodynamic treatment contributes to the effort to include more physical, chemical, and biological factors in the computer-aided analysis of metabolic pathways.

1. Introduction

1.1. Computational Context

The construction of metabolic pathways that meet desired design specifications has been addressed through an algorithm performing sequential constraint-satisfaction (Mavrovouniotis, 1992, Mavrovouniotis *et al.*, 1990); the algorithm combines biochemical reactions into partial pathways, and then partial pathways with each other, until it reaches pathways which satisfy all the constraints imposed. This article addresses a thermodynamic and design problem which is opposite from, and complementary to, pathway synthesis: Given a metabolic pathway, test its thermodynamic feasibility, and, if it is infeasible, construct the subpathways which cause and embody the thermodynamic infeasibility. We call this kind of subpathway a thermodynamic bottleneck of the pathway, and it can be thought of as that subpathway which, if modified thermodynamically, would make the whole pathway feasible.

From the viewpoint of design of metabolic pathways, the work presented here refers to a qualitative **analysis** step which complements the **synthesis** step and moves us further towards true design - which must include both analysis and synthesis. From the computational viewpoint, the algorithm presented here shares some of its most important constituent subprocedures with the pathway-synthesis algorithm. This should not be surprising, since in both cases we are constructing subpathways; here, of course, we seek (thermodynamically) infeasible subpathways, rather than the (stoichiometrically) feasible ones we pursued for pathway synthesis.

1.2. Basic Notation

Consider a general biochemical reaction pathway, involving A metabolites, designated as $a_1, a_2, ..., a_A$, and S bioreactions (pathway steps), designated as $s_1, s_2, ..., s_S$. Each step s_i accomplishes a specific biotransformation, denoted as $r_i=R(s_i)$. Let α_{ij} represent the stoichiometric coefficient of metabolite a_j in step s_i, with the usual convention that $\alpha_{ij}>0$ iff a_j is a product of s_i, $\alpha_{ij}<0$ iff a_j is a reactant (substrate) of s_i, and $\alpha_{ij}=0$ iff a_j does not participate in s_i. The stoichiometry of the biotransformation $R(s_i)$ accomplished by step s_i can then be written as:

$$r_i=R(s_i)=\sum_{j=1}^{A}\alpha_{ij}a_j \qquad (1)$$

1.3. Thermodynamic Quantities

For the evaluation of the thermodynamic feasibility of the reaction r_i, one needs the Gibbs Energy of Reaction, ΔG_i, which denotes the Gibbs Energy gained when the extent of reaction r_i increases by 1 mole, i.e., when the reaction causes the amount of each metabolite a_j to change by α_{ij} moles. Generally, ΔG_i depends on the conditions of the reaction, i.e., reactant and product concentrations, temperature, pressure, pH, and concentrations of other solutes present in the aqueous solution. The metabolite concentrations always have a significant effect on ΔG, but in the preliminary evaluation of the thermodynamic

feasibility of a reaction the concentrations are not known. A quantity independent of the actual reaction conditions is the Standard Gibbs Energy of Reaction, $\Delta G_i^{\circ\prime}$, which is defined in a prespecified standard state (25°C, 1 atm, pH=7, concentration units 1M, dilute aqueous solution). This standard Gibbs Energy can be estimated, if necessary, through a group-contribution method (Mavrovouniotis, 1991). When all actual conditions, except the metabolite concentrations, are the same as the standard conditions, we can relate ΔG_i to $\Delta G_i^{\circ\prime}$ with the expression:

$$\Delta G_i = \Delta G_i^{\circ\prime} + \sum_{j=1}^{A} [\alpha_{ij} RT \ln(\phi_j C_j)] \quad (2)$$

where R is the ideal-gas constant, T is the temperature, ln is the natural logarithm, C_j is the concentration of a_j (expressed in M), and ϕ_j is the activity coefficient – which is equal to 1 for ideal systems but may take different values dependent on the composition of the system.

1.4. Traditional Thermodynamic Evaluation

To evaluate the thermodynamic feasibility of a metabolic pathway, for a given system with known concentrations C_j and activity coefficients ϕ_j, we simply calculate ΔG_i for each bioreaction. The step s_i is thermodynamically feasible in the forward direction iff $\Delta G_i<0$. The pathway is therefore feasible iff $\Delta G_i<0$ for all i. In infeasible pathways, it is likely that most reactions are feasible and only some particular isolated reactions are infeasible, posing a **thermodynamic bottleneck**. Note that in this analysis each bioreaction is considered separately and classified as feasible or infeasible. Thus, bottlenecks are always localized to single bioreactions.

However, the C_j and ϕ_j are usually not fully prespecified. A common practice in this case is to examine instead of ΔG_i. If the absolute magnitude of $\Delta G_i^{\circ\prime}$ is large, then a negative $\Delta G_i^{\circ\prime}$ means that the forward reaction is favored, while a positive $\Delta G_i^{\circ\prime}$ means the reaction is favored only in the reverse direction and forms a bottleneck. If the absolute value of $\Delta G_i^{\circ\prime}$ is small, e.g., less than 4 kJ/mol (~1 kcal/mol), the reaction can take place in either direction.

