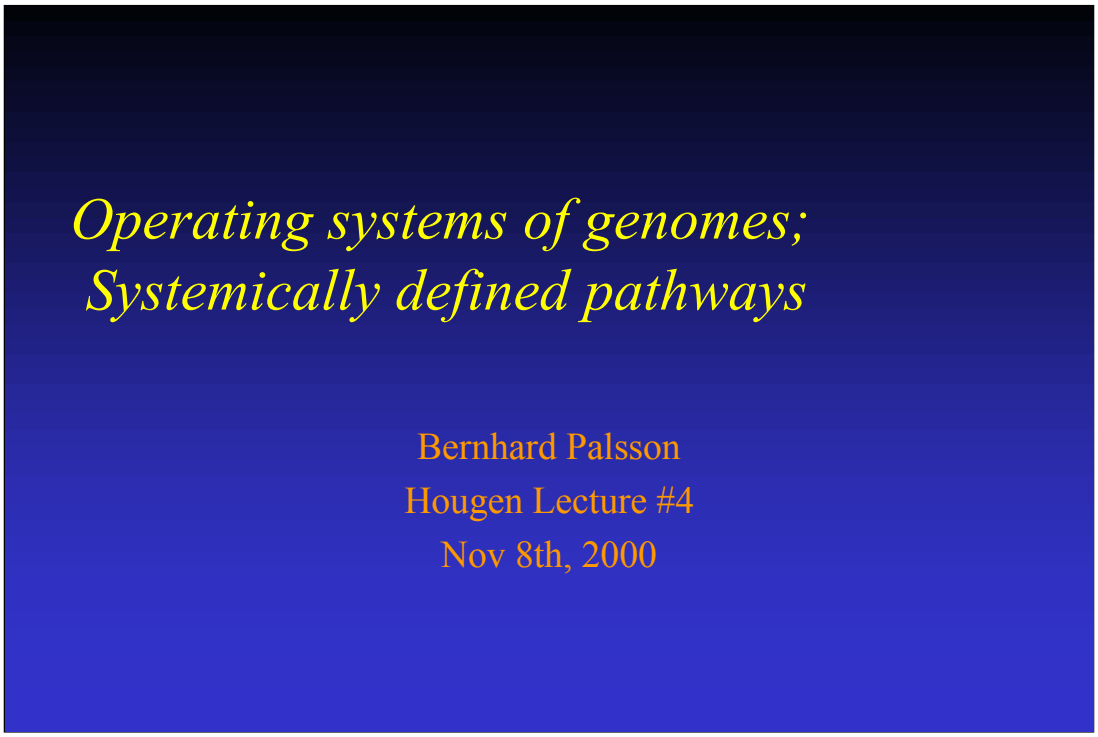
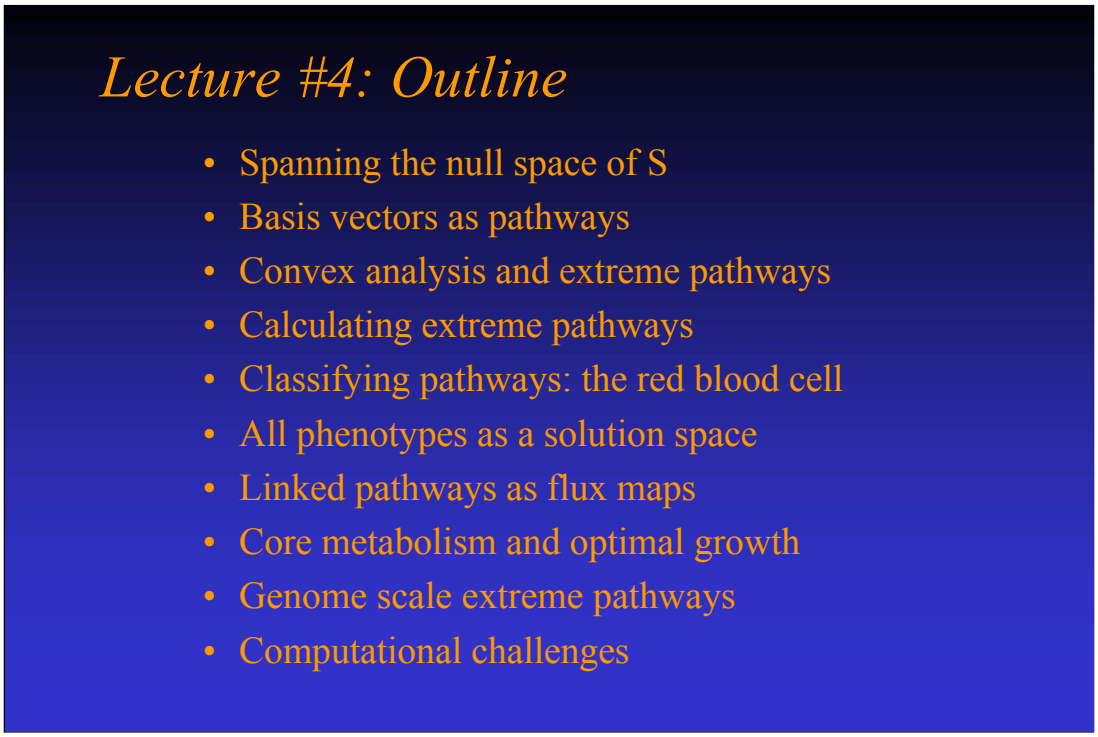


INTRODUCTION

In the previous three lectures we surveyed the world of genomics, how this information is giving us the biochemical reaction networks that operate in cells, and how we can approach the mathematical modeling of these networks and their simulation in a computer.

We now begin the mathematical modeling process in earnest and analyze the consequences of connectivity constraints and thermodynamics, i.e. the irreversibility of some reactions



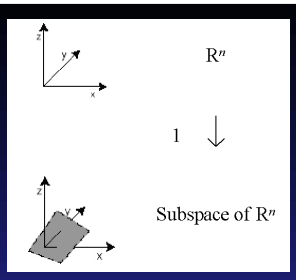
LECTURE #4

This lecture will cover the definition and use of systemic pathways. We begin with the vectors that span the null space of S. We show that these are biochemical pathways. However the basis for linear spaces are not unique but by imposing irreversibility constraints of thermodynamics we leave the domain of linear analysis and enter that of convex analysis. Now the solution space is conical in shape and the edges of the cone become the spanning vectors. These vectors are unique and are the 'extreme pathways.'

We then cover the algorithm that is used to calculate these extreme pathways, and compute them for red cell metabolism and investigate their biochemical significance. We then introduce linked outputs that respond to physiological functions such as growth and show that the extreme pathways are now metabolic maps.

We end the lecture by discussing some of the computational challenges associated with calculating these maps on a genome wide scale.

The Null Space of the Stoichiometric Matrix



- Contains all the solutions to **Sv=0**
- These are the steady state solutions to the dynamic mass balances
- The time constants of metabolic transients are typically very fast, i.e. shorter than about 1 to 5 minutes, especially in bacteria
- Thus for most practical purposes metabolism is in a steady state
- The null space contains all the steady state flux distributions and is thus of special importance to us
- The dimension of the null space is the number of columns in the matrix minus the number of independent rows (the rank of the matrix)

METABOLIC TRANSIENTS AND THE NULL SPACE

The concentrations of metabolites tend to be very low, in the order of miro-molar, or about 60,000 molecules per *E. coli* cell. Yet the metabolic fluxes are about 100,000 molecules per sec per cell. Thus, the average response time of a metabolic concentration is about 1 second. These transients are too fast for essentially all practical purposes.

Metabolites that are in higher concentrations, such as ATP, can have time constants that are on the order of minutes. Nevertheless, compared to the progression of an infection or bioprocessing this time is very short and metabolism is very fast and can be effectively considered to be in a steady state.

In some highly specialized mammalian cells, metabolic transients can be slower. For instance in the human red blood cell, the ATP turnover time is about a hour, and transient changes in 2,3DPG are on a 12 to 24 hr time scale. 2,3DPG binds to hemoglobin to regulate its affinity for oxygen. This time constant is responsible for the time that it takes us to adjust to higher altitudes.

Finding the basis for the null space

Any matrix A:

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

Can be row reduced using Gaussian elimination:

$$\begin{bmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{aligned} x_1 - 2x_2 - x_4 + 3x_5 &= 0 \\ x_3 + 2x_4 - 2x_5 &= 0 \\ 0 &= 0 \end{aligned}$$

FINDING A BASIS FOR THE NULL SPACE

A basis for the null space can be found by a so-called parameterization procedure. First any matrix A is row reduced by Gaussian elimination into the echelon form of the matrix (typically denoted by U). The pivot columns are identified (columns one and three in the example given). These are the columns with the pseudo-diagonal elements. These columns represent the fixed variables. The free variables are in the columns between the pivot columns (the second, fourth and fifth in the example given)

The matrix A has two pivot columns (1 and 3) and three free variables (2,4,5). All the variables can be expressed in terms of the free variables (parametric form)

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$

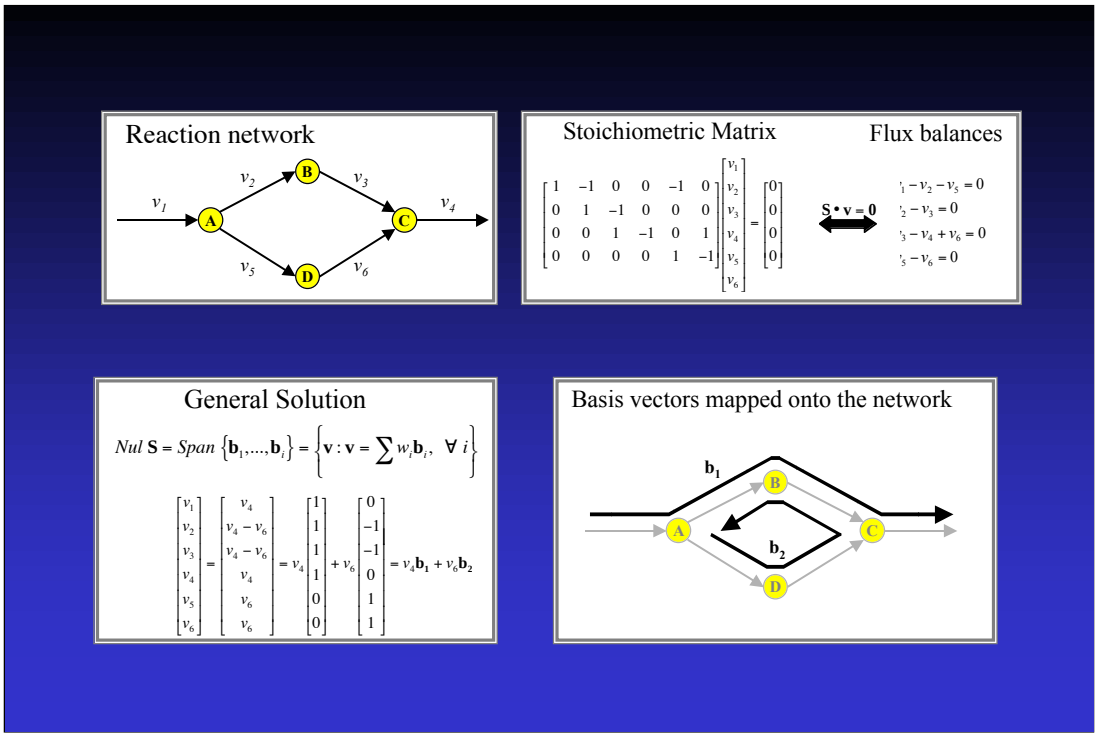
$$= x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w}$$

Then the vectors u, v, w span the 3-dimensional null space

FINDING A BASIS FOR THE NULL SPACE--CONTINUED

All the variables are then written in terms of the free variables. The equations are then written in vector form by factoring our the free variables individually. The columns that form constitute a spanning set--a basis--for the null space of **A**. The free variables can take on any numerical values to form additions of these basis vectors.

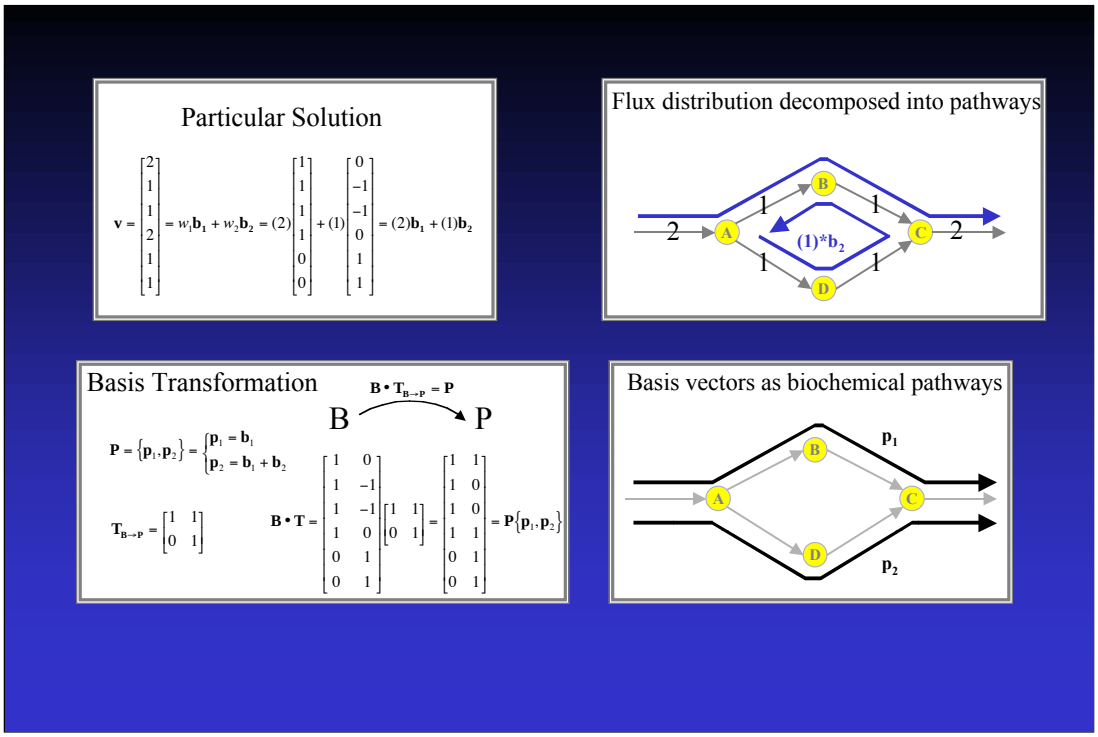
Verify that **Au=Av=Aw=0**, and that **u,v,w** is an independent set of vectors.



FINDING A BASIS FOR AN EXAMPLE STOICHIOMETRIC MATRIX

A simple reaction network is presented in the ULH panel. The corresponding stoichiometric matrix and its flux balances are written in the URH panel. The parameterization method, of the previous two slides, is applied to find a basis for this stoichiometric matrix, as shown in the LLH panel. These basis vectors can be graphically represented on the metabolic map (LRH Panel).