This approach is physically unrealistic because the standard conditions of $\Delta G^{\circ\prime}$ are arbitrary and do not conform to the particular system of interest, since $\Delta G_i^{\circ\prime}$ ignores the actual C_j and ϕ_j completely. The deficiencies are remedied in this paper by the definition of new thermodynamic quantities, which are used in algorithms that identify the localized or distributed thermodynamic bottlenecks of biochemical pathways.

1.5. Objective

The basic question that will be addressed here is the determination of whether a pathway is thermodynamically feasible, and, if it is not feasible, the determination of the pathway location where the infeasibility occurs. We assume that the available information includes a description of a metabolic pathway in the form of the stoichiometries α_{ij}; the thermodynamics of each bioreaction s_i, in the form of the standard Gibbs energy $\Delta G_i^{\circ\prime}$; and a partial description of the permissible concentrations of each metabolite a_j, in the form of an upper and a lower bound for $\phi_j C_j$.

Our approach will be based on a transformation of the thermodynamic quantities. We note that transformation of the Gibbs energy can be useful in using various types of information about the biochemical environment, such as the pH and pMg (Alberty, 1992a and 1992b, and references therein).

2. Transformation of the Problem

2.1. Bounds on Metabolite Concentrations

For each metabolite, we assume that we have an upper and lower bound either for the product $\phi_j C_j$ or for the quantities C_j and ϕ_j separately. We indicate these bounds with the superscripts ^{max} and ^{min}. The bounds are assumptions which define the spectrum of systems for which the analysis will be applicable. We will introduce parameters which will recast the thermodynamic requirements in a new form which takes into account the concentration limits of the metabolites.

First, we must accord special treatment to the set K of metabolites whose concentrations are fixed, i.e., ${\phi_j}^{max}{C_j}^{max} = {\phi_j}^{min}{C_j}^{min}$. These will commonly include $[H^{1+}]$ and $[OH^{1-}]$. For the purposes of thermodynamic analysis, these metabolites can be eliminated from the stoichiometry and all the algebraic expressions we will use, by rewriting ΔG_i in the form:

$$\Delta G_i = \Delta G_i^{\circ\prime} + \sum_{a_j \in K} [\alpha_{ij} RT \ln(\phi_j C_j)] + \sum_{a_j \notin K} [\alpha_{ij} RT \ln(\phi_j C_j)] \quad (3)$$

We then replace $\Delta G_i^{\circ\prime}$ by the first two terms in the right-hand side, and we redefine the stoichiometries and concentrations (and summations which involve them) to include only the species $a_j \notin K$. In the sequel, we will assume this step has been taken, so that we can assume that all (remaining) metabolites have concentrations varying within an interval of non-zero, finite span.

2.2. Scaling

For each metabolite a_j, we replace the concentration by a new scaled activity parameter f_j, defined as:

$$f_j = \left(\ln \frac{\phi_j C_j}{{\phi_j}^{min}{C_j}^{min}} \right) \Big/ \left(\ln \frac{{\phi_j}^{max}{C_j}^{max}}{{\phi_j}^{min}{C_j}^{min}} \right) \quad (4)$$

This activity parameter expresses the relative location of

the $\phi_j C_j$ in the interval $(\phi_j^{min} C_j^{min}, \phi_j^{max} C_j^{max})$, using a log scale because of the form of Eq. (2). For each bioreaction s_i, we replace the standard Gibbs energy by the new parameter g_i, defined as:

$$g_i = \frac{\Delta G_i^{o'}}{RT} + \sum_{j=1}^{A} [\alpha_{ij} \ln(\phi_j^{min} C_j^{min})] \quad (5)$$

Aside from a scaling by RT, this is merely the Gibbs energy in another standard state: Each metabolite is assumed to have its minimum permissible concentration rather than a predetermined value (such as the 1 M used in $\Delta G_i^{o'}$).

Finally, we transform the stoichiometric coefficients α_{ij} as follows:

$$w_{ij} = \alpha_{ij} \ln \frac{\phi_j^{max} C_j^{max}}{\phi_j^{min} C_j^{min}} \quad (6)$$

The scaled stoichiometries w_{ij} have the same signs as α_{ij} and they therefore preserve connotations such as $w_{ij}<0$ iff metabolite a_j serves as a substrate for step s_i. Furthermore, because the factor scaling α_{ij} to produce w_{ij} in Eq. (5) did not depend on i, if we choose to combine two reactions we can combine their w_{ij} the same way we would combine α_{ij}. In effect, w_{ij} has all the properties of reaction stoichiometries and can be viewed as nothing more than a redefinition of the mole unit of each compound. From the definitions of f_j and w_{ij}, note that:

$$\alpha_{ij} \ln \frac{\phi_j C_j}{\phi_j^{min} C_j^{min}} = w_{ij} f_j \quad (7)$$