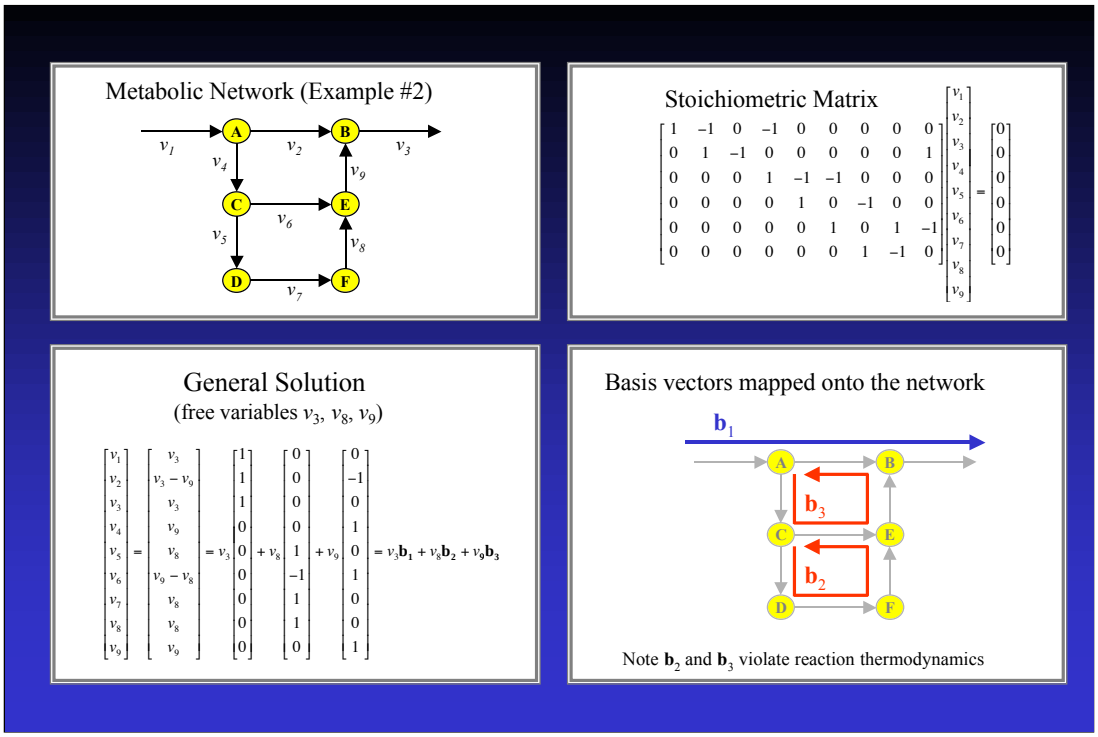
Note that the two basis vectors form a string of connected reactions--effectively pathways. The first basis vector, b_1, is a straight through pathway through the upper part of this small network. The second basis vector, b_2, is a circular path. It has steps in it that run opposite to the direction of two irreversible reactions. Although the basis vectors form mathematically acceptable pathway, biochemically they are not acceptable. However, as we saw previously, we can form another equivalent basis.



FINDING A BASIS FOR AN EXAMPLE STOICHIOMETRIC MATRIX--CONTINUED

•Every flux distribution, **v**, can be *uniquely* described by a combination of the particular set of basis vectors chosen to describe the null space. (Unique Representation Theorem). An example is given in the ULH panel and shown on the metabolic map in the URH panel

•A basis for a vector space imposes a coordinate system on the space. However, this coordinate system is not unique, which implies that other sets of vectors can be used as a basis for the same vector space. One basis can be transformed into another using a basis transformation, as shown in the LLH panel. We seek to find basis vectors whose elements are all positive. Such vectors will form biochemically acceptable pathways as shown in the LRH panel.



FINDING A BASIS FOR A STOICHIOMETRIC MATRIX

EXAMPLE #2

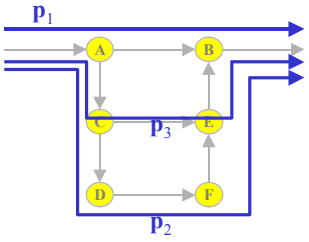
Basis Transformation (Example #2)

$$B \xrightarrow{B \bullet T_{B\to P} = P} P$$

$$P = \{p_1, p_2, p_3\} = \begin{cases} p_1 = b_1 \\ p_2 = b_1 + b_2 + b_3 \\ p_3 = b_1 + b_3 \end{cases} \qquad T_{B\to P} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$$B \bullet T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = P\{p_1, p_2, p_3\}$$

Basis vectors as biochemical pathways



All of the pathways now obey the thermodynamic constraints if they are positively weighted.

The selection of basis vectors is not unique. Therefore it is irrelevant that any flux distribution can be uniquely represented by a set of basis vectors. *We need to try and find a unique "basis" or set of pathways to describe the solution space.*

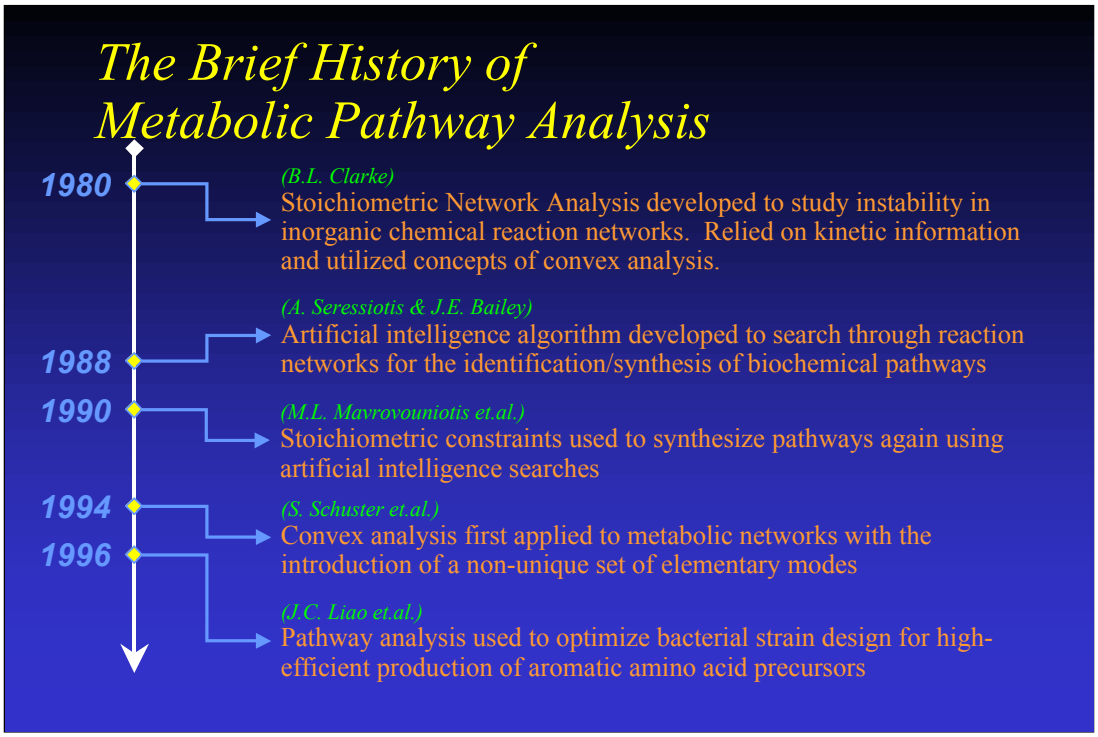
FINDING A BASIS FOR A STOICHIOMETRIC MATRIX
EXAMPLE #2--CONTINUED

The Null space of S

- The matrix has dimension of n columns, and m rows, representing the number of reactions and metabolites, respectively.
- Has dimensions of n-r (r is rank of S, r=m if matrix is full rank).
- Can be found by the parametric approach
- A linear basis is not unique
- An equivalent basis can be found by replacing a member of the spanning set with a linear combination of the other members of the set
- Basis for the null space may be found that contain only positive weights on the elements of v
- Such bases have members of the spanning set which are biochemically meaningful pathways

SOME FACTS ABOUT THE NULL SPACE OF S

We have now established that we can find vectors that span the null space of **S** that represent biochemically acceptable pathways. Note that these pathways are properties of the matrix itself, as they are the basis for one of its fundamental subspaces.



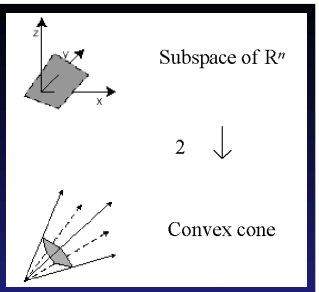
A BRIEF HISTORY OF THE FIELD OF PATHWAY ANALYSIS.

The first work on pathways can be traced back to 1980 with the development of SNA by Bruce Clarke. The theory was developed to study instability in inorganic chemical networks. This was the first attempt to apply convex analysis to reaction networks but was never extended to living systems. This was followed by some work using AI to search through reaction networks following along the lines of graph theory. This was taken another step by Mavro with the introduction of stoichiometric constraints. Both of these approaches lacked a sound theoretical basis.

In 1994 Schuster became the first to apply convex analysis to metabolic networks with the introduction of a non-unique set of elementary modes. This theory was applied a few years later by Liao to optimize bacterial strain design for the high-efficient production of aromatic amino acids.

So at this point in time pathway analysis is just beginning to be applied but still lacks a unified theoretical foundation, which is where the present work comes in.

Convex Analysis



- The study of systems of linear equations and inequalities

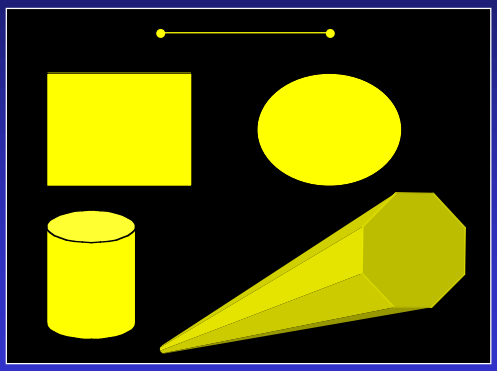
- Convex analysis is used to study metabolic networks where the linear equations are derived from the mass balances and the inequalities are generated from thermodynamic information on the reversibility of reactions.

- From linear algebra a null space is defined which contains all of the solutions to the set of linear homogenous equations. When we add inequality constraints (such as all variables must be positive) the solution space becomes restricted by these inequalities (the portion of the null space in the positive orthant)

What is Convexity?

Definition of a Convex Space: For every two points in the space, the line connecting these two points lies entirely within the space.

Convex Shapes



Non-Convex Shapes

Polyhedral Cones and Pathways

• Region determined by a linear homogeneous equation/inequality system is a convex polyhedral cone (***C***)

$$\mathbf{0} = \mathbf{S} \bullet \mathbf{v}, \qquad v_i \geq 0, \quad i = 1,\ldots,n$$

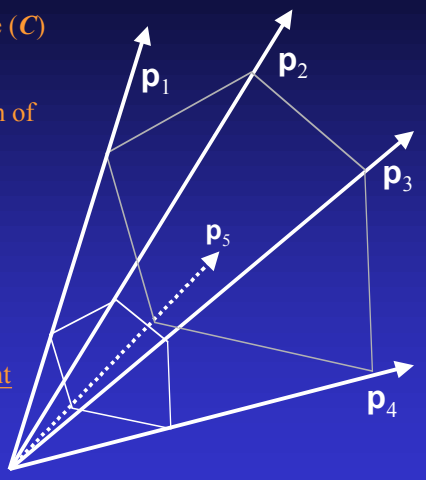
• Every point in the cone is a non-negative combination of the generating vectors (Extreme Pathways) of the cone

$$C = \left\{ \mathbf{v} \in \mathrm{R}^n \;\middle|\; \mathbf{v} = \sum_{i=1}^{k} \alpha_i \mathbf{p}_i, \quad \alpha_i \geq 0 \right\}$$

• The number of generating vectors can exceed the dimensions of the cone (i.e. linearly dependent)

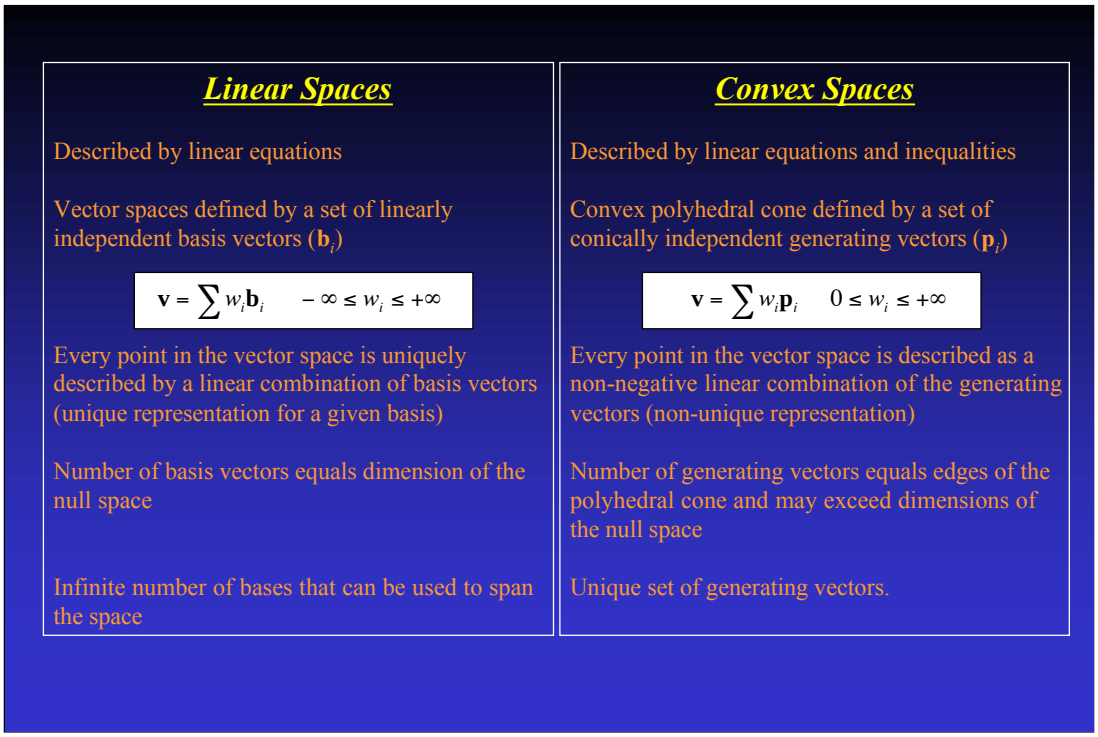
• Generating vectors represent systemically independent pathways which can theoretically be "switched" on or off

• Generating vectors are unique for a system



THE FLUX CONES

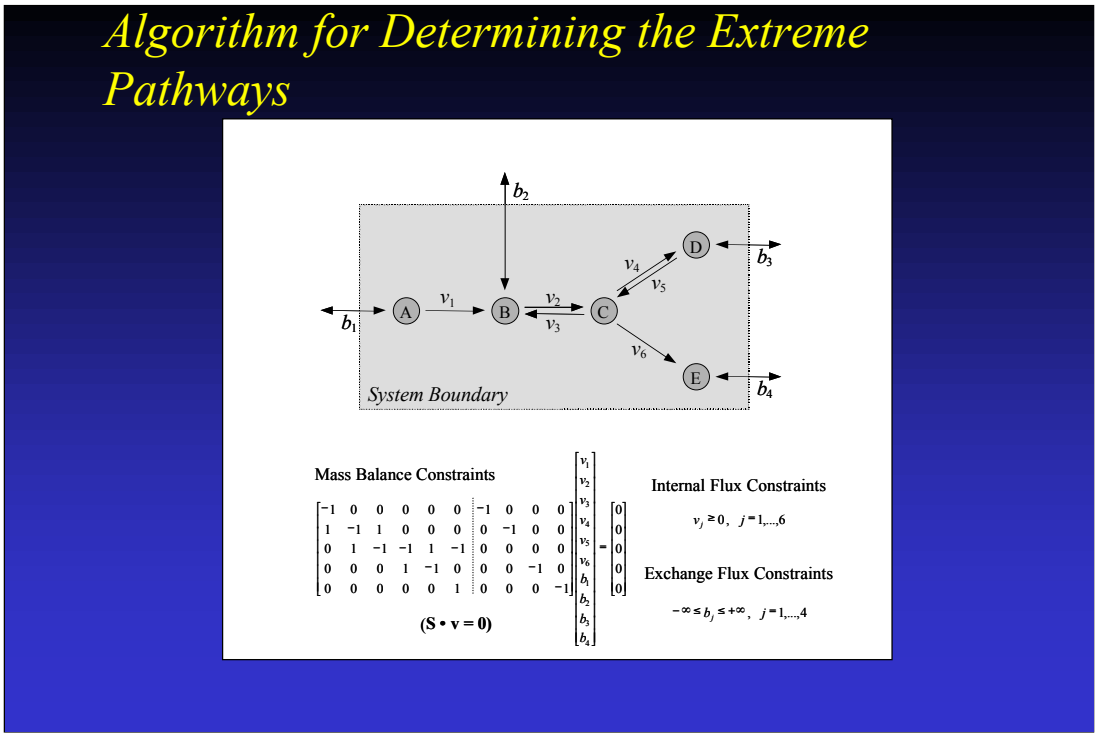
Through the principles of convex analysis it turns out that the shape of the null space for a set of linear equation with positive flux values such as the systems which we are concerned with is a convex polyhedral cone such as the one depicted here on the right. The perspective of the cone is supposed to look as if it is going into the plane of the slide. What is nice about cones is the condition that every point within the cone can be described as a non-negative combination of the generating vectors where the generating vectors are the edges of the cone. If we can determine these generating vectors which are biochemically feasible then we can describe every point within the cone. Additionally the number of generating vectors can exceed the dimensions of the cone which has the mathematical consequence that all of these pathways are not linearly independent. The best analogy for some of these concepts is to think of an Egyptian pyramid which has 4 edges and exists in three dimensional space. Algorithms exist for the determination of these generating vectors and the set of generating vectors represents what may be referred to as genetically independent pathways. This means that each pathway utilizes a unique set of reactions and gene products utilizing a different genotype. Also extremely important is the fact that the set of generating vectors is unique. So to best describe the null space and navigate through the metabolic map of an organism we have to determine this unique set of genetically independent pathways.



COMPARING LINEAR SPACES AND CONVEX ANALYSIS

The number of generating vectors can exceed the dimensions of the cone which has the mathematical consequence that all of these pathways are not linearly independent. The best analogy for some of these concepts is to think of an Egyptian pyramid which has 4 edges and exists in three dimensional space. While not linearly independent these pathways are systemically independent in that they cannot be decomposed into a combination of other pathways in a convex manner. An important fact is that the set of generating vectors is unique and below we represent algorithms to solve for these generating vectors. Thus, to describe the flux space and navigate through the metabolic map of an organism we have to determine that it is best to use this unique set of systemically independent pathways.

This set of pathways can be thought of as the "operating system" for a defined metabolic genotype, since the control over these pathways will enable the attainment of any state (phenotype) allowable by the constraints placed on the metabolic system.

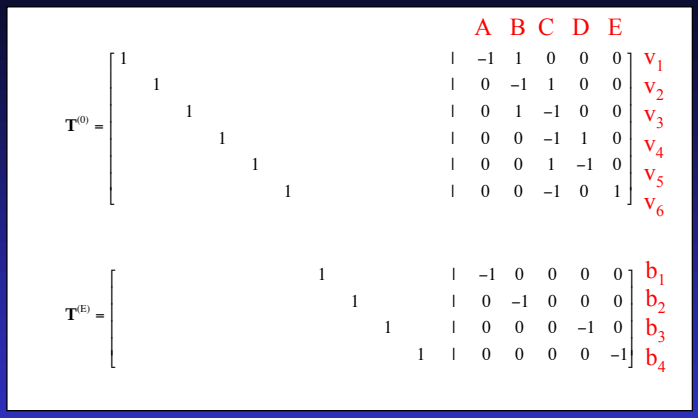


FINDING EXTREME PATHWAYS

Consider the example metabolic system shown above. The stoichiometric matrix is given and so are the constraints placed on the system. How do we now determine the extreme pathways?

The algorithm that is implemented to determine the set of extreme pathways for a reaction network follows the principles of algorithms for finding the external rays/generating vectors of convex polyhedral cones. This algorithm will give unique, biochemically feasible pathways which define the edges of the flux cone.

Initialization of the First Tableau



- Append identity matrix to S^T.
- Separate out unconstrained exchange fluxes into temporary $T^{(E)}$ matrix.

THE ALGORITHM

The algorithm begins with the formulation of an initial matrix consisting of an *n* x *n* identity matrix (**I**) appended to the transpose of the stoichiometric matrix, $\mathbf{S}^T$. Then we examine the constraints on each of the exchange fluxes. If the exchange flux is constrained to be positive nothing is done, however, if the exchange flux is constrained to be negative then we multiply the corresponding row of the initial matrix by –1. If the exchange flux is unconstrained then we move the entire row to a temporary matrix, $\mathbf{T}^{(E)}$. This completes the initialization of the first tableau, $\mathbf{T}^{(0)}$.

The above tableau has been constructed for the example network. In this network, all of the exchange fluxes are unconstrained hence they are all in the temporary matrix $\mathbf{T}^{(E)}$.

Formulating the Next Tableau: $T^{(1)}$

$$\mathbf{T}^{(1)} = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} & & & & & & \end{array} \\ \left[\begin{array}{ccccccc} 1 & & & & & & \\ & 1 & 1 & & & & \\ & 1 & & 1 & & & \\ & 1 & & & & & 1 \\ & & 1 & & 1 & & \\ & & & 1 & 1 & & \\ & & & & 1 & 1 & \end{array}\right. \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \left|\begin{array}{ccccc} -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{array}\right. \end{array} \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \\ \left.\begin{array}{c} v_1 \\ v_2+v_3 \\ v_2+v_4 \\ v_2+v_6 \\ v_5+v_3 \\ v_5+v_4 \\ v_5+v_6 \end{array}\right] \end{array}$$

- Copy all rows from $T^{(0)}$ into $T^{(1)}$ that have a zero in the column that corresponds to the first metabolite without an unconstrained flux (in this case C)
- Of the remaining rows, add together all possible combinations of rows which contain values of the opposite sign in the C column such that the addition produces a zero in the column and add to $T^{(1)}$.

Each element of the matrix is designated by T_{ij}. Starting with x equal to 1 and $\mathbf{T}^{(0)}$ equaling $\mathbf{T}^{(x-1)}$ the next tableau is generated in the following manner:

1. Identify all of the metabolites that do not have an unconstrained exchange flux associated with them. The total number of such metabolites is denoted by μ. In this example only metabolite C does not have such an unconstrained exchange flux so $\mu=1$.

2. Begin forming the new matrix $\mathbf{T}^{(x)}$ by copying all rows from $\mathbf{T}^{(x-1)}$ which contain a zero in the column of $\mathbf{S}^T$ that corresponds to the first metabolite identified in step 1, denoted by the index c. (This will be the third column of the transposed stoichiometric matrix, with only the first row containing zero in this column).

3. Of the remaining rows in $\mathbf{T}^{(x-1)}$ add together all possible combinations of rows which contain values of the opposite sign in column c, such that the addition produces a zero in this column. Given two rows, $\mathbf{r}_1$ and $\mathbf{r}_2$, whose elements will be denoted by $r_{1,j}$ and $r_{2,j}$ for $j=1,\ldots,(n+m)$, combine the rows using the following equation to generate a new row $\mathbf{r}'$ to be added to $\mathbf{T}^{(i)}$:

$$\mathbf{r}' = (|r_{2,c}| * \mathbf{r}_1) + (|r_{1,c}| * \mathbf{r}_2)$$

For the example, these steps result in the above $\mathbf{T}^{(1)}$ matrix.

Formulation Continued

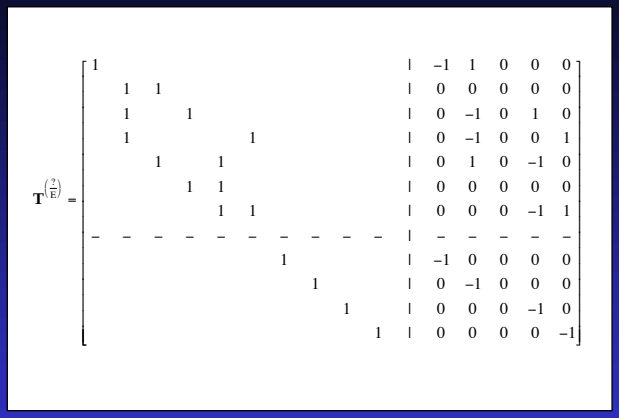
- Check for rows that are non-negative combinations of other rows and eliminate
- Repeat tableau formulation procedure and non-negative combination check for all unconstrained metabolites ending with $T^{(\mu)}$

In the end, the number of rows in $T^{(\mu)}$ will equal the number of extreme pathways

4. For all of the rows added to $\mathbf{T}^{(x)}$ in steps 2 and 3 check to make sure that no row exists that is a non-negative combination of any other sets of rows in $\mathbf{T}^{(x)}$. One method used is as follows: let $A(i)$ equal the set of column indices, j, for which the elements of row i equal zero. Then check to determine if there exists another row (h) for which $A(i)$ is a subset of $A(h)$.

5. With the formation of $\mathbf{T}^{(x)}$ complete repeat steps 2 through 4 for all of the metabolites that do not have an unconstrained exchange flux operating on the metabolite, incrementing x by one up to m. The final tableau will be $\mathbf{T}^{(\mathsf{m})}$. (In this example there is only one such metabolite so we do not need to iterate through steps 2-4 again. Therefore $\mathbf{T}^{(\mathsf{m})}$ equals $\mathbf{T}^{(1)}$ as in Eq.B.3.) Note that the number of rows in $\mathbf{T}^{(\mathsf{m})}$ will be equal to (k), the number of extreme pathways.

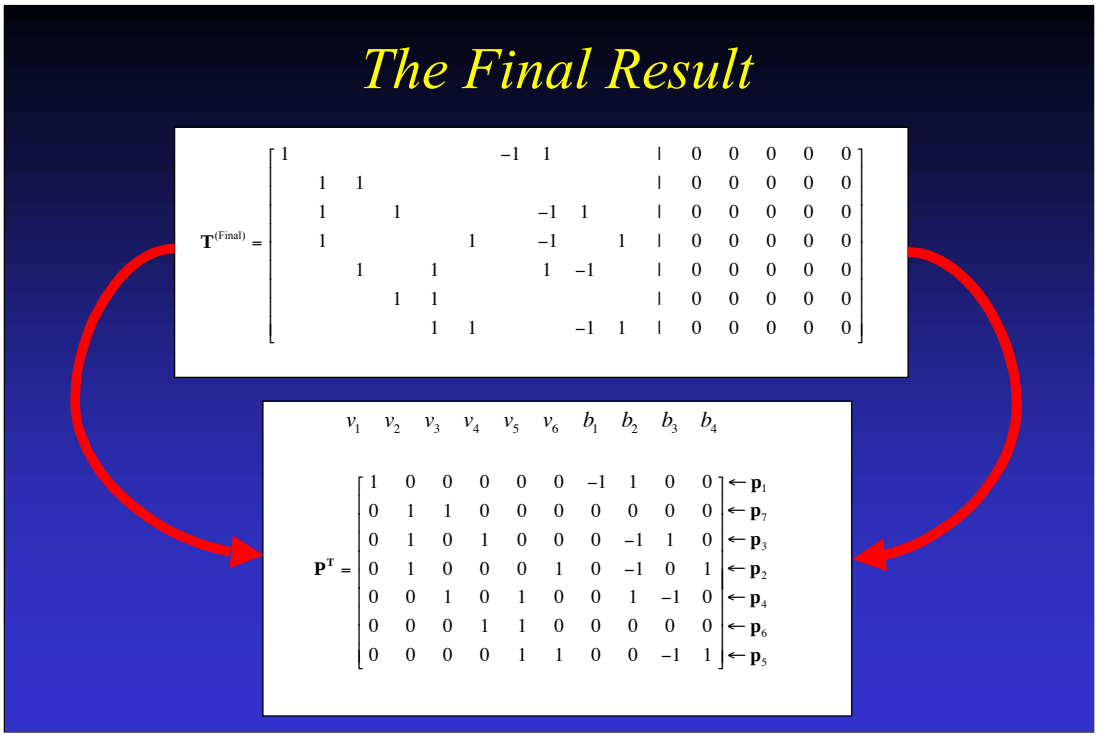
The Final Step



- Use $T^{(E)}$ to "zero" out the right hand side of $T^{(\mu)}$ by adding or subtracting rows from $T^{(E)}$.

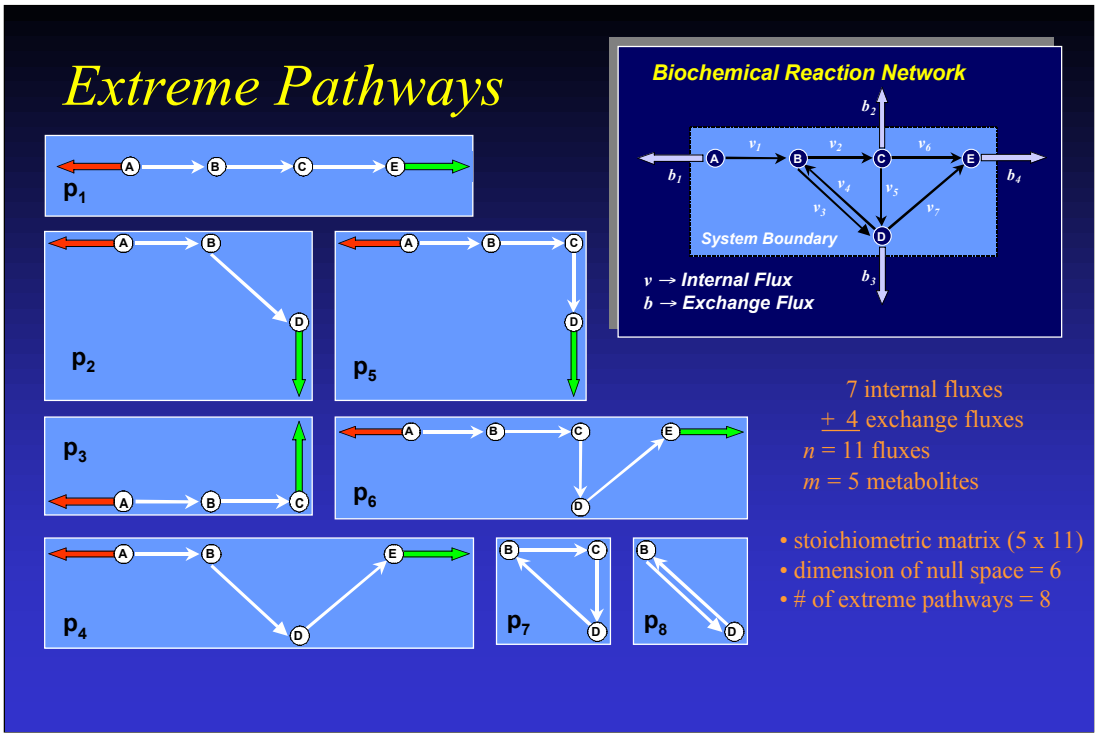
6. Starting in the n+1 column (or the first non-zero column of the right side), if $T_{i,(n+1)}$ does not equal zero, then add the corresponding nonzero row from $\mathbf{T}^{(E)}$ to row i so as to produce a zero in the (n+1) column. This is done by simply multiply the corresponding row in $\mathbf{T}^{(E)}$ by $T_{i,(n+1)}$ and adding this row to row i. Repeat this procedure for each of the rows in the upper portion of the tableau so as to create zeros in the entire upper portion of the (n+1) column. When finished remove the row in $\mathbf{T}^{(E)}$ corresponding to the exchange flux for the metabolite just balanced.