Given the standard Gibbs energies and the stoichiometries of all steps as well as the concentration bounds of all metabolites, the parameters g_i and w_{ij} can be uniquely computed. The scaled activity parameters f_j cannot be computed, but they are constrained by the permissible concentration values:

$$0 \le f_j \le 1 \quad (8)$$

2.3. Conditions in the Scaled Parameters

With these transformed variables, the condition that a pathway-step s_i is thermodynamically feasible (in the forward direction) iff the Gibbs energy of the reaction r_i is negative is equivalent to:

$$\frac{\Delta G_i^{o'}}{RT} + \sum_{j=1}^{A} [\alpha_{ij} \ln(\phi_j^{min} C_j^{min})] + \sum_{j=1}^{A} [\alpha_{ij} \ln \frac{\phi_j C_j}{\phi_j^{min} C_j^{min}}] < 0 \quad (9)$$

The first two terms of the above sum are the definition of g_i, while each term under the rightmost summation is equal to $w_{ij} f_j$, as was remarked earlier. Therefore, the thermodynamic condition becomes:

$$g_i + \sum_{j=1}^{A} w_{ij} f_j < 0 \quad (10)$$

Taking g and f to be column vectors of dimensions S and A respectively, and W a matrix of dimensions S×A with entries w_{ij}, we can write the condition (10) for all i in matrix form:

$$g + W f < 0, \quad f \ge 0, \quad f \le b \quad (11)$$

where $b=[1\ 1\ \ldots\ 1]^T$ is a column vector of dimension A. With W_i denoting row i of the matrix W, we define the function H for an isolated reaction r_i:

$$H(r_i, f) = g_i + W_i f = g_i + \sum_{j=1}^{A} w_{ij} f_j \quad (12)$$

We can then write the separate condition for the feasibility of each r_i:

$$H(r_i, f) < 0, \quad f \ge 0, \quad f \le b \quad (13)$$

2.4. Feasibility of a Single Reaction

The thermodynamic evaluation of a single reaction entails using the selected concentration bounds to compute the upper and lower bound of the Gibbs energy. The signs of these bounds reveal quantitatively the feasibility and reversibility of the reaction. The analysis can be done either directly with the unscaled parameters (Eq. 2), or after various stages of scaling.

While the scaling does not affect the sign of the Gibbs energy (and hence the signs of its upper and lower bounds), it does affect the interpretation of the **sensitivity** of the result, which is dependent on the magnitudes of the bounds. To obtain a useful scaled result, and to set the stage for the analysis of whole pathways, let $H_{max}(r_i)$ and $H_{min}(r_i)$ be defined, for any given reaction r_i, as the constrained optima:

$$H_{max}(r_i) = \max_f H(r_i, f), \text{ subject to } 0 \le f \le b \quad (14)$$

$$H_{min}(r_i) = \min_f H(r_i, f), \text{ subject to } 0 \le f \le b \quad (15)$$

If we are viewing one reaction in isolation, each variable f_j can vary between 0 and 1, completely independently from the others. Therefore, the maximum value of H is obtained by selecting $f_j=1$ whenever $w_{ij}>0$, and $f_j=0$ whenever $w_{ij} \le 0$ (the value of f_j is, of course, irrelevant when $w_{ij}=0$). Similarly, the minimum value of H can be obtained by selecting $f_j=1$ whenever $w_{ij}<0$, and $f_j=0$ whenever $w_{ij} \ge 0$. Thus:

$$H_{max}(r_i) = g_i + \sum_{j\ (w_{ij}>0)} w_{ij} \quad \text{and} \quad H_{min}(r_i) = g_i + \sum_{j\ (w_{ij}<0)} w_{ij} \quad (16)$$

A very simple criterion for the thermodynamic feasibility of an isolated metabolic step s_i is:

$$H_{min}(r_i) < 0 \quad (17)$$

The quantity $H_{min}(r_i)$ represents the scaled Gibbs energy when all concentrations take their most favorable values (for s_j). The scaling permits $H_{min}(r_i)$ to be interpreted as the most favorable distance from equilibrium that the reaction can attain. $H_{max}(r_i)$, on the other hand is the least favorable distance; if it is also negative, the reaction is thermodynamically feasible (and irreversible) for all concentrations in the specified bounds.

Note, however, that for a whole set of bioreactions to be feasible simultaneously (i.e., in the same system) the conditions $H(r_i,f)<0$ must be satisfied simultaneously using the same vector f. Therefore, testing the $H_{min}(r_i)$ of each reaction is a necessary but not sufficient condition.

2.5. Thermodynamic Bottlenecks

We can express the thermodynamic conditions in terms of reaction combinations as follows. Since the function H applies to isolated bioreactions, we also let it apply to a linear combination of reactions, er, where e is a **row** vector of dimension S, containing the combination coefficients used in forming the composite reaction.

$$H(er,f) = eg+eWf \quad (18)$$

where eW equals the stoichiometry of the reaction er. We will permit only non-negative entries in this combination vector e, because a negative entry has the effect of reversing a reaction; we will also exclude the trivial zero vector. These restrictions on e will apply throughout this paper, even when they are not explicitly mentioned.

$$e \geq 0 \text{ and } e \neq 0 \quad (19)$$

The sign of H(er,f) shows whether the linear combination of reactions er is feasible **when viewed as an overall transformation.** If $\forall s_j\ H(r_i,f)>0$, given that $c\geq 0$, we can clearly conclude that $H(er,f)>0$, since by Eq. (16) e cannot be the trivial zero vector; however the converse is **not** true.