7. Follow the same procedure in step 7 for each of the columns in the right portion of the tableau contain non-zero entries. (In this example we need to perform step 7 for every column except the middle column of the right side which corresponded to metabolite C).



THE FINAL RESULT

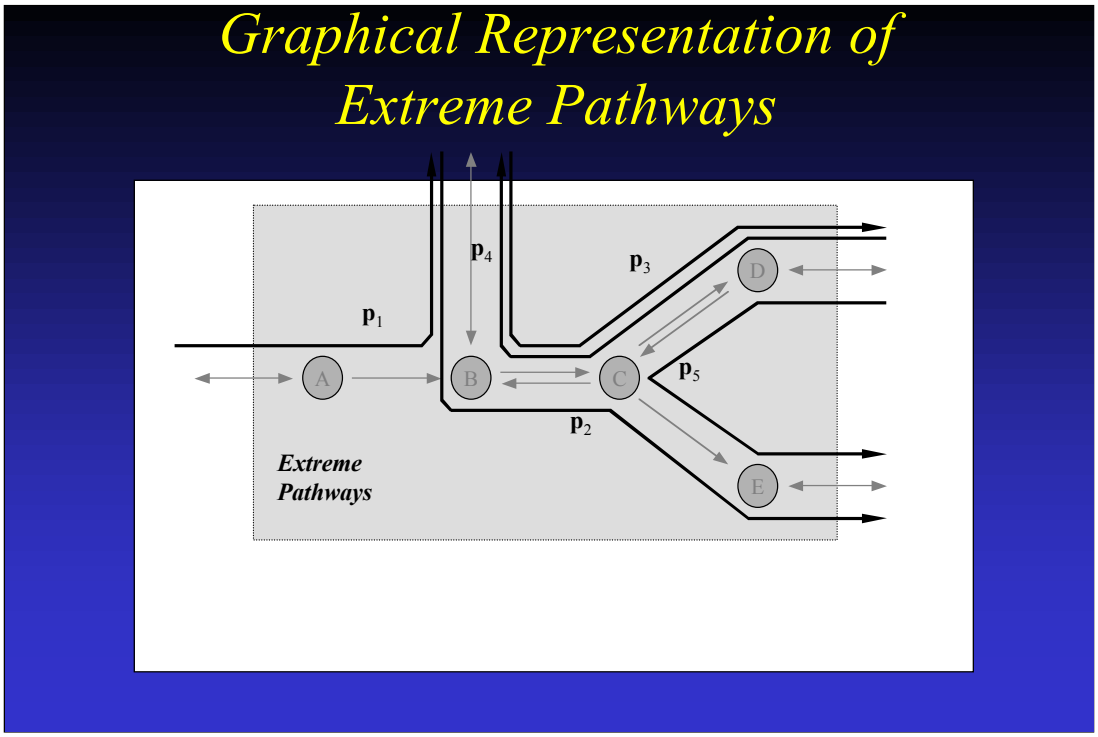
The final tableau, $\mathbf{T}^{(\text{Final})}$, will contain the transpose of the matrix **P** containing the extreme pathways in place of the original identity matrix.



WHAT DO THE EXTREME PATHWAYS LOOK LIKE?

Here is an example of what extreme pathways look like for the hypothetical reaction network shown above. In this case the stoichiometric matrix is 5 by 11 with the dimensions of the null space equaling 6 and the number of extreme pathways equaling 8. We can see here that 6 of these pathways are actually performing net reactions which consume a metabolite to produce another, however there are two pathways here that are only internal cycles within the network. So we see the necessity for a classification scheme for these pathways.

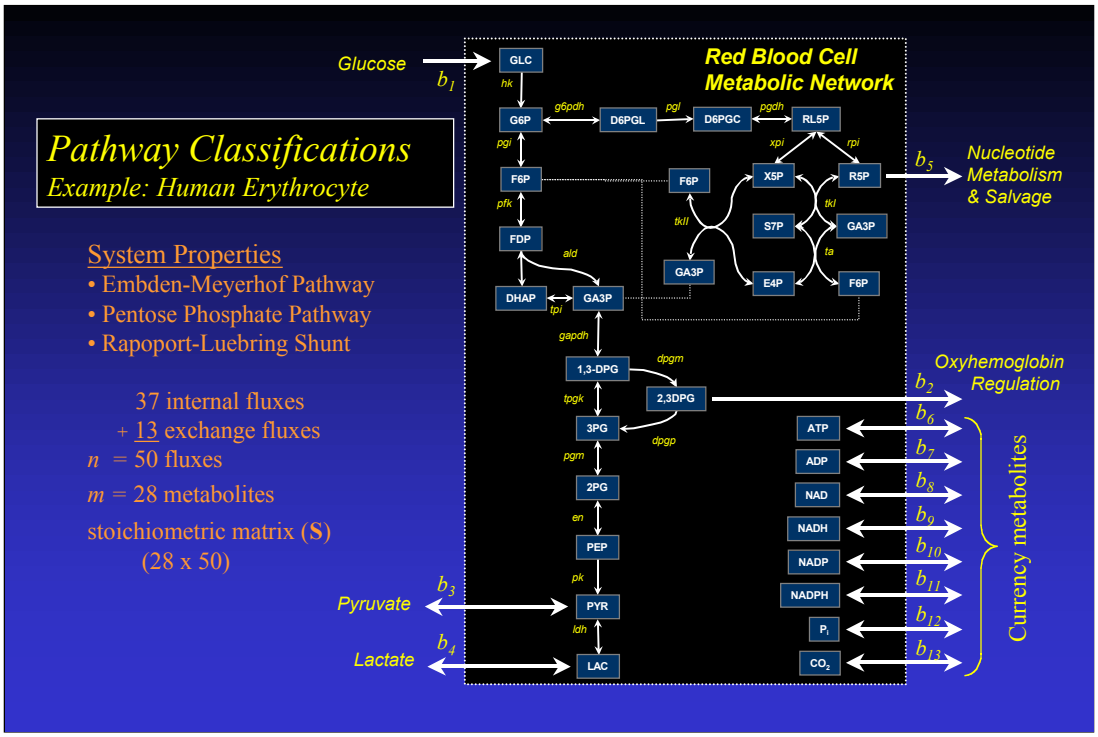
Compare these pathways to the linked output pathways that appear on a later slide.



GRAPHICAL REPRESENTATION OF EXTREME PATHWAYS

This slide shows the extreme pathways evaluated for a sample system. These pathway vectors will be the edges of a 7-dimensional cone. All admissible steady state solutions lie in this cone.

We need to work on software for good representation of these pathways.



THE RED BLOOD CELL

To illustrate the different classifications of pathways I will use the red blood cell as a limited but biologically realistic example. Here we have a partial metabolic reaction system for the red blood cell which is composed of the

• Embden-Meyerhof pathway,

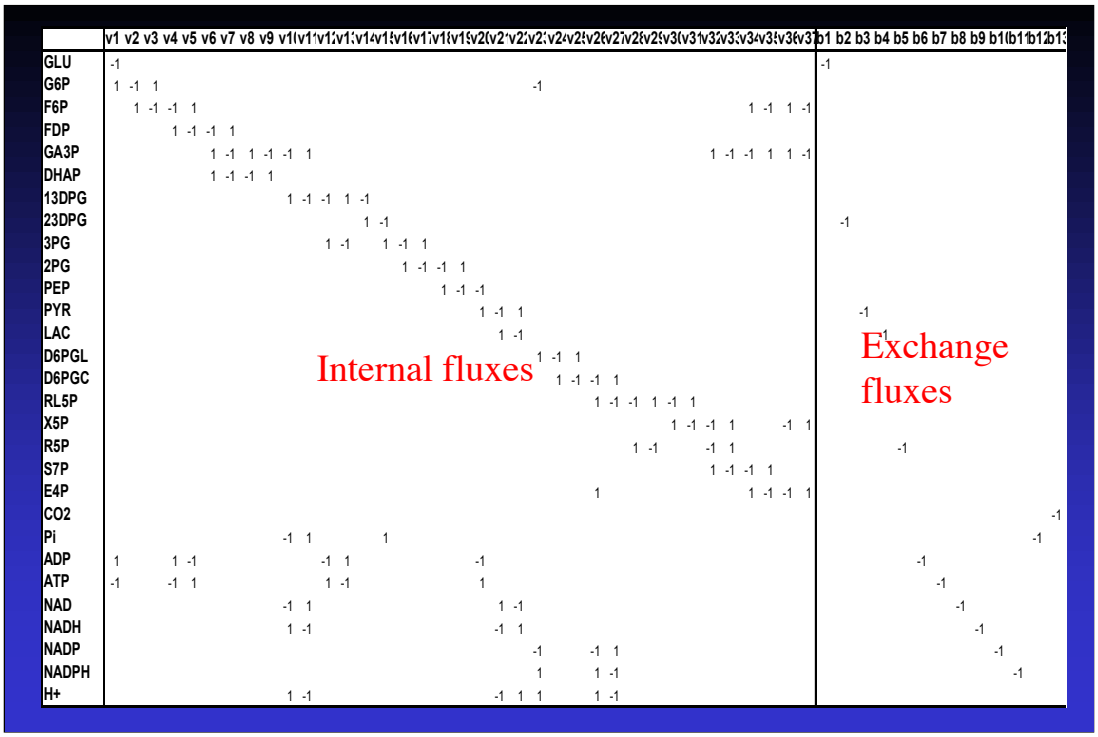
•Pentose phosphate pathway, and

•Rapoport-Luebring shunt.

The characteristics of this metabolic system are shown in this slide.

Note that there is a distinction made between primary metabolites and currency metabolites or those which are mainly involved in energy & redox exchange in the cell. So once we construct the stoichiometric matrix for this system we simply find the independent pathways which define the edges of the flux cone.

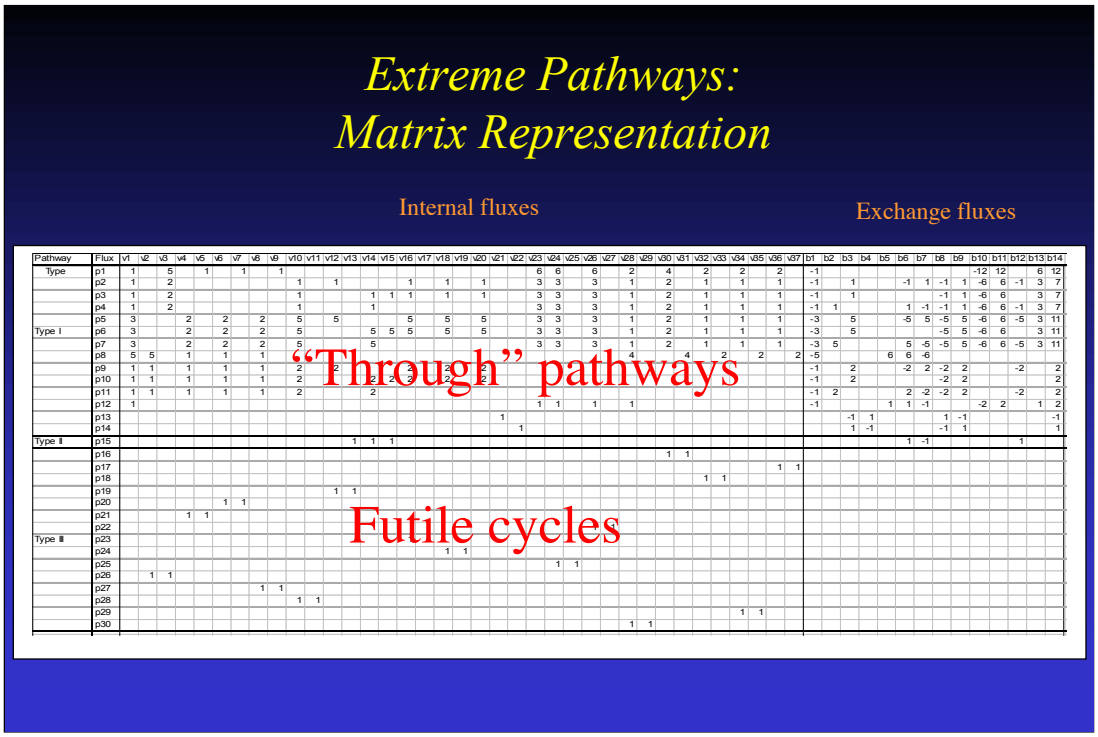
These extreme pathways then together can be used to describe every possible state which this system can operate in.



THE RED CELL STOICHOMETRIC MATRIX

There is the stoichiometric matrix for the red blood cell. Notice that the matrix is sparse and has only 1 and -1 non-zero entries.

It is partitioned based on the internal flux/exchange flux distinction as shown above.



EXTREME PATHWAYS FOR THE RED BLOOD CELL

The calculated extreme pathways for the red cell system are shown.

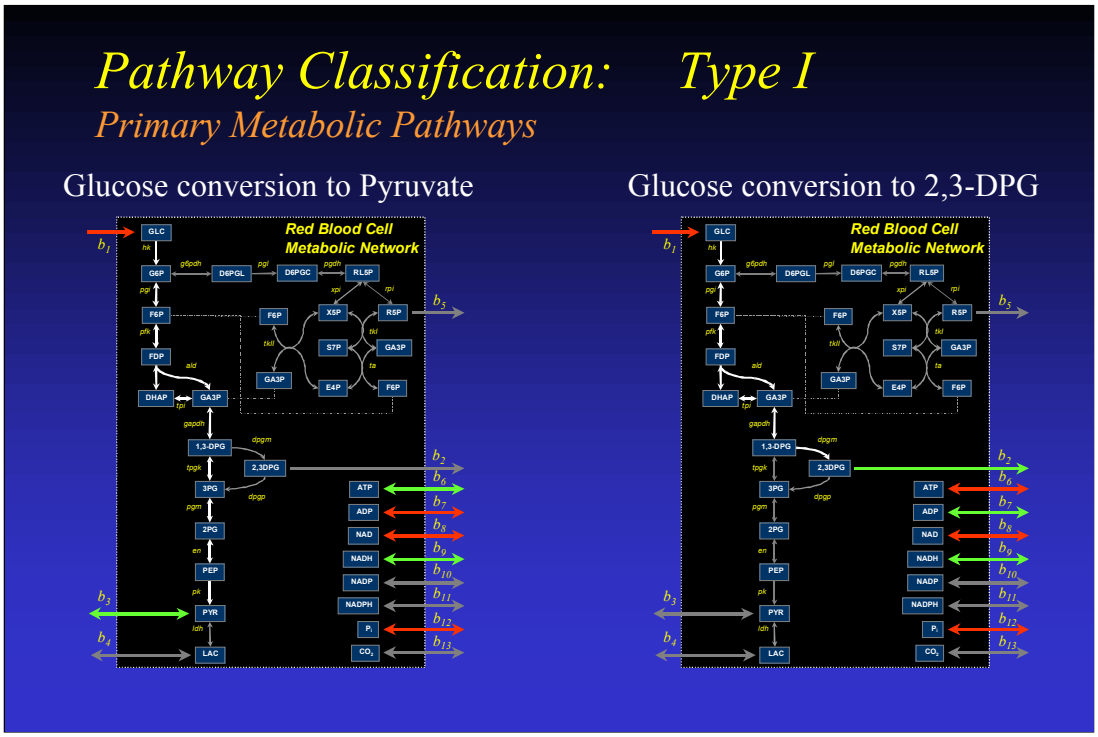
Note that there is a group of pathways (at the bottom) that have no exchange reactions. They are thus internal to the cell with no connections to the outside. We call these pathways type III pathways.

There is one pathway whose only exchange fluxes involve cofactors. This is an internal futile cycle. We call these type II pathways.

The rest are 'through' pathways, that connect an input to an output. These are the type I pathways.

Classifying Pathways

- Type I: Primary Metabolic Pathways
- Type II: "Futile" Cycles
  - only currency exchange fluxes active
- Type III: Reaction Cycling
  - no active exchange fluxes

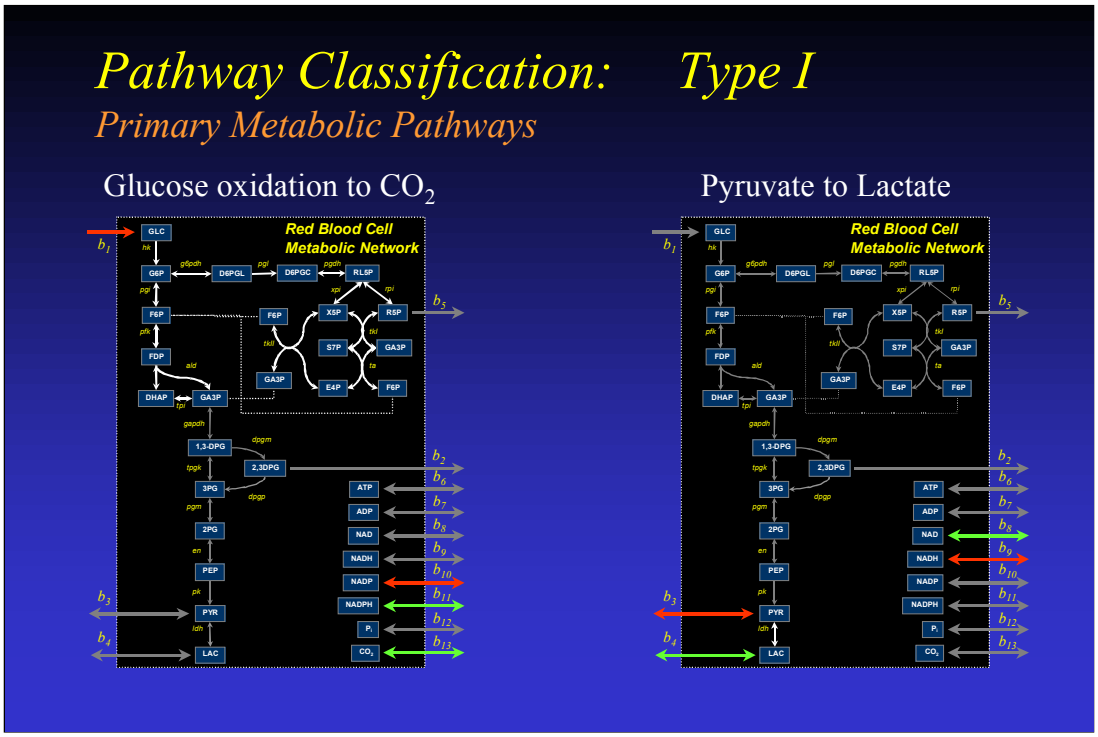


TYPE I PATHWAYS

The first type of pathways that are generated are what we refer to as primary pathways and these are the types of pathways that first come to mind when thinking about a metabolic map. These are simply pathways that connect an input to an output. The only requirement of the pathway is that one of the primary exchange fluxes must be active. Here are two examples of primary metabolic pathways that are extreme pathways on the cone. The green arrow denote the production of a metabolite by the pathway and the red arrows indicate the consumption of a metabolite while the white arrows indicate the internal fluxes which are operating.

The first example is simply the conversion of glucose to pyruvate using the glycolytic pathway . This is basically the glycolytic pathway that picks up glucose and secretes pyruvate, and producing both ATP and NADH.

The second pathway is the production of 2,3DPG from the Rapoport-Luebering shunt. This pathway becomes active when more 2,3DPG needs to be produced such as when one goes through changes in altitude and the oxygen binding characteristics of hemoglobin need to be changes. This will always be a low flux pathway.



MORE TYPE I PATHWAYS

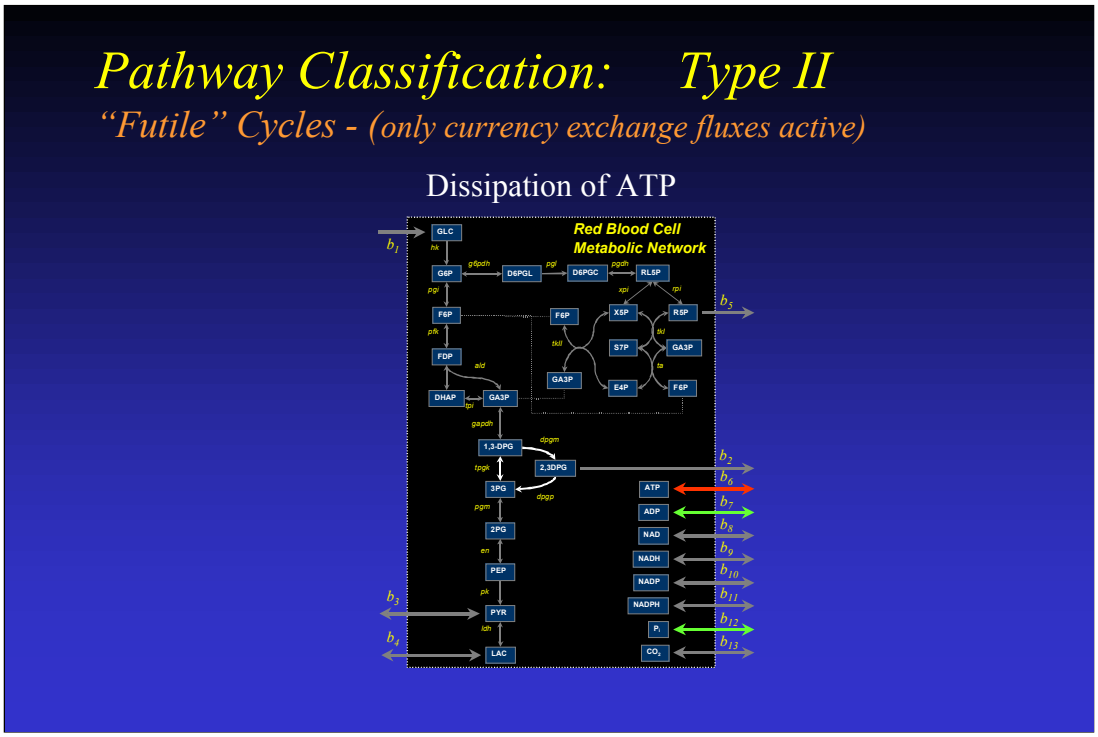
Here are two more examples of type I extreme pathways pathways:

•the complete oxidation of glucose to CO2 through the cycling of the pentose phosphate pathway producing NADPH and

•primary pathway which consists of only one reaction converting pyruvate into lactate used to balance the NAD/NADH ratio of the cell.

As you can see each of these primary metabolic pathways has a functional role in the cell and therefore these pathways can be used to interpret the functional attributes and activities of red cell metabolism in this case.

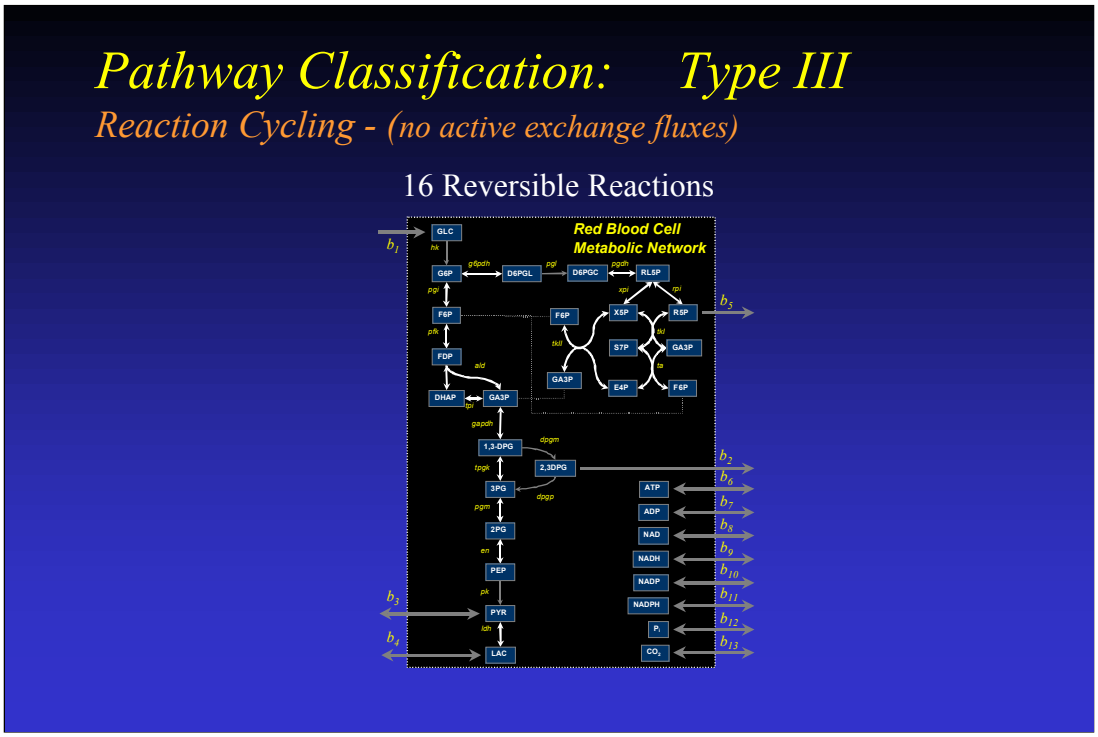
There are a total of 14 type I pathways in this simple red cell model and the others will not be discussed in detail.



TYPE II PATHWAYS

The second type of pathway is what is commonly referred to as a futile cycle in the truest sense of the word futile. In these pathways only the exchange fluxes for the currency metabolites are active. In this system there exists one futile cycle which occurs around the Rapoport-Luebering shunt. The net result of this pathway is the conversion of ATP into ADP and releasing inorganic phosphate which is obviously dissipating metabolic energy.

There is one futile cycle that operate in this system.



TYPE III PATHWAYS

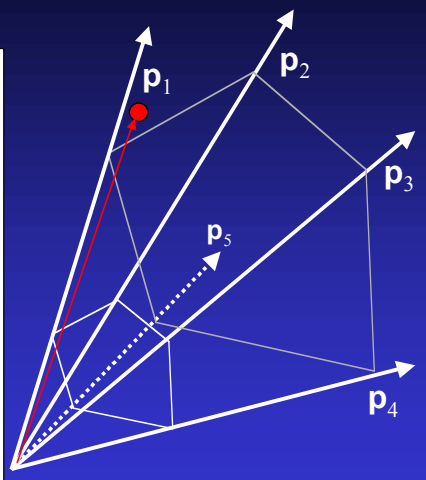
The third type of pathway consists of reversible cycles which are mainly the result of reversible reactions be characterized by a forward reaction and a separate reverse reaction. These pathways show no activity in any of the exchange fluxes. In this map all 16 reversible reactions are highlighted white. While these pathways are essentially generating vectors they can effectively be dismissed in any further analysis of the system as they have no net effect on the production capabilities of the system as they influence none of the exchange fluxes.

These pathways will become important later when we examine temporal decomposition of this system. A fast internal pathway leads to an 'equilibration' or the 'tying together' of two or more concentrations that then can be 'pooled' together to form an aggregate dynamic variable. Some simple examples of this feature were show in the last lecture.

Together all of the extreme pathways in a system fall under one of these three classifications of pathways and they all are edges of the cone determining the flux space.

Pathway Utilization in the Red Cell:
Geometric representation

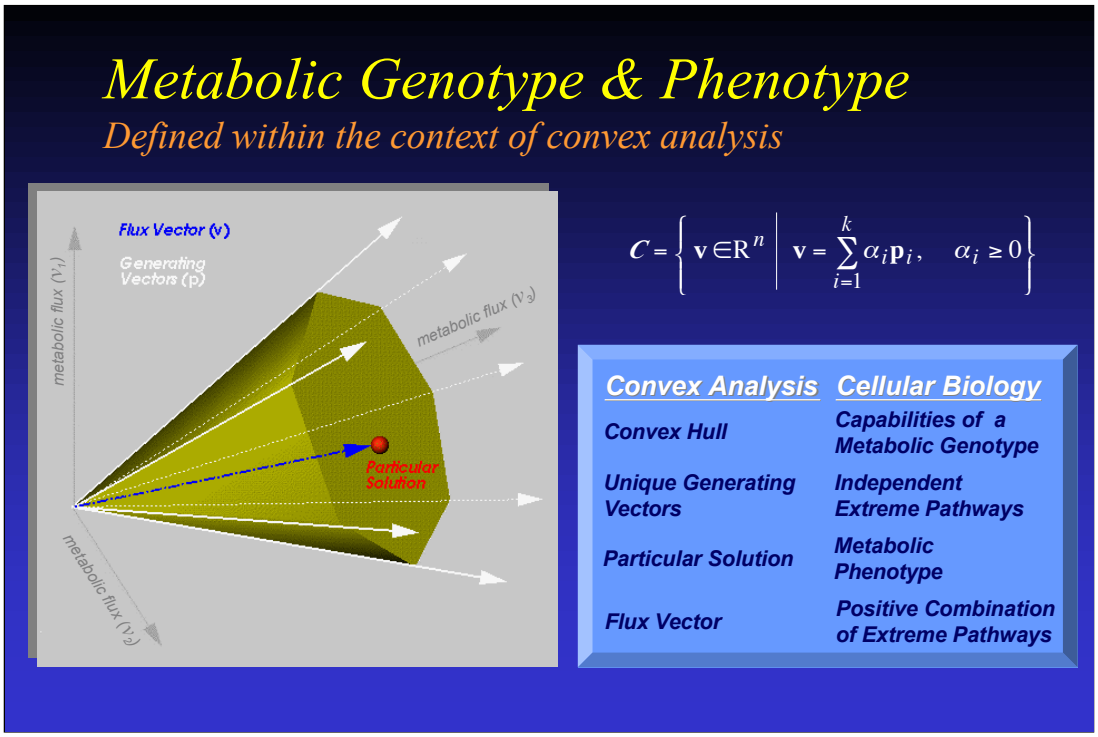
Basis Pathway	Net reaction Equation	Primary Functional Attribute
p1	glucose + 2 Pi +2 ADP -> 2 lactate +2 ATP	ATP energy production for metabolic energy
P2	glucose + 2 NAD+ -> 2 pyruvate + 2 NADH + 2 H+	NADH energy production for methemoglobin reduction
P3	glucose + 2 Pi + 2 ADP + 2 NAD+ -> pyruvate +2 ATP 2 NADH+ + 2 H+	ATP production for metabolic energy and NADH production for methemoglobin reduction
P4	glucose + 2 ATP +2 NAD+ -> 2 2,3DPG +2 Pi +2 ADP +2 NADH +2 H+	2,3DPG production for oxyhemoglobin modulation
P5	glucose + ATP + 2 NADP+ -> R5P +CO2 + ADP + 2 NADPH + 2 H+	R5P production for adenosine salvaging
P6	glucose +12 NADP+ -> 6 CO2 + 12 NADPH + 12 H+	NADPH energy production for glutathione reduction and subsequent antioxidant activity
p7-p22	(no net reaction)	