We now return to the question of thermodynamic bottlenecks, which we seek whenever a pathway is thermodynamically infeasible. In order to define bottlenecks, we require that the bounds on f always be satisfied and considering subsets of the set of inequalities $H(r_i,f)<0$. If the pathway as a whole is not feasible, then some of these subsets will be feasible (i.e., will be satisfiable by some f) and some will not. A set of steps B forms an infeasible subpathway iff for any f in the range $0\leq f\leq b$ at least one step in B is not feasible:

$$\forall f\ \ni\ 0\leq f\leq b:\ \exists s_i\in B\ \ni\ H(r_i,f)\geq 0 \quad (20)$$

If a subpathway contains (as a subset) an infeasible subpathway, then it is also infeasible. A **bottleneck** is a **minimal** infeasible subpathway, i.e., a set B which is infeasible and does not have any infeasible proper subsets. Thus, bottlenecks are defined by minimality (with respect to inclusion) in the class of infeasible subpathways. A bottleneck B is **localized** if it is a singleton set, and **distributed** if it involves two or more steps. The reason that distributed bottlenecks exist is that a number of inequalities of the form (1) may be unsatisfiable as a set, even though each inequality is separately satisfiable: The sets of vectors f that satisfy the inequalities separately are disjoint.

An infeasible pathway can have several bottlenecks, with different numbers of reactions, because a bottleneck is minimal only with respect to inclusion - not cardinality.

2.6. Conversion to the Dual Problem

Since for feasibility we require $H(r_i,f)$ to be negative for every i (with the same vector f), we can recast the condition equivalently as follows. There must exist a vector f ($0\leq f\leq b$) such that for all reaction combinations er (subject to Eq. 19), we have $H(er,f)<0$. To show that this follows from the condition (13) we simply multiply each $H(r_i,f)$ by e_i and sum up the results, noting that H(er,f) is a linear function with respect to e. Conversely, to show that the new condition leads back to (13), we simply use as e each of the unit-vectors $\hat{e}_i$ of the S-dimensional space of row vectors, $\hat{e}_1$=[1 0 0 0...], $\hat{e}_2$=[0 1 0 0 ...], $\hat{e}_3$=[0 0 1 0 ...], etc. With this explanation we write our reformulated condition mathematically, as:

$$\exists\ f\ (0\leq f\leq b):\ \forall e\geq 0\ \ H(er,f)<0 \quad (21)$$

The algorithm for the construction of the bottlenecks is based on the conversion of this condition to its equivalent dual form:

$$\forall e\geq 0\ \ \exists f\ (0\leq f\leq b):\ H(er,f)<0 \quad (22)$$

i.e., any reaction vector er in the convex cone spanned by the reactions must lead to a feasible overall transformation. We hasten to add that reversing the order of the two quantifiers is a significant transformation. To find an f such that all of the combinations er are feasible (with that fixed f) is conceptually a different operation from making each combination feasible by finding an f specific to that combination. The values of H(er,f) resulting from the two forms of the condition, for a given e, will not necessarily be the same. However, it can be shown through the theory of linear inequalities that the two conditions are equivalent.

Given any $e\geq 0$, we let v_j represent the stoichiometric coefficient of species a_j in the reaction er. Clearly:

$$v=eW \text{ and } H(er,f)=eg+vf \quad (23)$$

The definitions of H_{max} and H_{min}, applied to a combined reaction er are the same as those for an isolated reaction r_i. We have:

$$H_{max}(er)=\max_f H(er,f)=eg+\max_f (vf), \text{ subject to } 0\leq f\leq b \quad (24)$$

$$H_{min}(er)=\min_f H(er,f)=eg+\min_f (vf), \text{ subject to } 0\leq f\leq b \quad (25)$$

Again, for a fixed reaction er, the variables f_j can vary

independently between 0 and 1. Therefore, the value of H(er,f) is obtained by selecting $f_j=1$ whenever $v_j>0$, and $f_j=0$ whenever $v_j\leq 0$; the minimum of H(er,f) is obtained by selecting $f_j=1$ whenever $v_j<0$, and $f_j=0$ whenever $v_j\geq 0$. Thus:

$$H_{max}(er)=eg+\sum_{v_j>0} v_j \text{ and } H_{min}(er)=eg+\sum_{v_j<0} v_j \quad (26)$$

Using the definition of H_{min}, we can rewrite the dual form of the necessary and sufficient requirement for pathway feasibility as:

$$\forall e\geq 0: H_{min}(er)<0 \quad (27)$$

A pathway whose reaction vector is r is **not** feasible iff:

$$\exists e\geq 0: H_{min}(er)>0 \quad (28)$$

A pathway is irreversible (and feasible) over the entire space of permissible concentration iff:

$$\forall e\geq 0: H_{max}(er)<0 \quad (29)$$

2.7. Bottlenecks in the Dual Form

We now return to the definition of thermodynamic bottlenecks, which occur only if a pathway is thermodynamically infeasible. In the dual form, an infeasible subpathway is formed by the steps in a set B iff there is a combination e which is infeasible and contains only reactions from B; mathematically:

$$\exists e\geq 0: (s_i\notin B: e_i=0) \text{ and } H_{min}(er)>0 \quad (30)$$

B is a bottleneck if it is a minimal infeasible subpathway, i.e., there is no infeasible subpathway using a subset of the steps used by B.

It is convenient to declare the vector e itself as the bottleneck, rather than the associated set B, and we will do so in the rest of this paper. The step s_i of a pathway forms a localized thermodynamic bottleneck $\hat{e}_i$ iff $H_{max}(r_i)<0$ - in effect, the reaction is infeasible on its own. A distributed bottleneck occurs when e has more than one non-zero components.

3. Construction of Bottlenecks

3.1. Selection of Constraining Subpathways

The problem of deciding the feasibility of a pathway was reduced to examining all linear combinations of reactions, which is clearly impractical. In the proposed method, the active set of subpathways includes only pathways which are feasible. These are successively combined in pairs to produce an infeasibility. It turns out that most combinations can be ruled out of potential bottlenecks, based on the additivity properties of H. Suppose that for some e we have $H_{max}(er)<0$ - therefore for all acceptable f we must have:

$$H(er,f)<0 \quad (31)$$

Let also e' be a thermodynamically feasible combination, i.e., $H_{min}(er)<0$. If we form e"=e'+e, can this combination be a bottleneck?

We have, from the linearity of H:

$$H(e"r,f)=H(er,f)+H(e'r,f) \quad (32)$$

and using Eq. (31) we conclude:

$$H(e"r,f)<H(e'r,f) \Rightarrow H_{min}(e"r)<H_{min}(e'r) \quad (33)$$

Therefore, the combination of a subpathway for which $H_{max}(er)<0$ and a subpathway which is feasible cannot lead to an infeasibility. The consequence is that we can eliminate from the active set combinations that have $H_{max}(er)<0$ (which actually involve irreversible transformations).

To focus on a second important consideration, suppose that two reactions, r_1 and r_2, are linearly combined, with coefficients $t_1>0$ and $t_2>0$, into an aggregate reaction, $r_3=t_1r_1+t_2r_2$. Letting β_{ij} represent the scaled coefficient of metabolite a_j in reaction r_i, the coefficients of the three reactions are interrelated as follows: $\beta_{3j}=t_1\beta_{1j}+t_2\beta_{2j}$. Their scaled Gibbs parameters g_i are likewise interrelated as $g_3=t_1g_1+t_2g_2$. Therefore, through the definition of H, we have $H(r_3,f)=t_1H(r_1,f)+t_2H(r_2,f)$. The values of $H_{min}(r_3)$ and $H_{max}(r_3)$ depend on the choice of favorable values (0 or 1) for each f_j. The H_{min} and H_{max} of r_1 and r_2 will combine linearly (just like H), unless the reactions r_1 and r_2 share a metabolite in a way that contradictory choices for some f_J have to be made, i.e., unless some metabolite a_J is a product of one reaction and a reactant of the other. If this kind of shared metabolite exists, then $H_{min}(r_3)$ will be higher than the linear combination $t_1H_{min}(r_1)+t_2H_{min}(r_2)$, and $H_{max}(r_3)$ will be lower than $t_1H_{max}(r_1)+t_2H_{max}(r_2)$. Furthermore, the discrepancy will be most drastic if the coefficients t_1 and t_2 are selected so that $\beta_{3J}=0$ for the shared metabolite a_J; this choice of coefficients is actually $t_1=|\beta_{2j}|$ and $t_2=|\beta_{1j}|$.

3.2. Algorithm

The algorithm for the construction of the bottlenecks is based on the dual form of the criteria and the observations in the previous section. We maintain all the candidate bottlenecks in a list L, and the proven bottlenecks in a list L_B, and we introduce tests to avoid duplication of entries in these two lists. We represent each bottleneck as a vector e, whose overall reaction is er=R(es). The initial list of candidate bottlenecks consists of all individual reactions ($\hat{e}_i$).

The algorithm's operation is described in Figure 1, inthe simplest possible form of the algorithm, without worrying about efficiency-enhancing details of the implementation. For example, in the computation of H_{max} and H_{min} for all reactions one can postulate a number of improvements: The parameters can be

computed when a new combination e is constructed, from the corresponding parameters of the components used in the combination. The computation of H_{max} can be avoided if H_{min} turns out to be positive; or the computation of H_{min} can be avoided if H_{max} turns out to be negative; or we can compute both simultaneously. We are also not addressing the important issue of organizing the search (from its current breadth-first mode) to a more efficient arrangement in which the most promising leads are explored first.

```
Initialize L_B := ∅ and L:= {ê_i, i=1,2,...,S}
Repeat until L=∅:
  For each e∈L:
    Compute H_max and H_min
    If H_min(er)>0 then:
      Set L:=L-{e}
      If e is not a subpathway of a pathway in L_B then:
        Set L_B:=L_B∪{e}
    Else if H_max(er)<0, set L:=L-{e}
  Let L":=∅
  For each e∈L determine the corresponding v=eW
  For each j=1,...,A
    For each e∈L which has a coefficient v_j>0 do:
      For each e'∈L such that its v'_j<0 do:
        Let e"=(v_j e'-v'_j e)/(v_j-v'_j)
        If e" is not a subpathway of a pathway in L" then:
          Set L":=L"∪{e"}
  Set L:=L"
Output L_B (minimal bottlenecks)
```

Figure 1: The operation of the algorithm for the identification of bottlenecks.