NORMAL OPERATION OF THE RED CELL MATABOLIC NETWORK

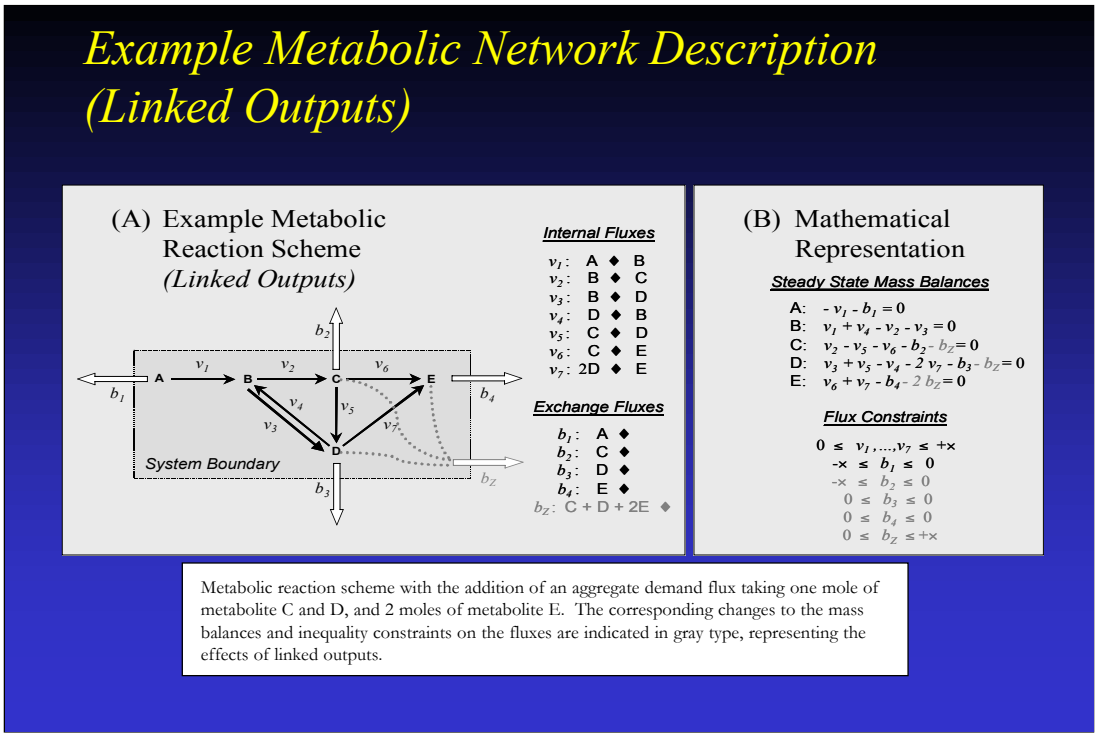
The nominal physiological steady state of the red cell metabolism is to produce ATP to run the Na/K pump to maintain the osmotic balance across the cells membrane. If other pathways are activated to produce, say 2,3 DPG with altitude change, two of these pathways would contribute to the flux map and the solution could 'creep' towards an edge of the flux cone.

This flux solution can be obtained from the full dynamic red cell model (downloadable in a MATHEMATICA form from http://gcrg.ucsd.edu) or from a flux balance model where the demands of the pump are stated and the uptake rate of glucose in minimized.



NIFTY INTERPRETATION OF THE FLUX CONE

So what does this all mean from a biological point of view. Here is the geometric interpretation of the flux cone in which every point described by the equation given. The entire flux cone actually corresponds to the capabilities of a reaction network and hence the defined metabolic genotype. What can the reconstructed network do, what can it not do? Each one of the generating vectors corresponds to an extreme pathway which the cell could theoretically control to reach every point in the flux cone. Now a particular point within this flux cone corresponds to a given flux distribution which represents a particular metabolic phenotype. The actual flux vector describing that point can be thought of as a positive combination of these extreme pathways. So you may think of these pathways as being theoretically turned on and off to reach a particular metabolic phenotype. Once again this means that every phenotype which the system can exhibit is a combination of these pathways which are then turned on or off. It's that simple. With these pathways we can describe all of the capabilities of the metabolic system and so we may say that these pathways represent the underlying pathway structure of the system.



TOWARDS PHYSIOLOGICAL FUNCTIONS: LINKED OUTPUTS

Under changing substrate/supply conditions metabolic networks are continuously faced with balanced sets of biosynthetic demands (i.e. production of amino acids, nucleotides, phospho-lipids, as well as metabolic energy and redox potential). Effectively this means that the network must generate a balanced rate through the exchange fluxes for the particular metabolites required to meet these demands. To assess the systemic performance of a network in meeting balanced biosynthetic demands, an exchange flux is introduced into a network. Additional constraints must also be added to the network to effectively close the material balances on the metabolites participating in the biosynthetic demand (or growth) flux. The introduction of a new flux and the associated restriction of existing fluxes will alter the mass balances and linear inequalities of the network, and subsequently alter the pathway structure. To distinguish between the two different forms (all material balances closed with a growth flux are included versus no growth flux and material balances not closed on biosynthetic precursors) we consider a system without a biosynthetic demand flux to have free outputs, and the consideration of balanced network demands defines a linked output system. For the example system, we introduce the growth exchange flux b_Z which is described in the drawing below. This flux must then be included into the mass balances. Additionally we change the constraints on the specific exchange fluxes for metabolites C, D, and E so as not to allow them to exit the system. All of the changes from the open to the closed system are highlighted in the figure.

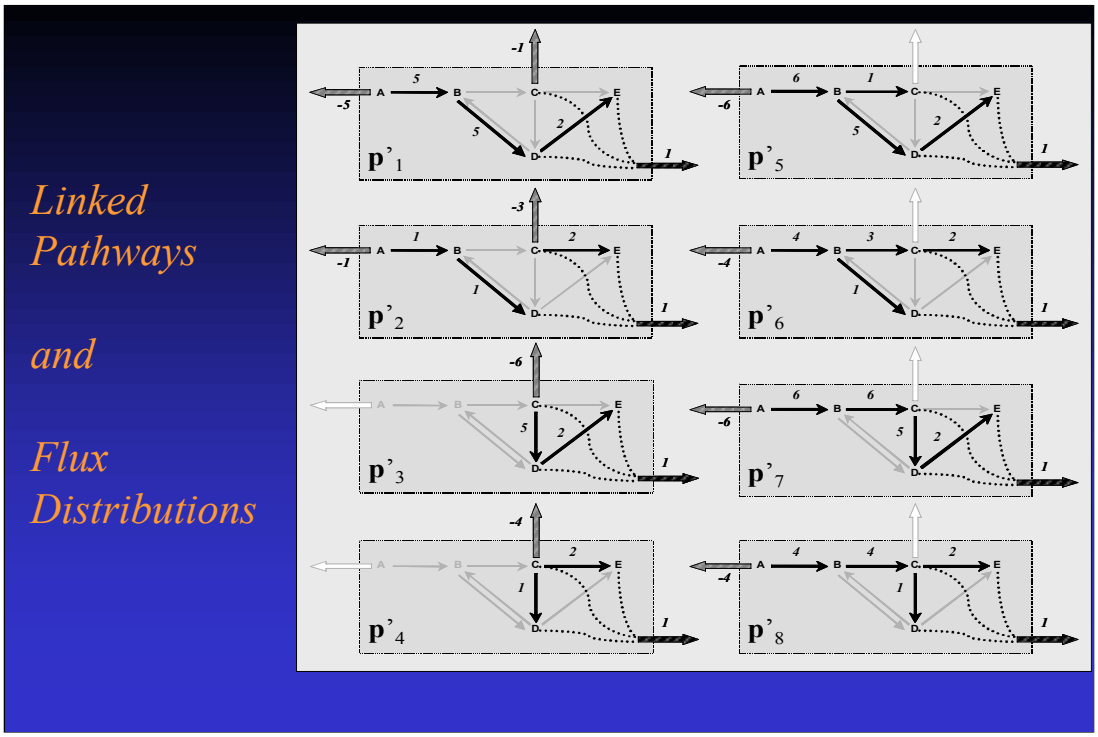
Extreme Pathways (Linked Outputs)

The 10 extreme pathway vectors for the linked output description of the example network. The first eight pathways correspond to type I pathways while the last two pathways are of type III. Pathway equivalencies between the free and linked output systems are provided for each pathway

Pathway	***Internal Fluxes***							***Exchange Fluxes***					***Pathway Equivalences***
Number	v_1	v_2	v_3	v_4	v_5	v_6	v_7	b_1	b_2	b_3	b_4	b_z	***Linked ~ Free***
$\mathbf{p'}_1$	5	0	5	0	0	0	2	-5	-1	0	0	1	$\mathbf{p'}_1 \sim \mathbf{p}_2 + 2\,\mathbf{p}_3$
$\mathbf{p'}_2$	1	0	1	0	0	2	0	-1	-3	0	0	1	$\mathbf{p'}_2 \sim \mathbf{p}_2 + 2\,\mathbf{p}_6$
$\mathbf{p'}_3$	0	0	0	0	5	0	2	0	-6	0	0	1	$\mathbf{p'}_3 \sim \mathbf{p}_4 + 2\,\mathbf{p}_5$
$\mathbf{p'}_4$	0	0	0	0	1	2	0	0	-4	0	0	1	$\mathbf{p'}_4 \sim \mathbf{p}_4 + 2\,\mathbf{p}_6$
$\mathbf{p'}_5$	6	1	5	0	0	0	2	-6	0	0	0	1	$\mathbf{p'}_5 \sim \mathbf{p}_1 + \mathbf{p}_2 + 2\,\mathbf{p}_3$
$\mathbf{p'}_6$	4	3	1	0	0	2	0	-4	0	0	0	1	$\mathbf{p'}_6 \sim 3\,\mathbf{p}_1 + \mathbf{p}_2 + 2\,\mathbf{p}_6$
$\mathbf{p'}_7$	6	6	0	0	5	0	2	-6	0	0	0	1	$\mathbf{p'}_7 \sim 6\,\mathbf{p}_1 + \mathbf{p}_4 + 2\,\mathbf{p}_5$
$\mathbf{p'}_8$	4	4	0	0	1	2	0	-4	0	0	0	1	$\mathbf{p'}_8 \sim 4\,\mathbf{p}_1 + \mathbf{p}_4 + 2\,\mathbf{p}_6$
$\mathbf{p'}_9$	0	0	1	1	0	0	0	0	0	0	0	0	$\mathbf{p'}_9 \sim \mathbf{p}_7$
$\mathbf{p'}_{10}$	0	1	0	1	1	0	0	0	0	0	0	0	$\mathbf{p'}_{10} \sim \mathbf{p}_8$

LINKED PATHWAYS ARE COMPINATIONS OF SINGLE OUTPUT PATHWAYS

For the linked output system there are 10 extreme pathways (8-type I and 2-type III pathways). The complete pathway vectors are provided in this table for pathway #1 through #8 (pathway #7 and #8 are type III pathways that exhibit no activity for the exchange fluxes, i.e. internal cycles).



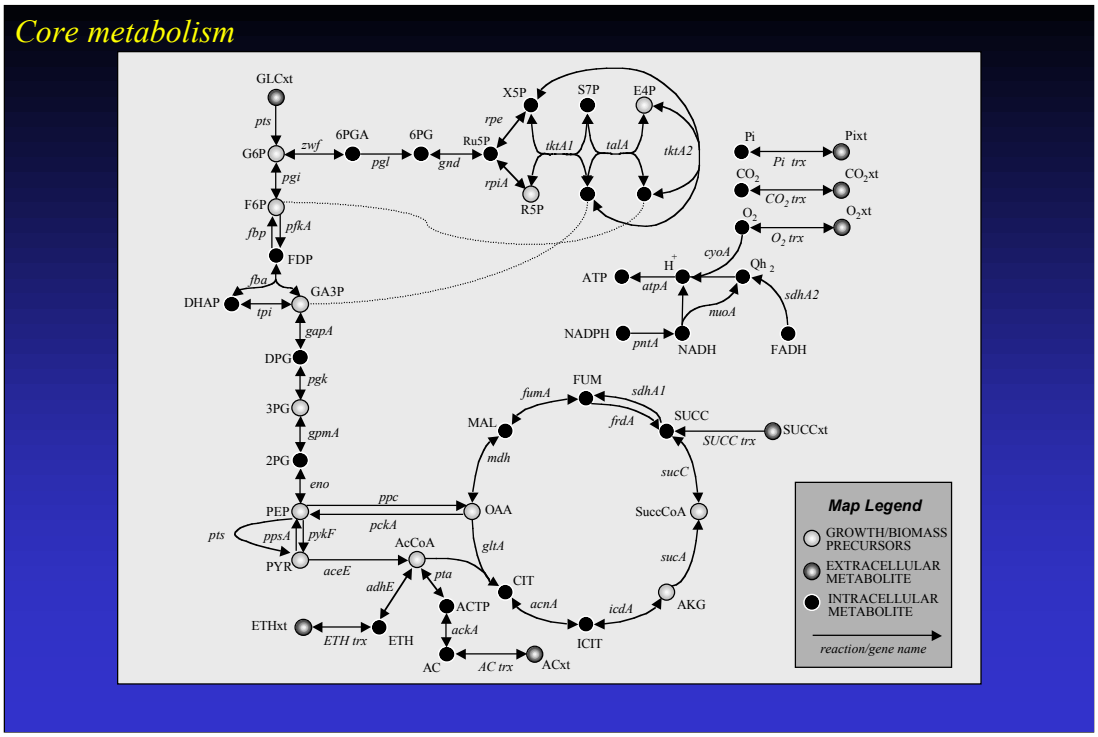
GRAPHICAL REPRESENTATION OF LINKED PATHWAYS

The pathway distributions are also illustrated graphically in this figure. Note that the extreme pathways for the linked outputs are systemic flux distributions that meet the balanced set of demands represented by the growth flux (b_z). These extreme pathways are non-negative combinations of the extreme pathways for the free output system. This leads to the definition of pathway equivalences that relate the free output system pathways to the linked output system.

The first two produce the required output using two inputs (A and C), the next two only from C, and the last four from A alone.

Compare these to the single output pathways shown on an earlier slide for the same network.

The linked pathways are no longer 'linear' or 'one-dimensional' entities, but actual flux maps.



A PSEUDO-REALISTIC METABOLIC NETWORK

A schematic of the central metabolic network of *E. coli* is depicted in this diagram. The network is comprised of glycolytic reactions, pentose phosphate shunt, and the tricarboxcylic acid cycle without the glyoxylate shunt along with the necessary transport reactions. The genes whose gene products carry out the reactions are used as the reaction names in most cases. The necessary electron transport chain reactions are included with the P/O ratio of 4/3. The system is comprised of 53 metabolites, 78 internal fluxes, and 8 exchange fluxes.

Note that this network does not completely describe central metabolism in *E. coli*. This representation has been chosen as a compromise between successfully representing the basic aspects of central metabolism and providing a useful example of the combined approach to study metabolic systems.

This system and its linked pathways are very insightful as we shall see on the ensuing slides.

Functional characteristics of the reduced set of 12 extreme pathways calculated for succinate as the sole carbon source for the *E. coli* metabolic network with linked outputs. Pathways are ordered based on the activity of the growth flux normalized by the succinate uptake. Pathway numbers coincide with the original numbers of the pathway vectors retained from the complete set. All of the exchange flux values are normalized to the succinate intake level (negative values are relative uptake ratios, positive values are production ratios). Exchange flux abbreviations: SUCC-succinate, ETH-ethanol, AC-acetate, PI-inorganic phosphate, CO_2-carbon dioxide, O_2-oxygen, GRO-biomass/growth flux .