4. An Example

We will analyze the glycolysis pathway (Figure 2) to illustrate the construction of localized and distributed thermodynamic bottlenecks. Table 2 shows the bioreactions and standard Gibbs energies, as given by Lehninger (1982).

We note that the traditional way of assessing thermodynamic feasibility, through the signs of the standard Gibbs energies in Table 1, would be quite misleading here: If all the positive $\Delta G^{\circ\prime}$ where interpreted as infeasibilities, then this central pathway would be ruled infeasible!

We assume, for simplicity, that T=298.15K (25°C), pH=7, and ϕ_j=1 for all j, and we focus on the concentration bounds. For the energy-currency metabolites (ADP, ATP, and Pi) we will assume constant concentrations and use the values given by Lehninger (1982, p. 373), for *E.Coli*: 0.82mM (AMP), 1.04mM (ADP), 7.9mM (ATP), 7.9mM (Pi). Lehninger's values correspond to an Energy Charge of approximately 0.87, which is consistent with the range stated by Ingraham *et al.* (1983, p. 166). For NAD and NADH, we use the range 0.03-0.07 given by Ingraham *et al.* (1983, p. 168) for the Catabolic Reduction Charge of growing *E.Coli*, and we assume the concentration of NAD to be approximately constant at 4mM. This gives the concentration of NADH in the range 0.12mM-0.28mM.

We will assume that the remaining metabolites share the same concentration range, $C_{min}<C_j<C_{max}$; we will examine a number of possibilities for this concentration range.

First Interval: $C_{min}=0.1\times10^{-3}$ **M**, $C_{max}=1\times10^{-3}$ **M.** The computations are shown in Table 2. Applying the algorithm, we first assume each individual bioreaction to be a candidate bottleneck; in other words, the subpathways that are potential bottlenecks are simply the unit vectors $\hat{e}_i$. We compute the $H_{min}(\hat{e}_i r)=H_{min}(r_i)$ and $H_{max}(\hat{e}_i r)=H_{max}(r_i)$. We then reject all subpathways that have $H_{max}(\hat{e}_i r)<0$. This immediately deletes subpathways 1, 3, 7, 10, 11. Physically, each of these bioreactions is always feasible and irreversible within the assumed concentration bounds; thus, not only is it not a bottleneck by itself, but it could also never be part of a distributed bottleneck.

In the same round, we identify as bottlenecks all those subpathways for which $H_{min}(\hat{e}_i r)>0$. This shows bioreactions 5 and 6 to be localized bottlenecks. Continuing, we examine the remaining subpathways (2, 4, 8, and 9) for which $H_{min}(\hat{e}_i r)<0<H_{max}(\hat{e}_i r)$, and we construct all combinations of two subpathways (from this set) such that one intermediate is eliminated. Only one such combination is possible, $\hat{e}_8+\hat{e}_9$. This completes one iteration of the algorithm; at this point, $L_B=\{\hat{e}_5, \hat{e}_6\}$ and $L=\{\hat{e}_8+\hat{e}_9\}$.

In the next iteration, we compute $g_8+g_9=2.533$ and $H_{min}(r_8+r_9)=0.23>0$. The combination r_8+r_9 is therefore a distributed bottleneck. There is for each of r_8 and r_9 a portion of the permissible concentration space that leads to a negative ΔG and make the reaction feasible. However, there is no overlap between these two portions of the concentration space. Our formulation addresses this by showing that the overall reaction r_8+r_9 (which is 3PG → PEP) is never feasible in the given concentration space.

The final result is $L_B=\{\hat{e}_5, \hat{e}_6, \hat{e}_8+\hat{e}_9\}$. Note that r_4, which had the largest (most unfavorable) $\Delta G^{\circ\prime}$, actually does not occur in a bottleneck at all!

Second Interval: $C_{min}=0.1\times10^{-3}$ **M**, $C_{max}=2\times10^{-3}$ **M.** With C_{max} relaxed to 2×10^{-3} M

(the new parameters shown in Table 2) the application of the algorithm follows the same route as before, and bioreactions 5 and 6 clearly remain localized bottlenecks. The only difference arises when we compute, for the combination r8+r9: $g_8+g_9=2.533$ and $H_{min}((\hat{e}_8+\hat{e}_9)r)=-0.463<0$. Thus, the subpathway $\hat{e}_8+\hat{e}_9$ cannot be classified as a bottleneck, but neither can it be rejected right away, because $H_{max}((\hat{e}_8+\hat{e}_9)r)=5.529>0$. We must therefore proceed to the algorithm phase that constructs combinations. However, no combinations can be constructed, and therefore the final result involves just two localized bottlenecks: $L_B=\{\hat{e}_5, \hat{e}_6\}$.

Third Interval: $C_{min}=0.02\times10^{-3}$ **M**, $C_{max}=4\times10^{-3}$ **M.** Relaxing both concentration bounds further we see that there are no longer any individual reactions with $H_{min}(r_i)>0$ (Table 2). Therefore, in the first iteration, no localized bottlenecks are uncovered, but $L=\{\hat{e}_2, \hat{e}_4, \hat{e}_5, \hat{e}_6, \hat{e}_8, \hat{e}_9\}$ remains for possible distributed bottlenecks. The combinations that are constructed are: $L''=\{(\hat{e}_8+\hat{e}_9), (\hat{e}_4+\hat{e}_5), (\hat{e}_4+\hat{e}_6), (\hat{e}_5+\hat{e}_6)\}$. For these can then compute:
$g_8+g_9=2.533$,
$H_{min}((\hat{e}_8+\hat{e}_9)r)=-2.765<0$, $H_{max}((\hat{e}_8+\hat{e}_9)r)=7.831>0$;
$g_4+g_5=1.944$,
$H_{min}((\hat{e}_4+\hat{e}_5)r)=-3.354<0$, $H_{max}((\hat{e}_4+\hat{e}_5)r)=12.54>0$;
$g_4+g_6=2.721$,
$H_{min}((\hat{e}_4+\hat{e}_6)r)=-2.577<0$, $H_{max}((\hat{e}_4+\hat{e}_6)r)=14.164>0$;
$g_5+g_6=6.957$, $H_{min}((\hat{e}_5+\hat{e}_6)r)=1.659>0$.

Thus, $\hat{e}_5+\hat{e}_6$ is a distributed bottleneck, while $\hat{e}_8+\hat{e}_9$, $\hat{e}_4+\hat{e}_5$, and $\hat{e}_4+\hat{e}_6$ must be examined further. But since no new combinations can be created from $\hat{e}_8+\hat{e}_9$ and $\hat{e}_4+\hat{e}_5$ they are rejected; note that although it is possible to combine $\hat{e}_4+\hat{e}_5$ and $\hat{e}_4+\hat{e}_6$ this combination already contains a bottleneck ($\hat{e}_5+\hat{e}_6$) and is therefore not new.

The relaxation of the concentration interval eliminated r_5 and r_6 as localized bottlenecks, but it gave rise to their combination, r_5+r_6, as a distributed bottleneck.

Fourth Interval: $C_{min}=0.004\times10^{-3}$ **M**, $C_{max}=5\times10^{-3}$ **M.** With these bounds, the three-reaction combination $\hat{e}_4+\hat{e}_5+2\hat{e}_6$ is found to be the only bottleneck:
$g_4+g_5+2g_6=8.069$, $H_{min}((\hat{c}_4+\hat{c}_5+2\hat{e}_6)r)=0.938>0$

This is the first time that r_4 appears in a bottleneck! The analysis suggests that the composite transformation $r_4+r_5+2r_6$, which yields FruDP+2 NAD+2 Pi → 2 PGP+2 NADH, is the most difficult thermodynamically. In order to take place, it requires FruDP to have the maximum and PGP the minimum possible concentrations. Furthermore, it points to the crucial role of the Catabolic Reduction Charge, which is the ratio NADH / (NAD+NADH).

Fifth Interval: $C_{min}=0.0025\times10^{-3}$ **M**, $C_{max}=5\times10^{-3}$ **M.** The elimination of all distributed bottlenecks is accomplished if C_{min} is lowered to the value 0.0025×10^{-3} M (i.e., 2.5μM). In this case, we obtain:
$H_{min}((\hat{c}_4+\hat{c}_5+2\hat{c}_6)r)=-0.003<0$

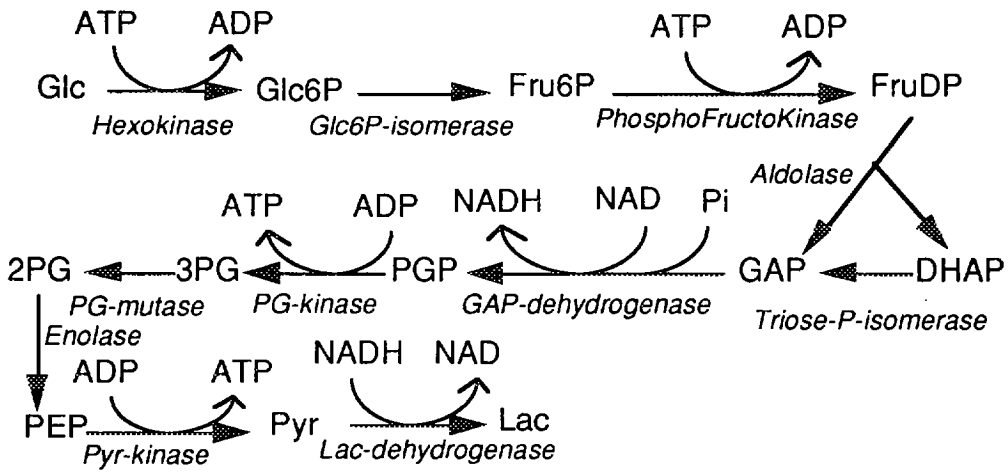

Figure 2: The glycolysis pathway.

Table 1: Indices, enzyme names, and stoichiometries for the biotransformations comprising glycolysis (also shown in Figure 2). Water, H^{1+} (or H_3O^{1+}) and OH^{1-} have been omitted from the stoichiometries. The biological standard Gibbs energies, as given by Lehninger (1982) are also shown.