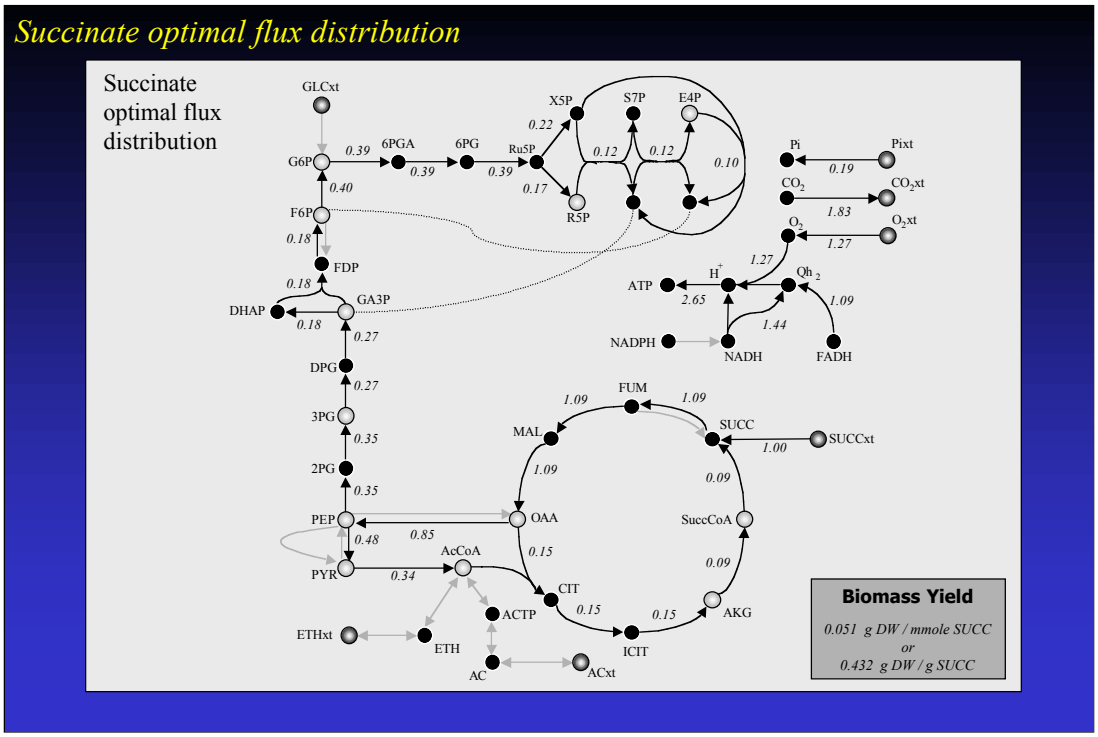
Pathway Number	*Exchange Fluxes* SUCCxt/ SUCCxt	ETHxt/ SUCCxt	ACxt/ SUCCxt	GRO/ SUCCxt	PIxt/ SUCCxt	CO2xt/ SUCCxt	O2xt/ SUCCxt	*Net Pathway Reaction Balance*
1	-1.000	0	0	0.051	-0.188	1.825	-1.267	SUCCxt + 0.188 PIxt + 1.267 O2xt ? 0.051 GRO + 1.825 CO2xt
10	-1.000	0	0	0.049	-0.182	1.895	-1.338	SUCCxt + 0.182 PIxt + 1.338 O2xt ? 0.049 GRO + 1.895 CO2xt
20	-1.000	0	0	0.047	-0.172	1.696	-1.142	SUCCxt + 0.172 PIxt + 1.142 O2xt ? 0.047 GRO + 1.696 CO2xt + 0.158 ACxt
3	-1.000	0	0.000	0.034	-0.125	2.553	-2.014	SUCCxt + 0.125 PIxt + 2.014 O2xt ? 0.034 GRO + 2.553 CO2xt
7	-1.000	0	0.000	0.033	-0.121	2.600	-2.062	SUCCxt + 0.121 PIxt + 2.062 O2xt ? 0.033 GRO + 2.6 CO2xt
12	-1.000	0	0	0.032	-0.117	2.644	-2.108	SUCCxt + 0.117 PIxt + 2.108 O2xt ? 0.032 GRO + 2.644 CO2xt
16	-1.000	0.000	0	0.031	-0.114	2.679	-2.144	SUCCxt + 0.114 PIxt + 2.144 O2xt ? 0.031 GRO + 2.679 CO2xt
19	-1.000	1	0.000	0.025	-0.092	1.837	-0.759	SUCCxt + 0.092 PIxt + 0.759 O2xt ? 0.025 GRO + 1.837 CO2xt + 0.549 ETHxt
23	-1.000	0.000	0.000	0.000	0.000	4.000	-3.500	SUCCxt + 3.5 O2xt ? 4.0 CO2xt
27	-1.000	0.000	1	0.000	0.000	2.000	-1.500	SUCCxt + 1.5 O2xt ? 2.0 CO2xt + 1.0 ACxt
31	-1.000	1.000	0	0.000	0.000	2.000	-0.500	SUCCxt + 0.5 O2xt ? 2.0 CO2xt + 1.0 ETHxt
35	-1.000	0.000	0	0.000	0.000	4.000	-3.500	SUCCxt + 3.5 O2xt ? 4.0 CO2xt

PATHWAYS FOR GROWTH ON SUCCINATE

The pathway analysis was performed with succinate as the sole carbon source for the system, generating the complete set of 66 extreme pathways (36 type I and 30 type III). To generate a reduced set of pathways that represents the full capabilities of the network, we retain only pathways from these sets that utilized the ATP drain flux instead of one of the three futile cycles; i.e. (*pfkA*/*fbp*), (*pckA*, *ppc*), (*pykF*,*ppsA*,*adk*). The type III pathways, which are mainly a consequence of the decomposition of reversible reactions into a forward and a reverse reaction, are also removed from consideration as they show no activity in the exchange fluxes. Following this simplification, a reduced set of 12 pathways is generated from the complete set.

The pathways in the table are ordered by the biomass yield that they generate (mg/mol Succinate). The best pathways produces 0.051 biomass units. This represents the optimal use of the network to produce biomass. Note that the next-best pathway produces 0.049 and is in general very similar to the best one. This is a feature that one observes. There are 'bundles' of extreme pathways located 'close' in this high dimensional conical space, which biologically is a reflection of the redundancy of the system.

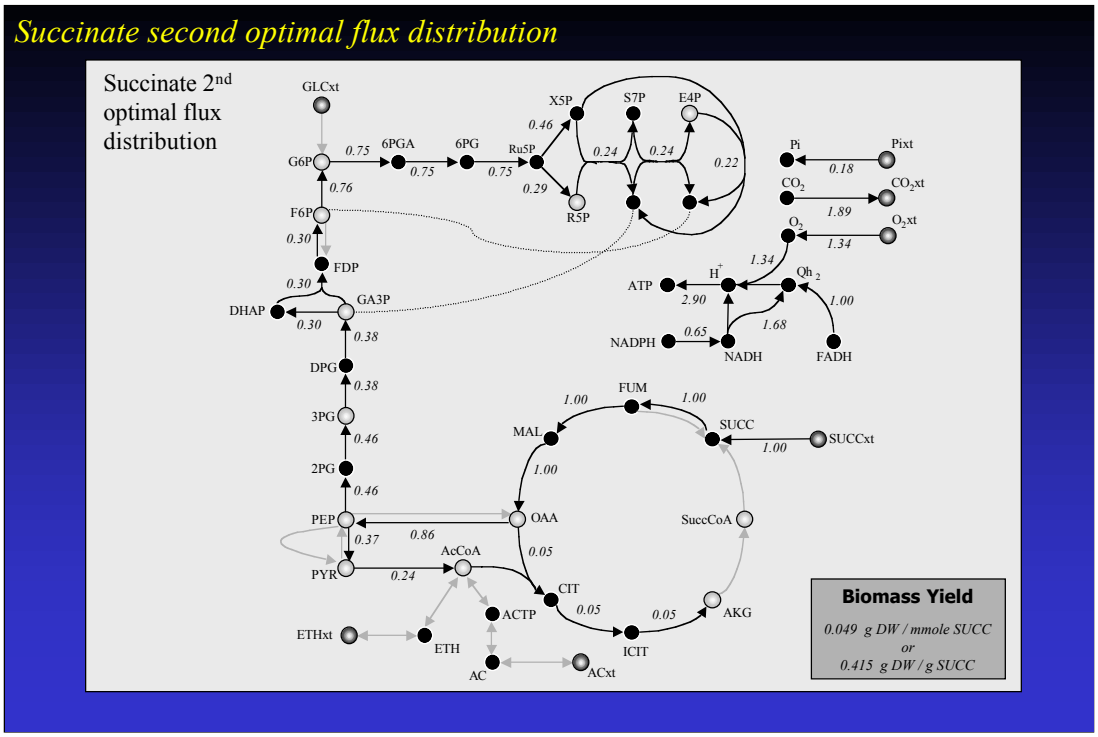
The third pathway represents partially aerobic growth and the secretion of acetate. As we shall see later, if oxygen is limiting, then the growth becomes a combination of this pathway and pathway #1. Note that there are two purely fermentative pathways producing acetate and ethanol respectively.



THE OPTIMAL PATHWAY (FLUX MAP)

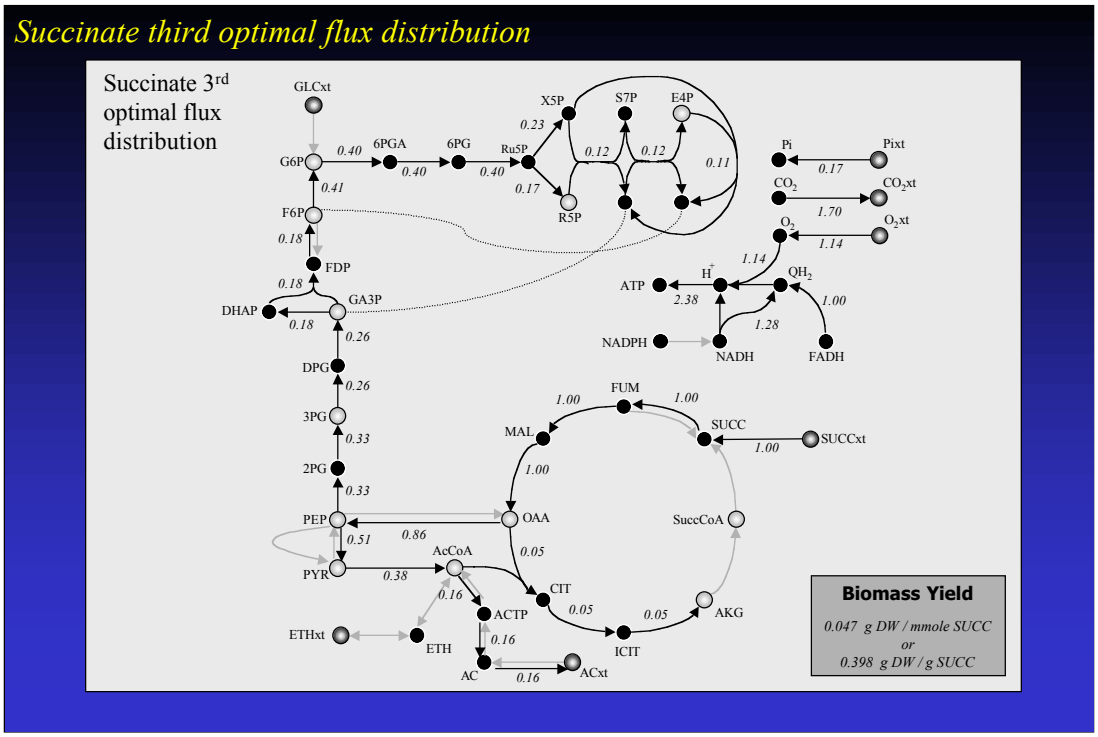
Flux balance analysis can be used to quantitatively examine the system with linked outputs. Geometrically, the constraints imposed on the input values of the exchange fluxes will bound the flux cone by the extreme pathways into a bounded polyhedron. Optimal solutions within this space will then lie on a vertex of the polyhedron. These are the bounded feasible solutions of the linear programming problem.

The flux distributions for growth are calculated on succinate normalized to 1 mmol of substrate. The optimal flux distributions are illustrated in this figure. The optimal biomass yield is 0.051 g DW/mmol succinate, which is identical to the optimal yield calculated from the pathways (pathway #1). This result reveals that the optimal solution lies directly on the vertex of the polyhedron that is defined by the extreme pathway.



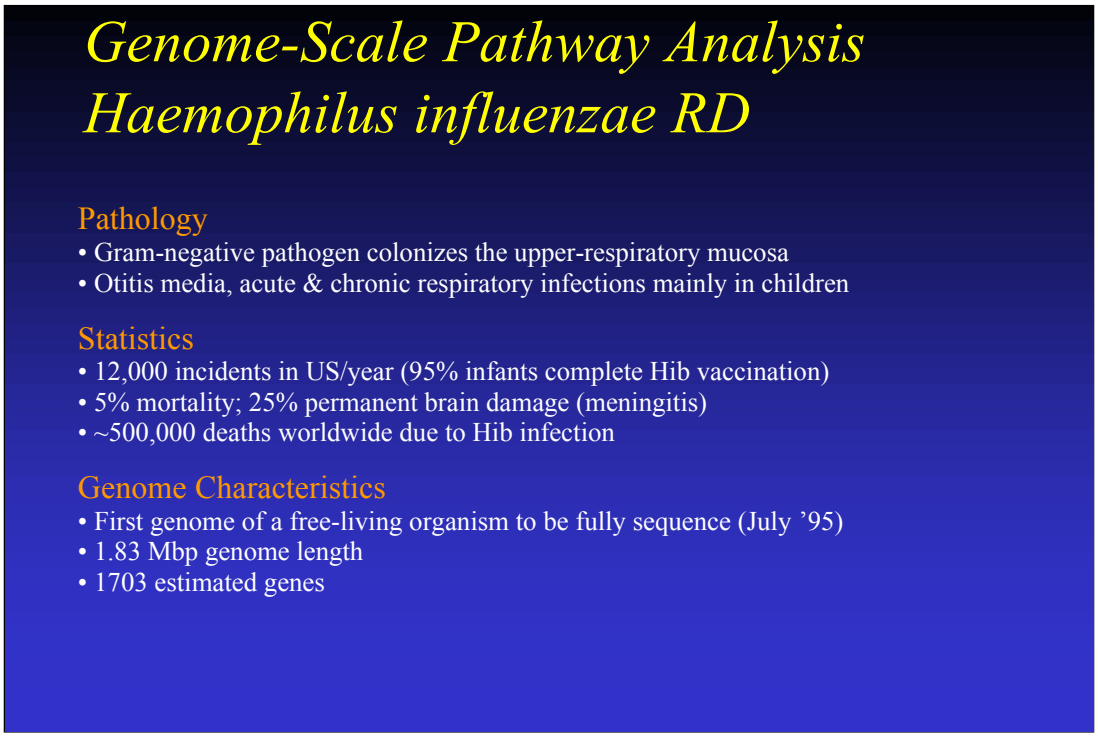
A PSEUDO-OPTIMAL PATHWAY (FLUX MAP)

The second highest optimal flux distribution may also be depicted graphically (pathway #10 in the table). The fluxes in the 2nd optimal flux distribution are identical to the optimal flux distribution as shown in the previous diagram except for the reactions catalyzed by the following enzymes: 2-ketoglutarate dehydrogenase (converting AKG to SuccCoA), succinyl-CoA synthetase (SuccCoA to Succ), and pyridine nucleotide transhydrogenase.The flux values of glycolytic and pentose phosphate pathways are higher and tricarboxylic acid cycle fluxes are lower than the optimal flux distribution.