Index	Stoichiometry	$\Delta G^{\circ\prime}$ (kcal/mol)	Name
1	Glc+ATP → Glc6P+ADP	- 4.00	Hexokinase
2	Glc6P → Fru6P	0.40	Glucose-6-phosphate Isomerase
3	Fru6P+ATP → FruDP+ADP	- 3.40	Phosphofructokinase
4	FruDP → DHAP+GAP	5.73	Fructose Diphosphate Aldolase
5	DHAP → GAP	1.83	Triose-phosphate Isomerase
6	GAP+NAD+Pi → PGP+NADH	1.50	Glyceraldehyde-phosphate Dehydrogenase
7	PGP+ADP → 3PG+ATP	- 4.50	Phosphoglycerate Kinase
8	3PG → 2PG	1.06	Phosphoglycerate Mutase
9	2PG → PEP	0.44	Enolase
10	PEP+ADP → Pyr+ATP	- 7.50	Pyruvate Kinase
11	Pyr+NADH → Lac+NAD	- 6.00	Lactate Dehydrogenase

Table 2: Computations of individual reaction parameters, for various bounds C_{min} and C_{max}.

	First Interval: $C_{min}=0.1\times10^{-3}$ M, $C_{max}=1\times10^{-3}$ M			**Second Interval:** $C_{min}=0.1\times10^{-3}$ M, $C_{max}=2\times10^{-3}$ M			**Third Interval:** $C_{min}=0.02\times10^{-3}$ M, $C_{max}=4\times10^{-3}$ M		
Index	g_i	$H_{min}(r_i)$	$H_{max}(r_i)$	g_i	$H_{min}(r_i)$	$H_{max}(r_i)$	g_i	$H_{min}(r_i)$	$H_{max}(r_i)$
1	-8.781	-11.083	-6.478	-8.781	-11.777	-5.785	-8.781	-14.079	-3.483
2	0.675	-1.627	2.978	0.675	-2.320	3.671	0.675	-4.623	5.974
3	-7.768	-10.070	-5.465	-7.768	-10.764	-4.772	-7.768	-13.066	-2.470
4	0.464	-1.839	5.069	0.464	-2.532	6.455	-1.146	-6.444	9.451
5	3.090	0.787	5.392	3.090	0.094	6.085	3.090	-2.209	8.388
6	3.867	1.564	7.017	3.867	0.871	7.710	3.867	-1.432	10.012
7	-5.570	-7.872	-3.267	-5.570	-8.565	-2.574	-5.570	-10.868	-0.271
8	1.790	-0.513	4.092	1.790	-1.206	4.785	1.790	-3.509	7.088
9	0.743	-1.560	3.045	0.743	-2.253	3.739	0.743	-4.555	6.041
10	-10.635	-12.937	-8.332	-10.635	-13.630	-7.639	-10.635	-15.933	-5.336
11	-6.623	-9.773	-4.321	-6.623	-10.466	-3.628	-6.623	-12.769	-1.325

5. Concluding Remarks

The methodology presented in this paper clarifies the thermodynamic concepts relevant to the feasibility and reversibility of biochemical reactions. Furthermore, it allows the analysis of biochemical pathways, identifying both localized and distributed bottlenecks. The analysis can be used to reject proposed pathways for the production of a bioproduct, when they are shown to be infeasible. It can also be used to compare alternative pathways, to select the one that presents fewer or less serious bottlenecks. Identification of the bottlenecks may also be the first step in a search for improvements that would bypass the thermodynamic difficulties.

Acknowledgments

This work was supported in part by the National Institutes of Health (National Library of Medicine grant 5 R29 LM 05278-02), and by the Institute for Systems Research (NSF grant CDR-8803012). The author is grateful to Gregory Stephanopoulos (M.I.T.) and Michael Liebman (Amoco Technology Co.) for their comments and encouragement.

References

Alberty, R.A. Equilibrium Calculations on Systems of Biochemical Reactions at Specified pH and pMg. *Biophysical Chemistry* **42**, 117-131, 1992a.

Alberty, R.A. Calculation of Transformed Thermodynamic Properties of Biochemical Reactants at Specified pH and pMg. *Biophysical Chemistry* **43**, 239-254, 1992b.

Ingraham, J.L., Maaløe, O., and Neidhardt, F.C. *Growth of the Bacterial Cell.* Sinauer Associates, Sunderland, Mass., 1983.

Lehninger, A.E. *Biochemistry*, 2nd ed. Worth Publishers, New York, 1975.

Mavrovouniotis, M. L., Stephanopoulos, G., and Stephanopoulos, G. Computer-Aided Synthesis of Biochemical Pathways. *Biotechnology and Bioengineering*, **36**, 1119-1132, 1990.

Mavrovouniotis, M. L. Estimation of Standard Gibbs Energy Changes of Biotransformations. *Journal of Biological Chemistry*, **266**, 14440-14445, 1991.

Mavrovouniotis, M. L. Identification of Qualitatively Feasible Metabolic Pathways. In *Artificial Intelligence and Molecular Biology*, L. Hunter (editor), AAAI Press / MIT Press, in press, 1993.