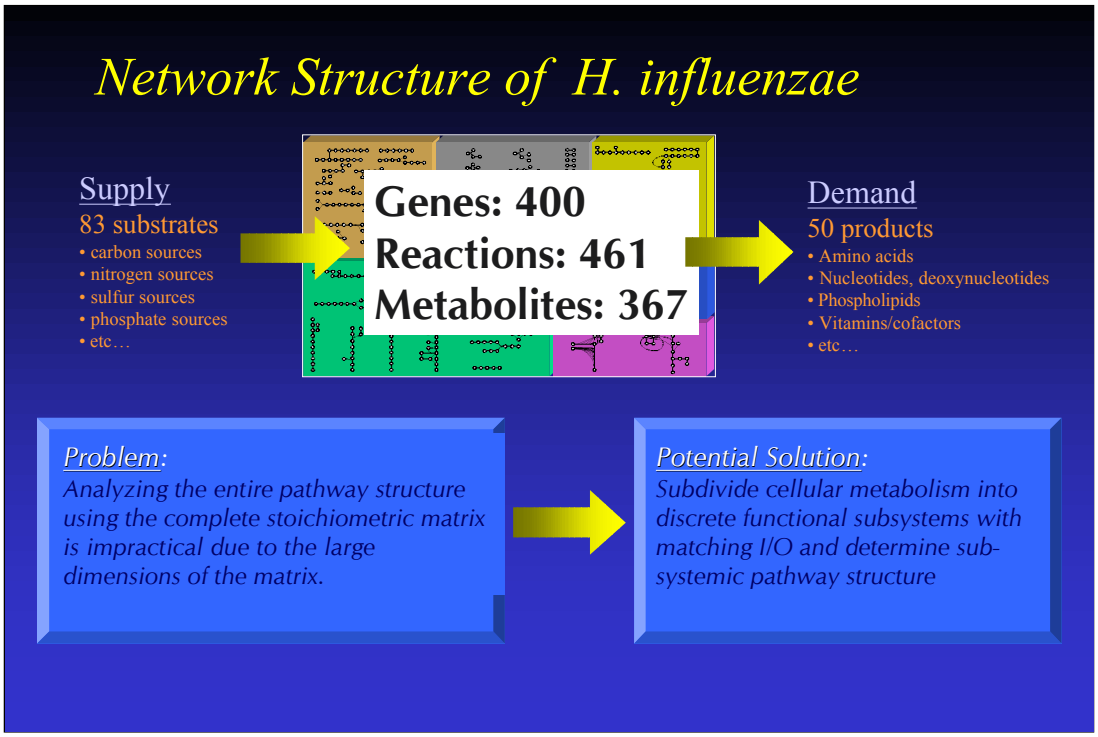


A PATHWAY FOR PARTIALLY AEROBIC GROWTH (ACETATE SECRETION)

The third optimal flux distribution of the core metabolic network of *E. coli* with succinate as the sole carbon source is illustrated on this figure (corresponding to pathway #20). In comparison with the optimal flux distribution, acetate is secreted here and enzymatic activities of 2-ketoglutarate dehydrogenase and succinyl-CoA synthetase are reduced to zero.



H. Influenzae is a gram-negative pathogen which colonizes the upper respiratory track and leads to acute and chronic respiratory infections primarily in children. While this pathogen used to be a serious threat to the health of children, the implementation of effective vaccination programs has significantly reduced the incidences of H. influenzae infections. Although its prominence as a pathogen has decreased, it gained recognition in 1995 as being the first free living organism to have its genome completely sequenced. There genome itself is ~1.8Mbp and contains over 1700 genes. We used the genome along with biochemical and physiological data on the organism to reconstruct metabolism and determine the pathway structure of the metabolic network so as to assess the organism's capabilities and fitness under various simulated conditions.

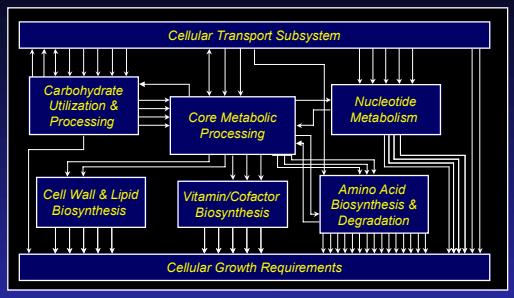


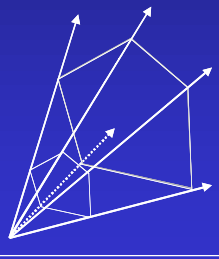
Using the algorithm to determine the stoichiometric matrix and hence metabolic genotype of an organism we assembled the metabolic network for H.influenzae and some of the important numbers are listed here. The network is supplied by 83 potential substrates and requires 50 products to be generated using a series of 461 reactions. As you can see the number of reactions and metabolites which exist within the system is quite large to no surprise. If we were to apply an algorithm to determine all of the genetically independent pathways operating in such systems the number of pathways would be on the order of tens of thousands. Obviously this quite impractical and so a potential solution is to divide and conquer. We can divide cellular metabolism into discrete function subsystems with matching inputs and outputs and determine the pathway structure of each subsystem.

Pathway Structure of H.influenzae

Subsystems of Cellular Metabolism:

- (T) Transport [Electron Trans, Assim./Dissim.]
- (A) Amino Acid Metabolism
- (N) Nucleotide Metabolism
- (V) Vitamin & Cofactor Biosynthesis
- (C) Central Metabolic Processing
- (L) Lipid and Cell Wall Biosynthesis





Subsystem	System Characteristics: Reactions	Intr. Fluxes	Exch. Fluxes	Extreme Pathways: Type I	Type II
T	144	207	152	150	16
A	87	111	59	44	1
C	34	53	35	67	3
N	84	124	45	702	13
V	58	63	39	15	1
L	56	74	34	16	2

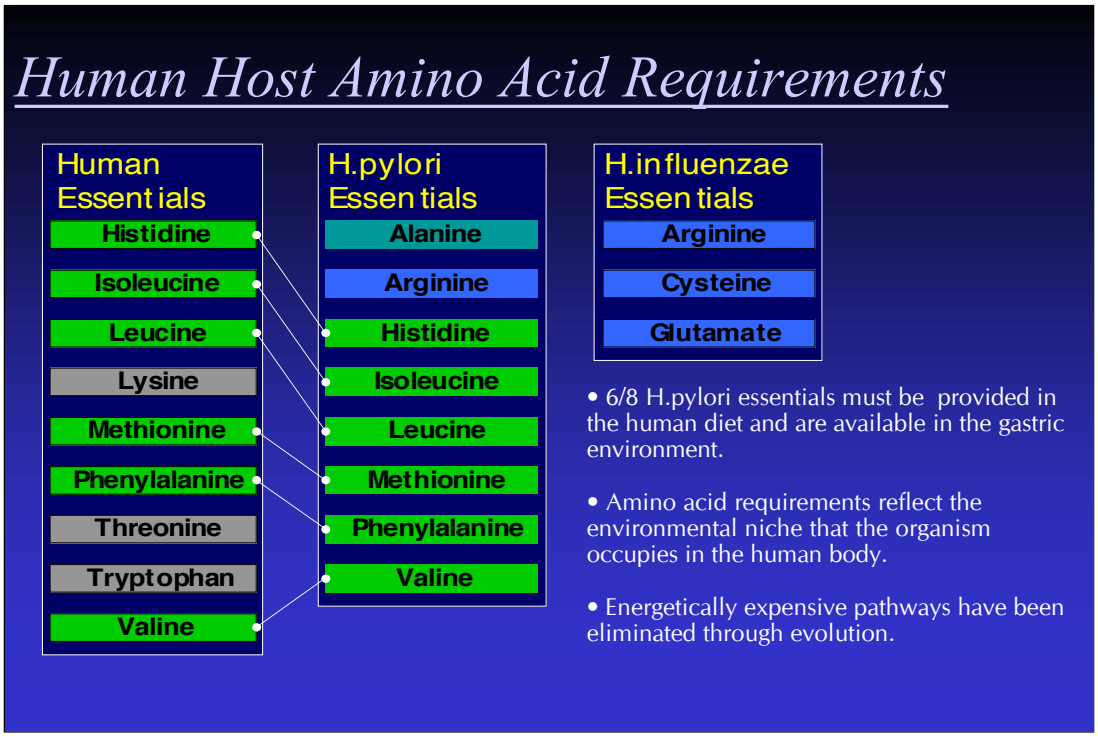
Using the computer algorithms to determine these generating vectors or extreme pathways for the system we can find all of the pathways in each subsystem and classify them. The table above indicates the number of pathways calculated in each subsystem along with the systems characteristics of each of the subsystems. The same exact analysis was performed on Helicobacter pylori which is comprised of a metabolic network that is roughly the same size as the H. influenzae model.

Uses of Pathway Analysis

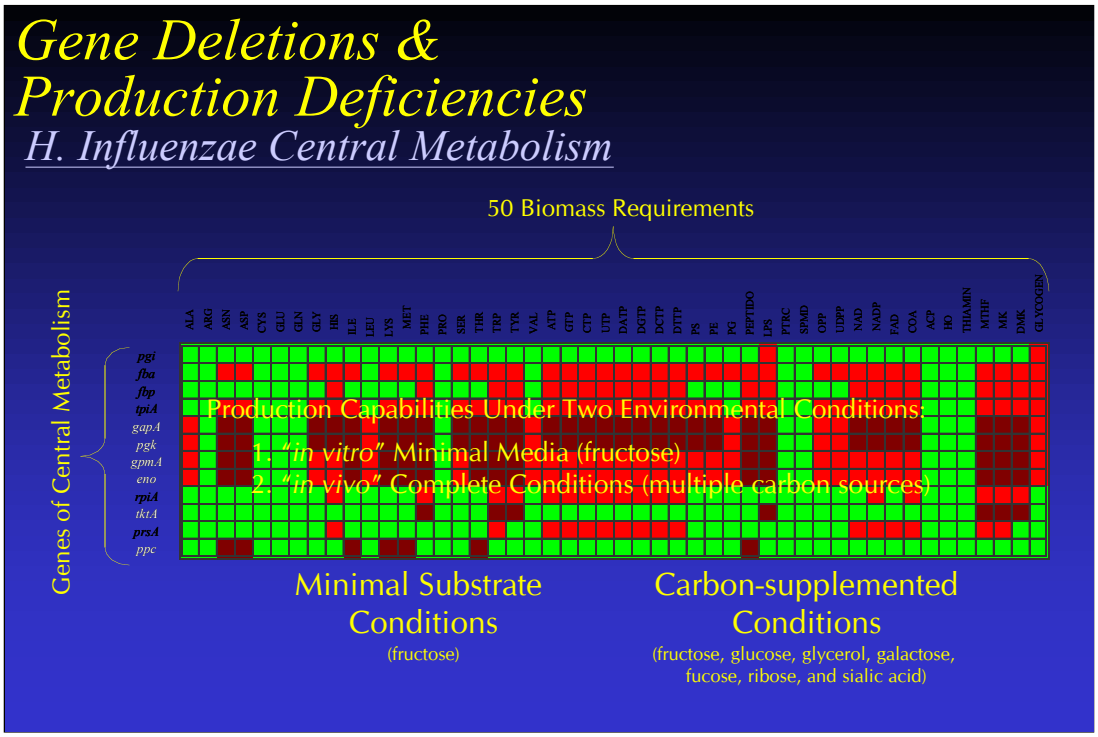
- Complete network divided into six subsystems and extreme pathways calculated in each system
- Applied to H. pylori & H. influenzae Metabolism

- Reaction subsets imply limited regulation
- Minimal Substrate Requirements
- Essential Amino Acid Requirements
- Gene deletions and loss of capability
- 7 Global Entry Points into central metabolism

From the detailed analysis of these sets of pathways a number of interesting results can be generated. Reactions that do not appear in any of the pathways can be used to reconcile possible gaps in the genome annotation. Enzyme subsets can be identified which indicate groups of reactions that always occur in the same pathways in the same flux ratios indicating potential regulons. Also minimal substrate requirements and alternative substrate can be identified by assessing which pathways can be combined under different conditions to produced the required demands on the system.



For Helicobacter pylori the model reveals that there are 8 essential amino acid which the organism must acquire from the environment. Of these 8, 6 of them are essential amino acids which are required by the human host, meanwhile in H. influenzae there is no overlap between required amino acids and human essentials. This leads to the conclusion that H. pylori has removed expensive biosynthetic pathways for amino acid production in favor of acquiring the amino acid from the gastric environment where amino acids should be plentiful as this is the site of proteolysis in the human digestive system.



Using the pathways we can begin to identify potential antimicrobial drug targets by examining the consequences of removing genes under different environmental conditions. In H. influenzae we found 12 genes in central metabolism which were essential to the networks ability to produce the biomass requirements. The red boxes indicate those requirements which could not be produced when the gene on the left was deleted and the minimal substrate conditions were presented. The dark red boxes indicate products which could not be produced even with the addition of all the possible carbon sources made available to the network.

Computation of genome-scale pathways for H. pylori

- The metabolic network for H. pylori consists of 583 reactions and 381 metabolites
- Currently computing the pathways on the San Diego supercomputer cluster (alpha machines)
  - Algorithm can not be parallelized easily and requires a fast processor with large memory
  - Currently calculation of the full pathway structure is infeasible (time and memory requirements are too large)
  - Can restrict the outputs of the metabolic network to smaller subsets for useful studies

Computation of pathways, continued

- Instead of looking at pathway structure of entire network with all the outputs, restrict the network to subsets (amino acid production, nucleotide production)
- Can examine a number of different issues for this limited set of outputs
  - Correlation of pathways
  - Biochemical yields
  - How these pathways correspond to physiological function
  - Degree of robustness and duplication of the network

Immer immer…..

in dem computer zimmer

Jim Rawlings, 1982

Summary

- Basis vectors of the null space are pathways
- Convex analysis by using positive fluxes only
- Extreme pathways as edges of cones--there are three basic types
- These pathways give much physiological insight
- Linked outputs lead to flux distributions
- Linked pathways cannot yet be calculated on a genome scale

References

- Clarke, B.L., "Stability of complex reaction networks," *Advances in Chemical Physics* **43**: 1-215 (1980).
- Gilbert Strang, Linear Algebra and Its Applications, Academic Press, New York, 1981.
- Seressiotis, A., and Bailey, J.E., "MPS: An artificially intelligent software system for the analysis and synthesis of metabolic pathways," *Biotech. And Bioeng.* **31**: 587-602 (1988).
- Mavrovouniotis, M.L. Stephanopolous, G. and Stephanopoulous, G.,"Computer-aided synthesis of biochemical pathways," *Biotech. And Bioeng.*, **36**: 1119-1132 (1990).
- Schuster, S., and Hilgetag, C., "On elementary flux modes in biochemical reaction systems at steady state," *J. Biological Systems* **2**: 165-182 (1994).
- Reinhart Heinrich and Stefan Schuster, The Regulation of Cellular Systems, Chapman and Hall, New York, 1996.
- Liao, J.C., Hou, S-Y., and Chao, Y-P., "Pathway analysis, engineering, and physiological considerations for redirecting central metabolism," *Biotech. and Bioeng.*, **52**: 129-140 (1996).
- David Lay, Linear Algebra and its Applications, Addison-Wesley, Menlo Park, 1997.
- C.H. Schilling and B.O. Palsson, "The underlying pathway structure of biochemical reaction networks," *Proc. Natl. Acad. Sci (USA),* **95**: 4193-4198, (1998)
- C.H. Schilling, S. Schuster, B.O. Palsson, and R. Heinrich, "Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-Genomic Era," *Biotechnology Progress*, **15**: 296-303 (1999).

References

- C.H. Schilling, D. Letscher, and B.O. Palsson, "Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective," *J. Theoret. Biol.*, **203:** 229-248 (2000)**.**
- C.H. Schilling and B.O. Palsson, "Assessment of the Metabolic Capabilities of *Haemophilus influenzae* Rd through a Genome-Scale Pathway Analysis," *J. Theoret. Biol.*, **203:** 249-283 (2000).
- Schwikowski, B., Uetz, P., and Fields, S., "A network of protein-protein interactions in yeast," *Nature Biotechnology*, **402**: 1257-61 (2000).
- Eisenberg, D.,Macotte, E.M., Xenarios, I., and Yeates, T.O., "Proteomics in the post-genomic era," *Nature,* **405**: 823-826 (2000).